Applied Intelligence (2023) 53:27483–27504
https://doi.org/10.1007/s10489-023-04928-3

Fairness-aware fake news mitigation using counter information propagation

Akrati Saxena[1,2] · **Cristina Gutiérrez Bierbooms**[1] · **Mykola Pechenizkiy**[1]

Accepted: 29 July 2023 / Published online: 12 September 2023
© The Author(s) 2023

Abstract
Given the adverse impact of fake news propagation on Social media, fake news mitigation has been one of the main research directions. However, existing approaches neglect fairness towards each community while minimizing the adverse impact of fake news propagation. This results in the exclusion of some minor and underrepresented communities from the benefits of the intervention, which can have important societal repercussions. This research proposes a fairness-aware truth-campaigning method, called FWRRS (**F**airness-aware **W**eighted **R**eversible **R**eachable **S**ystem), which focuses on blocking the influence propagation of a competing entity, in this case, with the use case of fake news mitigation. The proposed method employs weighted reversible reachable trees and maximin fairness to achieve its goals. Experimental analysis shows that FWRRS outperforms fairness-oblivious and fairness-aware methods in terms of both total outreach and fairness. The results show that in the proposed approach, such fairness does not come at a cost in efficiency, and in fact, in most cases, it works as a catalyst for achieving better effectiveness in the future. In real-world networks, we observe up to ∼10% improvement in the saved nodes and ∼57% improvement in maximin fairness as compared to the second best-performing baseline, which varies for each network.

Keywords Fake news mitigation · Influence blocking · Algorithmic fairness

1 Introduction

Long gone are the days of people getting their news information from news channels or newspapers. Consumers continue to shift away from these traditional sources and are moving more towards social media for their news consumption [45]. Social media has attracted over 2.4 billion internet users by easing access to not only textual content but also multimedia and external websites. In a perfect world, everything reported would be entirely based on facts so that one could trust that the consumed media is reliable; however, this is not the case in the real world [10, 14, 21, 57]. Misinformation and rumors are becoming highly prevalent [2, 79], and such ease of access and fast dissemination of information on social media has unfortunately been exploited for spreading fake news [80].

A famous example of the impact of fake news on society is the US elections of 2016. Facebook provided a testimony that 126 million Americans were exposed to Russian-backed, politically-oriented fake news via Facebook during the presidential election campaign of 2016 [10, 46]. In the aftermath of these elections, as the extent of fake news could be observed, three scientists from the Ohio State University [61] - Gunther, Beck, and Nisbet- explored whether people might have changed their vote due to fake news. They interviewed a group of people who voted for Obama in 2012 and looked at their votes in the 2016 election. The voters were asked how much they believed in three statements that, according to an independent analysis, had been promoted by fake news while being false. These statements were: (i) Hilary was in poor health due to serious illness, (ii) Pope Francis endorsed Trump, and (iii) during Hilary's time as Secretary of State, she approved weapon sales to Islamic Jihadists like ISIS. Although the majority of people did not believe these

Akrati Saxena and Cristina Gutiérrez Bierbooms contributed equally to this work.

✉ Akrati Saxena
a.saxena@liacs.leidenuniv.nl

[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

[2] Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

assertions to be true, there was a very substantial correlation between their beliefs and voting behavior. Among those voters who did not believe in any of the three statements, 89% did cast their vote for Hilary in 2016; among those who believed in just one of the items 61% voted for her, but among those who believed in at least two of these fake news only 17% supported Hilary Clinton. However, these results must be taken with consideration, as this correlation does not imply the causality of Hilary's defeat. This shows how big the impact of fake news was on the election and how misinformed voters could change their votes. Besides elections, the adverse impact of misinformation propagation has also been observed during COVID-19 [18, 69]. Bermes [6] investigated the link between information overload and sharing of fake news during the COVID-19 pandemic using the stressor-strain-outcome model. The findings suggested that perceived information overload has a negative impact on users' psychological strain, leading to a higher likelihood of sharing fake news. Misformation propagation can be used to manipulate users for political agenda, marketing, riots, rumor spreading, or misinformed behavior.

In literature, different types of methods exist to mitigate fake news, including immunizing/blocking nodes, truth-campaigning, and fact-checking assistance tools [80]. It has long been pointed out that actively broadcasting the true counter information is much more effective in minimizing the inverse impact of fake news than merely immunizing some nodes, also known as user blocking [81, 98, 99]. Immunizing the users by requesting them not to spread the fake news when encountered might have feasibility problems, as the effectiveness of the method will rely on the willingness of the selected users to follow the recommendation. The truth campaigning approach, on the other hand, might offer higher feasibility as users are more likely to accept and follow the recommendation as a result of the provided knowledge about the real fact. Studies have shown that once users have both fake and true information, they are more prone to explore further the event or topic and believe in true information [49, 63, 81]. In truth-campaigning, given a set of source nodes of the fake news propagation in an online social network, we aim to find a set of nodes, called truth-campaigners, to propagate counter-true information to minimize the impact of the fake news [75, 78, 80]. A small example of truth-campaigning is shown in Fig. 1, where node no. 6 and 14 are fake news spreaders, node no. 11 and 13 are truth-campaigners, and light red and light blue nodes are the nodes that accepted fake and counter-true information, respectively. It explains how truth-campaigning reduces the impact of fake news on the network.

In a social network, different groups or communities are unequally represented, and typically, minority groups are disproportionately absent from advantageous positions that create a diversity gap [38, 47]. The impact of network inequalities on the fairness of social interventions for influence maximization or awareness spread is well-known [87, 89]. There have been proposed various techniques to combat fake news over social media; most are collected in these survey [44, 80, 82]. However, none of the proposed methods considers the structural biases of social networks and fairness in fake news mitigation. The state-of-the-art methods aim to minimize the total number of users who believe in fake information and do not consider the minor and underrepresented communities in the network. This could result in the exclusion of these communities from the benefits of the intervention, which can have important societal repercussions. In this work, we highlight this issue in influence blocking and propose a fairness-aware truth-campaigning method to combat fake news.

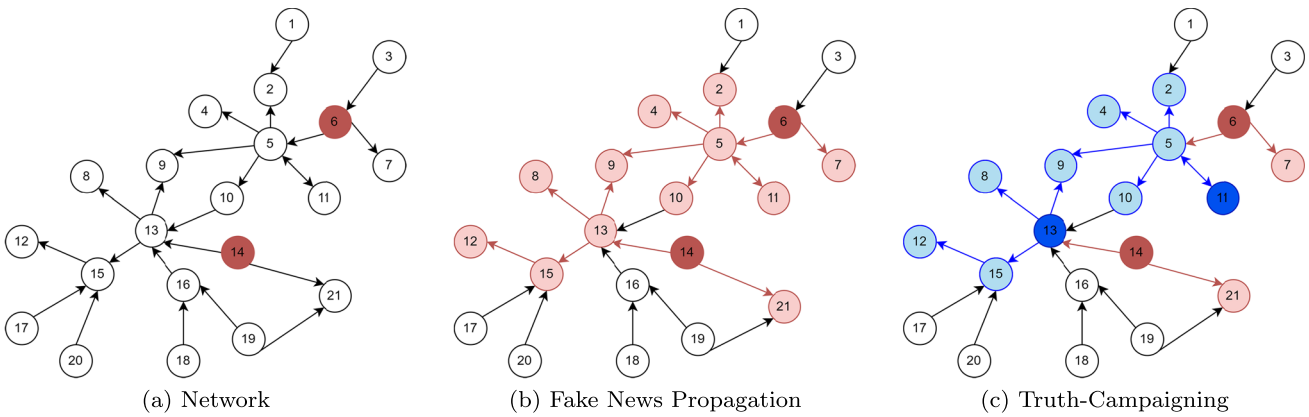


Fig. 1 An example to show truth-campaigning. (a). shows the network, where nodes in red color (node no. 6 and 14) are fake news spreaders, (b). shows fake information propagation, where the nodes in pink color received fake information, and (c). shows truth-campaigning process, where nodes shown in dark blue color (node no. 11 and 13) are seed nodes to spread counter-true information, and light blue nodes are the saved nodes as they accepted counter-true information and would have received fake information in the absence of truth-campaigners

We aim that the outcome of the truth-campaign should be fair for all communities irrespective of their size or any other characteristics, also known as group fairness [28, 73]. The fairness is considered using the Maximin fairness constraint, which strives to maximize the saved nodes for the community having a minimum fraction of saved nodes. This notion of fairness relates to the legal notion of disparate impact, which states that a community has been unfairly treated if its "success rate" under a policy is substantially lower than that of the other communities. The proposed solution aims to maximize a fairness-aware objective function while identifying the truth-campaigners to minimize the spread of fake information. The proposed method, called FWRRS (Fairness-aware Weighted reversible reachable system), relies on the weighted reversible reachable tree structure [93] to approximate the blocking power of each node in order to select those that achieve the best and fairest outcome. We verify the proposed method on real-world and synthetic networks, and observe that the proposed method is fair as well as saves the maximum number of nodes as compared to all considered baselines. We also observe that considering fairness while selecting early seed nodes reduces the price of fairness for later stages and makes the method more effective. Therefore, it provides higher efficiency as compared to baselines as the number of seed nodes increases.

To the best of our knowledge, this is the first work that has considered fairness while minimizing the negative spread of information propagation. The main contributions of the paper are briefly discussed below.

1. We propose a fair truth-campaigning method, called FWRRS, to minimize fake news propagation (explained in Section 5).
2. We evaluate the proposed method and compare it with state-of-the-art baselines on real-world social networks and synthetic networks. The experimental evaluation also highlights the long-term effectiveness of using the proposed fairness-aware method.
3. We also propose two fair baseline methods - (i) Fair-CMIA-O truth-campaigning method that extends CMIA-O (heuristic method based on Campaign- oblivious independent cascade model, Maximum Influence In-Arborescence and Maximum Influence Out- Arborescence) method [94] using maximin fairness constraint (explained in Appendix A), and (ii) Parity Seeding that extends fair Influence maximization method proposed in [86] for truth-campaigning.
4. We propose a network generating model, called HICH-BA (HIgh Clustering Homophily Barabási-Albert), that can be used to create synthetic networks having multiple communities of the given distribution with the desired homophily and clustering coefficient (explained in Appendix C).

The rest of the paper is structured as follows. In Section 2, we discuss related work to fake news and fairness-aware social network analysis. Section 3 explains the information propagation model used in the study. In Section 4 and 5, we discuss problem formulation and the proposed method, respectively. In Section 6 and 7, we explain the experimental setup and results of the proposed method on real-world and synthetic networks, respectively. The paper is concluded in Section 8 with future directions.

2 Related work

In this section, we discuss related work to detect fake news, influence blocking, and algorithmic fairness in social network analysis.

2.1 Fake news detection

Fake news detection has been a challenging problem and aims to automatically identify the fake news [34]. To classify a piece of news (also called meme or microblog) as fake or not, there has been used different features, such as (i) content-based features, (ii) user-based features, and (ii) network-based features [79]. Content-based features are the linguistic features of the post, such as the sentiment of the text, stance of the post, subjectivity and readability of text, usage of function words, external links (URLs), hashtags, lexical and syntactic features [13]. User-based features focus on the users who posted or reacted to the post by liking, sharing, commenting, or replying. These features include users' profile features, such as followers and friends count, the total number of posts by the user, user's bio, user's account age, user's verification, and so on [9, 91]. Network-based features consider users' connections in the network and the propagation network of the news on social media [53, 83].

Fake news detection methods focus on classifying the features of fake news from real news. [9, 33, 62] studied that the linguistic features of fake news are different from true (real) news. The other main differences are fake news might contain deceptive language [71], lack objectivity (might be biased towards an ideology) [66], the users involved with fake news might have lower credibility [84, 88], or it might contain false multimedia images or videos [36, 67], According to research by Vosoughi et al. [92], fake news spread differently and quicker than real news, with a wider reach. For detecting fake news based on its propagation, methods take into account social context information, such as who shared the news, how it spread through the network, and the connections between engaged users [23, 53, 83]. Given the complexity of fake news detection models, in recent years, the focus has also been on designing explainable fake news detection methods [17, 37, 58].

2.2 Influence blocking maximization

The problem of fake news mitigation is also referred to as *influence blocking maximization* that aims to maximize the blocking of the spread of the given fake information. The influence blocking maximization problem is NP-hard, and its objective function has been proven under most propagation models to be monotone sub-modular [54]. This implies that there exists a **greedy algorithm** that offers a solution with an approximation ratio of $(1 - 1/e - \epsilon)$ to the optimal solution, where ϵ depends on the accuracy of the influence range of each node, estimated using Monte Carlo (MC) simulations. However, this greedy method is very slow as a consequence of numerous and time-consuming required simulations. It is not feasible to apply them in real life, especially for large-scale networks encountered in social media. Therefore, several fast and scalable approaches for the IBM problem have been proposed. Most proposed methods can be grouped into two categories: (i) heuristic and (ii) approximation methods. There have also been other proposed methods based on reinforcement learning [30, 35], GCN [31], edge-blocking [97], and swarm intelligence [15].

