Fairness-Aware Engagement Loss Reduction in Rumor Mitigation

Jiajie Fu, Xueqin Chang, Xiangyu Ke, Lu Chen
Zhejiang University
{jiajiefu, changxq, xiangyu.ke, luchen}@zju.edu.cn

***Abstract*—The rapid spread of negative information, such as rumors, poses significant challenges for social network operations and has considerable economic consequences. Traditional rumor mitigation approaches focus on minimizing the number of rumor-affected users, often overlooking the varying social engagement levels and fairness concerns among user groups. To address these gaps, we introduce the Fairness-aware Temporal Loss-based Rumor Mitigation (FAIR-TLRM) problem, which aims to identify a set of truth spreaders that maximize the expected reduction in economic loss while satisfying fairness constraints. FAIR-TLRM problem is NP-hard, monotone, and non-submodular. To solve this, we propose a dual-objective optimization algorithm, Fair-Greedy, for selecting truth spreaders by balancing loss reduction and group fairness. Given the #P-hardness of computing mitigation rewards, we further devise a scalable algorithm, FWS-RM, that leverages group-aware weighted reverse influence sampling and a sandwich approximation technique. We empirically explore the trade-offs between rumor-induced losses and fairness concerning the network characteristics and introduce an effective Joint-Greedy framework. Compared with existing RM solutions, FWS-RM reduces economic losses by an average of 2×, improves fairness by up to 8×, and runs 5× faster.**

***Index Terms*—Rumor Mitigation, Fairness, Social Networks**

I. Introduction

Social networks have become vital platforms for information dissemination, playing a critical role in applications such as recommendation systems [1], network monitoring [2], and viral marketing [3]. These applications have generated substantial revenue, contributing billions to their owners. However, the rapid growth of social networks has also facilitated the spread of harmful content, including rumors and disinformation [4], leading to significant financial losses[1].

Existing rumor mitigation (RM) approaches can be generally categorized as blocking and clarification approaches. Blocking-based methods [6]–[8] focus on removing detected rumor accounts or relevant connections from networks to stop the rumor spread, while clarification-based ones [9]–[11] seek to disseminate factual information to counteract rumors. However, banning users or connections may significantly disrupt the network structure and raise ethical concerns [12]. Therefore, this study follows the clarification-based manner.

Active user engagement is crucial for social network platforms, as their stock prices and advertising revenue are closely tied to user vitality [13], [14]. However, the rapid spread of rumors on online social networks (OSNs) can significantly erode user trust, leading to a decline in engagement and, consequently, a drop in platform revenue [5]. Existing clarification-based methods primarily aim to reduce the number of users exposed to rumors, yet they often overlook the economic consequences of user disengagement. In this work, we argue that to truly address rumor spread, it is essential to *prioritize mitigating engagement loss*. Intuitively, users with higher engagement levels—such as those with substantial interaction histories—represent more significant potential losses for platforms, as their trust and activity are more likely to diminish after encountering rumors [15] (§ III-B).

In addition to the economic implications, there is growing concern about the *social impact* of rumors, particularly *their targeting of specific, often marginalized, groups* [16], [17]. Research indicates that many rumors disproportionately affect certain communities, such as racial minorities or diasporic populations. For example, communities of color are frequently the focus of malign influence campaigns, particularly during elections and public health crises[2]. During the COVID-19 pandemic, rumors spread $5\times$ broader among Latino and African American communities, undermining trust and increasing vulnerability [18]. This disproportionate targeting exacerbates social inequalities [19], yet many existing rumor mitigation strategies fail to address these fairness concerns adequately.

Building on the discussions above, we introduce a novel problem formulation called Fairness-aware Temporal Loss-based Rumor Mitigation (FAIR-TLRM). Given a rumor seed set S_F, our objective is to identify a budget-limited truth seed set S_T, such that the expected user engagement loss reduction on OSNs is maximized after S_T disseminates the truth, while ensuring the welfare for the *least advantaged* group to reach a predetermined threshold (§ III-C). Practically, this fairness consideration often goes overlooked when focusing solely on engagement loss reduction for more active users, but it is critical for achieving equitable and sustainable rumor mitigation, as demonstrated below.

Case Study. The real-world Email-EuAll social network [20] captures email communications, where each user is associated with one of 42 departments based on their job occupation. We evaluate three simple greedy rumor mitigation algorithms with different truth seed selection criteria: (1) Rumor Exposure Mitigation (REM) [21]: minimizes the number of rumor-affected users; (2) Engagement Loss Mitigation (ELM) (§ V-A): minimizes the user engagement loss; and (3) ELM+ (§ V-A): further introduces a fairness constraint to ELM.

[Settings] We impose a simple fairness constraint: the least advantaged group must achieve at least 90% of its optimum loss reduction. 10 nodes are randomly selected as rumor seeds, and the results are averaged on 5 repeats. Individual user loss is measured based on network topology[3]. Only departments with total loss exceeding 0.2 are displayed. Besides the losses

[1]False news can easily result in a loss of 2.11 million USD in equity value over a ten-day period on Twitter [5]

[2]https://mediaengagement.org/research/the-impact-of-disinformation-targeted-at-communities-of-color

[3]Please refer to § III-B for the computation details and more choices.

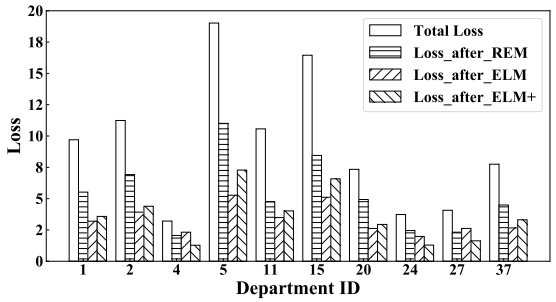


Fig. 1. The loss of different departments. The **total loss** represents the initial loss within each department before any intervention.

TABLE I
THE LOSS AFTER APPLYING THREE ALGORITHMS

Metric \ Method	REM	ELM	ELM+
Total Loss	216.62	216.62	216.62
Loss After Algorithm	127.05	72.16	84.71
Save Ratio	41.35%	66.69%	60.89%

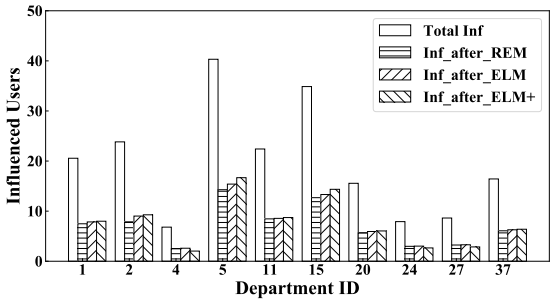


Fig. 2. The count of influenced users of different groups

TABLE II
THE INFLUENCE AFTER APPLYING THREE ALGORITHMS

Metric \ Method	REM	ELM	ELM+
Total Influence	459.28	459.28	459.28
Influence After Algorithm	167.23	175.45	178.57
Save Ratio	63.59%	61.79%	61.12%

and the number of influenced users across different departments, we present the *save ratio*: the proportion of the initial loss mitigated by each algorithm, to measure the effectiveness.

[Economic Impact and Fairness] In Figure 1, both ELM and ELM+ achieve more significant economic loss reduction than REM (i.e., lower bar; see Table I for detailed numbers). This demonstrates that minimizing the number of affected users alone, as done in [10], [21], fails to account for the varying individual rumor-induced engagement losses across users.

However, a noticeable drawback of REM and ELM is their *significant fairness imbalance*. For instance, in department 4, the save ratio for REM is only 34.78%, much lower compared to the 55.15% achieved in department 11. The fairness issue is even more pronounced for ELM, where the save ratio in department 4 dropped to just 27.73%, compared to a high of 72.29% in department 5. Such discrepancy stems from the fact that users with high engagement tend to be clustered within the same group, which worsens the fairness imbalance. Upon applying the ELM+, we observe a significant improvement in fairness: The save ratio in department 4 increases to 60.12%, and even in department 37—typically the department with the lowest ratio—the percentage rises to 57.29%. As shown in Table I, the sacrificed economic loss for ensuring fairness is minimal, e.g., only reduced by less than 6% in terms of save ratio and is still much higher than the 41.35% of REM.

[Compatibility with Maximizing Affected User Count] We evaluate three algorithms under the traditional rumor mitigation problem, which focuses on minimizing the number of users affected by rumors. While REM results in a slightly smaller number of rumor-affected users compared to ELM and ELM+ (Figure 2), Table II demonstrates that the performance of ELM and ELM+ is very close to that of REM, achieving nearly 96% of its effectiveness. This demonstrates that our algorithms perform well not only on the Fair-TLRM problem but are also *highly compatible* with the traditional RM objective of *minimizing rumor exposure*.

Challenges and Our Contributions. To the best of our knowledge, we are the first to approach rumor mitigation from the perspective of simultaneously reducing economic losses and maintaining group fairness. We prove that our FAIR-TLRM problem is **NP**-hard and that computing the loss reduction achieved by any truth seed set is #**P**-hard (§ IV). This dual focus introduces several challenges that are not addressed by existing literature. Among existing works, *Fair Targeted-IM* [22] is the most similar to ours, as it considers both fairness and targeted users in the *influence maximization (IM)* setting. However, several key differences make this approach unsuitable for RM problems: (1) In Targeted-IM [23]–[25], the target users are known beforehand, which allows for the application of efficient *Reverse Influence Sampling (RIS)* technique [26]. In RM, however, the rumor-affected users are unknown, making RIS impractical in this context. (2) It does not account for individual loss reductions or the group structure of the users, which leads to suboptimal seed selection in terms of reducing engagement losses. (3) *Fair Targeted-IM* is submodular [22], which allows the use of a greedy approach with approximation guarantees, whereas our FAIR-TLRM is non-submodular, thus requiring more sophisticated techniques to find optimal solutions.

To tackle these challenges, we adopt the *Sandwich Approximation* strategy [27], a well-established framework for solving non-submodular maximization problems. A key challenge here is *devising accurate and tight submodular bounds*, as the effectiveness of the Sandwich strategy hinges on how closely these bounds approximate the objective function. Based on this framework, we propose Fair-Greedy, a greedy-based *dual-objective* optimization approach that consists of two stages to choose the truth seed set. Specifically, it selects the minimum number of nodes required to meet the fairness criterion and utilizes the remaining budget to maximize the loss reduction. As shown in Table I and Table II, compared to existing methods [10], [21] that simply maximize the number of users reached by truth, ours can increase the economic loss saving by almost 20% under a fairness constraint, with minor sacrifice in rumor exposure ($< 2.5\%$). For better efficiency, we designed a novel RR-Set generation method specially tailored for the unknown rumor victim and varying engagement level issues

posed by our problem and propose FWS-RM, achieves up to 3 orders of magnitude speedup over Fair-Greedy empirically (§ VII). Finally, we propose the *Joint-Greedy* framework, which balances algorithm performance and fairness more effectively, ensuring that both objectives are met without compromising effectiveness (§ VI).

Contributions and Roadmap.

- We formulate the novel *Fair-TLRM* problem that mitigates rumor-induced losses while considering fairness issues (§ III), and characterize the hardness (§ IV).
- We design effective submodular and monotonic bounding functions for sandwich approximation to work on (§ IV).
- We propose a greedy-based dual-objective optimization algorithm, augmented with customized RIS sampling, which maintains a $(1 - 1/e - \epsilon)$-approximation guarantee and achieves up to 3 orders of magnitude speedup (§ V).
- We conduct several real-world case studies for insights on balancing the loss reduction and maintaining fairness, and propose a *Joint-Greedy* framework accordingly (§ VI).
- We evaluate our solutions on five real-world social networks. The results demonstrate that our algorithm reduces economic losses by an average of $2\times$, improves fairness by up to $8\times$, and runs $5\times$ faster than RM baselines. (§ VII).

II. Related Work

Influence Maximization. The Influence Maximization (IM) problem was first formulated as a discrete optimization problem by Kempe et al. [28], who demonstrated that it is **NP**-hard and proposed a Monte Carlo-based greedy algorithm with $(1-1/e)$-approximation. Since then, various efficient solutions and extensions of IM have been explored [25], [29]–[31]. Among the many variants, targeted IM [23], [32] and fairness-aware IM [22], [33], [34] are most relevant to our work. However, key differences—such as the submodular nature of their objective functions—make these solutions unsuitable for our problem. Specifically, targeted IM focuses on maximizing coverage for a specific group, without considering fairness across all groups. In contrast, fairness-aware IM approaches either lack theoretical guarantees and rely heavily on machine learning techniques [34], or suffer from scalability [33].

Rumor Mitigation. Budak et al. [21] first introduced Rumor Mitigation (RM) problem under the IC model, which aims to limit the spread of rumors in social networks. Since then, substantial follow-up research has sought efficient methods to address RM and its variant [35]–[37]. In addition to launching truth campaigns as mentioned above, blocking strategies, which involve banning certain nodes or connections, serve as another common method for RM [38]–[40].

Nevertheless, few of these works consider the fact that the propagation speed of truth and rumor often differs. The recently proposed TCIC model [11] begins to address this gap by accounting for this difference, but the focus is not on reducing economic losses yet. Moreover, existing studies fail to provide methods for assessing the individual economic losses (e.g., drop in user engagements) incurred by social platforms when users encounter rumors. Furthermore, fairness considerations are rarely integrated into these approaches.

In this paper, we address the novel problem of maximizing the reduction of rumor-induced losses in OSNs while incorporating fairness constraint (Fair-TLRM). We propose an efficient approach that combines a customized Reverse Influence Sampling (RIS) technique with the sandwich method, providing valuable insights into balancing the trade-off between economic losses from rumors and fairness.

III. Preliminaries

In this section, we introduce the TCIC model (§ III-A) and present various methods for measuring individual user losses when influenced by rumors (§ III-B). Building on this foundation, we formally define the Fairness-aware Temporal Loss-based Rumor Mitigation (Fair-TLRM) problem (§ III-C).

A. The TCIC Model

Consider a directed graph $G = (V, E, p)$, where V and $E \subseteq V \times V$ represent the sets of vertices and edges, respectively. Let $|V| = n$ and $|E| = m$. Each edge $e = (u, v)$ is associated with an influence weight $p_{u,v} \in$ [0,1], which quantifies the impact strength of u on v.

The Temporal Competitive Independent Cascade (TCIC) model depicts a diffusion process where two mutually exclusive campaigns, rumor (F) and truth (T), vie for propagation [11]. Unlike the traditional CIC [41], [42] and IC [28] models, which assume uniform propagation rates, TCIC introduce a a key distinction: a *meeting delay* to account for varying propagation speeds of different types of information[4]. Additionally, TCIC incorporates the concept of an *activation window (AW)*, a period during which users contemplate which information to adopt [43]. If both campaigns reach a user in this window, a *proportional tie-breaking*[5] dictates her decision.