The **heuristic-based approaches** rely on the characteristics of each node to make the selection of the k most suitable nodes for the truth campaign. Most of these methods identify seed nodes using different centrality measures [76]. Yao et al. [95] proposed selecting truth campaigners based on *eigenvector centrality (EVC)* [96]. A high EVC value for a node means it is connected to many nodes who themselves also have high EVC values. They show that under the Competitive Linear Threshold Model (CLTM), it performs better than the state-of-the-art algorithms. Other heuristics have also been applied, such as degree, clustering coefficient [26], betweenness centrality [11], or percolation centrality [64]. The performance of these metrics and some others, for e.g., random, were evaluated by Erd et al. [25] under two cost functions. One of these considered a degree cost penalty making highly-connected nodes more expensive, and the other one considered equal cost across all the nodes in the network. It was shown that under the Multi-Campaign Independent Cascade Model (MCICM), degree, betweenness, and percolation centrality offered the best performance. However, they required expensive nodes for their success; thus, these methods under-performed with the degree penalty cost function.

Under this category, methods exploiting the network structure have also been proposed. These rely on the community structure of the graph and will be considered as **community-based heuristic methods**. Lv et al. [54] proposed a method that allocated resources to each community proportional to the fraction of negative seed nodes within that community and their power of infection, i.e., computed using Monte Carlo simulation. Similarly, the work of Arazkhani et al. [3] proposed to select the nodes with the largest degree, betweenness, and closeness centrality measures from the largest k communities as truth campaigners. The community-based heuristic methods aim to allocate resources to communities based on their risk of infection; this could lead to fairer outcomes as resources are fairly divided, though the approach might be unfair toward indirectly infected communities.

In **approximation algorithms**, the aim is to approximate the influence-blocking power of a node for selecting the top-k influence minimizers. He et al. [32] proposed the CLDAG algorithm that restricts the influence computation of a node v to its local area to approximate the influence blocking power. The proposed method reduces computation cost by carefully selecting a local graph structure for v to allow efficient and accurate influence computation. Wu et al. [94] also proposed two very similar methods, CMIA-H and CMIA-O, that exploit the characteristics of the multi-campaign independent cascade model (MCICM) with high-effectiveness property (assuming the probability of positive transmission is always 1) and campaign-oblivious independent cascade model (COICM), respectively. Their research showed that these methods outperform the considered heuristic approaches, such as degree or proximity, while remaining fast.

Despite the good efficiency of the CMIA-O, Lin and Dai [52] showed that its performance is far from optimal. The method relies on the construction of numerous subgraphs and the consideration of a vast amount of nodes as candidates for the positive seed, which is time-consuming. To improve this further, they proposed the Block Influence Out Graph (BIOG) approach, which claims to be significantly faster than *CMIA-O* while performing similarly. Song et al. [85] proposed TIB-solver method, in which, first, the set of nodes that might be infected by the rumor is computed, and then the threat of each of these nodes (threat of v is the expected number of nodes that can be infected by v) is calculated. Then they use weighted Reverse Reachable (WRR) trees to greedily select top-k nodes that save the most nodes in the given deadline. They use WRR trees to predict the probability of reaching a node with positive information, and use a different structure for the negative one, namely a DAG. In our proposed method, WRR trees are used to estimate the spread of both involved information types. The main drawback of the existing methods is that they simplify the network structure, such as removing self-loops, for estimating the blocking power of a node, which affects their effectiveness in reality. The proposed method, called FWRRS (Fairness-aware Weighted Reversible Reachable System), does not simplify the network structure; however, we exploit the structure of weighted reversible reachable (WRR) trees to estimate the blocking power.

2.3 FairSNA

To the best of our knowledge, fairness has not been considered for solving the Influence Blocking Maximization problem using immunizing nodes or truth-campaigning. In the extensive surveys on fairness in graph mining [22, 73], no existing work in this area was included. However, fairness has been considered in other related areas of information diffusion, such as the influence maximization (IM) problem. The IM problem was first formulated by Kempe et al. [39] and focuses on identifying a set of k initial adopters for maximizing the spread of certain information within a network. In the past three years, researchers have highlighted the bias in social influence maximization methods and have proposed fairness-aware influence maximization methods [1, 86].

Ali et al. [1] proposed to greedily choose the node that achieves the highest value for an objective function that balances both the total influence and fairness. In this work, fairness was defined using equality that aims to achieve an equal fraction of influenced nodes across all the communities. Thus, the objective function would punish nodes that would cause the distance between the most and least influenced communities to be large. They showed that guaranteeing such fairness comes at the cost in effectiveness. Following this work, Stoica [86, 87] showed that when the seed set size is large enough, promoting parity in the seed results leads to better parity in the outreach. Additionally, this diverse seed set taps into inactivated communities that are hard to reach from central nodes and thus leads to a better outreach, proving the benefits of fairness beyond a balance in the output. Halabi et al. [24] empirically demonstrated that fair solutions are often nearly optimal on real-world scale-free networks, and therefore the price of fairness is significantly low (less than 15%). Additionally, they proved that the algorithms that do not impose fairness constraints introduce significant bias.

Fairness-aware solutions have also been studied for other social network analysis problems, such as link prediction [41, 51, 55, 72, 74] and centrality ranking [90]. In this work, we define fairness for influence blocking maximization that aims that the impact of negative information should not be unfair to any community, especially small and less connected communities.

3 Preliminaries

3.1 Independent cascade model (ICM)

ICM is a stochastic model for simulating information diffusion on a network. Nodes in the network can be in one of two states: (i) active, if it has been influenced by the information, or (ii) inactive, if the node is unaffected by the information or unaware of it. Each edge (u, v) has an influence probability that represents the likelihood of node u influencing node u. In ICM, the diffusion starts from a set of seed nodes whose state is changed to active. When a node becomes active, it attempts to activate its neighboring nodes with the influence probability of the connecting edge. This process continues until no new nodes are activated, as each infected node attempts to impact its neighbors once.

3.2 Multi-campaign ICM (MCICM)

MCICM extends the classic ICM by incorporating multiple information types that diffuse simultaneously [12]. For our analysis, let's assume that there are two information types - fake news (negative) and counter-true information (positive). The negative information consists of the one we want to limit and the positive one is propagated to limit that. In this model, the nodes can be in any of the three states (i) negative active, (ii) positive active, or (iii) inactive, unaware of all diffusion taking place. Once a node changes its state to negative active or positive active, it remains in that state. Each edge has two associated probabilities of transferring positive and negative information. In the influence propagation, there are two sets of seed nodes for fake news and counter-true news, which have negative active and positive active states, respectively. When a node is affected by one of the information types, it will propagate to its neighbors with the respective influence propagation probability. When a node is reached simultaneously by both types, it is assumed that there is a natural order to the two campaigns, and in this work, the positive one will take effect. In our work, we use the Campaign-Oblivious Independent cascade Model (COICM)[12], where the probabilities of positive and negative dissemination are arbitrary and shared over the network. This means that both information propagates using the same cascade.

4 Problem formulation

Let's assume that the given network is represented as $G = (V, E)$, where V is the set of nodes, and E is the set of edges. Each edge (u, v) is associated with an influence propagation probability that shows the ability of node u to infect or influence node v. The negative seed nodes, represented by a set $S_N \subseteq V$, spread the misinformation in the network. The positive seed nodes, represented by a set $S_P \subseteq V \setminus S_N$, spread the counter-true information in the network. At a given time, the state of a node u (s_u) can be in any of the following three states $[-1, 0, 1]$, where -1 shows that the node is negative active, 1 shows that the node is positive active, and 0 shows that the node is inactive and has not been influenced by any of the propagating information. $s^-(G, S_N, S_P)$ is the number of nodes that are negative active in G if S_N spreads misinformation and S_P spreads the counter truth. The influ-

ence blocking power of positive seed set S_P is computed as follows:

$$\mathbb{B}(G, S_N, S_P) = s^-(G, S_N, \phi) - s^-(G, S_N, S_P)$$

The influence blocking power is also referred to as 'saved nodes,' where saved nodes are those that would have been "infected" by the fake news if it was not for the truth campaign.

In the IBM problem, given a graph $G(V, E)$, a set of fake news spreaders S_N, and a positive integer k representing the available budget for the mitigation campaign, the aim is to find the positive seed set $S_P \subseteq V \setminus S_N$ of size at most k that minimizes the expected number of negatively activated nodes under the $COICM$ propagation model. It aims to maximize the following optimization function.

$$\underset{S_P}{\text{maximize}} \quad \mathbb{B}(G, S_N, S_P) \tag{1a}$$

subject to

$$|S_P| <= k, \ S_P \subseteq V \setminus S_N \tag{1b}$$

Fairness-aware Truth-Campaigning. In **Fairness-aware Influence Blocking Maximization** (FIBM), the goal is not only to maximize the blocking power of the positive seed set but also to ensure that the outcome of the truth campaign is fair. The fairness of the campaign will be measured using the maximin fairness constraint, choosing S_P to be the positive seed that maximizes the minimum proportion of nodes saved in each community. This relates to the legal notion of disparate impact, which states that a community has been unfairly treated if its "success rate" under a policy is substantially lower than that of the other communities [5]. Maximin aims to avoid the scenario in which some communities are disproportionately neglected from help concerning the remaining population. Thus, trying to maximize the minimum help all communities receive. Let $\mathbb{M}(G, S_N, S_P)$ define the maximin value achieved by S_P in the graph G having a set of communities C and a set of fake news spreader S_N. The maximin value is computed as follows:

$$\mathbb{M}(G, S_N, S_P) = \min_{c \in C} \frac{\mathbb{B}_c(G, S_N, S_P)}{s_c^-(G, S_N, \phi)}$$

where $\mathbb{B}_c(G, S_N, S_P) = s_c^-(G, S_N, \phi) - s_c^-(G, S_N, S_P)$ denotes the blocking power of the positive seed S_P within a community $c \in C$.

Problem Statement. Given a graph $G(V, E)$ having communities C, a set of misinformation spreaders S_N, and a given budget k for the mitigation campaign, we aim to identify positive seed set S_P using the following objective function.

$$\underset{S_P}{\text{maximize}} \quad \mathbb{B}(G, S_N, S_P) \tag{2a}$$

subject to

$$S_P \in \arg\max_{S'_P} \mathbb{M}(G, S_N, S'_P), \tag{2b}$$

$$|S'_P| <= k, \ S'_P \subseteq V \setminus S_N \tag{2c}$$

In the proposed problem, an additional constraint, namely (2b), guarantees the fairness of the approach. Therefore, the solution set S_P shall not only guarantee that the truth camping is as effective as possible by saving the maximum number of nodes from believing the fake news but also guarantee that the maximin value has also been maximized.

5 The proposed method

In the proposed method, given the network G and S_N, we first identify the set of nodes S that can be infected by the given rumor starters and estimate the probabilities of their infection. Next, we generate weighted reversible reachable (WRR) trees from set S (the process to generate WRR is explained in Algorithm 2). For each node, WRR trees aim at identifying all those other nodes which can reach it with their information. These WRR tree structures can be generated very fast, and therefore, allowing the method to generate a large pool of WRR trees that resemble how information could diffuse in the network. Thus, if a path of diffusion is to be considered, a WRR tree is expected to cover it. To generate this sample of WRR trees, we rely on the Dynamic Stop and Stare algorithm (D-SSA) [60], which offers an error guarantee on the approximation as well as a good performance, as it is 1200 times faster than the next best sampling method; as claimed by the authors and also we observed in our work.

Next, the generated WRR trees are used to estimate the blocking power of each node with respect to different communities. Now the optimization function defined in (2a) is maximized for maximin fairness constraint, and iteratively, the node with the highest influence blocking power providing maximum fairness is selected until k truth campaigning nodes are chosen. For the process of choosing top-k positive seed nodes, we use six selection steps explained in building Blocks 3.B (in Section 5.1).

The proposed method is presented in Algorithm 1. This takes six parameters as input, network $G(V, E)$ having communities C, the negative seed set S_N, the budget k, and three variables γ, ϵ, and $iterations$. The parameter γ represents

the infection probability threshold to add a node to S, which is the set of nodes that are expected to get infected. Next, the parameter ϵ, which is one of the input parameters of the *D-SSA* sampling algorithm; we set it to 0.1. This parameter will be the approximation guarantee of the probabilities in the WRR trees returned by $D\text{-}SSA$. Further is δ, which determines the number of node copies allowed in the WRR trees. Last is $iterations$ consisting of the number of simulations of ICM that will be utilized to approximate the infection probabilities.

5.1 Building blocks

Here, we explain the steps of Algorithm 1 in detail.

1. Identify the subset of nodes S with expected infection probability (Line 1). The set S consists of nodes that will be influenced by the fake news propagated from nodes S_N with a probability of at least γ in the absence of a truth-campaign. To obtain these probability values, monte-carlo simulation of ICM is performed $iterations$ times from S_N. The probability of infection of a node v is set to be the ratio of simulations in which it was included and stored in $probs[v]$. The nodes having a $probs$ value higher than γ are then included in S. Additionally, the variable $paths_v$ keeps track of simulation paths for each node $v \in S$.