The propagation dynamics of the TCIC model are outlined as follows: **(1)** Seeds S_F for rumor and S_T for truth are activated at timestamp 0, while other nodes remain inactive; **(2)** Each newly activated node has one chance to influence its inactive neighbors after a specified *meeting delay*, with the likelihood of success determined by the edge's influence probability. **(3)** Once activated, a node enters an *activation window*, after which it adopts either F, T, or uses *tie-breaking rules* if influenced by both. After adoption, we assume the user remains committed to their decision. **(4)** Propagation continues until no further nodes can be activated.

Example 1. *In Figure 3, v_F is the rumor seed and v_T is the truth seed. All edges share a same propagation probability of* 1. *The blue numbers on the edges represent the meeting delays of truth. We assume a constant delay of 1 for rumor. The red number beside the node represents the activation window, with unspecified AW lengths set to 0* [6]*. We exemplify the influence propagation as follows:*

- *$t = 0$, node v_F and v_T activate.*

[4]Rumors typically propagate $6\times$ faster than truth on average [4].

[5]TCIC usually applies a weighted random choice policy based on the number of activated neighbors. Specifically, a node u adopts rumor F with probability $p_F(u) = |N_F^-(u)|/|N^-(u)|$, where $|N_F^-(u)|$ is the number of neighbors who propagated F, and $|N^-(u)|$ is the total number of neighbors exposed to either F or T. The same logic applies to the truth campaign.

[6]Unless otherwise specified, in this work we set the activation window to 0 for simplicity, though our algorithm is applicable to all possible values.

- $t = 1$, *rumor from* v_F *reaches* v_1, *opening* v_1*'s AW.*
- $t = 2$, *truth from* v_T *reaches* v_1 *at the end of its AW. Following the tie-breaking rule,* v_1 *adopts* F *or* T *with equal probability. Let us assume* v_1 *adopts* F.
- $t = 3$, v_3 *is influenced by* v_T *and* v_2 *by* v_1. *As both AWs are 0,* v_3 *adopts* T *while* v_2 *adopts* F, *and the propagation concludes.*

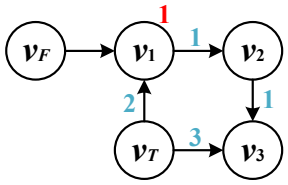


Fig. 3. The propagation process under TCIC model

The rumor mitigation reward (i.e., the loss reduction) only accounts for nodes that would have adopted the rumor in the absence of the truth set but switch to adopting the truth once the truth set is introduced. These nodes are considered to be "successfully rescued". A similar approach is outlined in [10].

B. Individual user losses measurement method

In this section, we detail two distinct methods for assessing user engagement. These methods quantify the impact of being affected by rumors, relying on either network topology or user historical behavior data. In addition, we keep another basic method of assigning random weights for experimental purposes in later sections (§ VII).

- ***Topology-based:*** Lurker ranking methodologies [44] assign users scores that reflect their level of activity within a social network, especially the passive ones. The intuition is that, in social networks, a user u's in-degree $in(u)$ often represents her information reception, while the out-degree $out(u)$ signifies her contribution to information sharing, both of which provide a robust indicator of user activity. The *user engagement score* based on network topology is **recursively defined** as $w(\cdot)$ in Equation 1, with the assumption that a node v with infinite in/out-degree ratio is trivially regarded as a lurker. For more details, please refer to [44]:

$$s(v) = d[s_{in}(v)(1 + s_{out}(v))] + (1 - d)p(v) \quad (1)$$

where $s_{in}(v)$ is the in-neighbors-driven score function, the idea behind it is that the strength of v's lurking status is proportional to the influential status of the v's in-neighbors, and it's further enhanced by including a factor that is inversely proportional to the v's out-degree. $p(v)$ is a personalized value, which is set to $1/|V|$ by default and d is a damping factor ranging within [0,1]:

$$s_{in}(v) = \frac{1}{out(v)} \sum_{u \in N^{in}(v)} \frac{out(u)}{in(u)} s(u) \quad (2)$$

and $s_{out}(v)$ is the out-neighbors-driven score function, it follows the principle of *Non-authoritativeness of the information produced*, i.e., the strength of v's lurking status is proportional to the lurking status of the v's out-neighbors:

$$s_{out}(v) = \frac{in(v)}{\sum_{u \in N^{out}(v)} in(u)} \sum_{u \in N^{out}(v)} \frac{in(u)}{out(u)} s(u) \quad (3)$$

This user engagement score is then logarithmically transformed and normalized to obtain a value between $[0, 1]$, representing the users' silence weights. To derive the user's **node weight** $w(v)$, we subtract the silence weight from 1.

- ***Historical-data-based:*** A user's historical behavior data is often the most direct indicator of her engagement level, such as app usage time, purchase records, or review counts. We use the most widely-used logarithmic normalization method [45] to measure *user engagement score* as below:

$$s(v) = \frac{log(1 + count(v))}{log(1 + max_{u \in V}\, count(u))} \quad (4)$$

Where V is the vertex set in G and $count(v)$ could be any frequency-based historical data associated with user v. The user's node weight $w(v)$ is equal to $s(v)$ in this setting.

C. Problem Definition

Given a known rumor seed set S_F, our goal is to identify a truth seed set S_T that maximizes the loss reduction for OSNs while maintaining a desired level of fairness. Specifically, we aim to save the nodes that would otherwise adopt the rumor in the absence of S_T, maximizing the *total losses* (i.e., node weights) of the saved nodes, all within the fairness constraint.

Definition 1. (Loss Measure Function). *Given a graph* G, *a rumor seed set* S_F *and a truth seed set* S_T, *the loss measure function* $\mathbb{L}(G, S_F, S_T)$ *represents the expected losses caused by rumor on graph* G *under* TCIC *model, formally,*

$$\mathbb{L}(G, S_F, S_T) = \sum_{v \in MAS} w(v) \quad (5)$$

where MAS represents the nodes affected by rumors.

When given the specific group C_i, the expected losses on the i^{th} group of G can be written as $\mathbb{L}_i(G, C_i, S_F, S_T)$. Then, expected loss reduction $\mathbb{R}(\cdot)$ can be calculated as follows:

$$\mathbb{R}(G, S_F, S_T) := \mathbb{L}(G, S_F, S_T) - \mathbb{L}(G, S_F, \emptyset) \quad (6)$$

Definition 2. (Maxmin Fairness). *Given a graph* G, *a rumor seed set* S_F, *a budget* b, *and a group division* $\mathcal{C}$ *of nodes in* G, *the Maxmin-Fairness* $opt(G, S_F, b, \mathcal{C})$ *of* G *is:*

$$\max_{\substack{S_T \subseteq V \setminus S_F, \\ |S_T| \leq b}} \min_{C_i \in \mathcal{C}} \frac{\mathbb{L}_i(G, C_i, S_F, \emptyset) - \mathbb{L}_i(G, C_i, S_F, S_T)}{\mathbb{L}_i(G, C_i, S_F, \emptyset)}$$

Unlike *Equality Fairness* and *Diversity Fairness* [46] that often struggle to accommodate balanced outcomes, the *Maximin Fairness* aligns with John Rawls' well-known philosophical "Theory of Justice" [47]: *Societal systems should be structured to maximize the welfare of the least advantaged group*. In social analysis, policies often aim to reduce inequality by improving conditions for the worst-off rather than just ensuring equal treatment across all groups. Maximin fairness prioritizes the most affected individuals (e.g., those most seriously impacted by rumors), making it a straightforward yet powerful tool for evaluating and promoting equitable interventions, such as the dissemination of factual information.

Problem 1. (FAIR-TLRM). *Given a social graph* G, *a rumor seed set* S_F, *a budget* b, *and a fairness parameter* α.

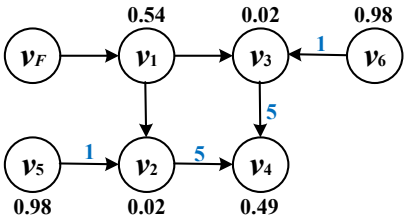


Fig. 4. Example to the proof of non-submodularity

*The Fairness-aware Temporal Loss-based Rumor Mitigation (*FAIR-TLRM*) problem finds a truth spread set S_{T^*} of at most b nodes that maximizes the expected loss reduction while achieving at least α of the maxmin fairness. Formally:*

$$S_{T^*} := \arg \max_{\substack{S_T \subseteq (V \setminus S_F), \\ |S_T| \leq b}} \mathbb{R}(G, S_F, S_T)$$
$$s.t. \ \min_i \frac{\mathbb{L}_i(G, C_i, S_F, \emptyset) - \mathbb{L}_i(G, C_i, S_F, S_T)}{\mathbb{L}_i(G, C_i, S_F, \emptyset)} \geq \alpha * opt$$

IV. PROBLEM CHARACTERISTICS

In this section, we analyze the computational complexity and mathematical properties of the FAIR-TLRM problem. First, we establish the **NP**-hardness of the FAIR-TLRM problem. Then, we prove that the FAIR-TLRM is both *monotone* and *non-submodular*.

NP-Hardness. The Influence Minimization (IM) problem under the CIC model has been proven to be **NP**-hard [21]. The CIC-based IM can be viewed as a special case of FAIR-TLRM problem, where the meeting delay $m(u, v)$ for all nodes is set to 1, all node weights are equal 1, the activation window (AW) is zero, truth-dominant tie-breaking[7] is applied, and the fairness threshold α is set to 0. Given these conditions, the FAIR-TLRM problem inherits the **NP**-hardness of the IM problem under the CIC model.

Monotonicity. Our FAIR-TLRM problem remains *monotonically non-decreasing* under TCIC model, as shown in the theorem below.

Theorem 1. *Given a graph G and a rumor seed set S_F, the function $\mathbb{R}(G, S_F, S_T)$ is monotonically non-decreasing under the* TCIC *model.*

Proof. Let S_z and $S_{z'}$ be two truth seed sets such that $S_z \subseteq S_{z'} \subseteq (V \setminus S_F)$. Since adding more truth seeds can only reduce the number of nodes influenced by rumors, we must have $\mathbb{R}(G, S_F, S_{z'}) \geq \mathbb{R}(G, S_F, S_z)$. Furthermore, if S_z has satisfied the fairness constraint, adding any node $v \in S_{z'} \setminus S_z$ to S_z will not violate the fairness constraint, since the group with the minimum gain has already met the fairness requirement. This completes the proof. □

Non-Submodularity. We illustrate the non-submodularity of the FAIR-TLRM problem using a counterexample.

Theorem 2. *The* FAIR-TLRM *problem is non-submodular under the* TCIC *model.*

Proof. Consider the graph in Figure 4, where v_F represents the rumor seed. Assume the rumor propagation delay d_e^F takes 1 for each edge, while those for truth d_e^T are given by the **blue** numbers. The weight beside each node is calculated through network topology. The fairness constraint parameter α is set to 0, i.e., any truth seed set simply satisfies the fairness constraint. We use t_v^F and t_v^T to represent the timestamp that v is reached by F or T. For convenience, we use $\mathbb{R}(S_T)$ to replace $\mathbb{R}(G, S_F, S_T)$ in following works.

$\{v_1, v_2, v_3, v_4\}$ are originally affected by the rumor. Consider two truth seed sets: $A = \emptyset \subseteq B = \{v_5\}$. Truth seed v_5 can reach two victims v_2 and v_4. v_2 can be successfully rescued as $t_{v_2}^T = 1 < t_{v_2}^F = 2$. For v_4, there still exists a faster rumor pathway $v_F \to v_1 \to v_3 \to v_4$, leading to $t_{v_4}^F = 3 < t_{v_4}^T = 6$. Therefore, v_4 cannot be saved and the cost reduction is only $\mathbb{R}(B) = w(v_2) = 0.02$.

Now, consider adding $u = v_6$ to both sets: as the toy graph is symmetric, we can easily obtain $\mathbb{R}(A \cup \{u\}) = w(v_3) = 0.02$.

Moreover, securing v_3 cuts off the aforementioned rumor pathway to v_4, leading to a significant increase in loss reduction, i.e., $\mathbb{R}(B \cup \{u\}) = w(v_2) + w(v_3) + w(v_4) = 0.02 + 0.02 + 0.49 = 0.53$. Hence, we have $\mathbb{R}(B \cup \{u\}) - \mathbb{R}(B) > \mathbb{R}(A \cup \{u\}) - \mathbb{R}(A)$ and $A \subseteq B$, indicating the non-submodularity of our FAIR-TLRM problem. □

Sandwich Approximation [27]. Sandwich approximation method is a well-established approach to provide a data-dependent and quality-bounded solution to non-submodular problems (Lemma 3). Challenges lie in devising a pair of submodular lower ($\underline{\mathbb{R}}$) and upper bounds ($\overline{\mathbb{R}}$) respectively. Next, we present our design and demonstrate the practicality.

We adopt the *Possible World* semantic [28] to facilitate the illustration. Given a uncertain graph G, a possible world $X = (V, E_X)$ is generated by randomly removing each edge $e = (u, v)$ in G with $1 - p(u, v)$ probability. The probability of any possible world X occurring is $P(X) = \prod_{(u,v) \in E_X} p_{u,v} \prod_{(u,v) \in E \setminus E_X} (1 - p_{u,v})$. This yields 2^m distinct possible worlds, denoted as $\mathcal{X}$. The meeting delay on each edge in E_X and the activation time of users are assumed to be pre-determined: The former is sampled through the geometric distribution separately for F and M while the latter can be set to zero for simplicity in the discussion below.

Lower Bound. In any fixed possible world X, the loss reduction for a node $v \in MAS$, denoted by $r_X(S_T \to v)$, depends on whether S_T rescues v. It takes either $p_T(v) \cdot w(v)$ or 0, where $p_T(v) = 1$ if T rescues v without proportional tie breaking, or $p_T(v)$ is the probability that T wins the tie-breaking when both campaigns reach v. The expected loss reduction for v across all possible worlds, denoted by $r(S_T \to v)$, is the weighted sum over all possible worlds, i.e., $r(S_T \to v) = \sum_{X \in \mathcal{X}} Pr(X) \cdot r_X(S_T \to v)$. Hence, we define the lower bound $\underline{\mathbb{R}}(S_T)$ by considering only the nodes in S_T that provide the largest loss reduction to each v and summing these values for all v. Since the loss reduction from the entire truth seed set S_T is evidently not smaller than that from any individual node in S_T, intuitively, we have $\underline{\mathbb{R}}(S_T) \leq \mathbb{R}(S_T) = \sum_v r(S_T \to v)$.

Lemma 1. *$\underline{\mathbb{R}}(S_T) = \sum_v \underline{r}(S_T \to v)$ is a submodular lower bound of $\mathbb{R}(S_T)$, where $\underline{r}(S_T \to v) = max_{u \in S_T} r(\{u\} \to v)$.*

[7]In truth-dominant tie-breaking, when both campaigns reach a node, the node always chooses the truth, which is a simplified form of proportional tie-breaking mentioned in § III-A and does not affect the NP-hardness.