2. Generate a pool of WRR trees (Line 2). The pool of WRR trees is constructed using an extended version of the D-SSA sampling algorithm, called *D-SSA_WRRS* [60]. The general idea of the algorithm is to recursively double the number of WRR trees until two conditions are met, in terms of coverage of the selected set and the error in the coverage approximation. The D-SSA algorithm is extended to contain a third condition, which ensures that the generated WRR trees shall be sufficient to guarantee with high probability a good approximation of the influence power of S_N, i.e., necessary to accurately predict how to block such influence. Additionally, to speed the process of meeting this last condition, the roots are chosen with the infection probabilities computed in the previous step, namely $probs$. Each WRR is generated by calling $generate_WRR(G, S, probs, S_N)$, presented in Algorithm 1. First, a root r is chosen at random from the set S using the probability distribution in $probs$ (Line 2). Including these probabilities helps boost the performance of the method, as the $FWRRS$ method is interested in those WRR trees that include a negative seed node. From r, we begin to visit G in a breadth-first search manner. From the node r, all of its predecessors v are added to WRR with probability $p_{v\to u}$, and the process continues until no new node is added or WRR reaches to S_N as no nodes at a distance larger than this will be able to

Algorithm 1 FWRRS (G, S_N, k, γ, ϵ, $iterations$,δ)

Input: G, S_N, k, γ, ϵ, $iterations$, δ
Output: S_P

1: $S, probs, paths \leftarrow compute_S(G(V, E, C), S_N, \gamma, iterations)$
2: $WRRS \leftarrow D\text{-}SSA_WRRS(G, S, probs, paths, S_N, k, \epsilon, \delta)$
3: $WRRS_{inf} \leftarrow WRR \in WRRS$ if $WRR.has_S_N = True$
4: $init_inf \leftarrow$ list of length $|C|$ consisting of 0s
5: **for** $v \in S$ **do**
6: $\quad count_v \leftarrow \#WRRs$ in $WRRS$ with root v
7: $\quad prob_inf_v \leftarrow \frac{count_inf_v}{\#WRRtrees \in WRRS_{inf} \text{ with root } v}$
8: $\quad init_inf[comm_v] += prob_inf_v$
9: $C_inf \leftarrow [c \in C$ if $init_inf[c] > 0]$
10: $pct_sav_c \leftarrow 0$ for $c \in C_inf$, $exp_inf \leftarrow init_inf$
11: $B_v \leftarrow$ list of length $|C|$ consisting of 0s, for all $v \in V$
12: $candidates \leftarrow \phi$, $S_P \leftarrow \phi$, $Mm \leftarrow 0$
13: **for** $WRR \in WRRS_inf$ **do**
14: $\quad r \leftarrow WRR.root$
15: $\quad$ **for** $v \in WRR \setminus S_N$ **do**
16: $\quad\quad B_v[comm_r] = +1/count_r$ and add v to $candidates$
17: **while** $|S_P| < k$ **do**
18: $\quad M_v \leftarrow \min_{c \in C_inf} \frac{B_v[c]}{init_inf[c]}$ for all $v \in candidates$ — Selection Step 1
19: $\quad opts_1 \leftarrow [v \in candidates$ if $M_v \in (1 \pm \epsilon * \Phi_1) * \max_{w \in V} M_w]$
20: $\quad total_sav_v \leftarrow \sum_{c \in C_inf} B_v[c]$ for $v \in opts_1$
21: $\quad list_total_sav \leftarrow$ list of values $total_sav_v$ for $v \in opts_1$
22: $\quad C_no_help \leftarrow [c \in C_inf$ if $pct_sav_c = 0]$ — Selection Step 2
23: $\quad$ **if** $|C_no_help| > 0$ **then**
24: $\quad sav_no_help_v \leftarrow \sum_{c \in C_no_help} B_v[c]$ for $v \in opts_1$
25: $\quad$ **if** $\frac{\sum_{c \in C_no_help} exp_inf[c]}{\sum_{c \in C_inf} exp_inf[c]} \geq \frac{\epsilon * |C_no_help|}{|C_inf|}$ or $\exists_{c \in C_no_help}$ disconnected from S_P **then**
26: $\quad\quad opts_1 \leftarrow [v \in opts_1$ if $sav_no_help_v > 0]$
27: $\quad avg_help \leftarrow \sum_{c' \in C_inf} \frac{pct_sav_{c'}}{|C_inf|}$ — Selection Step 3
28: $\quad C_be_avg \leftarrow [c \in C_inf$ if $pct_sav_c \leq avg_help]$
29: $\quad wgt_sav_v \leftarrow \sum_{c \in C_be_avg} B_v[c] * (1 - pct_sav_c)$ for $v \in opts_1$
30: $\quad opts_2 \leftarrow [v \in opts_1$ if $wgt_sav_v \in (1 \pm \epsilon) * \max_{w \in opts_1} wgt_sav_w + \Phi_2]$
31: $\quad C_min_help \leftarrow [c \in C_inf$ if $pct_sav_c \in (1 \pm \epsilon) * Mm]$ — Selection Step 4
32: $\quad sav_min_v \leftarrow \sum_{c \in C_min_help} B_v[c]$ for $v \in opts_2$
33: $\quad opts_3 \leftarrow [v \in opts_2$ if $sav_min_v \in (1 \pm \epsilon) * \max_{w \in opts_2} sav_min_w + \Phi_3]$
34: $\quad$ **if** $\max_{v \in opts_3} total_sav_v \leq \max list_total_sav * (1 - \epsilon)$ **then** — Selection Step 5
35: $\quad\quad$ remove $v \in opts_3$ if $total_sav_v \notin (1 \pm \epsilon) * \max_{v \in options_3} total_sav_v$
36: $\quad u \leftarrow \arg\min_{v \in opts_3} \sum_{w \in Neigh_out_v \cup \{v\}} prob_inf_w$ for $v \in opts_3$ — Selection Step 6
37: $\quad$ add u to S_P, remove u from $candidates$
38: $\quad exp_inf \leftarrow init_inf - B_u$, $Mm \leftarrow M_u$, $pct_sav_c \leftarrow \frac{B_u[c]}{init_inf[c]}$
39: $\quad$ **for** $WRR \in WRRS_inf$ with $u \in WRR$ **do**
40: $\quad\quad r \leftarrow WRR.root$
41: $\quad\quad prob_inf_r -= 1/count_r$
42: $\quad\quad$ **for** $v \in candidates \setminus WRR$ **do** $B_v[comm_r] += 1/count_r$
43: $\quad\quad$ remove WRR from $WRRS_inf$
44: Return S_P

save r from getting infected. For every node $v \in WRR$, the probability of the path towards the root is stored in p_v (Lines 7-8). If v is already in WRR, a copy of v is added (Lines 14-20), unless any copy of v is already present in the path between v and r, as in the MCICM a node can not activate twice (line 13). Finally, the nodes with a lower probability of reaching r than any negative seed in the WRR are removed (Lines 26-28). **Efficiency Improvement.** The list $paths$ was introduced to boost the performance of this algorithm when the probabilities of infection are high. In this case, the size of the WRR trees can increase that significantly slows the method. The number of copies of nodes grows significantly with the number of paths through which information diffuses. However, not all paths need to be considered. If a node is present in a WRR it is assumed that its information will reach the root node, so having multiple copies of this node does not provide any additional information. However, including a copy of this node might be necessary to cover a path through which the root can get infected; the variable $paths_r$ is used to avoid this. If a new node v is to be added to a WRR tree, where v already has numerous copies present, $paths_r$ is used as follows. If the path being considered from the copy of v to the root, namely $des_u \cup \{v\}$, does not lead to the infection of the root, it will not be considered any further. This is assumed to be the case if it is not a subset of any path in $paths_r$. This allows for the pruning of the WRR trees significantly as the consideration of numerous unnecessary paths is avoided (Line 15). The maximum number of copies, δ, is set to 10 in all our experiments, considering at least ten different paths through which any node v can inform r. A larger maximum number of copies will not have a significant improvement for blocking power on the evaluated data sets (also observed experimentally). In fact, it is rarely reached for lower influence probabilities, and the limit comes into significant use when the probabilities become higher

3. **Choose Truth-Campaigners (line 3- 44)**. This has the following three steps.

A. Initialisation (Line 3- 16). In Algorithm 1, after generating the pool of WRR trees, the infection probability for each node $v \in S$ is computed. For this, only the WRR trees that include a node in S_N are considered and stored in $WRRS_{inf}$(Line 3). A node's probability of infection is initialized in line 4 and set to the ratio of WRR trees rooted at v that includes a node in S_N (Lines 5-8). After computing the infection probability for all nodes in S, the expected number of initial infections per community $init_inf$ is computed (Line 8). The value of $init_inf$ in the community c is set to be the sum of the infection probabilities of all nodes $v \in c$. Further, the variable

Algorithm 2 generate_WRR(G, S, $probs$, $paths$, S_N, δ)

Input: G, S, $probs$, $paths$, S_N, δ
Output: WRR

1: $WRR \leftarrow$ empty graph, $max_depth \leftarrow \infty$
2: $r \leftarrow$ random node in S chosen using the probabilities in $probs$
3: $WRR.has_S_N \leftarrow False$, $WRR.root \leftarrow r$, $A \leftarrow [r]$
4: $des_r \leftarrow \phi$, $original_r \leftarrow r$, $s_r \leftarrow 0$, $copies_r \leftarrow 0$, $prev_r \leftarrow \phi$
5: **while** A is not empty **do**
6: $u \leftarrow$ take first element out of A
7: **if** $prev_v = \phi$ **then** $p_v \leftarrow 1$
8: **else** $w \leftarrow prev_v$, $p_v \leftarrow p_{original_w \rightarrow original_v} * p_w$
9: **if** $original_u \notin S_N$ **then**
10: **if** $s_u \leq max_depth$ **then**
11: **for** $v \in V$ such that the edge $(v, original_u) \in E$ **do**
12: add_v is True with probability $p_{v \rightarrow original_u}$ else False
13: **if** add_v and $v \notin des_u$ **then**
14: **if** $v \in WRR$ **then**
15: **if** $copies_v \leq \delta$ or $(\exists_{p \in paths_r} des_u \cup \{v\} \subseteq p)$ **then**
16: $copies_v \leftarrow copies_v + 1$, $i \leftarrow copies_v$
17: Add node v_copy_i and edge (v_copy_i, u) to WRR
18: $des_{v_copy_i} \leftarrow des_u \cup \{u\}$
19: $s_{v_copy_i} \leftarrow s_u + 1$, $prev_{v_copy_i} \leftarrow u$
20: add v_copy_i to A
21: **else**
22: Add node v and edge (v, u) to WRR
23: $des_v \leftarrow des_u \cup \{u\}$, $s_v \leftarrow s_u + 1$, $prev_v \leftarrow u$
24: add v to A
25: **else** $max_depth \leftarrow s_u$, $WRR.has_S_N \leftarrow True$
26: **for** $v \in WRR$ **do**
27: **if** $p_v < \max_{v \in WRR \cap S_N} p_v$ **then**
28: remove v from WRR
29: **return** WRR

C_inf contains the communities where infections take place (Line 9). The percentage of saved nodes per community pct_saved is initialized to 0, and the number of expected infections, exp_inf, is initialized to be equal to the initial ones (Line 10). The blocking power of a node v per community, B_v, for all $v \in V$ is initialized to 0 (Line 11), as well as the maximin value Mm (Line 12). In lines 13-16, the WRR trees in $WRRS_{inf}$, which are those with $WRR.has_S_N = True$, are iterated over. If any node $v \in WRR$ were to be added to S_P, the root r of this WRR would be saved from being infected within such a tree, as v would reach it with its information before any node in S_N. Thus, the number of WRR trees in which r gets infected reduces by 1, and the value of $B_v[comm_r]$ is increased by (1/number of WRR trees rooted at r). This logic is then applied to compute the blocking power in each community for any node. Additionally, to avoid iterating over the entire set of nodes of the network, a set $candidates$ is created to contain the nodes with some blocking power.

B. Selection Steps (Line 18-37). Next, we iteratively select k truth-campaigners, and in each iteration, the following six selection steps are performed. In brief, steps 1 and 2 maximize optimization function 2b, steps 3-5

maximize optimization function 2a, and step 6 is used to break the ties given the budget.

1. First, optimize 2b, which maximizes maximin fairness, and store candidate nodes in $opts_1$.
2. If there is any community that is infected significantly (according to the input parameter ϵ) and not reachable from S_P, remove candidate nodes that will not reach any of these communities.
3. Choose candidate nodes that maximize the weighted blocking power in the communities with a below-average saving ratio (based on already chosen S_P) and store them in $opts_2$.
4. Pick out the candidate nodes that maximize blocking power in the communities having the lowest fraction of saved nodes (referred to as maximin community) and store them in $opts_3$.
5. If there is a cost of fairness due to the selection steps 2-4, choose the nodes from $opts_3$ with the highest blocking power (This is done as it will not impact the fairness as all the nodes chosen in $opts_3$ provide a similar maximin value, though it will improve the blocking power).
6. Break the ties, if any, by selecting the node with the maximum weighted out-degree, where a node's weight is set to be its infection probability under the current S_P.