Proof: In any possible world X, it is sufficient to demonstrate submodularity for each affected users, i.e., $v \in MAS_X$, due to closed property of summing submodular functions [48]. Denote $\underline{r_X}(S_T \to v) = max_{u \in S_T} r_X(\{u\} \to v)$. Consider any node $v \in MAS_X$ that satisfies the following condition:

$$r_X(B \cup \{w\} \to v) > r_X(B \to v) \tag{7}$$

where B is a truth set and $w \in V \backslash (B \cup S_F)$. To satisfy Inequation 7, the following two cases must occur:

Case(i): v is influenced by F under B, but influenced by T under $B \cup \{w\}$ without a tie-breaking. Since we only consider the node that brings the max loss reduction in T, under B, $t_v^F < t_v^T$, and under $B \cup \{w\}$, $t_v^T < t_v^F$, the shortest path from $B \cup \{w\}$ to v must begin with w; otherwise, it contradicts Equation 7. Therefore, w contributes to the loss reduction.

Case(ii): v is influenced by F under B, but influenced by T under $B \cup \{w\}$ with a tie-breaking. If F and T arrive at v simultaneously, the probability of T winning the tie-breaking increases if more of v's in-neighbors adopt T. Let N_B^+ denote the number of v's in-neighbors that adopt T under B. To satisfy Equation 7, under B, $t_v^F \leq t_v^T$, and under $B \cup \{w\}$, $t_v^T = t_v^F$ and $N_{B \cup \{w\}}^+ > N_B^+$. The shortest path from $B \cup \{w\}$ to v's in-neighbors must begin with w; otherwise, it contradicts Equation 7. Therefore, w contributes to the loss reduction.

From the above two cases, we could conclude that $\underline{r_X}(B \cup \{w\} \to v) = \underline{r_X}(\{w\} \to v)$, since node w is solely responsible for the increase in loss reduction achieved at v. Let $A \subseteq B$, similar to the situation of B, we can get $\underline{r_X}(A \cup \{w\} \to v) = \underline{r_X}(\{w\} \to v) = \underline{r_X}(B \cup \{w\} \to v)$. Besides, due to the monotonicity of $\underline{r}$, it must satisfy that $\underline{r_X}(A \to v) \leq \underline{r_X}(B \to v)$, therefore, $\underline{r_X}(A \cup \{w\} \to v) - \underline{r_X}(A \to v) \geq \underline{r_X}(B \cup \{w\} \to v) - \underline{r_X}(B \to v)$. Since $\underline{r}(S_T \to v)$ is the weighed aggregation of $\underline{r_X}(S_T \to v)$ and $\underline{\mathbb{R}}$ is the linear aggregation of $\underline{r}$, thus $\underline{\mathbb{R}}$ is also submodular, the proof ends. □

Uppper Bound. Under TCIC model, the propagation of truth typically lags behind the rumor. To adjust the upper bound function, we reduce all truth meeting delays to 1, means that for every e in graph G, the meeting delay for both F and T is 1. We denote such revised possible world as $X^{'}$. This faster truth spread allows truth to reach rumor-affected nodes sooner, increasing the potential loss reduction. For the bounding function $\overline{\mathbb{R}}$, similar to the lower bound, we only consider the node in S_T that offers the greatest loss reduction for each v. However, with proportional tie-breaking, focusing on a single node may decrease the value of $r_X(S_T \to v)$, since $r_X(S_T \to v) = p_T(v) \cdot w(v)$. Under this situation $p_T(v) = N_u^+ \leq N_{S_T}^+$, lowering $p_T(v)$. To counter this, we revise the loss reduction function to $r^*(\cdot)$, where the loss reduction for v always equals $w(v)$ when it encounters a tie-breaking, regardless of which side v ultimately adopts.

Lemma 2. *Denote $\overline{r}_X(S_T \to v) = max_{u \in S_T} r_{X'}^*(\{u\} \to v)$, where $r_X^*(S_T \to v)$ is defined as follows:*

$$r_X^*(S_T \to v) = \begin{cases} 0 & \textit{if v is only reached by } S_F \\ w(v) & \textit{otherwise.} \end{cases} \tag{8}$$

$\overline{\mathbb{R}}(S_T) = \sum_v \overline{r}(S_T \to v) = \sum_v \sum_{X \in \mathcal{X}} Pr(X) \cdot \overline{r}_X(S_T \to v)$ is submodular upper bound of $\mathbb{R}(S_T)$.

Proof: In any modified possible world $X^{'}$, there exist two cases: **(i)** v **is reached by truth earlier** under $B \cup \{w\}$, meaning v is still influenced by the rumor with truth set B but switches to truth under $B \cup \{w\}$. Thus, the shortest path to v must contain w; adding w helps reduce the loss for v. **(ii)** v **is contested by both rumor and truth** under $B \cup \{w\}$, that is, v is still influenced by the rumor under truth set B, but compete with truth under $B \cup \{w\}$. By modifying r to r^*, the loss reduction is always $w(v)$, regardless of the tie-breaking. In this case, the shortest path to v also begins with w. In both scenarios, w is responsible for reducing the loss at v. By monotonicity and aggregation over $X \in \mathcal{X}$ and $v \in MAS$, the loss reduction function $\overline{\mathbb{R}}(S_T)$ is submodular. □

Tighten the Upper Bound. Consider Figure 4, where neither v_5 nor v_6 alone can act as a truth seed to rescue v_4, but together they can. We call this *combination effect*, which arises because their paths to v_4 cannot block (i.e., *intersect* with) all the shortest paths from S_F to v_4 individually [8]. We introduce a indicator function a to determine whether the paths from $S_F(\mathcal{P}_v^F)$ and $S_T(\mathcal{P}_v^T)$ to node v intersect; if no overlap exist, it equals to 0. When this happens, $r_X(S_T \to v) = max_{u \in S_T} r_X(\{u\} \to v)$. To improve the tightness of our lower bound, we revise $\overline{r}_X(S_T \to v)$ as follows:

$$\overline{r}_X(S_T \to v) = \begin{cases} max_{u \in S_T} r_X(\{u\} \to v) & \text{if } a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0 \\ max_{u \in S_T} r_{X'}^*(\{u\} \to v) & \text{otherwise.} \end{cases} \tag{9}$$

Proof: When $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0$, which means there is no overlapped path from S_T and S_F to v. This implies that $r_X(S_T \to v) = max_{u \in S_T} r_X(\{u\} \to v)$. By modifying the reduction function and changing tie-breaking to the truth-dominant one, we obtain:

$$\overline{r}_X(S_T \to v) = \begin{cases} max_{u \in S_T} r_X^{'}(\{u\} \to v) & \text{if } a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0 \\ max_{u \in S_T} r_{X'}^{'}(\{u\} \to v) & \text{otherwise.} \end{cases} \tag{10}$$

Here, $r^{'}$ indicates the propagation follows a truth-dominant policy. It can be proved that $\overline{r}_X(S_T \to v)$ under (10) is submodular. However, after removing all MLs, even without truth-dominance, the reduction we gain still forms an upper bound. The reason is that the *combination effect* comes from the meeting delay of truth, consider the graph in Figure 4, after removing all the MLs, v_5 itself can "save" v_4 without the help of v_6. Due to the fact that multiple nodes in S_T and S_F may arrive at v at the same time, with the existence of *proportional tie-breaking*, only consider the node with the maximum reduction is wrong, the reason is that if the meeting delay of truth is the same as rumor, then only consider the node u with maximum reduction will decrease the reduction at the node v who may encounter a tie-breaking with $N_{S_T}^+ > N_u^+$. Therefore, we modify r to r^*. Now consider the graph in

[8] In aforementioned upper bound, we have consider the case of establishing even shorter path for truth.

Figure **??**[9] where $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 1$ and there exists a special case of tie-breaking: with $\{v_2\}$ as the truth set, the paths for rumor set $\{v_1\}$ to reach $\{v_3, v_6 ... v_7\}$ overlap with that of $\{v2\}$ to these nodes (the overlapped node is v_5, and there is a tie-breaking on v_5 – truth and rumor both reach v_5 at $t = 1$), by switching the r to r^*, the loss reduction of v_5 is always $w(v_5)$ no matter v_5 finally adopts which side, increasing the expected loss reduction. This observation justifies modifying the $\overline{r}_X(S_T \to v)$ to:

$$\overline{r}_X(S_T \to v) = \begin{cases} max_{u \in S_T} r_X(\{u\} \to v) & \text{if } a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0 \\ max_{u \in S_T} r^*_{X'}(\{u\} \to v) & \text{otherwise.} \end{cases} \tag{11}$$

Fix a possible world X, in order to demonstrate submodularity, it is sufficient to prove submodularity for each $v \in MAS_X$ by leveraging the closed property of submodular functions [48]. Consider a truth set B and a node $w \in V \setminus (S_F \cup B)$:

When $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0$, in order to satisfy (7), the loss reduction at v must increase with the inclusion of node w. Consider the following two cases: (i) v is switching to adopt T under $B \cup \{w\}$ w/o tie-breaking and (ii) v believed in T under $B \cup \{w\}$ w/ tie-breaking.

Case(i): If v is influenced by F under B, but influenced by T under $B \cup \{w\}$ without a tie-breaking, for v to satisfy Inequality 7, under B, $t_v^F < t_v^T$, and under $B \cup \{w\}$, $t_v^T < t_v^F$. Therefore, the shortest path from $B \cup \{w\}$ to v must begin with w; otherwise, it contradicts the condition. w contributes to the reduction.

Case(ii): If v is influenced by F under B, but by T under $B \cup \{w\}$ with a tie-breaking, when F and T arrived at v simultaneously. The probability of winning a tie-breaking increases if more of v's in-neighbors adopt T. To satisfy (7), $t_v^F \leq t_v^T$ under B, and $t_v^T = t_v^F$ and $N^+_{B \cup \{w\}} > N^+_B$ under $B \cup \{w\}$. Hence, the shortest path from $B \cup \{w\}$ to v's in-neighbors must begin with w, otherwise it contradicts the condition. Therefore, w also contributes to the reduction.

When $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 1$, in order to satisfy (7), the loss reduction at v must increase with the inclusion of node w. Consider the following two cases: (i) v is not reached by F under $B \cup \{w\}$ and (ii) v is reached by F under $B \cup \{w\}$.

Case(i): Since v is not reached by F under $B \cup \{w\}$, e.g., truth reached v earlier, it must be the situation that under B, $t_v^F < t_v^T$, and under $B \cup \{w\}$, $t_v^T < t_v^F$. Therefore, the shortest path from $B \cup \{w\}$ to v must begin with w; otherwise, it contradicts the condition. w contributes to the loss reduction.

Case(ii): By modifying r to r^*, even F and T arrived at v simultaneously, the obtained loss reduction is $w(v)$ regardless of the tie-breaking. To satisfy (7), $t_v^F < t_v^T$ under B, and $t_v^T = t_v^F$ under $B \cup \{w\}$. Hence, the shortest path from $B \cup \{w\}$ to v must begin with w, otherwise it contradicts the condition. Therefore, w contributes to the loss reduction.

The above cases concluded that $\overline{r}_X(B \cup \{w\} \to v) = \overline{r}_X(\{w\} \to v)$. Let $A \subseteq B$, similar to the situation of B, we can get $\overline{r}_X(A \cup \{w\} \to v) = \overline{r}_X(\{w\} \to v) = \overline{r}_X(B \cup \{w\} \to v)$. Besides, the monotonicity of $\overline{r}$ implies that $\overline{r}_X(A \to v) \leq \overline{r}_X(B \to v)$, therefore, $\overline{r}_X(A \cup \{w\} \to v) - \overline{r}_X(A \to v) \geq \overline{r}_X(B \cup \{w\} \to v) - \overline{r}_X(B \to v)$. Since $\overline{r}(S_T \to v)$ is the weighed aggregation of $\overline{r}_X(S_T \to v)$ and $\overline{\mathbb{R}}$ is the linear aggregation of $\overline{r}$, thus $\overline{\mathbb{R}}$ is submodular, the proof ends.

□

Lemma 3 ([27]). *$\mathbb{R}(S_T) \geq \beta \cdot (1 - 1/e) \cdot \mathbb{R}(S_T^o)$, where $\beta = \max\{\frac{\mathbb{R}(S_T^U)}{\overline{\mathbb{R}}(S_T^U)}, \frac{\mathbb{R}(S_T^o)}{\overline{\mathbb{R}}(S_T^o)}\}$, S_T^o is the optimal solution and S_T^U is the solution derived using the upper bounding function.*

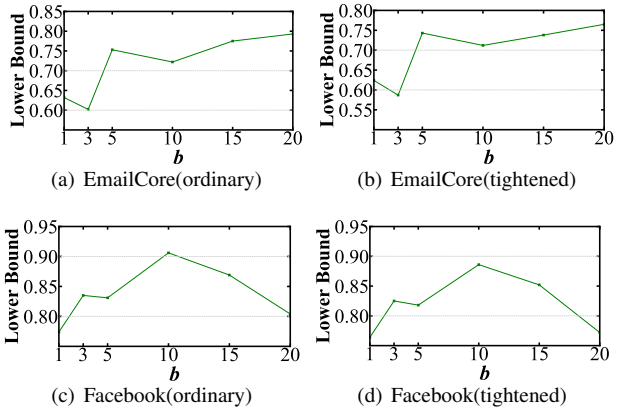


Fig. 5. The Lower Bound of β

Empirical Analysis. Figure 5 shows the lower bounds of β for two real-world datasets. It is evident that for all datasets, $\beta > 0.6$, demonstrating the tightness of our bound. Additionally, we compare the tightened upper bound with the standard bound presented in Lemma 2. The results indicate that the new bound improves the approximation by approximately 5% on average, validating its effectiveness.

V. FAIR-TLRM ALGORITHMS

Traditional truth maximization approaches [10], [21] for rumor mitigation often overlook the differing individual losses among users and do not adequately address fairness concerns (§ I). To tackle these issues, we introduce Fair-Greedy, a dual-objective optimization algorithm tailored for FAIR-TLRM problem (§ V-A). Recognizing the high computational demands of Fair-Greedy on large networks, we propose FWS-RM (§ V-B), a scalable alternative that utilizes a novel group-aware weighted reverse influence sampling method alongside the aforementioned sandwich approximation technique.

A. Greedy Baseline

Revisiting Saturate algorithm. Krause et al. [49] proposed the Saturate algorithm to identify a set that satisfies the Maxmin constraint. The core idea is to greedily select candidates that yield the largest marginal gain for all groups, i.e., "saturate", by truncating the objective function. Specifically, it establishes a fairness threshold opt', ensuring that each group's influence is bounded by this value, thereby facilitating the identification of candidates that maximize the overall influence. In our context, the key is to truncate the maximum ratio of users each group can save to opt' (where opt' is

[9]The number beside the node is its node weight, the blue number represents the MLs of truth, MLs not shown are 1, the red number over the edge is the influence probability, default set is 1.

Algorithm 1 FAIR-Greedy

Input: G, S_F, b, α
Output: S_T

1: Initialize $S_T = \emptyset$
2: Get $S_T^{'}$ and $opt^{'}$; the solution satisfies the $opt^{'}$-fairness constraint and the Maxmin-Fairness score
3: Define $f_\alpha(S_T) := \frac{1}{c}\sum_{i=1}^{c} \min\{1, \frac{\mathbb{R}_i(S_T)}{\alpha \cdot opt^{'} \cdot \mathbb{L}_i(G, C_i, S_F, \emptyset)}\}$
4: **while** $|S_T| < b$ and $f_\alpha(S_T) < 1$ **do**
5: $\quad v^* \leftarrow \arg\max_{v \in V \setminus (S_F \cup S_T)} f_\alpha(S_T \cup \{v\}) - f_\alpha(S_T)$
6: $\quad S_T \leftarrow S_T \cup \{v^*\}$
7: **if** $f_\alpha(S_T) < 1$ **then**
8: $\quad$ **Return** $S_T^{'}$
9: **while** $|S_T| < b$ **do**
10: $\quad v^* \leftarrow \arg\max_{v \in V \setminus (S_F \cup S_T)} \mathbb{R}(S_T \cup \{v\}) - \mathbb{R}(S_T)$
11: $\quad S_T \leftarrow S_T \cup \{v^*\}$
12: **Return** S_T

determined through bisection in $[0, 1]$ and serves as the fairness constraint). We then employ a set-cover algorithm to assess the effectiveness of $opt^{'}$ until the optimal solution is found.

Fair-Greedy. Algorithm 1 presents our basic Fair-Greedy. We start by leveraging Saturate [49] to find the truth set $S_T^{'}$ that satisfies the fairness constraint, with the Maxmin-Fairness score $opt^{'}$ (Line 1-2). Using this $opt^{'}$, we define the truncated objective function $f_\alpha(S_T)$ as mentioned in Saturate (Line 3). We then iteratively choose the node v^* that maximizes the $f_\alpha(S_T)$ until the objective reaches 1 (Line 4-6). If the size of S_T has reached b, but $f_\alpha(S_T)$ does not meet the requirements yet, we return $S_T^{'}$ as the result (Line 7-8). However, if the fairness criteria are fulfilled and the size of S_T is still below b, we will continue to greedily add nodes that maximize the loss reduction until the size of S_T equals b, where the loss reduction could be either measured by $\overline{\mathbb{R}}(S_T)$ or $\underline{\mathbb{R}}(S_T)$ so as to apply Sandwich technique (Line 9-12).

B. Scalable Optimization via Reverse Sampling

Algorithm 1 requires extensive computations of loss reduction to identify candidates for the truth campaign that maximize the increase in $\mathbb{R}(S_T)$ while satisfying the fairness constraint. Calculating the exact influence spread $\sigma(S_T)$ under the CIC model is known to be #P-hard [50], and determining the loss reduction under the TCIC model is even more challenging due to the intricate tie-breaking policies involved.

In the traditional RIS approach, a node v is sampled uniformly at random from G, and a reverse BFS from v is performed to generate a RR-Set, which consists of the nodes reachable from v. The influence spread of a seed set is then estimated by its coverage across multiple RR-Sets. However, these works cannot be extended to solve ours easily. First, traditional RIS starts by sampling a random node from G, whereas the TLRM problem requires sampling nodes affected by rumors. Unlike the well-known targeted IM [24] variant, we do not know in advance which users are influenced by rumors, making it challenging to initiate reverse propagation. Second, in standard RIS, a seed set's expected influence is proportional to the number of RR-Sets it covers, treating all RR-Sets equally. However, in our settings, where nodes have varying weights, applying this approach without adjustment could yield suboptimal truth seed sets. Finally, the fact that nodes belong to different groups introduces additional complexity. Traditional RR-Set indexing methods, which do not consider group structure, are unsuitable for maintaining fairness constraints. While a naive solution might involve building an index for each group, this would introduce unnecessary memory consumption and reduce efficiency.

Algorithm 2 FWS-RM

Input: G, S_F, b, α, ϵ, δ
Output: S_T

1: Initialize $S_T = \emptyset$
2: $\delta^{'} \leftarrow \frac{\delta}{9}$, $\epsilon^{'} \leftarrow \frac{\epsilon}{2}$, $\Delta \leftarrow \frac{8\delta}{9}$
3: Set N_{max} according Eq.(12), $N_0 = N_{max} \cdot \epsilon^2 \frac{LB \cdot ln(1/\delta)}{4n}$
4: Compute $INF_F^{'}$; an $(\epsilon^{'}, \delta^{'})$-approximation of INF_F
5: Generate two sets $\mathcal{R}_1$ and $\mathcal{R}_2$ of random RR-sets, with $|\mathcal{R}_1| = |\mathcal{R}_2| = \theta_0$
6: $i_{max} \leftarrow \lceil \frac{N_{max}}{N_0} \rceil$
7: **for** $i \leftarrow 1$ **to** i_{max} **do**
8: $\quad S_T^* \leftarrow$ WeightedMaxCover($\mathcal{R}_1$,b,n)
9: $\quad$ Compute $\mathbb{R}^l(S_T^*)$ and $\mathbb{R}^u(S_T^o)$ according to (14) and (13), where $\delta_1 = \delta_2 = \Delta/(3i_{max})$
10: $\quad z \leftarrow \mathbb{R}^l(S_T^*)/\mathbb{R}^u(S_T^o)$
11: $\quad$ **if** $z \geq (1 - 1/e - \epsilon)$ **or** $i = i_{max}$ **then**
12: $\quad\quad$ **Break**
13: $\quad$ Double the size of $\mathcal{R}_1$ and $\mathcal{R}_2$ with new RR-Sets
14: Get $S_T^{'}$ and $opt^{'}$; the solution satisfies the $opt^{'}$-fairness constraint and the Maxmin-Fairness score
15: Define $f_\alpha(S_T) := \frac{1}{c}\sum_{i=1}^{c} \min\{1, \frac{INF_F^{'} \cdot \mathbb{E}_i[Y(S_T,R)]}{\alpha \cdot opt^{'} \cdot \mathbb{L}_i(G, C_i, S_F, \emptyset)}\}$
16: **while** $|S_T| < b$ and $f_\alpha(S_T) < 1$ **do**
17: $\quad v^* \leftarrow \arg\max_{v \in V \setminus (S_F \cup S_T)} f_\alpha(S_T \cup \{v\}) - f_\alpha(S_T)$
18: $\quad S_T \leftarrow S_T \cup \{v^*\}$
19: **if** $f_\alpha(S_T) < 1$ **then**
20: $\quad$ **Return** $S_T^{'}$
21: **if** $|S_T| < b$ **then**
22: $\quad$ **for** $i \leftarrow 1$ **to** b **do**
23: $\quad\quad$ **if** i-th item of S_T^* **Not** in S_T **then**
24: $\quad\quad\quad v^* \leftarrow i$-th item of S_T^*
25: $\quad\quad\quad S_T \leftarrow S_T \cup \{v^*\}$
26: $\quad\quad$ **if** $|S_T| = b$ **then**
27: $\quad\quad\quad$ **Break**
28: **Return** S_T

Weighted RR-Set Generation for Grouped Nodes. To tackle the limitations of existing methods, we designed a novel RR-Set generation scheme that is effective and efficient for our *Fair-TLRM* problem. The process begins with a forward propagation simulation starting from the rumor source S_F, identifying the nodes influenced by the rumors (denoted as Inf_F), and recording the time t_F when the rumors reach each node. The start node is then randomly selected from Inf_F, along with the time it reached by F. Next, reverse Dijkstra is applied to trace back and find nodes capable of "saving" the rumor-affected node, adding them to the RR-set. We accelerate this process using the SUBSIM technique [51]. Finally, the loss reduction associated with this RR-set, along with the group to which it belongs, is then packaged into a tuple and stored. For a given user set U and a random RR set R, we define a random variable $Y(U, R)$ such that $Y(U, R) = w_i$ if $U \cap R_i \neq \emptyset$, and $Y(U, R) = 0$ otherwise, where w_i is the loss reduction brought by R_i. The expected influence $\mathbb{R}(S_T)$ under the TCIC model is equal to $INF \cdot \mathbb{E}[Y(U, R)]$, where INF is the expected influence of S_F.

FWS-RM. Building on Fair-Greedy and the group-aware

Weighted RR-Set generation way, we propose the FWS-RM algorithm (Fairness-aware Weighted-RIS-Based Solution to Rumor Mitigation). Like Fair-Greedy, FWS-RM first ensures fairness by selecting the minimum number of nodes necessary to meet the fairness constraint. It then iteratively picks nodes that maximize marginal loss reduction (i.e., mitigation reward). As mentioned earlier, computing the exact value of $\mathbb{R}(S_T \cup v) - \mathbb{R}(S_T)$ is computationally intractable in polynomial time, so we leverage the RIS technique to approximate this value. By generating a sufficient number of RR-Sets, we can guarantee a near-optimal expected mitigation reward is obtained from S_T. Besides, as demonstrated by previous work [11], [29], with a properly designed threshold N_{max}, the seed set S_T is a $(1-1/e-\epsilon)$-approximate solution with at least $1-\delta$ probability if the number of RR-Sets exceeds that threshold, where the N_{max} in our work is defined as Eq. (12) with $\Delta = \frac{8}{9}\delta$.

$$N_{max} = \frac{8n(3+\epsilon^{'})(1-1/e)[ln\frac{27}{4\Delta} + ln(\frac{n}{k})]}{3LB[\epsilon(1+\epsilon^{'}) - 2\epsilon^{'}(1-1/e)]^2} \tag{12}$$

Meanwhile, different from the lower bound (LB) in IM works that can be easily replaced by k, in RM problem, a node v must first be reached by S_F for S_T to gain any loss reduction. Therefore, to estimate the LB, we calculate the probability that a randomly selected node is influenced by S_F. This probability vector is computed via BFS, and the lower bound is given by $LB \geq r \cdot \sum_{v \in MAS} pro(v)$, where $pro(v)$ represents the probability that v is reached by S_F, and r is the smallest node weight in MAS.

Algorithm 2 provides the pseudo-code for FWS-RM. The algorithm operates within an OPIM framework, utilizing two sets of RR-Sets: $\mathcal{R}_1$ to generate the truth seed set S_T, and $\mathcal{R}_2$ to evaluate its effectiveness. Both the lower and upper bounds on $\mathbb{R}(S_T)$ are derived from these sets. Formally,

$$\mathbb{R}^u(S^o) = (\sqrt{\frac{\Lambda_1(S^*)}{1-1/e} + ln(1/\delta_1)} + \sqrt{ln(1/\delta_1)})^2 \frac{n}{\theta_1(1-\epsilon^{'})} \tag{13}$$

$$\mathbb{R}^l(S^*) = ((\sqrt{\Lambda_2(S^*) + \frac{25ln(1/\delta_2)}{36}} - \sqrt{ln(1/\delta_2)})^2 - \frac{ln(1/\delta_2)}{36}) \frac{n}{\theta_2(1+\epsilon^{'})} \tag{14}$$

Specifically, $\Lambda_1(S^*)$ (resp. $\Lambda_2(S^*)$) is the mitigation reward covered by S^* in $\mathcal{R}_1$ (resp. $\mathcal{R}_2$).

Theorem 3. (Theoretical guarantee of FWS-RM) *Algorithm 2 returns a $(1-1/e-\epsilon)$-approximation solution for our bounding functions with at least $1-\delta$ probability.*

Proof Sketch: The loop of FWS-RM runs at most i_{max} times, by setting $\delta_1 = \delta_2 = \Delta/(3i_{max})$, the approximation (i.e., z) is incorrect with at most $2\Delta/3$ probability with the union bound of the first $i_{max}-1$ rounds. When $i = i_{max}$, by setting N_{max} to Eq. (12) and $\epsilon^{'} = \frac{\epsilon}{2}$, through Chernoff Inequalities [52], the final error of S_T^* can be bounded by $(1-1/e-\epsilon)$ with at least $1-\Delta/3$ probability. Combining the failure probability lies in $INF_F^{'}$, the probability that FWS-RM returns an incorrect solution is $2\Delta/3 + \Delta/3 + \delta^{'} = 8/9 \cdot \delta + \delta/9 = \delta$. Meanwhile, as demonstrated in [53], when solving the dual-objective optimization IM-problem, a hybrid greedy framework that first uses local search to find nodes that satisfy the constraints, then uses a greedy approach to maximize the submodular objective function with the remaining budget, the final result is a $(1-1/e)$-approximation of the optimal solution under the situation that certain constraint is satisfied. Putting them together, S_T is a $(1-1/e-\epsilon)$-approximation solution of the optimal solution S_T^o, the proof completes. □

Time Complexity of FWS-RM. Since the main time cost of the FWS-RM comes from the generation of RR-Sets, the time complexity of FWS-RM is $O(\sum_{R \in \mathcal{R}_1 \cup \mathcal{R}_2}(INF_F + \frac{m}{n} \cdot (INF_1)^2)$, where INF_F is the expected influence spread of S_F and INF_1 is the largest reverse expected influence spread of the node r in G.

Sandwich Algorithm. Drawing upon the discussed findings, we incorporate the sandwich method in Algorithm 3. This yields the optimal seed set $S_T^* \in \{S_T^L, S_T^U\}$ that results in the maximum objective value with respect to our original goal.

Algorithm 3 Sandwich Algorithm

Input: G, S_F, b, $\underline{\mathbb{R}}$, $\overline{\mathbb{R}}$
Output: S_T^*

1: $S_T^U \leftarrow$ FWS-RM($G, S_F, b, \overline{\mathbb{R}}$)
2: $S_T^L \leftarrow$ FWS-RM($G, S_F, b, \underline{\mathbb{R}}$)
3: $S_T^* = \arg\max_{S_T \in \{S_T^U, S_T^L\}} \mathbb{R}(S_T)$
4: **Return** S_T^*

VI. TRADE-OFF: FAIRNESS AND REWARD

In this section, we present several insightful case studies as a complementary contribution. Our algorithm aids online social network owners in maximizing loss reduction through truth campaigns while ensuring fairness. However, our experimental explorations reveal that the performance of FWS-RM varies significantly across different networks, particularly when networks are partitioned into distinct groups based on attributes such as age or gender. For instance, in niche groups with weak internal connections, applying FWS-RM may not yield the desired outputs. To illustrate this, we applied algorithm 2 on a real-world social network *LastFM*[10], a social network established via public API in March 2020. This dataset categorizes users into 18 groups primarily based on location information. We randomly selected *20* nodes as fake seeds for our experiment, setting the fairness parameter α to *0.9*. Subsequently, we launched a truth campaign of size *20*.

Fig. 6. The structure of relations in group 4

In group 4 of the *LastFM* dataset, nodes exhibit very limited internal connections, as illustrated in Figure 6. While there are

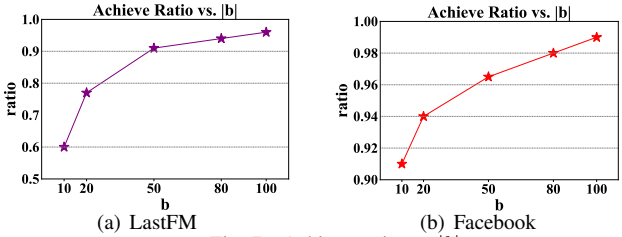
[10]https://snap.stanford.edu/data/feather-lastfm-social.html

some external connections from group 4 nodes to other groups, a seed from group 4 struggles to impact users within its own group directly. Conversely, seeds from other groups can only reach a small portion of users in group 4, rendering fairness considerations impractical in this scenario. When comparing the performance of FWS-RM with the fairness-agnostic algorithm WSRM, we found that the loss reduction achieved by FWS-RM was only **60.17%** of the mitigation reward obtained by WSRM. This observation prompts us to identify several factors that hinder the effectiveness of the fairness constraint in practical applications, thereby assisting algorithm users in making informed decisions about its implementation.