This way, in each step, we select the node that maximizes the fairness objective without causing a significant cost in the effectiveness of the campaign.

C. Recompute and Repeat (Lines 38-44). Once a new node u is added to S_P, the expected infections, maximin value, and ratio of saved nodes per community are modified (Line 38). Next, the blocking power of impacted nodes is updated by iterating over those $WRR \in WRRS_{inf}$ in which the root r is expected to be reached by u (Lines 39-43). In the WRR trees that include the node u, the root will remain saved, and the expected infections of all nodes not present in this WRR are reduced by $1/(\text{number of } WRR \text{trees rooted at } r)$. Further, this WRR is removed from the set of WRR trees. The process finalizes when the size of S_P is k, and the resulting set of truth campaigners is returned.

The flowchart of the proposed method is shown in Fig. 2.

6 Experimental set-up

In this section, we discuss datasets, baselines, and experimental settings.

6.1 Datasets

We evaluate the proposed method on six real-world datasets that are summarized in Table 1. The communities are identified using the Louvain community detection method [8]; this is briefly explained in Appendix B. Facebook and Twitter are a snapshot taken from online social networking websites Facebook and Twitter, respectively. PHY, HEPT, and ASTRO are co-authorship networks extracted using publications in the category of Physics, High Energy Physics, and Astro Physics at ArXiv, respectively. Enron is an email communication network constructed for the Enron organization.

6.2 Baselines

We compare the proposed method with fairness-oblivious and fairness-aware baselines to choose truth-campaigners. The fairness-oblivious baselines are described below.

- **Degree:** nodes with the highest degrees are selected.
- **CMIA-O** [94]: compute the blocking power of each node using the Maximum Influence In-Arborescence (MIA) structure. These structures are used to compute the blocking power of each node and greedily choose the node with the highest value and recompute.
- **BIOG** [52]: This is similar to *CMIA-O*; however, it uses a Maximum Influence Out-Graph (MIOG) structure to approximate the influence region. It also reduces the set of candidates for S_P to those nodes directly connected to S_N or adjacent to these.
- **TIB-Solver** [85]: It first computes the threat of each node that can be reached by the rumor. The threat of a node is set to be the number of nodes that will be infected by it. Then WRR trees are generated to estimate the number of reachable nodes, and greedily the top-k nodes that best block the rumor are chosen.

As mentioned before, there is no work on fairness-aware influence blocking maximization. Therefore, we extend two existing approaches for the IBM problem to achieve a fairer outcome. The fairness baselines used are the following.

- **Parity Seeding** [7]: Stoica et al. [86] proposed a fairness-aware influence maximization method. We extend the method to the influence blocking maximization problem where the nodes with the highest degree while maintaining parity are chosen to be truth-campaigners, and their ratio in different communities is proportional to S_N.
- **Fair-CMIA-O** [7]: Fair-CMIA-O is an extension of CMIA-O [94] method using maximin fairness constraint that follows the same structure as CMIA-O, but the nodes

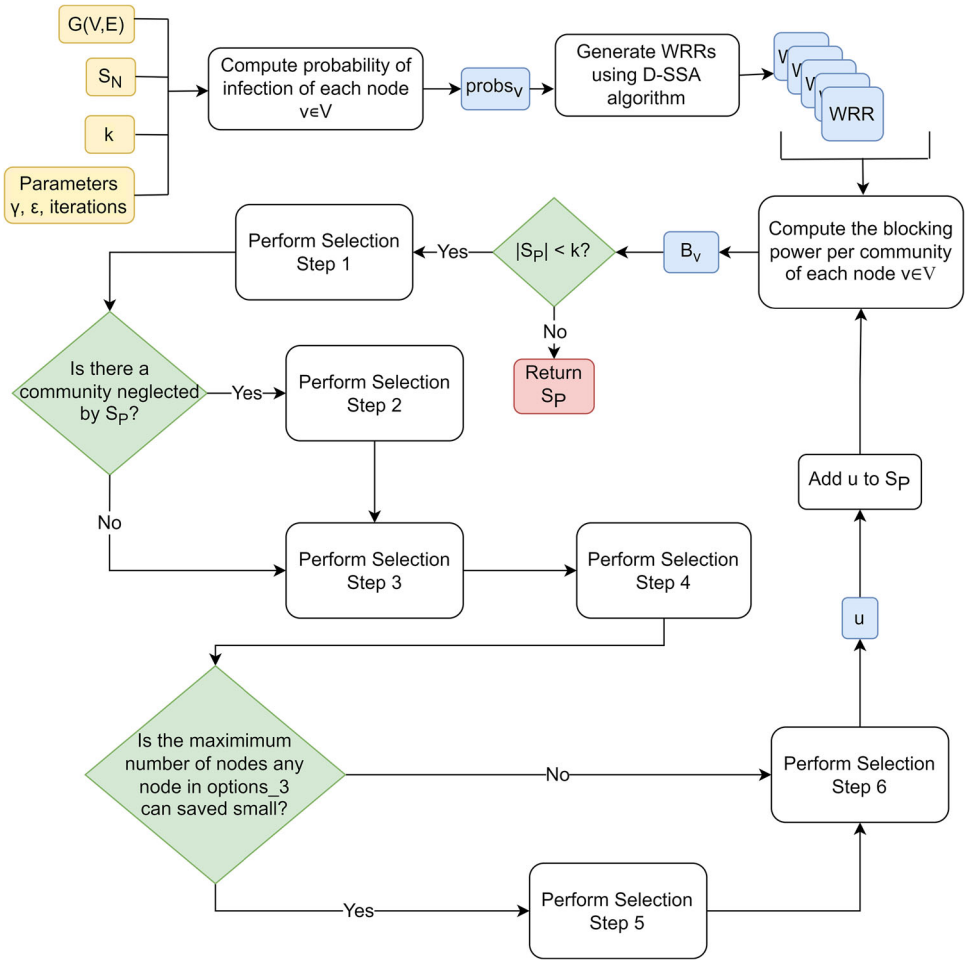


Fig. 2 Flowchart of FWRRS method

resulting in the highest maximin value are chosen. Ties are broken by choosing the node with the highest blocking power within those communities that still need help. The detailed algorithm is explained in Appendix A.

6.3 Evaluation metric

The proposed method is evaluated using the following two metrics that consider both efficiency as well as fairness.

1. Fraction of Saved Nodes: It shows the fraction of saved nodes from the total negatively influenced nodes in the absence of a truth-campaign. It is computed as,

$$\frac{\mathbb{B}(G, S_N, S_P)}{s^-(G, S_N, \phi)}$$

2. Maximin: It represents the fairness of the proposed method, and it is computed as,

$$\min_{c \in C} \frac{\mathbb{B}_c(G, S_N, S_P)}{s_c^-(G, S_N, \phi)}$$

Table 1 Characteristics of datasets

Datasets	#Nodes	#Edges	Avg. Degree	#Communities	Ref
Facebook	4039	176468	43.69	16	[56]
Twitter	81309	1768149	21.74	74	[56]
PHY	37149	231584	9.38	2445	[16]
HEPT	15235	62776	4.12	1821	[16]
ASTRO	18772	396160	21.10	322	[48]
Enron	36692	367662	10.02	1245	[43]

6.4 Implementation details

The influence propagation is modeled using COICM, and the probability $p_{u\to v}$ for each edge (u, v) is set to be $1/inDegree(v)$, where $inDegree(v)$ is the number of incoming edges to v. The parameters are set as, $\gamma = 0$, $\epsilon = 0.01$, and $iterations = 1000$. In each iteration, uniformly at randomly 50 nodes are chosen as rumor spreaders, and each experiment is repeated 1000 times, and the average is taken to compute the number of saved nodes, as well as fairness.

The code was written in Python, and experiments were carried out on a server with the following characteristics: CPU 1x Intel Xeon E5-2698v4 @ 2.2GHz, RAM 256GB, and GPU 4x Nvidia Tesla V100. Our implementation is available at Github repository[1].

7 Results

We first compare the fraction of saved nodes by the proposed method as compared to baselines. The results are shown in Fig. 3. We observe that the proposed method, FWRRS, provides high efficiency. The performance of the considered baselines is inconsistent on all datasets, and we observe that some baselines provide good results on some datasets and some on others. The top three baselines are Parity seeding, Degree, and Fair-CMIA-O, where Degree is fairness-oblivious, and the other two are extended fair methods for IBM. In Fig. 3a, b, e, and f, choosing S_P using higher degree nodes seems to be a powerful strategy. For these datasets, the best-performing method, after the proposed *FWRRS*, is generally *Parity seeding*. The high value of the degree of the most connected nodes results in high coverage of the positive information. Consequently, most paths of diffusion of negative information are intercepted. The maximum degree in $Facebook$, $Twitter$, $ASTRO$, and $Enron$ is significantly above the average degree, and the average degree is also quite large, resulting in information spreading through numerous paths. Hence, reducing the network to maximum probability paths, like the $CMIA\text{-}O$ method, is inefficient, as, in reality, information diffuses throughout many other paths. However, in the $HEPT$ and PHY networks, the next best performing methods are *CMIA-O* and *Fair-CMIA-O*. These methods benefit from the low average degree, which causes a reduction in the number of paths through which information diffuses. Blocking the maximum probability paths results in the blocking of the most propagation paths. In these networks, *degree* and *Parity seeding* do not perform as well as in other networks.

[1] Refer to https://github.com/akratiiet/FWRRS.

Next, we compare the fairness of the proposed method, and the results are shown in Fig. 4. The results show that the *FWRRS* method consistently provides the highest fairness value given that it also provides the highest influence blocking. The plots also highlight the unfairness of fairness-oblivious IBM methods, including the highest degree. Existing methods tend to neglect minor communities, showing the need for fairness-aware methods. Fairness-aware methods mostly achieve a maximin value higher than 0 on all datasets, and it shows that there is no community that has been completely ignored while blocking the negative influence. For the $Facebook$ dataset, in Fig. 4a, other baselines, including $CMIA\text{-}O$ and $BIOG$, also achieve a maximin value higher than 0 for smaller k as this network has a small number of communities and a high average degree that makes spreading the true information to all communities more accessible for any method. As a result, all communities that get infected are reached and saved to some extent. Other networks have many small-size communities, and they are most likely to get infected by randomly chosen negative seed sets. However, the existing methods are more likely to neglect such minor communities for the larger good, a higher percentage of saved nodes, and provide low fairness.

Despite the clear impact of the network characteristics in the considered baselines' effectiveness, the proposed method *FWRRS* seems to perform best over all networks. The fairness-aware objective function does not negatively impact the number of saved nodes. In fact, if the seed nodes are chosen fairly, then for a sufficiently large budget k, it works as a catalyst to improve the effectiveness. Enforcing the true information to reach all relevant communities guarantees that the intervention reaches communities that otherwise would have been neglected. This helps overcome the overlapping influence spheres that the nodes being selected by the other methods have. For low-budget values, the performance is similar or superior to the next best-performing method. It is only in Fig. 3e and f that some cost of fairness is observed, as *degree* outperforms *FWRRS* for $k \leq 20$. However, the cost of such fairness is very low, only 2% in the worst case. This cost comes from trying to reach all communities in need with a low number of truth spreaders, but this initial effort becomes valuable for larger k values, where the *FWRRS* achieves better performance than all other baselines. The FWRRS method is the fairest compared to all the evaluated methods while consistently achieving the highest percentage of saved nodes.

7.1 Sensitivity analysis

In this section, we examine the effect of various parameters on the performance of different methods, and the results are shown on the PHY network. We would like to mention that there is no specific reason to pick this network. It is a medium

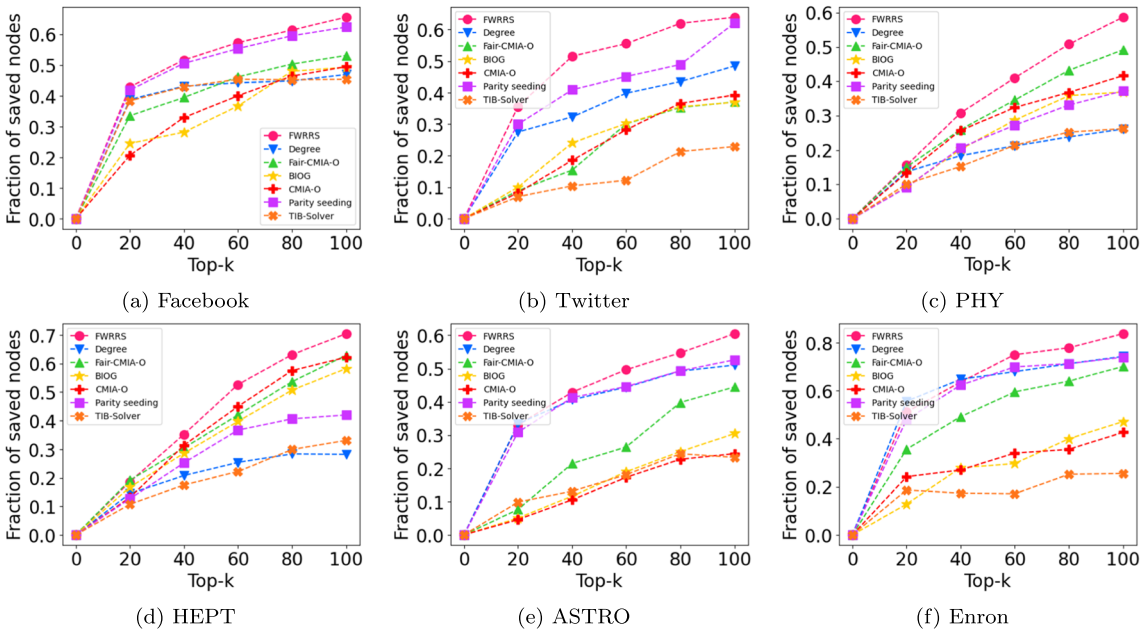


Fig. 3 Percentage of saved nodes versus k (the number of chosen truth-campaigners)

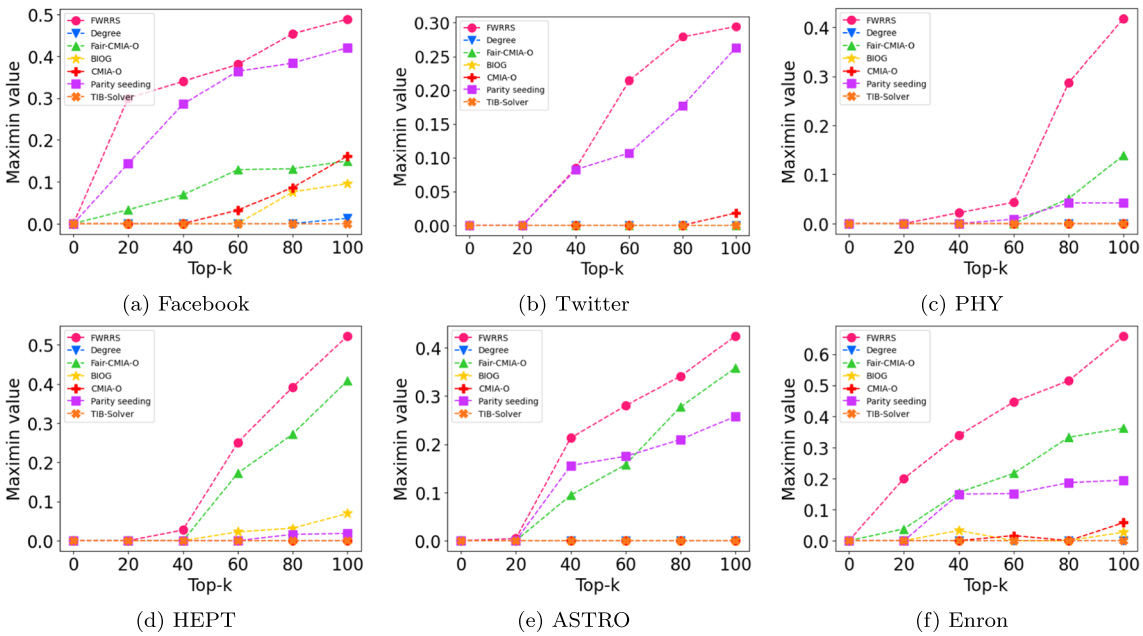
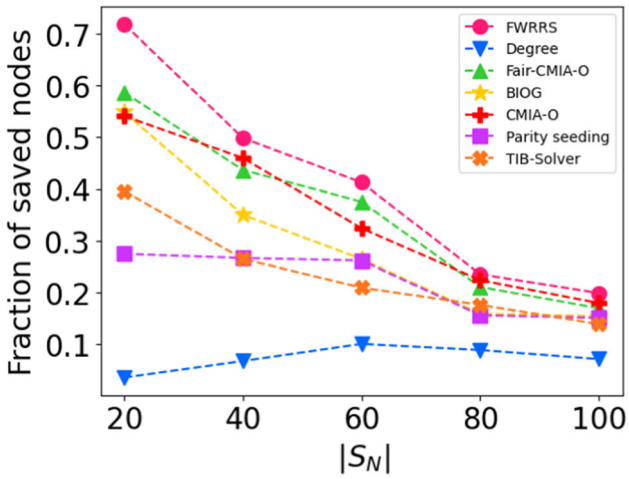
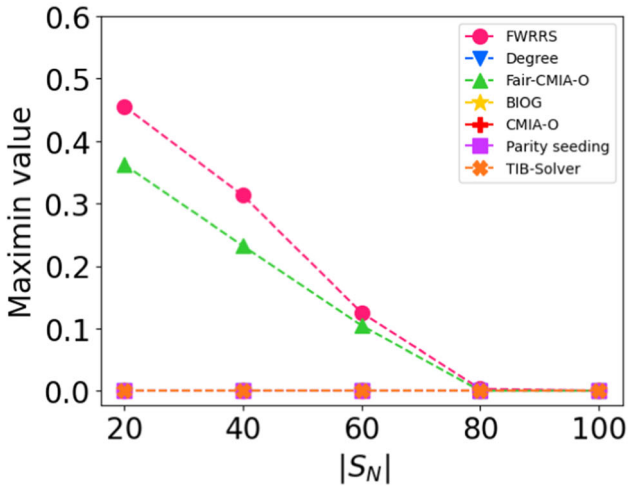


Fig. 4 Fairness versus k (the number of chosen truth-campaigners)



(a) Fraction of saved nodes



(b) Maximin value

Fig. 5 Experiments in PHY under varying sizes of S_N

size network, and therefore, the results are shown on this. The results are generalizable to all considered datasets.

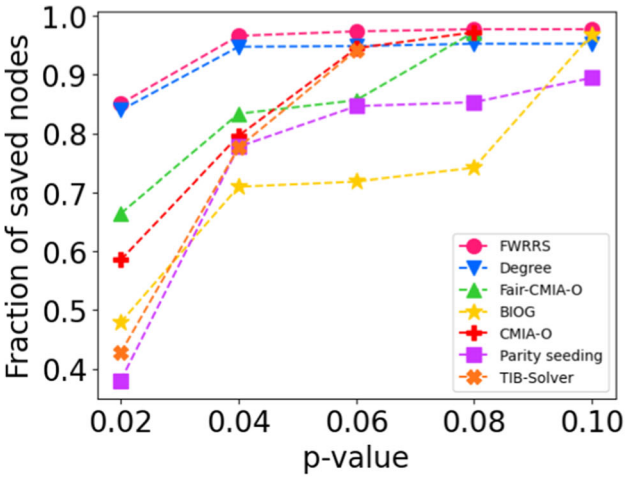
7.1.1 Effect of varying $|S_N|$

In all previous experiments, the number of rumor starters was fixed to 50. We evaluate the impact of the size of the rumor starter set that varies in the interval [20, 100], and the size of the positive seed set is 50. The experiment results are displayed in Fig. 5.

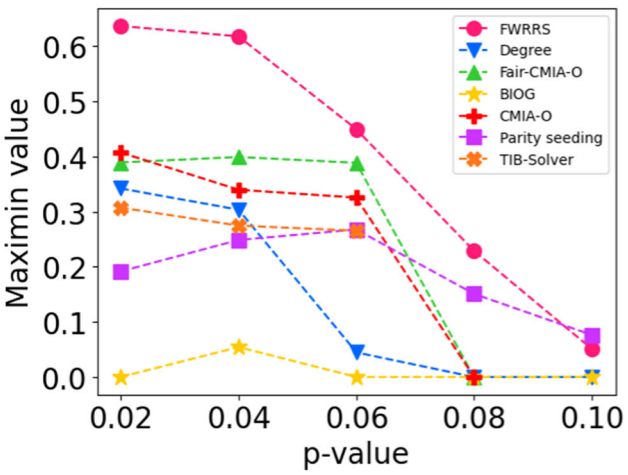
In Fig. 5a, we observe that the fraction of saved nodes is negatively affected with the increase in $|S_N|$, as the methods' effectiveness becomes more limited. However, we observe that this reduction remains quite linear for all methods. In Fig. 5b, we observe that as the number of initially infected nodes increases, the more difficult it becomes for methods to remain relatively fair. The consequential increase in the number of communities being infected challenges the task of saving all communities with a limited budget, especially when the number of rumor nodes is more than the truth-campaigners, and only fairness-aware methods achieve relative fairness.

7.1.2 Impact of influence probability

We further evaluate the proposed method to study the impact of edge influence probabilities. We evaluate all methods under the COICM model with uniform probabilities, where all edge probabilities have been set to a fixed p value. In our experiments, we consider various p values in the interval [0.02, 0.1], and $|S_N| = 50$ and $|S_P| = 50$. The results are shown in Fig. 6 for the PHY Network.



(a) Fraction of saved nodes



(b) Maximin value

Fig. 6 Experiments on PHY under the Uniform model with random S_N

We observe that the FWRRS method saves the maximum number of nodes, and for higher p, the method is more effective.

In this experiment, we observe a similar reduction in fairness, in Fig. 6b, as the one in Fig. 5b. Because the negative information propagates strongly as the p-value grows, the number of communities infected also increases, afflicting the fairness of the methods, which have a limited budget of 50. The increase in p-values also causes the positive information to propagate stronger, leading to an increase in the effectiveness of the methods, observed in Fig. 6a. But despite this effect, the proposed $FWRRS$ method remains fairest and consistently achieves the largest percentage of saved nodes.

7.2 Robustness analysis

7.2.1 Impact of community detection method

To show that the proposed method is not dependent on a specific community detection method, we use different community detection methods to identify communities and evaluate the fairness and effectiveness of the methods. The selection of community detection methods covers different types of methods, including greedy modularity [19], Walktrap [65], Infomap [70], label propagation [68] as well as Louvain [8] community detection methods; these methods are explained in Appendix B. The results on the Physics coauthorship network are shown in Fig. 7.

In Fig. 7, the fraction of nodes saved from infection of each method is barely altered (maximum difference of 0.01) by the change in the community detection method. Slight variations are observed; for example, the fraction of saved nodes of the *Fair-CMIA-O* method drops 1%, compared to the *Louvain* algorithm, under the greedy modularity algorithm and increases by 1% in the label propagation case.

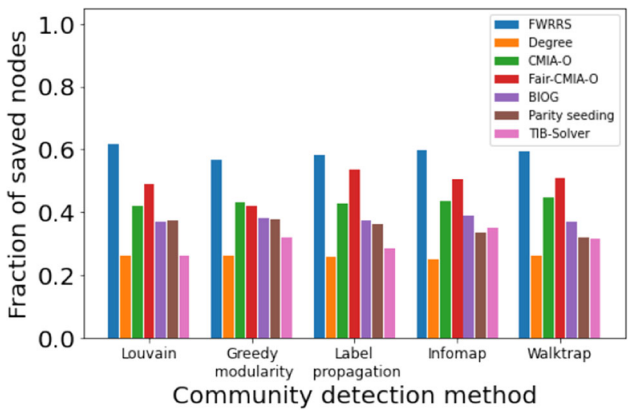


Fig. 7 Experiments on PHY under multiple community detection methods

These variations do not necessarily come from the change in community detection method, as they could be inaccuracies in the measurements caused by the limit of 1000 simulations of the COICM model being used, especially given the small magnitude of this variation. Such variations are also observed in methods that do not take as input the communities, and therefore, shall not be affected by them. Overall, the *FWRRS* method remains the best performing in terms of saved nodes and fairness under all scenarios, and we can conclude that the performance of the proposed method is not dependent on a community detection method.

7.3 Other fairness metrics

We further measure the performance of all methods using other fairness metrics that are defined below.