A. Empirical Studies and Suggestions

We conducted several case studies across different real-world social networks and presented the results about two identified key factors: **(i)** the impact of the truth set budget, **(ii)** the effect of the relative propagation speeds of rumor and truth on the algorithm's mitigation reward. These help characterize the better-performing social networks for more effectively mitigating losses caused by rumors while maintaining fairness constraints. Although our tests were conducted on all our datasets (§ VII), we only present the results on two representative ones, Facebook (connections are dense) and LastFM (featuring dispersed connections).

Impact of Budget Size. Figure 7 shows that as b increases, the achieved ratio (the ratio of the mitigation reward obtained by applying FWS-RM to that achieved by WSRM) improves across both datasets, simply because a larger budget gives the algorithm more flexibility to balance the trade-off between loss reduction and fairness. Additionally, nodes contributing to fairness can create the "combination effect" (see Section § IV), enhancing overall performance when b is larger. Notably, in the *LastFM* dataset, the achieved ratio remains low when b is small—a scenario common in practice due to financial constraints limiting the size of the seed set. However, the *Facebook* dataset does not observe this effect. This difference likely stems from the structural characteristics of the networks. In practice, for datasets with sparse connections like *LastFM*, *allocating less than 0.5% of the total users as seed nodes can lead to a significant negative impact on loss reduction when fairness is factored in*.



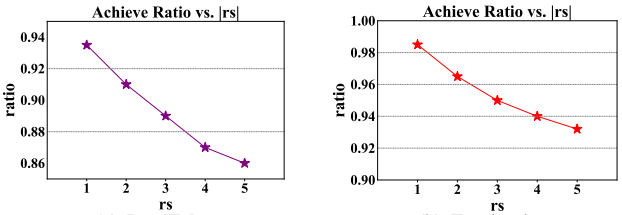
(a) LastFM (b) Facebook

Fig. 7. Achieve ratio vs. $|b|$

Impact of the Relative Propagation Speed of Rumor and Truth. In practical applications, slowing down the spread of rumors, such as by muting certain users, can alter the relative spread speed of rumor and truth. As shown in Figure 8, where the rs represents the relative speed ratio (i.e., $\mathbb{E}[d_e^T/d_e^F]$), our empirical study reveals that as the relative speed of rumors decreases, the achieve ratio increases. This suggests that when rumors spread more slowly, truth campaigns can intervene more effectively, allowing for a better balance between maximizing loss reduction and maintaining fairness. *Thus, identifying the rumor source quickly and imposing the in-time interventions can improve the performance of fairness-aware algorithms in real-world applications.*

TABLE III
HARMONIC CENTRALITY OF THE GROUP THAT HAS THE GREATEST IMPACT ON LOSS REDUCTION IN LASTFM AND FACEBOOK.

Metric \ Datasets	LastFM	Facebook
HCW	0	0.145
HCIE	0.156	0.166



(a) LastFM (b) Facebook

Fig. 8. Achieve ratio vs. $|rs|$

Characteristics of Better-performing Social Networks. We analyzed node characteristics in both the *LastFM* and *Facebook* datasets. Specifically, we calculated various metrics, including harmonic centrality [54] (which reflects connectivity between nodes), out-degree, in-degree, node weight, and rumor reach probability (the likelihood that a path exists from S_F to v). The results represent the average values within each group. Specifically, for the other four metrics, there is no obvious difference between the group with a large impact on loss reduction and the group with a small impact. Our observations reveal that harmonic centrality has the greatest influence on fairness, groups with higher harmonic centrality tend to exhibit better fairness performance. This indicates that *tightly connected nodes within a group can more easily propagate truth, allowing fairness to be achieved with minimal sacrifice in reward*. Table III presents the harmonic centrality of the group that has the greatest impact on the mitigation reward in both *LastFM* and *Facebook*. HCW represents the harmonic centrality of the nodes within the group while HCIE means the harmonic centrality between the internal nodes and external nodes. In *LastFM*, the worst-performing group had no internal connections and very few external links, which significantly lowered the mitigation reward when fairness constraint was posed, especially when b was small. In contrast, *Facebook* did not exhibit this issue.

Joint-Greedy Framework. Building on the insights from the experiments, we introduce a *Joint-Greedy Framework* tailored for datasets like *LastFM*. This framework uses a reward-fairness balance factor, p, to determine whether to apply FWS-RM or WSRM. In the Joint-Greedy framework, a random number r is uniformly sampled from the interval $[0, 1]$. If $p \geq r$, the framework selects FWS-RM, detailed in Algorithm 2. Otherwise, it opts for WSRM, a simplified

Algorithm 4 JOINT-Greedy

Input: G, S_F, b, α, p, ϵ, δ
Output: S_T

$r \leftarrow rand(0, 1)$
if $p \geq r$ **then**
  $S_T \leftarrow$ FWS-RM$(G, S_F, b, \alpha, \epsilon, \delta)$
else
  $S_T \leftarrow$ WSRM$(G, S_F, b, \epsilon, \delta)$
Return S_T

version of FWS-RM that focuses solely on maximizing the platform's economic loss reduction and does not consider fairness constraint. This approach balances the trade-off between mitigation reward and fairness by adapting based on the value of p. It is straightforward to show that the Joint-Greedy framework, with a probability distribution over policies, can improve the expected value of the reward. The full details of the Joint-Greedy framework are provided in Algorithm 4.

VII. EXPERIMENTS

We empirically evaluate our proposed algorithms in this section. All methods are implemented in C++ and executed on a Linux server with 2.20 GHz CPU and 128GB of RAM.

TABLE IV
DATASET STATISTICS

Dataset	n	m	d_{avg}	$\lvert\mathcal{C}\rvert$	Type
EmailCore	1,005	25,571	25.4	42	Directed
Facebook	1,216	42,443	69.8	4	Undirected
Digg	68,634	1,242,541	18.1	6	Directed
Twitch	168,114	6,797,557	5.3	21	Undirected
Weibo	1,787,443	83,135,996	46.5	2	Directed

A. Experimental Settings

Datasets. We evaluate our algorithms with five real-world datasets, namely, *EmailCore* [20], *Twitch* [20], *Facebook* [55], *Digg* [56], and *Weibo* [57]. Detailed in Table IV. We use Weibo to test the scalability of our algorithm in § VII-C. Following [29], [31], for each dataset, we use the *weighted-cascade model* [28] to determine the edge-weight, i.e., $p_{u,v} = 1/|N_{in}(v)|$, where $|N_{in}(v)|$ is the number of v's in-neighbors.

Baselines. We compare a set of algorithms in experiments.

- **WSRM**: a simplified version of FWS-RM without considering fairness and solely maximizing the loss reduction.
- **Fair-Greedy**: the baseline greedy method we proposed, considering both individual losses and fairness constraints.
- **NAMM** [11]: algorithm specially designed for RM problem under TCIC model, without considering both issues.
- **MOIM** [22]: a dual-objective optimization algorithm under target IM, its constraints can be easily transformed into the fairness constraints in our problem. We modified this algorithm to make it suitable for our problem and take fairness factors into consideration.
- **Weighted Out-neighbors (Weighted-ON)**: a heuristic method selects top-k nodes with the largest individual losses from S_F's out-neighbors as the truth seed set.

Default Parameters. We evaluate the performance of algorithms on different parameters, including the budget b, the fairness threshold α, and the sampling error factor ϵ. We set the default $b = 50$ for *EmailCore* and *Facebook*, $b = 100$ for *Digg*, *Twitch* and *Weibo*, and $\alpha = 0.8$ for all datasets. Following [58], [59], we set $\epsilon = 0.2$ for small datasets like *EmailCore* and *Facebook*,and $\epsilon = 0.3$ for *Digg*, *Twitch* and *Weibo* as the default values. For MOIM, we use 10K Monte Carlo simulations to determine the target users. We terminate an algorithm if it cannot finish in 48 hours and use OOT (i.e., out of time) to represent the result.

Evaluation Metrics. We employ the losses reduction, running time and fairness score (FS) as evaluation metrics in experiments. Specifically, the losses reduction is the decrease in loss after we launch the truth set. The fairness score is used to measure the fairness of the algorithms, as computed in the following equation:

$$FS = \min_{C_i \in \mathcal{C}} \frac{\mathbb{L}_i(G, C_i, S_F, \emptyset) - \mathbb{L}_i(G, C_i, S_F, S_T)}{\mathbb{L}_i(G, C_i, S_F, \emptyset) \cdot opt}$$

Following [11], we use 20K Monte Carlo to simulate the loss reduction of final solution. We run each method three times and report the average results.

B. Evaluation of FWS-RM Algorithm

In this set of experiments, we evaluate the performance of FWS-RM algorithm. Due to space limitations, we only show the results that use network topology to calculate the individual losses.

Varying the Budget b. We study the effect of budget b on the loss reduction, running time, and fairness score. We vary b from 10 to 100 for small datasets, and 20 to 200 for large datasets. As shown in Figure 9-11, when b grows, the loss reductions across six algorithms increase, the reason is that as b increases, the number of spreaders disseminating the truth grows, leading to a corresponding increase in loss reduction. Moreover, the FWS-RM, WSRM and Fair-Greedy significantly outperform MIOM, NAMM and Weighted-ON in terms of loss reduction, and MIOM usually gains less mitigation reward and spends more time comparing with NAMM since it puts more consideration in fairness constraint and finding target users. It is observed that the loss reduction yielded from FWS-RM and WSRM exhibit negligible differences in the Digg, Weibo and Facebook datasets, with the former achieving a staggering 95% percentage of the latter's reduction value. In the remaining two datasets, when b is small, WSRM performs relatively better compared with FWS-RM. The reason can be found in § VI. It is worth mentioning that FWS-RM significantly improves fairness, and a runtime analysis reveals that satisfying fairness does not incur a significant time overhead, typically only about 1% of the total runtime.

Varying the Error Factor ϵ. We study the impact of error factor on loss reduction, running time and fairness score by varying ϵ from 0.1 to 0.3. Since only FWS-RM, WSRM, NAMM and MOIM provide a theoretical guarantee, we show the result of these four algorithms. In Figure 12-14, it can be observed that as ϵ increases, the reduction loss decreases slightly, this is because when ϵ is small, the number of RR-Sets is of great largeness, the difference between the upper

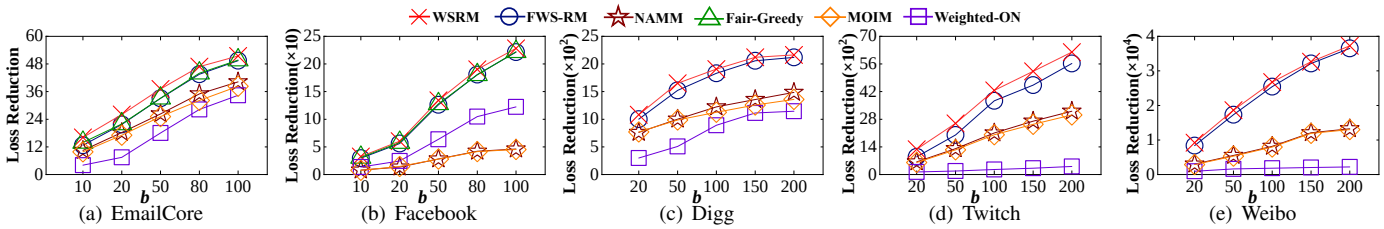


Fig. 9. Reduction Losses vs. b

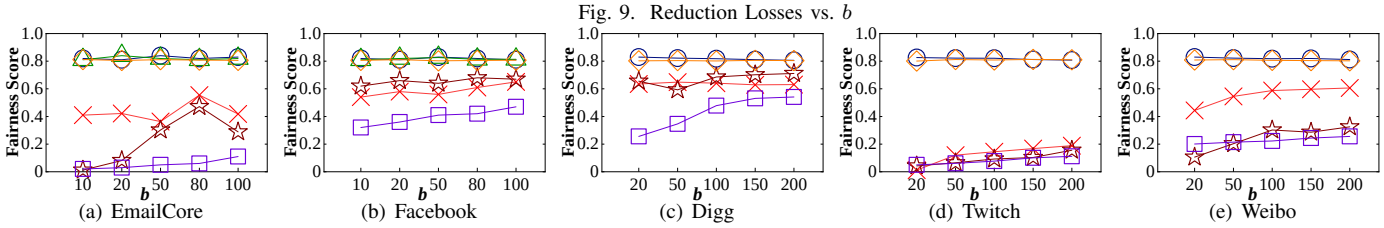


Fig. 10. Fairness Score vs. b

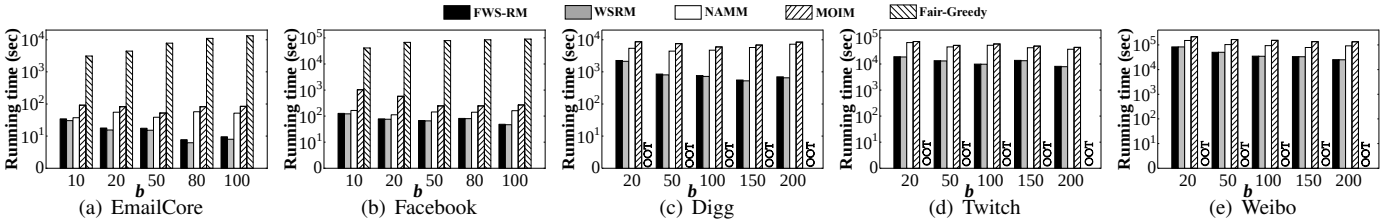


Fig. 11. Running time vs. b

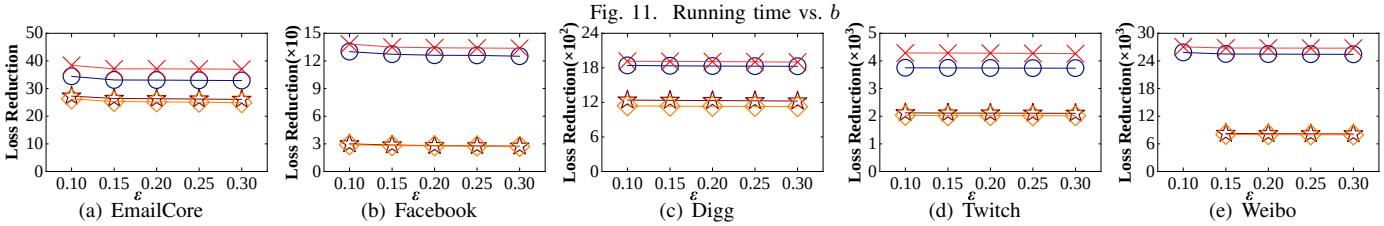


Fig. 12. Reduction Losses vs. ϵ

and lower bounds is smaller, that is, the bounds are tighter and the results are closer to actual reward that with more loss reduction, thus the calculated value is larger. Meanwhile, as ϵ increases, the running times of the three algorithms decreases and their respective fairness scores increase, the reason is that as ϵ increases, the error of the algorithm increases, and the denominator decreases accordingly when calculating FS, causing FS to rise slightly. Another observation is that TLRM problems often require shorter running times than RM problems. Although they both use the OPIM framework, TLRM tends to meet the accuracy requirements in fewer rounds, which allows it to generate fewer RR-Sets and reduce the time cost, that is why FWS-RM and WSRM usually run faster than the other two algorithms.