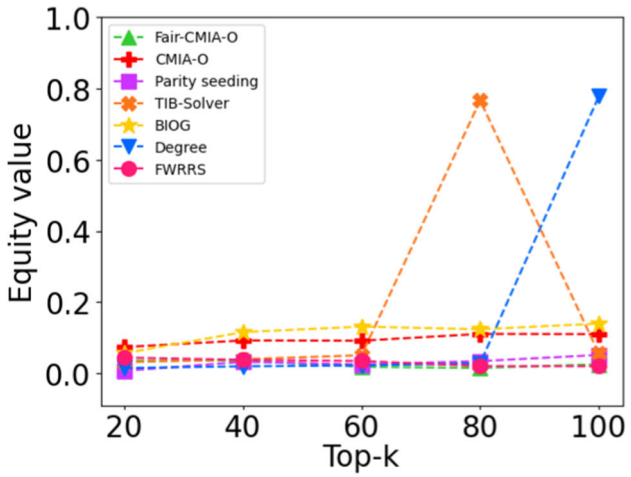
– Equity: The Equity fairness constraint focuses that the number of saved nodes in each community should be proportional to the number of infected nodes [27]. It is computed by taking the variance over the fraction of saved nodes in all communities, defined as,

$$var\left\{\frac{\mathbb{B}_{C_i}(G, S_N, S_P)}{s_{C_i}^{-}(G, S_N, \phi)} : \forall C_i \in C\right\}$$

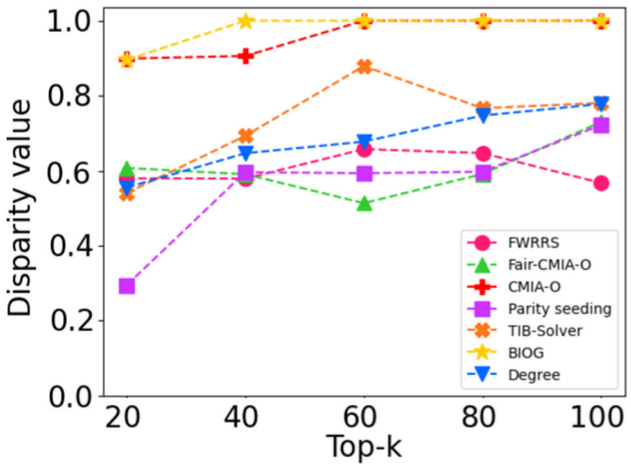
– Disparity [1]: The disparity of a method is computed by the maximum disparity in the normalized saved nodes across all communities. It is defined as,

$$\max_{i,j\in\{1,2,\ldots,|C|\}}\left|\frac{\mathbb{B}_{C_i}(G, S_N, S_P)}{s_{C_i}^{-}(G, S_N, \phi)} - \frac{\mathbb{B}_{C_j}(G, S_N, S_P)}{s_{C_j}^{-}(G, S_N, \phi)}\right|$$

Figure 8 shows Disparity and Equity on the Physics network for $S_N = 50$. In Fig. 8a, we observe that FWRRS performs better for Equity as maximin fairness aims to maximize the saved nodes for the minimally influenced community so that the saved nodes are balanced among all communities. All fairness-oblivious methods (except Degree in some cases) have high Equity value, showing that the mitigation has huge variance between different communities. In Fig. 8b, we observe that fairness-oblivious methods have a higher disparity than fairness-aware methods. The disparity is low for FWRRS, Fair-CMIA-O, and Parity seeding methods, though if we refer to Fig. 3c, the performance for saved nodes is poor for Fair-CMIA-O and Parity seeding methods. Another important thing to note is that the disparity is consistent for our method, and it improves to be best as k increases; the reason is that FWRRS has been optimized for maximin that also reduces disparity over time as we also maximize the fraction of saved nodes.



(a) Equity



(b) Disparity

Fig. 8 Evaluation of different fairness measures on NetPhy

7.4 Analysis on synthetic networks

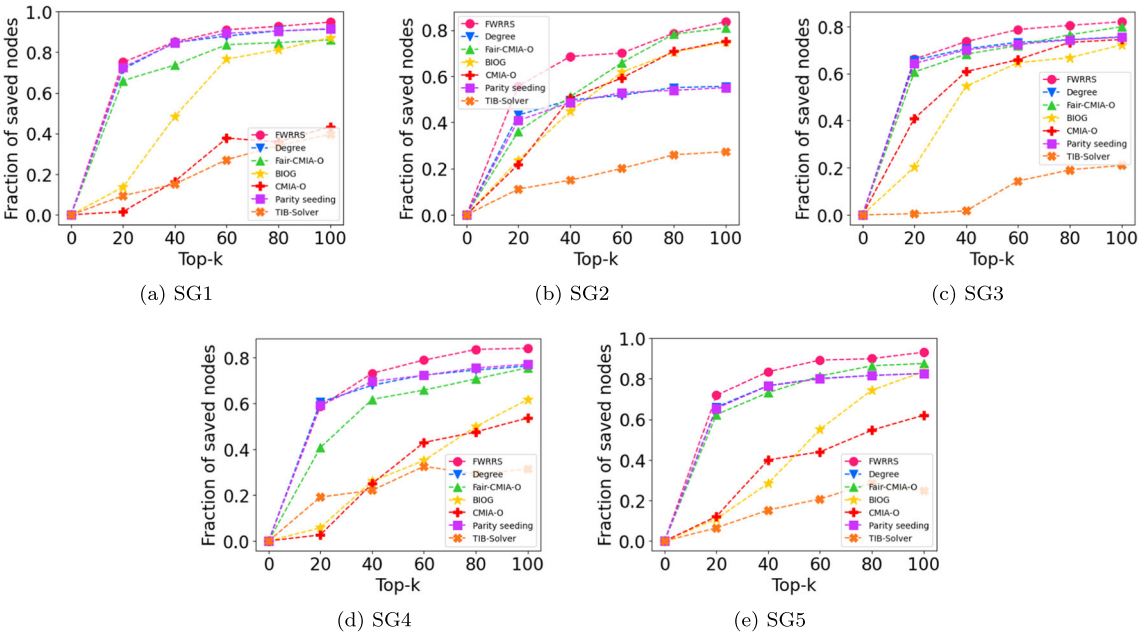
To better understand the performance of the proposed method on different types of networks, we evaluate the performance on synthetic networks. We propose a synthetic network generation model, called HIgh Clustering Homophily Barabási-Albert (HICH-BA) model, that generates scale-free networks better resembling the characteristics of OSNs, having multiple communities and a high clustering coefficient. The model is explained in Appendix C. All generated networks have 10, 000 nodes and 5 communities of diverse sizes.

We generate synthetic networks having 10,000 nodes and five communities of varying sizes, where two communities are of large size and three are of comparatively small size. The parameters used to generate the networks and their characteristics are summarized in Table 2, where the first column is the name assigned to the generated network based on the given setting, n is the number of nodes, m is the number of edges, h is the homophilic index, p_t is the probability of completing a triad edge to increase the clustering coefficient, and p_{PA} is the probability with which new edges are established using preferential attachment instead of at random, that are followed by average degree, assortativity coefficient [20], and clustering coefficient.

Next, we briefly discuss the choices of the selected parameters and results shown in Figs. 9 and 10. The **SG1** network contains all favorable conditions for the *Degree* or *Parity seeding* methods, as was observed in the real datasets experiments that these methods benefit from a high average degree. Further, in SG1, the highest degree is comparatively a very high degree and more powerful in information coverage, and lastly, it has a high clustering coefficient that results in information propagating strongly in the local area of the selected nodes. In these conditions, methods based on estimating influence power using shortest paths, such as *CMIA-O*, underperform, as information spreads through many paths other than the maximum probability ones. Therefore in the SG1 network (refer Fig. 9a), after the *FWRRS* method, the next best-performing methods are *degree* and *Parity seeding*. The proposed method *FWRRS* achieves the largest maximin value as observed in Fig. 10a, and therefore, is the fairest and best-performing method. Further, despite *CMIA-O* performing sub-optimally, its fair alternative performs well. Spreading the information fairly over all communities seems

Table 2 Synthetic datasets characteristics

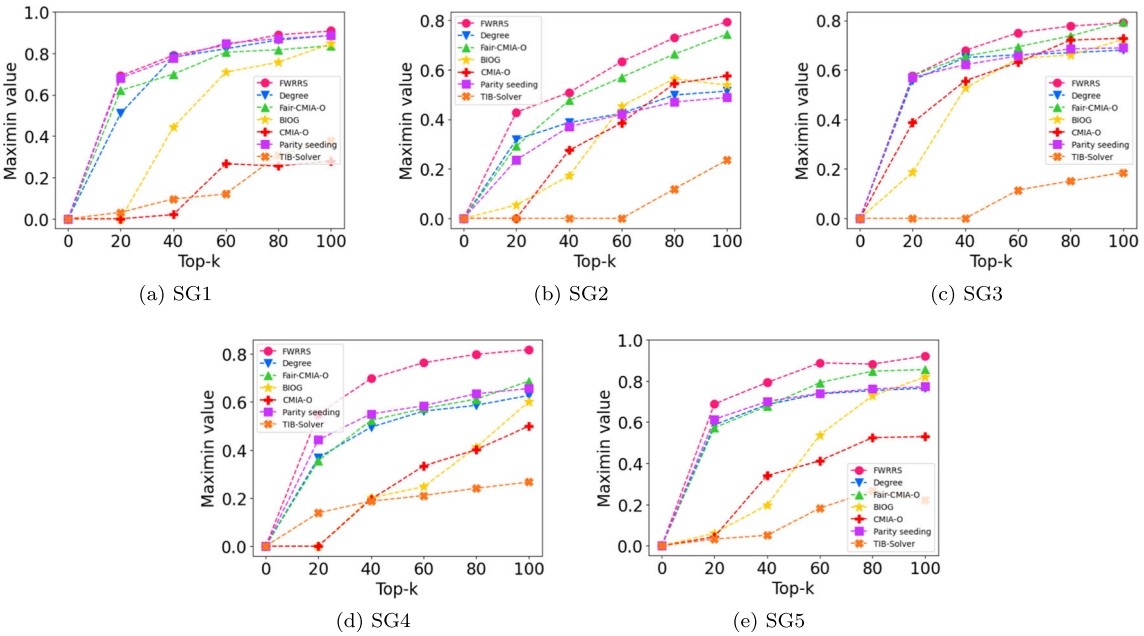
Dataset	p_N	h	p_t	p_{PA}	m	Avg. Degree	Assortativity Coef.	Clustering Coef.
SG1	1/40	0.9	0.9	0.9	101,990	20.398	0.714	0.626
SG2	1/10	0.9	0.9	0.9	26,695	5.339	0.955	0.351
SG3	1/20	0.9	0.9	0.1	48,985	9.797	0.760	0.379
SG4	1/20	0.9	0.2	0.9	48,295	9.659	0.788	0.177
SG5	1/20	0.2	0.9	0.9	44,840	8.968	0.356	0.388



(a) SG1 (b) SG2 (c) SG3

(d) SG4 (e) SG5

Fig. 9 Percentage of saved nodes versus k (the number of chosen truth-campaigners)



(a) SG1 (b) SG2 (c) SG3

(d) SG4 (e) SG5

Fig. 10 Fairness versus k (the number of chosen truth-campaigners)

to cover a larger portion of the propagation paths, leading to better blocking the negative information.

The **SG2** network, on the other hand, favors the conditions where *CMIA-O* and *Fair-CMIA-O* perform better. This is the case when information does not propagate through many paths, achieved with a low average degree. The remaining parameters of the model were left unchanged; however, this reduction in average degree has some indirect impact on the clustering coefficient and homophily. The negative information propagated throughout fewer paths, and thus, blocking the maximum probability paths is very effective. Further, *Degree* and *Parity seeding* perform well for lower budgets, but for higher budgets, they become outperformed by other methods, as the selected nodes do not reach as many nodes as in the previous network due to their lower degree. The change in the methods' performance is also seen on fairness in Fig. 10b. *Fair-CMIA-O* achieves a higher maximin value than in the previous case, and *Degree* and *Parity seeding* are more unfair. The proposed method *FWRRS* shows to be unaltered by this change in the network characteristics and consistently achieves the highest number of saved nodes as well as maximin value, adapting its selection of seeds to this change in the average degree.

Next, the impact of randomness in the graph was evaluated, and for this **SG3** was generated. In SG3, 90% of the edges were established at random while still maintaining the homophilic preference. In Fig. 9, for the SG3 network, we observe that most methods tend to perform similarly, especially for larger values of k. In Fig. 9, it is observed that the randomness in the graph encourages information to spread randomly and reach all communities, eventually leading to achieving comparatively fairer outcomes for all methods.

The fourth synthetic network **SG4** is a graph in which the value of p_t is significantly smaller to evaluate the impact of a lower clustering coefficient. In Fig. 9, for SG4, *Degree* and *Parity seeding* are better for lower k, that is due to the scale-free structure of the network. Since the highest-degree nodes reach to many nodes and avoid many infections. Despite this power of high degree nodes in sparse, scale-free networks with low clustering coefficient, *FWRRS* still performs better than *Degree* and *Parity Seeding*, achieving a portion of saved nodes close to 10% higher. The low clustering coefficient of the graph causes information within communities to spread less powerfully. This means that the positive information that was easily reachable to small size communities now reaches lesser nodes in these groups. For achieving a higher portion of saved nodes in these communities, more effort is required than just reaching a community from highly connected nodes. In such cases, the *FWRRS* method is a lot more fair than other methods as it considers the maximin value explicitly in the objective function and guarantees that the counter true information reaches to all small-size affected communities.

Lastly, we create **SG5** network with low homophily. In this graph, the associativity coefficient has a low value of 0.356. In Fig. 9, one can observe that on this network, methods have very similar effectiveness to the one observed for the SG4 network. Since the number of inter-edges increases in more heterophilic networks, the number of paths between communities also increases. This means information spreading through other paths than the maximin probability ones, leading to the poor effectiveness of *CMIA-O* and *BIOG*, leading too to an unfair outcome. Explicit consideration of the fairness objective is then required, as *FWRRS* and *Fair-CMIA-O* achieve the highest fairness and efficiency for all k.

In summary, in this evaluation, we were able to observe the impact of network characteristics on the fairness and performance of the methods. High average degree and reach of the highest degree nodes are crucial aspects for methods, such as *Degree* and *Parity seeding*, to achieve good results. On the other hand, for the *CMIA-O* method, the opposite is required, as it benefits from fake news spreading through a small number of paths. We can also conclude that the *TIB-Solver* method does not perform as well as claimed in most cases. It only works when information spreads strongly through the network, like under the uniform probabilities for the COICM propagation model as described in Fig. 6. Overall we show the usefulness of the proposed method *FWRRS*, as it consistently achieves a higher percentage of saved nodes as well as the highest fairness. This method achieves such a successful outcome for the intervention while being the fairest in all the considered scenarios.

7.5 Running time analysis

The scalability of the different methods included in our analysis is evaluated on a family of synthetic networks generated using the HICH-BA model. The networks generated increased ten times in size in each iteration, having 1,000, 10,000, 100,000, and 1,000,000 nodes, and average degree 10. All these graphs consisted of five communities and maintained a high clustering coefficient, as $p_{PA} = 0.9$, and a high homophily value, as $h = 0.9$. In the experiment, 50 nodes at random were established as rumor starters, and each method would then select 100 truth campaigners.

The running time for all methods is shown in Fig. 11. We observe that the proposed $FWRRS$ method performs similarly to the $CMIA\text{-}O$ and its fair version, which is four magnitudes faster than the greedy algorithms. It is still slower than $BIOG$ and the heuristic algorithms $degree$ and $parityseeding$. Further, we observe that $FWRRS$

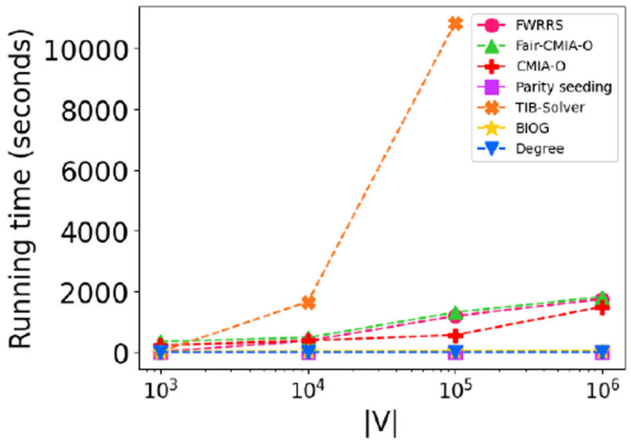


Fig. 11 Scalability of different methods on synthetic networks, where x-axis is the number of nodes and y-axis is running time in seconds

has a sub-linear relationship with the size of the network, similar to $CMIA\text{-}O$ and its fair version [94]. This is because the $FWRRS$ method does not consider all nodes in the network, only those that will be reached by the rumor with an imputed probability. The TIB method is significantly negatively affected by the increase in network size and was not completed on 1M node network within 24 hour time.

8 Conclusion

In this paper, we propose a fairness-aware method to choose truth-campaigners, called FWRRS, to minimize the influence of fake news spreaders in the network. The proposed method first identifies the nodes that are highly likely to get infected by the fake news and then uses weighted reversible reachable trees to identify the most influential truth campaigners by maximizing the fairness-aware optimization function. We compared the performance and fairness of the proposed method with fairness-oblivious and fairness-aware baselines on real-world and synthetic networks using various case scenarios and observed that the proposed method provides state-of-the-art accuracy and the highest fairness values. The experimental analysis showed the unfairness of existing methods as they carry the inequalities of the network structure into the outcome, as well as the performance of these methods was not consistent and strongly depended on the characteristics of the network. However, the proposed method *FWRRS* provides consistent results for graphs having different characteristics and performs best overall in both fairness and total saved nodes. The methods were also evaluated under these controlled networks, showing under which circumstances each model performs best. This evaluation also confirmed the instability of the performance of baseline methods, and showed that the proposed method *FWRRS* consistently performs best, disregarding the network characteristics. The results showed that fairness always does not come at a cost, and in fact, in most cases, it is beneficial for the future effectiveness of the mitigation campaign. We observed that enforcing fairness like in the *FWRRS* method might catalyze efficiency by exploiting the community structure of the social networks.

As per the best of our knowledge, this is the first work that addresses fairness in influence blocking and shall motivate the research in this area, showing its benefits beyond moral aspects. However, there are some limitations of this work, such as it assumes that the network structure is known, and one can further work on designing a method where the network is dynamic or the entire network structure is not available. Besides this, we have studied the problem using the COICM model, where the probabilities of sharing fake and its counter-true information are the same. However, in the real-world this might not be the case, and it will be interesting to study fairness-aware mitigation for different propagation models, such as shortest path model [42], penta-level spreading model [29, 77], trust-based latency-aware independent cascade [59], conformity-aware cascade model [50], continuous-time markov chain- independent cascade model [100], and Linear Threshold model [40]. In real life, some users might have strong beliefs in some opinion, and therefore, will be more biased towards sharing a particular type of information [4]. Fairness-aware fake news mitigation for highly polarized networks is also an interesting open direction. In the future, we would also like to work on algorithmic fairness on time-constrained fake news propagation, where it is crucial to mitigate the impact within the given time limit. One can further aim to propose a feature-blind fair influence blocking method where the whole network structure is not known, or applying the community detection method is an infeasible option.

Appendix

A Fair-CMIA-O method

We extend the CMIA-O [94] method that considers fairness by having $maximin$ as an objective function instead of only maximizing the total number of saved nodes, and it is referred to as Fair-CMIA-O. The remaining computations stayed the same. This method exploits the maximum probability arborescence structure to compute the blocking power of each node per community. It iteratively selects the node that would cause the largest increase in maximin. Once a node is added to S_P, the blocking power per community of all nodes is updated if needed. The process repeats until k nodes have been selected. The pseudo-code of the Fair-CMIA-O is presented in Algorithm 3.

Algorithm 3 Fair-CMIA-O(G(V,E), S_N, k, θ)

$S_P \leftarrow \phi$, $NegS \leftarrow \phi$
$DecInf_c(v) \leftarrow 0$ for each node $v \in V$ and $c \in C$
$ExpInf_c \leftarrow 0$ for each community $c \in C$
for $u \in S_N$ **do**
  construct MIOA(u,θ)
  $NegS \leftarrow NegS \cup (MIOA(u,\theta) \setminus S_N)$
for $u \in NegS$ **do**
  construct MIIA(u,θ)
  compute $ap^N(u, S_N, S_P, MIIA(u,\theta))$
  $ExpInf_c += ap^N(u, S_N, S_P, MIIA(u,\theta))$
  for $v \in MIIA(u,\theta)$ **do**
    compute $ap^N(u, S_N, S_P \cup \{v\}, MIIA(u,\theta))$
    $c \leftarrow comm_u$
    $DecInf_c(v) += ap^N(u, S_N, S_P, MIIA(u,\theta)) - ap^N(u, S_N, S_P \cup \{v\}, MIIA(u,\theta))$
for $i = 1$ to k **do**
  $u \leftarrow \underset{v \in V \setminus (S_P \cup S_N)}{\arg\max} \min_{c \in C} \frac{DecInf_c(v)}{ExpInf_c}$
  construct MIOA(u,θ)
  for $v \in$ MIOA(u,θ) **do**
    $c \leftarrow comm_v$
    for $w \in MIIA(v,\theta)$ **do**
      $DecInf_c(w) -= ap^N(v, S_N, S_P, MIIA(v,\theta)) - ap^N(v, S_N, S_P \cup \{w\}, MIIA(v,\theta))$
  $S_P \leftarrow S_P \cup \{u\}$
  for $v \in$ MIOA(u,θ)$\setminus\{u\}$ **do**
    $c \leftarrow comm_v$
    compute $ap^N(v, S_N, S_P, MIIA(v,\theta))$
    for $w \in MIIA(v,\theta)$ **do**
      compute $ap^N(v, S_N, S_P \cup \{w\}, MIIA(v,\theta))$
      $DecInf(w) += ap^N(v, S_N, S_P, MIIA(v,\theta)) - ap^N(v, S_N, S_P \cup \{w\}, MIIA(v,\theta))$
return S_P

B Community detection methods

We have used the following community detection methods in the analysis.

1. Louvain Community Detection Method [8]: It is a hierarchical clustering algorithm that recursively moves nodes across the communities until the modularity has been maximized.
2. Greedy modularity [19]: This method initially sets each node to belong to a different community and iteratively chooses the pair of nodes that if belong to the same community, cause the largest increases in modularity and join them. This is repeated until the modularity can not be further increased.
3. Walktrap [65]: The fundamental premise of this approach is that the short random walks tend to get trapped in densely connected regions, corresponding to the communities in a network.
4. Infomap [70]: The procedure followed by this method is very similar to that of Louvain, with the difference that the objective function is the map equation instead of modularity. The method considers the probability flow of random walks as a proxy for information flows in a network and uses it further to identify communities.
5. Label propagation [68]: In the label propagation method, each node carries a label that represents the community to which they belong. Throughout the iterations, each node will update its label according to the labels of its neighbor nodes by choosing the label with the highest frequency within its neighbors. The process repeats until no further label changes are made, or the maximum number of iterations is reached.

C HICH-BA synthetic model

We propose the HICH-BA network generator model by extending the homophily BA model [38] using triad formation and randomness to create networks having different properties that iteratively add edges and nodes to the graph until the desired inputted parameters are achieved. The proposed model guarantees the desired community distribution and guarantees that these are strongly connected and controls the clustering coefficient using triad connections.

The $HICH\text{-}BA$ model uses the following parameters, (i) n, i.e., the desired number of nodes, (ii) p_N, i.e., the probability of adding a node to the graph, with probability $1 - p_N$ an edge is added, (iii) r, list where each entry r_i corresponds to the probability of a new node belonging to the community i, (iv) h, i.e., the homophily factor and represents the probability of a node establishing an intra-community connection, (v)p_t, the probability of forming a close triad connection, and (vi) p_{PA}, the probability with which a new edge will be established using the preferential attachment (PA).

The network starts with a seed graph having n_0 nodes as in the other models [38]. At each time step:

a) With probability p_N, a new node v is added to the network. This node then chooses the community $c \in C$ to which it will belong and connect, according to the probabilities in r. The node v will connect to a node $u \in c$, selected using the preferential attachment with probability p_{PA} or at random with probability $1 - p_{PA}$. The probability of selecting such node u is defined as,

$$p_u = \begin{cases} \frac{deg_u}{\sum_{w \in c \setminus \{v\}} deg_w} & \text{with probability } p_{PA} \\ \frac{1}{|c|} & \text{otherwise} \end{cases} \quad (3)$$

b) With probability $1 - p_N$, the network self-grows, and a new edge is added. Under this case, we differ two scenarios: (i) triad connection, and (ii) random connection.

(i) With probability p_t, the new edge will complete a triad connection. We chose a node v_1 u.a.r., and we

make a set of second distance neighbors that are not directly connected with v_1, called $opts$. Now, the triad connection will be formed with a node $v_2 \in opts$ using homophily factor and preferential attachment as follows,

$$p_{v_2} = \begin{cases} \frac{deg_{v_2}}{\sum_{w \in c \cap opts} deg_w} & \text{with probability } h * p_{PA} \\ \frac{1}{|c \cap opts|} & \text{with probability } h * (1 - p_{PA}) \\ \frac{deg_{v_2}}{\sum_{w \in opts \setminus c} deg_w} & \text{with probability } (1-h) * p_{PA} \\ \frac{1}{|opts \setminus c|} & \text{otherwise} \end{cases}$$

(ii) With probability $1 - p_t$, the edge will connect two existing nodes. A node v_1 is chosen u.a.r., and the node v_2 is chosen from $G \setminus \{v_1\}$ using the following probabilities.

$$p_{v_2} = \begin{cases} \frac{deg_{v_2}}{\sum_{w \in c} deg_w} & \text{with probability } h * p_{PA} \\ \frac{1}{|c-1|} & \text{with probability } h * (1 - p_{PA}) \\ \frac{deg_{v_2}}{\sum_{w \in G \setminus c} deg_w} & \text{with probability } (1-h) * p_{PA} \\ \frac{1}{|G \setminus c|} & \text{otherwise} \end{cases}$$

The proposed model can create homophilic scale-free and homophilic random networks, as well as networks having higher and lower clustering coefficients by controlling triad formation links.