Varying Threshold α. We evaluate the impact of α (i.e., the threshold we use to satisfy the fairness constraint) on loss reduction and running time. Since only FWS-RM uses the fairness parameter α, we compare the loss reduction (i.e., effectiveness) and running time (i.e., efficiency) of FWS-RM by varying α from 0.1 to 0.9 and the result is displayed in Figure 15. We can observe that as α increases, loss reduction decreases within the range of 5%-9%, which suggests that FWS-RM enhance the fairness of the algorithm at the cost of a slight reward reduction. Moreover, as α increases, the running time rises slightly. This is due to the additional time required by the algorithm to meet the stricter fairness constraints.

C. Scalability Analysis

We test the scalability of FWS-RM with the increase of graph size. The *Weibo* dataset has about 80 million edges, we select 10M, 20M, 30M, 40M, 50M, and 60M edges uniformly at random to generate 6 graphs, and apply our algorithm on them. Figure 16(a) and Figure 16(b) demonstrate that the running time increase linearly with the number of edges, which confirms good scalability of our algorithm.

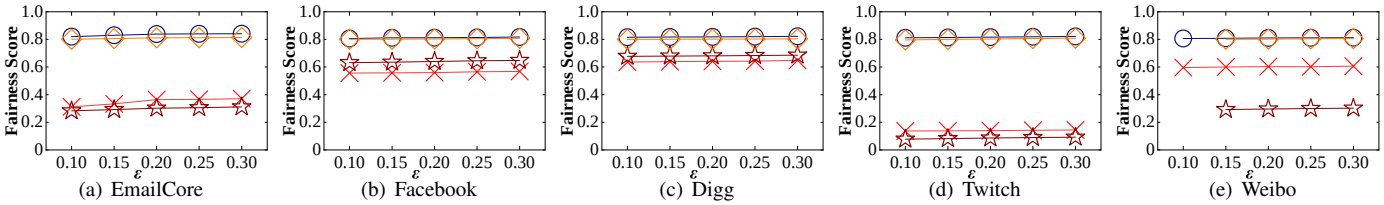
D. Additional Tests

TABLE V
THE OVERLAPS OF S_T BETWEEN TLRM AND RM

Node-weight set way	Topology			Historical			Random		
Datasets / $\|b\|$	Digg	Twitch	Weibo	Digg	Twitch	Weibo	Digg	Twitch	Weibo
50	20	16	22	19	20	22	18	12	21
100	44	38	43	40	36	41	38	37	42
200	56	49	63	50	47	53	48	43	54
300	100	89	111	105	98	109	87	82	91

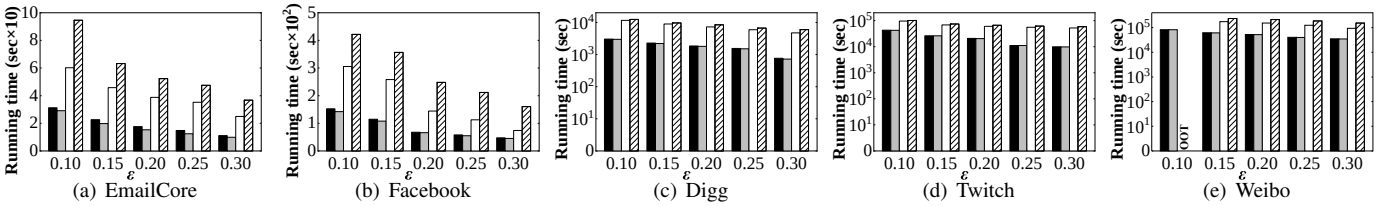
In this part, we calculate the number of overlaps in the truth seed sets selected by TLRM and RM (which only minimizes the number of users affected by the rumor) to highlight the differences between the two problems.

In the experiment, we set the number of rumor seed nodes as **50**. In order to mitigate the effects of stochasticity, we conduct three separate tests in all node-weight set methods. Notably, for random node-weight set function, the individual



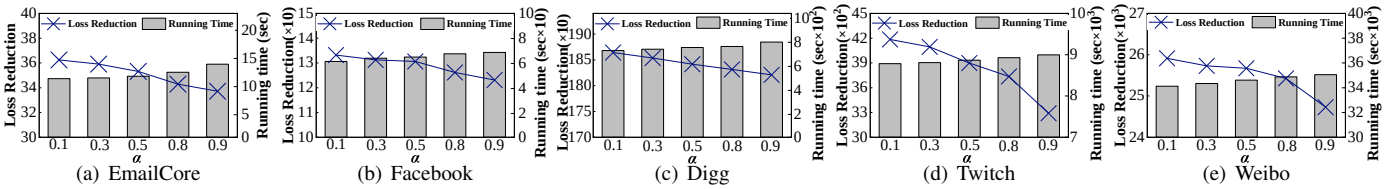
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 13. Fairness Score vs. ϵ



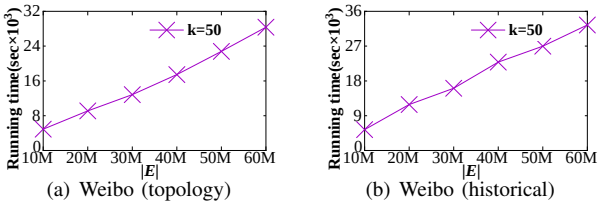
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 14. Running time vs. ϵ



(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 15. The impact of α on loss reduction and running time



(a) Weibo (topology) (b) Weibo (historical)

Fig. 16. Running time vs. Edges Number

losses of each user are randomized and selected from the set $\{0.1, 0.5, 0.8\}$. The average value derived from these tests served as the final experimental outcome. The table delineates the extent of overlap between truth sets chosen by TLRM and RM. Note that once the budget surpasses 300, the increase in loss reduction of the truth set becomes negligible and the truth set parallels the one where the budget is 300. Hence, results beyond this point have not been included for consideration.

Table V shows insights garnered from the experimental findings. It can be seen that, when varying budgets and individual loss set methods, the overlap ratio between the truth sets selected by TLRM and RM among three datasets ranges from 30% to 40% and averages around 1/3, suggesting that the overlap rate of the truth sets consistently maintains at a low frequency. This emphasizes that our problem is significantly different from existing RM problems.

VIII. Conclusion

In this work, we address a novel problem, Fair-TLRM, which aims to maximize the expected economic loss reduction for OSNs while jointly considering fairness. We develop a dual-objective optimization algorithm, Fair-Greedy, and introduce a scalable version, FWS-RM, with a data-dependent approximation guarantee by employing a novel group-aware weighted RIS and sandwich technique. Additionally, we examine the impact of fairness on algorithm behavior through several case studies, drawing key insights and proposing a new framework, Joint-Greedy, to enhance algorithm performance. Experiments on five real-world datasets demonstrate the effectiveness and efficiency of our proposed approach.

REFERENCES

[1] M. Ye, X. Liu, and W.-C. Lee, "Exploring social influence for recommendation: a generative model approach," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 671–680.

[2] S. Lee, K. Levanti, and H. S. Kim, "Network monitoring: Present and future," *Computer Networks*, vol. 65, pp. 84–98, 2014.

[3] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 57–66.

[4] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[5] S. Velichety and U. Shrivastava, "Quantifying the impacts of online fake news on the equity value of social media platforms–evidence from twitter," *International Journal of Information Management*, vol. 64, p. 102474, 2022.

[6] L. Sun, X. Rui, and W. Chen, "Scalable adversarial attack algorithms on influence maximization," in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 2023, pp. 760–768.

[7] J. Xie, F. Zhang, K. Wang, X. Lin, and W. Zhang, "Minimizing the influence of misinformation via vertex blocking," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2023, pp. 789–801.

[8] H. T. Nguyen, A. Cano, T. Vu, and T. N. Dinh, "Blocking self-avoiding walks stops cyber-epidemics: a scalable gpu-based approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 7, pp. 1263–1275, 2019.

[9] M. A. Manouchehri, M. S. Helfroush, and H. Danyali, "Temporal rumor blocking in online social networks: A sampling-based approach," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 7, pp. 4578–4588, 2021.

[10] C. Song, W. Hsu, and M. L. Lee, "Temporal influence blocking: Minimizing the effect of misinformation in social networks," in *2017 IEEE 33rd international conference on data engineering (ICDE)*, 2017, pp. 847–858.

[11] M. Simpson, F. Hashemi, and L. V. Lakshmanan, "Misinformation mitigation under differential propagation rates and temporal penalties," *Proceedings of the VLDB Endowment*, vol. 15, no. 10, pp. 2216–2229, 2022.

[12] S. Wen, J. Jiang, Y. Xiang, S. Yu, W. Zhou, and W. Jia, "To shut them up or to clarify: Restraining the spread of rumors in online social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3306–3316, 2014.

[13] H. A. Voorveld, G. Van Noort, D. G. Muntinga, and F. Bronner, "Engagement with social media and social media advertising: The differentiating role of platform type," *Journal of advertising*, vol. 47, no. 1, pp. 38–54, 2018.

[14] K. Simon, "Digital trends 2020: Every single stat you need to know about the internet," *TNW: Blog Podium: Post Marc*, vol. 4, p. 2021, 2021.

[15] X. Ding, X. Zhang, R. Fan, Q. Xu, K. Hunt, and J. Zhuang, "Rumor recognition behavior of social media users in emergencies," *Journal of Management Science and Engineering*, vol. 7, no. 1, pp. 36–47, 2022.

[16] S. T. Malamut, M. Dawes, and H. Xie, "Characteristics of rumors and rumor victims in early adolescence: Rumor content and social impact," *Social Development*, vol. 27, no. 3, pp. 601–618, 2018.

[17] Y. R. Velez, E. Porter, and T. J. Wood, "Latino-targeted misinformation and the power of factual corrections," *The Journal of Politics*, vol. 85, no. 2, pp. 789–794, 2023.

[18] "Testimony: "a growing threat: The impact of disinformation targeted at communities of color," 2022. [Online]. Available: https://mediaengagement.org/research/the-impact-of-disinformation-targeted-at-communities-of-color

[19] E. Lee, F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic, "Homophily and minority-group size explain perception biases in social networks," *Nature human behaviour*, vol. 3, no. 10, pp. 1078–1087, 2019.

[20] J. Leskovec and A. Krevl, "Snap datasets: Stanford large network dataset collection," 2014, http://snap.stanford.edu/data.

[21] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 665–674.

[22] S. Gershtein, T. Milo, and B. Youngmann, "Multi-objective influence maximization," in *Proceedings of the 24th International Conference on Extending Database Technology (EDBT)*, 2021, pp. 145–156.

[23] C. Song, W. Hsu, and M. L. Lee, "Targeted influence maximization in social networks," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 1683–1692.

[24] Y. Li, D. Zhang, and K.-L. Tan, "Real-time targeted influence maximization for online advertisements," 2015.

[25] X. Ke, A. Khan, and G. Cong, "Finding seeds and relevant tags jointly: For targeted influence maximization in social networks," in *Proceedings of the 2018 international conference on management of data*, 2018, pp. 1097–1111.

[26] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, "Maximizing social influence in nearly optimal time," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2014, pp. 946–957.

[27] W. Lu, W. Chen, and L. V. Lakshmanan, "From competition to complementarity: comparative influence diffusion and maximization," *Proceedings of the VLDB Endowment*, vol. 9, no. 2, pp. 60–71, 2015.

[28] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.

[29] J. Tang, X. Tang, X. Xiao, and J. Yuan, "Online processing algorithms for influence maximization," in *Proceedings of the 2018 International Conference on Management of Data*, 2018, pp. 991–1005.

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

[31] Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: A martingale approach," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015, pp. 1539–1554.

[32] T. Cai, J. Li, A. Mian, R.-H. Li, T. Sellis, and J. X. Yu, "Target-aware holistic influence maximization in spatial social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 4, pp. 1993–2007, 2020.

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

[34] Y. Feng, A. Patel, B. Cautis, and H. Vahabi, "Influence maximization with fairness at scale," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 4046–4055.

[35] M. Simpson, V. Srinivasan, and A. Thomo, "Reverse prevention sampling for misinformation mitigation in social networks," *arXiv preprint arXiv:1807.01162*, 2018.

[36] C. V. Pham, Q. V. Phu, and H. X. Hoang, "Targeted misinformation blocking on online social networks," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2018, pp. 107–116.

[37] G. Tong, W. Wu, L. Guo, D. Li, C. Liu, B. Liu, and D.-Z. Du, "An efficient randomized algorithm for rumor blocking in online social networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 845–854, 2017.

[38] L. Nie, X. Song, and T.-S. Chua, *Learning from Multiple Social Networks*. Morgan & Claypool Publishers, 2016.

[39] B. Wang, G. Chen, L. Fu, L. Song, and X. Wang, "Drimux: Dynamic rumor influence minimization with user experience in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2168–2181, 2017.

[40] C. Sun, H. Liu, M. Liu, Z. Ren, T. Gan, and L. Nie, "Lara: Attribute-to-feature adversarial learning for new-item recommendation," in *Proceedings of the 13th international conference on web search and data mining*, 2020, pp. 582–590.

[41] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Internet and Network Economics: Third International Workshop, WINE 2007, San Diego, CA, USA, December 12-14, 2007. Proceedings 3*. Springer, 2007, pp. 306–311.

[42] P. Cheng, X. Lian, L. Chen, and S. Liu, "Maximizing the utility in location-based mobile advertising," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 776–788, 2020.

[43] M. Gabielkov, A. Ramachandran, A. Chaintreau, and A. Legout, "Social clicks: What and who gets read on twitter?" in *Proceedings of the*

2016 ACM SIGMETRICS international conference on measurement and modeling of computer science, 2016, pp. 179–192.
[44] A. Tagarelli and R. Interdonato, "Lurking in social networks: topology-based analysis and ranking methods," *Social Network Analysis and Mining*, vol. 4, pp. 1–27, 2014.
[45] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.
[46] G. Farnad, B. Babaki, and M. Gendreau, "A unifying framework for fairness-aware influence maximization," in *Companion Proceedings of the Web Conference 2020*, 2020, pp. 714–722.
[47] J. Rawls, "Atheory of justice," *Cambridge (Mass.)*, 1971.
[48] S. Fujishige, *Submodular functions and optimization*. Elsevier, 2005.
[49] A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta, "Robust submodular observation selection." *Journal of Machine Learning Research*, vol. 9, no. 12, 2008.
[50] Y. Zhu, D. Li, and Z. Zhang, "Minimum cost seed set for competitive social influence," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.
[51] Q. Guo, S. Wang, Z. Wei, W. Lin, and J. Tang, "Influence maximization revisited: Efficient sampling with bound tightened," *ACM Transactions on Database Systems (TODS)*, 2022.
[52] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.
[53] F. Xiang, J. Wang, Y. Wu, X. Wang, C. Chen, and Y. Zhang, "Rumor blocking with pertinence set in large graphs," *World Wide Web*, vol. 27, no. 1, pp. 1–26, 2024.
[54] M. Marchiori and V. Latora, "Harmony in the small-world," *Physica A: Statistical Mechanics and its Applications*, vol. 285, no. 3-4, pp. 539–546, 2000.
[55] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 251–260.
[56] K. Lerman, R. Ghosh, and T. Surachawala, "Social contagion: An empirical study of information spread on digg and twitter follower graphs," *arXiv preprint arXiv:1202.3162*, 2012.
[57] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, "Social influence locality for modeling retweeting behaviors," in *Twenty-third international joint conference on artificial intelligence*. Citeseer, 2013.
[58] K. Han, B. Wu, J. Tang, S. Cui, C. Aslay, and L. V. Lakshmanan, "Efficient and effective algorithms for revenue maximization in social advertising," in *Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data*, 2021, pp. 671–684.
[59] C. Aslay, F. B. L. V. Lakshmanan, and W. Lu, "Revenue maximization in incentivized social advertising," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, 2017.

APPENDIX

This section shows the complete experiment results that are omitted in Section VII due to space constraints (experiment results that use random and historical-activity-based individual loss weight). Specifically, in the random weight set function, we randomly selected a value from the set $\{0.1, 0.5, 0.8\}$ as the user's loss weight for three times and averaged the results. It is worth noting that EmailCore and Facebook datasets do not contain any user historical activity information, therefore, for the historical weight set way, we only conducted experiments on the Digg, Twitch, and Weibo datasets and report their results.

Varying the Budget b. The correlation between budget and factors like loss reduction, running time, and fairness score are explored through the manipulation of b. As manifested in Figure 17, 20-22, 29-31, an escalating trend in loss reduction across the five algorithms is concurrent with the upswing of b. The algorithms of FWS-RM, WSRM and Fair-MC notably outperform the other two methods in terms of loss reduction. Upon scrutinizing the results, it becomes apparent that the loss reduction yielded from FWS-RM and WSRM exhibit negligible differences in the context of the Digg, Weibo, and Facebook datasets, with the former achieving a staggering 95% approximation of the WSRM rewards. In the remaining two datasets, when b is small, FWS-RM performs relatively ordinary comparing with WSRM. The underlying reason can be found in Section VI.

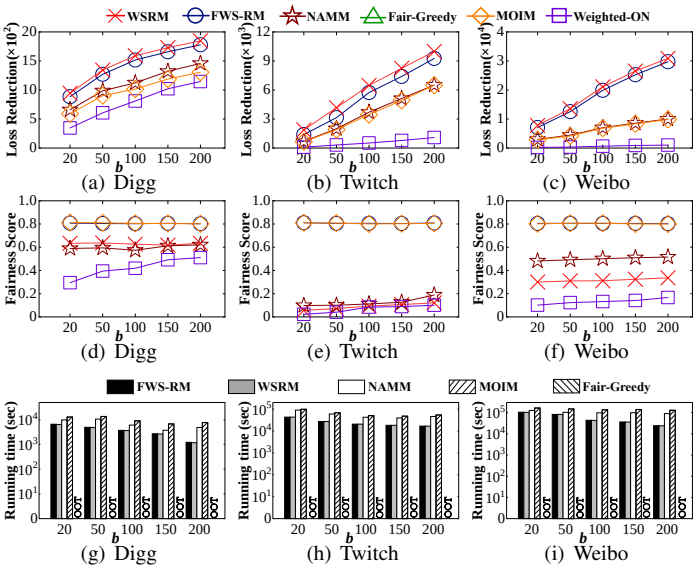


Fig. 17. Reduction Losses / FS / Running Time vs. b (historical)

Varying the error factor ϵ. We study the impact of error factor on loss reduction, running time and fairness score by varying ϵ from 0.1 to 0.3. Since only FWS-RM, WSRM and EIL provide a theoretical guarantee, we show the result of these three algorithms. In Figure 18, 23-25, 32-34, it can be observed that as ϵ increases, the reduction loss decreases slightly, this is because when ϵ is small, the number of RR-Sets is of great largeness, the difference between the upper and lower bounds is smaller (in another words, the bounds are tighter and the results are closer to actual reward that with more loss reduction), thus the calculated value is larger. Meanwhile, an upswing of ϵ is linked with not just diminished implementation durations for all three algorithms, but also a gradual elevation in their respective fairness scores. It is worth claiming that FWS-RM significantly improves fairness, and through running time analysis, it can be found that meeting fairness does not bring a lot of time overhead. Another point worth mentioning is that TLRM problems often require shorter running times than RM problems. Although they both use the OPIM framework, TLRM often meets the accuracy requirements in fewer rounds because it generates fewer RR-Sets and takes less time to run, this accounts for the circumstances that FWS-RM usually runs faster than EIL. Last but not least, as shown in Figure 18(g), the EIL takes a shorter time in Digg. This underlying reason is that in Digg with historical set function, the weights of most users are very close to 1, thus the problem can be regarded as a more complex RM problem, therefore FWS-RM takes more time.

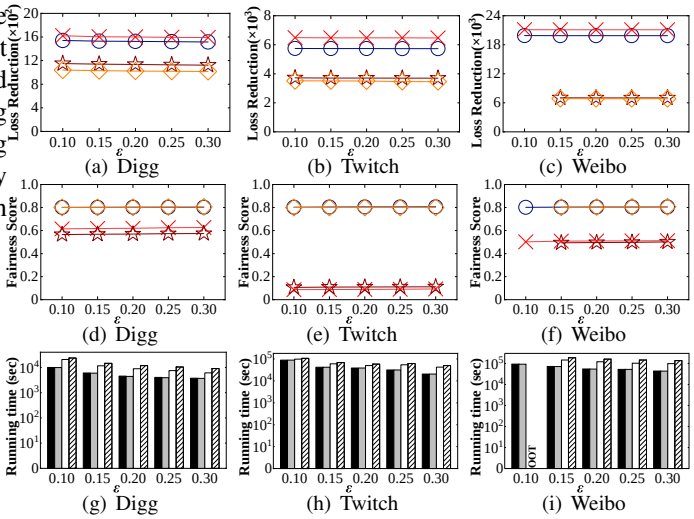
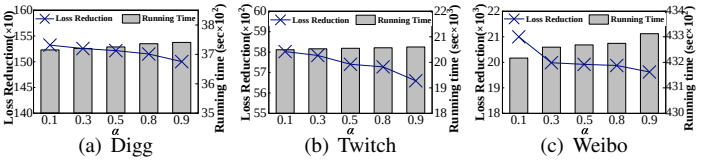


Fig. 18. Reduction Losses / FS / Running Time vs. ϵ (historical)

Varying the α. Since only FWS-RM uses the fairness parameter α, we compare the loss reduction (i.e., effectiveness) and running time (i.e., efficiency) of FWS-RM by varying α which is displayed in Figure 19, 26, 35. Observations reveal that as α increases, loss reduction decreases within the range of 5%-9%, this suggests that FWS-RM can enhance the fairness of the algorithm at the cost of a slight reward reduction. Concurrently, a minor increase in running time is seen, attributed to the additional time required by the algorithm to comply with stricter fairness constraints.

Lower Bound. Figure 27 shows the lower bounds of β, which is the data-dependent approximation guarantee (Lemma 3) achieved by FWS-RM. It is easy to find that among all datasets $\beta > 0.6$, proving the tightness of our bound. Besides, we also compared our lower bound with the bound mentioned in Lemma 2, as show in Figure 28. Our tightness improved by about 5% on average, proving that our new bound has a better approximation effect.



(a) Digg (b) Twitch (c) Weibo

Fig. 19. Reduction Losses / Running Time vs. α (historical)

A. DETAILED PROOF OF LEMMA 2

Lemma 4. *$\overline{r}_X(S_T \to v)$ under (10) is submodular.*

Proof. By applying truth-dominant policy $r^{'}$ and set all the MLs to 1 for campaign T will definitely increase the reward, therefore, it forms a upper bound. Since under (10), truth always dominates, thus the cases reduce to whether $a(\mathcal{P}_v^T, \mathcal{P}_v^F)$ equals to 0 or not.

Now fix a possible world X. For any node $v \in MAS_X$, when a node is added to S_T, v will either remain activated in F or switch to activating in T. To prove submodularity, it is sufficient to prove submodularity for each $v \in MAS_X$ due to the closed property of submodular functions [48]. Consider a truth set B and a node $w \in V \setminus (S_F \cup B)$:

Case(i): When $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 0$, the combination effect does not exist, therefore, $r_X^{'}(S_T \to v) = max_{u \in S_T} r_X^{'}(\{u\} \to v)$. If v is influenced by F under B, but influenced by T under $B \cup \{w\}$ w/o a tie-breaking, for v to satisfy Inequality 7, under B, $t_v^F < t_v^T$, and under $B \cup \{w\}$, $t_v^T < t_v^F$. Therefore, the shortest path from $B \cup \{w\}$ to v must begin with w; otherwise, it contradicts the condition. w contributes to the reward.

Case(ii): When $a(\mathcal{P}_v^T, \mathcal{P}_v^F) = 1$, by set all the MLs to 1 for campaign T, the node in S_T gain the same power like *combination effect* since the *combination effect* itself comes from the meeting delay of truth, this strength is further enhanced by truth-dominant policy. If a node v is influenced by F under B, but influenced by T under $B \cup \{w\}$ regardless of whether the truth-domination policy is used, for v to satisfy Inequality 7, it must be that under B, $t_v^F < t_v^T$, and under $B \cup \{w\}$, $t_v^T \leq t_v^F$. Therefore, the shortest path from $B \cup \{w\}$ to v must begin with w; otherwise, it contradicts the condition. w also contributes to the reward.

From the above two cases, we could conclude that $\overline{r}_X(B \cup \{w\} \to v) = \overline{r}_X(\{w\} \to v)$. Let $A \subseteq B$, similar to the situation of B, we can get $\overline{r}_X(A \cup \{w\} \to v) = \overline{r}_X(\{w\} \to v) = \overline{r}_X(B \cup \{w\} \to v)$. Besides, the monotonicity of $\overline{r}$ implies that $\overline{r}_X(A \to v) \leq \overline{r}_X(B \to v)$, therefore, $\overline{r}_X(A \cup \{w\} \to v) - \overline{r}_X(A \to v) \geq \overline{r}_X(B \cup \{w\} \to v) - \overline{r}_X(B \to v)$, the proof ends. □

B. DETAILED PROOFS OF THE UNBIASED ESTIMATE $\mathbb{R}(S_T)$

Conclusion. $\mathbb{R}(S_T) = INF \cdot \mathbb{E}[Y(S_T, R)]$

Proof. Let $\Pr_X(S_F \to v)$ represent the probability that v is affected by S_F in a possible world X, and $\Pr_X(S_T \to v = w(v))$ indicates the probability that S_T can propagate to v before S_F, or that v believes the truth through tie-breaking in a possible world X. When the root v of R_i is selected uniformly at random from MAS_X:

$$\begin{aligned}
\mathbb{R}(S_T) &= \sum_v [Pr_X(S_F \to v) \cdot Pr_X(S_T \to v = w(v))] \\
&= \sum_v [Pr_X(S_F \to v) \cdot Pr_X(S_T \cap R_i(v) = w(v))] \\
&= \sum_v Pr_X(S_F \to v) \cdot [Pr_X(S_T \cap R_i(v) = w(v))] \\
&= INF \cdot [Pr_X(S_T \cap R_i(v) = w(v))] \\
&= INF \cdot \mathbb{E}[Y(S_T, R)]
\end{aligned}$$

□

C. PROOFS ABOUT TIME COMPLEXITY

Conclusion. The time complexity of FWS-RM is $O(\sum_{R \in \mathcal{R}_1 \cup \mathcal{R}_2}(INF_F + \frac{m}{n} \cdot (INF_1)^2)$, where INF_F is the influence spread of S_F and INF_1 is the largest reverse expected influence spread of any size-one node set in G.

Proof. Since the main time cost of the FWS-RM comes from the generation of RR-Sets, we can get the algorithm time complexity of FWS-RM is $O(\sum_{R \in \mathcal{R}_1 \cup \mathcal{R}_2} EPT)$, where EPT is the expected complexity of generating a RR-Set. During the generation process of RR-Sets, we first perform BFS of the S_F, the time complexity of which is INF_F. Then we did a reverse Dijkstra, it has a running time of $n^* log n^* + m^*$, where n^* and m^* are the number of nodes and edges inspected by the randomly chosen node, solving the tie-breaking need a time cost of $n^* m^*$. Therefore, the EPT can be calculated as follows:

$$EPT = INF_F + \mathbb{E}[n^* log n^* + m^* + n^* m^*]$$

The size of n^* is bounded by INF_1 due to the fact that INF_1 is the largest reverse expected influence spread of the a node r in G, using the conclusion from [31], we can get $m^* \leq \frac{m}{n} INF_1$. Combining all the above conclusions, we can get

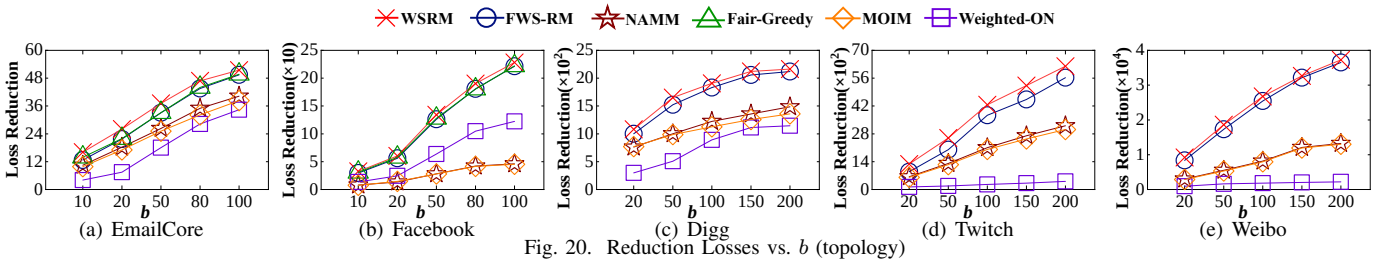
$$\begin{aligned}
EPT &= INF_F + \mathbb{E}[n^* \log n^* + m^* + n^* m^*] \\
&\leq INF_F + \mathbb{E}[2 \cdot n^* m^*] \\
&= O(INF_F + \frac{m}{n} \cdot (INF_1)^2)
\end{aligned}$$

Thus, the time complexity of FWS-RM is $O(\sum_{R \in \mathcal{R}_1 \cup \mathcal{R}_2}(INF_F + \frac{m}{n} \cdot (INF_1)^2)$, the proof ends. □

TABLE VI
METRIC OF THE GROUP THAT HAS THE GREATEST IMPACT ON LOSS REDUCTION IN LASTFM AND FACEBOOK.

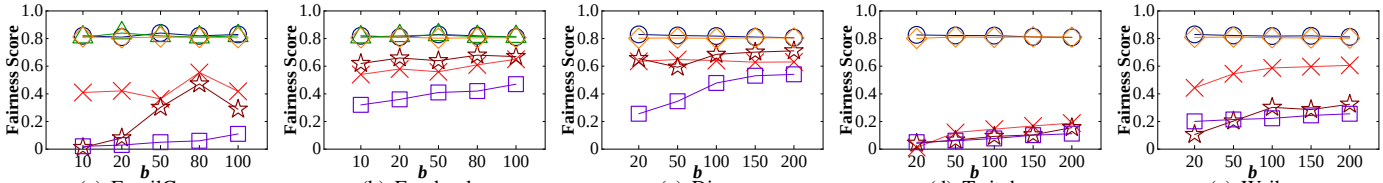
Metric \ Datasets	LastFM	Facebook
In-degree	6.9 (The second least)	8.8 (The least)
Out-degree	7.5 (The least)	9.2 (The second least)
Rumor reach probability	0.459	0.411

As show in VI, both the rank of In-degree, Out-degree and Rumor reach probability in the worst performance group are the last or the second last, and there are minor differences between them comparing with Harmonic Centrality, emphasizing the importance of Harmonic Centrality in good algorithm-performing.



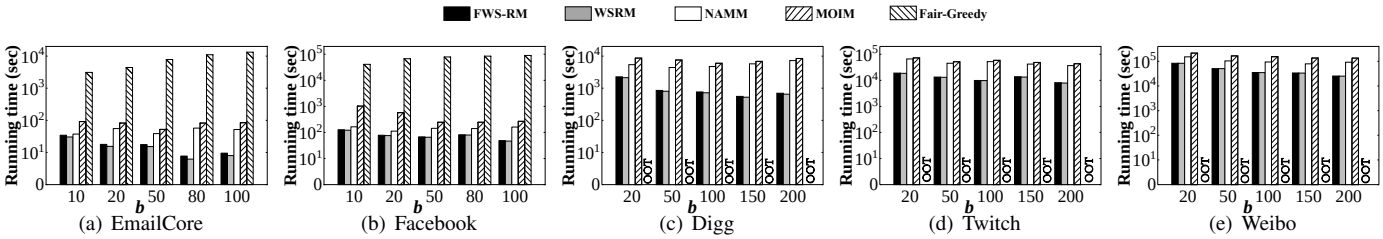
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 20. Reduction Losses vs. b (topology)



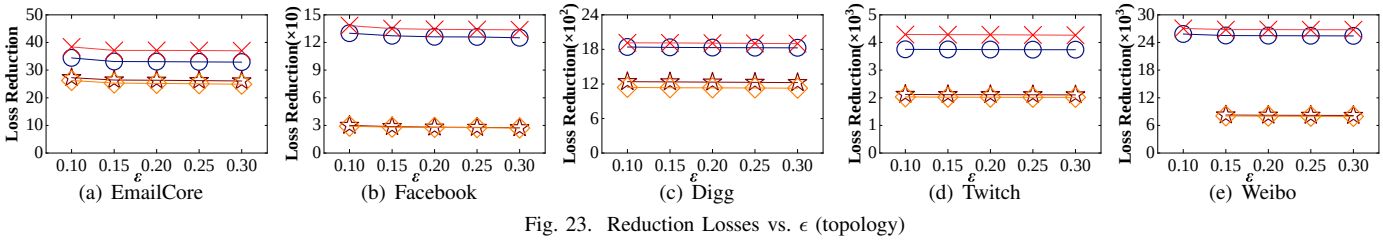
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 21. Fairness Score vs. b (topology)



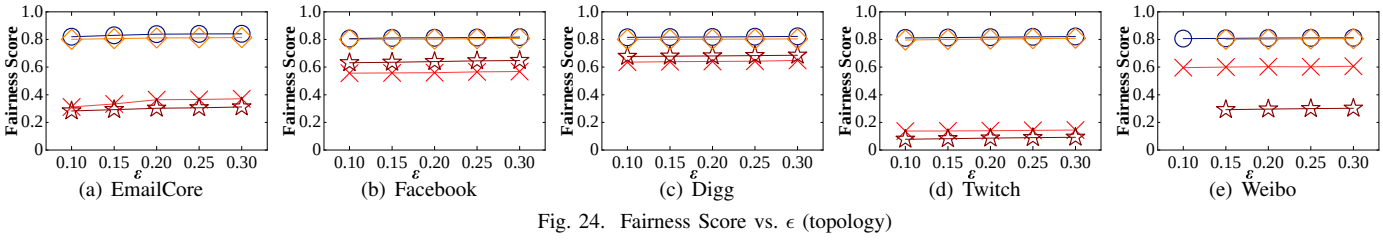
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 22. Running time vs. b (topology)



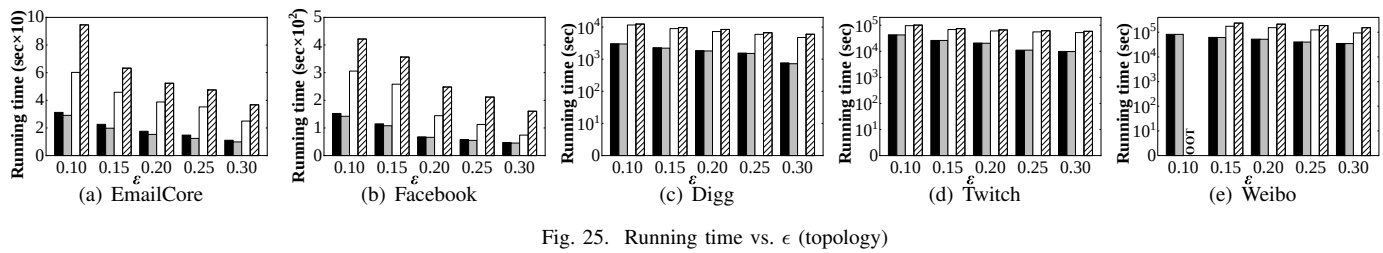
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 23. Reduction Losses vs. ϵ (topology)



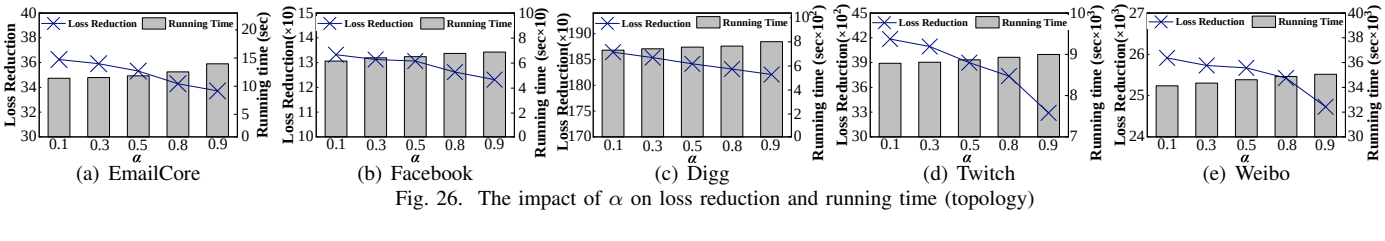
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 24. Fairness Score vs. ϵ (topology)



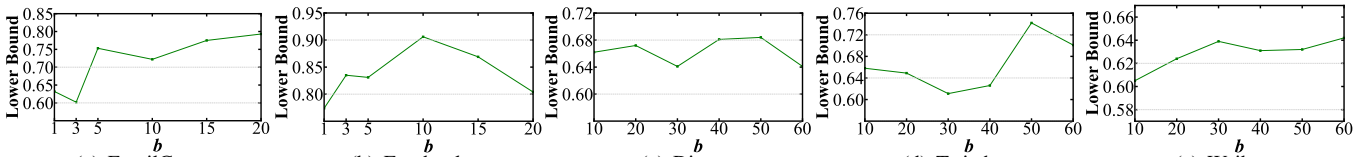
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 25. Running time vs. ϵ (topology)



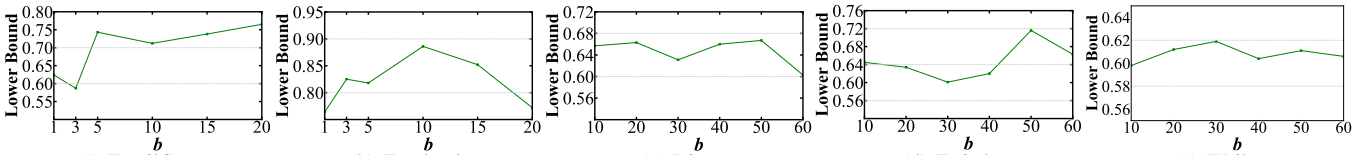
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 26. The impact of α on loss reduction and running time (topology)



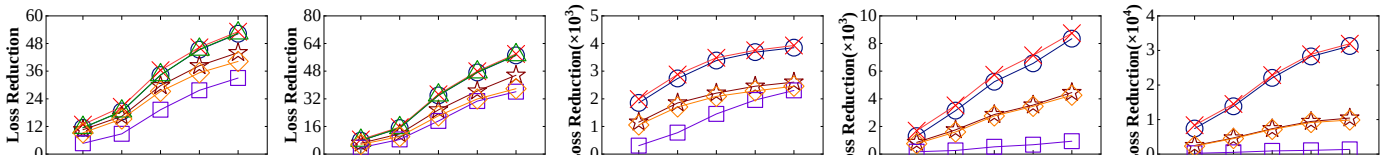
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 27. The lower bound of β (tightened)



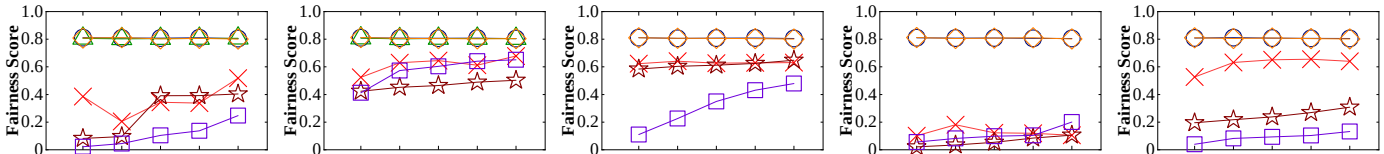
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 28. The lower bound of β (ordinary)



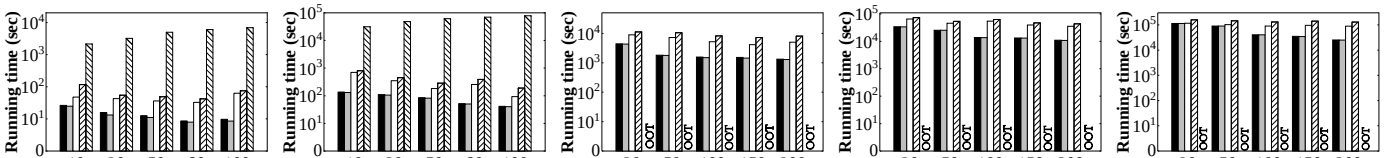
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 29. Reduction Losses vs. b (random)



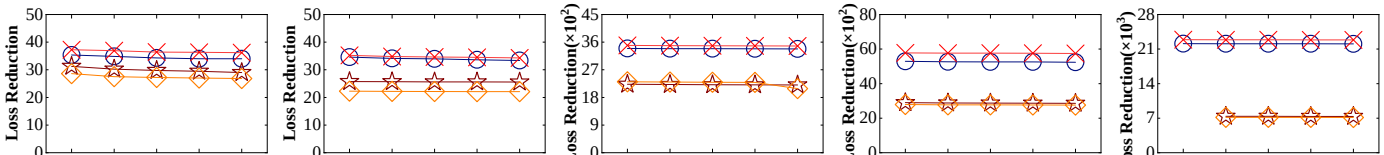
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 30. Fairness Score vs. b (random)



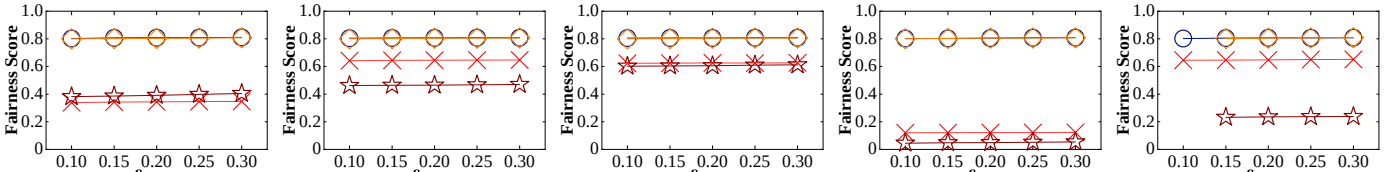
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 31. Running time vs. b (random)



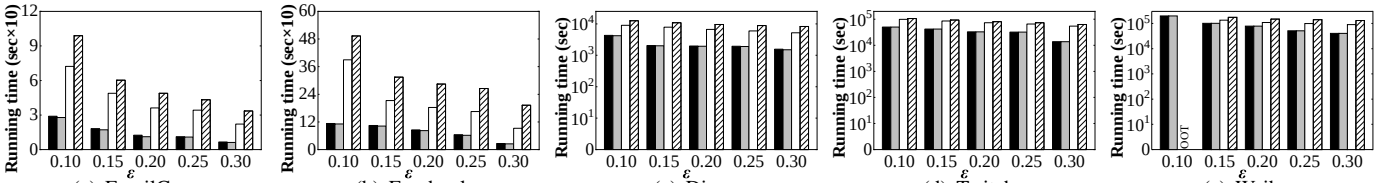
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 32. Reduction Losses vs. ϵ (random)



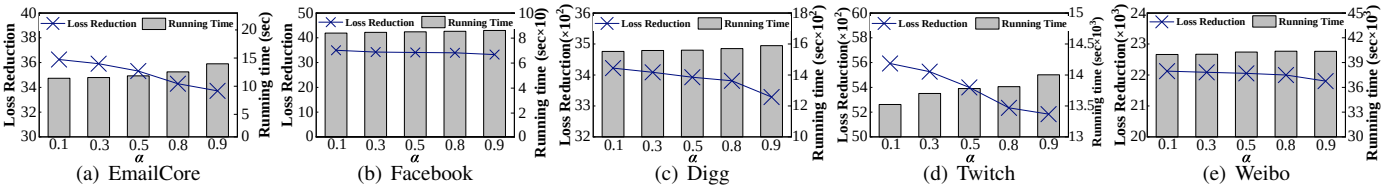
(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 33. Fairness Score vs. ϵ (random)



(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 34. Running time vs. ϵ (random)



(a) EmailCore (b) Facebook (c) Digg (d) Twitch (e) Weibo

Fig. 35. The impact of α on loss reduction and running time (random)