Declarations

Conflicts of interest The authors declare that there is not any conflict of interest in this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Ali J, Babaei M, Chakraborty A, Mirzasoleiman B, Gummadi K, Singla A (2021) On the Fairness of Time-Critical Influence Maximization in Social Networks. IEEE Trans Know Data Eng. https://doi.org/10.1109/TKDE.2021.3120561
2. Allcott H, Gentzkow M, Yu C (2019) Trends in the diffusion of misinformation on social media. Res Polit 6(2):2053168019848554
3. Arazkhani N, Meybodi MR, Rezvanian A (2019) Influence Blocking Maximization in Social Network Using Centrality Measures. 2019 IEEE 5th Conference on Knowledge Based Engineering and Innovation, KBEI. 2019 p 492–497. https://doi.org/10.1109/KBEI.2019.8734920
4. Azzimonti M, Fernandes M (2023) Social media networks, fake news, and polarization. Eur J Political Econ 76:102256
5. Barocas S, Selbst AD (2018) Big Data's Disparate Impact. SSRN Electron J 671:671–732. https://doi.org/10.2139/ssrn.2477899
6. Bermes A (2021) Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during covid-19. J Retail Consum Serv 61:102555
7. Bierbooms CG (2022) Fairness-aware influence blocking maximization for combating fake news
8. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):1–12. https://doi.org/10.1088/1742-5468/2008/10/P10008
9. Boididou C, Papadopoulos S, Zampoglou M, Apostolidis L, Papadopoulou O, Kompatsiaris Y (2018) Detection and visualization of misleading content on twitter. Int J Multimed Inf Retr 7(1):71–86
10. Bovet A, Makse HA (2019) Influence of fake news in twitter during the 2016 us presidential election. Nat Commun 10(1):7
11. Brandes U (2010) The Journal of Mathematical A faster algorithm for betweenness centrality. J Math Sociol 2012:37–41
12. Budak C, Agrawal D, Abbadi AE (2011) Limiting the spread of misinformation in social networks. Proceedings of the 20th International Conference on World Wide Web, WWW 2011 pp. 665–674. https://doi.org/10.1145/1963405.1963499
13. Capuano N, Fenza G, Loia V, Nota FD (2023) Content based fake news detection with machine and deep learning: a systematic review. Neurocomputing
14. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web. p 675–684
15. Chen BL, Jiang WX, Yu YT, Zhou L, Tessone CJ (2022) Graph embedding based ant colony optimization for negative influence propagation suppression under cost constraints. Swarm Evol. Comput 72:101102
16. Chen W Accessed on 20-Apr-2023 http://research.microsoft.com/enus/people/weic/graphdata.zip
17. Chien SY, Yang CJ, Yu F (2022) Xflag: Explainable fake news detection model on social media. Int. J. Hum.-Comput. Interact. 38(18–20):1808–1827
18. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F, Scala A (2020) The covid-19 social media infodemic. Sci Rep 10(1):1–10
19. Clauset A, Newman ME, Moore C (2004) Finding community structure in very large networks. Phys Rev E - Stat Phys Plasm Fluids Rel Interdisc Topics 70(6):6. https://doi.org/10.1103/PhysRevE.70.066111
20. Coleman JS (1964) Introduction to mathematical sociology. Free Press of Glencoe, New York
21. David F, Guimarães N, Figueira Á (2023) A webapp for reliability detection in social media. Procedia Comput. Sci. 219:228–235
22. Dong Y, Ma J, Wang S, Chen C, Li J (2023) Fairness in graph mining: A survey. IEEE Trans Knowl Data Eng
23. Dou Y, Shu K, Xia C, Yu PS, Sun L(2021) User preference-aware fake news detection. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. p 2051–2055

24. El Halabi M, Mitrović S, Norouzi-Fard A, Tardos J, Tarnawski JM (2020) Fairness in streaming submodular maximization: Algorithms and hardness. Adv. Neural Inf. Process 33:13609–13622
25. Erd FC, Vignatti AL, Silva MV (2021) The generalized influence blocking maximization problem. Soc Netw Anal Min 11(1):1–17
26. Fagiolo G (2007) Clustering in complex directed networks. Phys Rev E - Stat Nonlinear Soft Matter Phys 76(2):1–8. https://doi.org/10.1103/PhysRevE.76.026107
27. Farnad G, Babaki B, Gendreau M (2020) A Unifying Framework for Fairness-Aware Influence Maximization. The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020 (1), 714–722. https://doi.org/10.1145/3366424.3383555
28. Gajane P, Pechenizkiy M (2017) On formalizing fairness in prediction with machine learning. arXiv:1710.03184
29. Gupta Y, Saxena A, Das D, Iyengar S (2016) Modeling memetics using edge diversity. In: Complex networks VII, p 187–198. Springer
30. He Q, Lv Y, Wang X, Huang M, Cai Y (2022) Reinforcement learning-based rumor blocking approach in directed social networks. IEEE Syst J 16(4):6457–6467
31. He Q, Zhang D, Wang X, Ma L, Zhao Y, Gao F, Huang M(2022) Graph convolutional network-based rumor blocking on social networks. IEEE Trans Comput Soc Syst
32. He X, Song G, Chen W, Jiang Q(2012) Influence blocking maximization in social networks under the competitive linear threshold model. Proceedings of the 12th SIAM International Conference on Data Mining, SDM 2012 (October). 463–474. https://doi.org/10.1137/1.9781611972825.40
33. Horne BD, Adali S(2017) This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Eleventh International AAAI Conference on Web and Social Media
34. Hu L, Wei S, Zhao Z, Wu B(2022) Deep learning for fake news detection: A comprehensive survey. AI Open
35. Jiang J, Chen X, Huang Z, Li X, Du Y(2023) Deep reinforcement learning-based approach for rumor influence minimization in social networks. Appl Intell 1–18
36. Jin Z, Cao J, Guo H, Zhang Y, Luo J(2017) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on Multimedia. p 795–816
37. Kapadia P, Saxena A, Das B, Pei Y, Pechenizkiy M(2023) Co-attention based multi-contextual fake news detection. In: Complex Networks XIII: Proceedings of the 13th Conference on Complex Networks, CompleNet 2022. Springer, p 83–95
38. Karimi F, Génois M, Wagner C, Singer P, Strohmaier M (2018) Homophily influences ranking of minorities in social networks. Sci Rep 8(1):11077
39. Kempe D, Kleinberg J(2003) Maximizing the spread of influence through a social network. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p 137–146
40. Kempe D, Kleinberg J, Tardos É(2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, p 137–146
41. Khajehnejad A, Khajehnejad M, Babaei M, Gummadi KP, Weller A, Mirzasoleiman B (2022) Crosswalk: Fairness-enhanced node representation learning. Proceedings of the AAAI Conference on Artificial Intelligence 36:11963–11970
42. Kimura M, Saito K(2006) Tractable models for information diffusion in social networks. In: Knowledge Discovery in Databases: PKDD 2006: 10th European Conference on Principles and Practice of Knowledge Discovery in Databases Berlin, Germany, September 18-22, 2006 Proceedings 10. Springer, p 259–271
43. Klimt B, Yang Y(2004) Introducing the Enron Corpus. Machine Learning
44. Kumar P, Devi PR, Sai NR, Kumar SS, Benarji T (2021) Battling fake news: A survey on mitigation techniques and identification. In: 2021 5th international conference on trends in electronics and informatics (ICOEI). IEEE, p 829–835
45. Kwan R(2023) Breaking news: Social media changed the way we consume news. Accessed on 20-Apr-2023
46. Lee CE, Kent JL(2023) Facebook says russian-backed election content reached 126 million americans. Accessed on 20-Apr-2023
47. Lee E, Karimi F, Wagner C, Jo HH, Strohmaier M, Galesic M (2019) Homophily and minority-group size explain perception biases in social networks. Nat Hum Behav 3(10):1078–1087
48. Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. ACM Trans Know Disc Data 1(1):2–es
49. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction: Continued influence and successful debiasing. Psychol Sci Public Int 13(3):106–131
50. Li H, Bhowmick SS, Sun A, Cui J (2015) Conformity-aware influence maximization in online social networks. VLDB Journal 24:117–141
51. Li Y, Wang X, Ning Y, Wang H (2022) Fairlp: Towards fair link prediction on social network graphs. Proceedings of the International AAAI Conference on Web and Social Media 16:628–639
52. Lin KS, Dai BR (2019) BIOG: An Effective and efficient algorithm for influence blocking maximization in social networks, vol 1071. Springer, Singapore
53. Lu YJ, Li CT(2020) Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, p 505–514
54. Lv J, Yang B, Yang Z, Zhang W (2019) A community-based algorithm for influence blocking maximization in social networks. Cluster Comput 22(s3):5587–5602
55. Masrour F, Wilson T, Yan H, Tan PN, Esfahanian AH(2020) Bursting the filter bubble: Fairness-aware network link prediction. AAAI 2020 - 34th AAAI Conference on Artificial Intelligence. p 841–848. https://doi.org/10.1609/aaai.v34i01.5429
56. Mcauley J, Leskovec J(2012) Learning to Discover Social Circles in Ego Networks. NIPS, p 1–9
57. Mendoza M, Poblete B, Castillo C(2010) Twitter under crisis: Can we trust what we rt? In: Proceedings of the first workshop on social media analytics. p 71–79
58. Mishima K, Yamana H (2022) A survey on explainable fake news detection. IEICE Trans Inf Syst 105(7):1249–1257
59. Mohamadi-Baghmolaei R, Mozafari N, Hamzeh A (2015) Trust based latency aware influence maximization in social networks. Eng. Appl. Artif. Intell 41:195–206
60. Nguyen HT, Thai MT, Dinh TN(2016) Stop-and-Stare: Optimal sampling algorithms for viral marketing in billion-scale networks. Proc. ACM SIGMOD Int. Conf. Manag Data 26-June-20, 695–710. https://doi.org/10.1145/2882903.2915207
61. Nisbet EC, Beck P, Gunther R(2018) "Trump May Owe His 2016 Victory to 'Fake News', New Study Suggests,". The Conversation
62. O'Brien N, Latessa S, Evangelopoulos G, Boix X(2018) The language of fake news: Opening the black-box of deep learning based detectors
63. Pennycook G, Rand DG(2018) Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. SSRN Electron J
64. Piraveenan M, Prokopenko M, Hossain L(2013) Percolation Centrality: Quantifying Graph-Theoretic Impact of Nodes during Percolation in Networks. PLoS ONE 8(1) https://doi.org/10.1371/journal.pone.0053095

65. Pons P, Latapy M (2006) Computing communities in large networks using random walks. J. Graph Algorithms Appl. 10(2):191–218. https://doi.org/10.7155/jgaa.00124
66. Potthast M, Kiesel J, Reinartz K, Bevendorff J, Stein B(2018) A stylometric inquiry into hyperpartisan and fake news. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). p 231–240
67. Qi P, Cao J, Yang T, Guo J, Li J(2019) Exploiting multi-domain visual information for fake news detection. In: 2019 IEEE International Conference on Data Mining (ICDM). IEEE , p 518–527
68. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E - Stat Nonlin Soft Matter Phys 76(3):1–12. https://doi.org/10.1103/PhysRevE.76.036106
69. Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD(2021) The impact of fake news on social media and its influence on health during the covid-19 pandemic: A systematic review. J Public Health 1–10
70. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 105(4):1118–1123. https://doi.org/10.1073/pnas.0706851105
71. Rubin VL, Conroy NJ, Chen Y(2015) Towards news verification: Deception detection methods for news discourse. In: Hawaii International Conference on System Sciences. p 5–8
72. Saxena A, Fletcher G, Pechenizkiy M (2021) Hm-eiict: Fairness-aware link prediction in complex networks using community information. J. Comb. Optim 1–18
73. Saxena A, Fletcher G, Pechenizkiy M(2022) Fairsna: Algorithmic fairness in social network analysis.arXiv:2209.01678
74. Saxena A, Fletcher G, Pechenizkiy M (2022) Nodesim: node similarity based network embedding for diverse link prediction. EPJ Data Sci 11(1):24
75. Saxena A, Hsu W, Lee ML, Leong Chieu H, Ng L, Teow LN (2020) Mitigating misinformation in online social network with top-k debunkers and evolving user opinions. Companion proceedings of the web conference 2020:363–370
76. Saxena A, Iyengar S(2020) Centrality measures in complex networks: A survey. arXiv:2011.07190
77. Saxena A, Iyengar S, Gupta Y(2015) Understanding spreading patterns on social networks based on network topology. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, p 1616–1617
78. Saxena A, Saxena H, Gera R(2020) k-truthscore: Fake news mitigation in the presence of strong user bias. In: International Conference on Computational Data and Social Networks. Springer, p 113–126
79. Saxena A, Saxena P, Reddy H(2022) Fake news detection techniques for social media. In: Principles of Social Networking. Springer, p 325–354
80. Saxena A, Saxena P, Reddy H(2022) Fake news propagation and mitigation techniques: A survey. In: Principles of Social Networking. Springer, p 355–386
81. Seifert CM(2002) The continued influence of misinformation in memory: What makes a correction effective? In: Psychology of learning and motivation, vol. 41. Elsevier, p 265–292
82. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: A survey on identification and mitigation techniques. ACM Trans. Intell. Syst. Technol. 10(3):1–42
83. Shu K, Cui L, Wang S, Lee D, Liu H(2019) defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. p 395–405
84. Shu K, Wang S, Liu H(2019) Beyond news contents: The role of social context for fake news detection. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. p 312–320
85. Song, C., Hsu, W., Lee, M.L(2017) Temporal influence blocking: Minimizing the effect of misinformation in social networks. Proceedings - International Conference on Data Engineering p 847–858 https://doi.org/10.1109/ICDE.2017.134
86. Stoica AA, Chaintreau A (2019) Fairness in social influence maximization. The Web Conference 2019 - Companion of the World Wide Web Conference. WWW 2019:569–574. https://doi.org/10.1145/3308560.3317588
87. Stoica AA, Han JX, Chaintreau A(2020) Seeding Network Influence in Biased Networks and the Benefits of Diversity. The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020 2, 2089–2098. https://doi.org/10.1145/3366423.3380275
88. Tacchini E, Ballarin G, Della Vedova ML, Moret S, de Alfaro L, et al (2017) Some like it hoax: Automated fake news detection in social networks. In: CEUR Workshop Proceedings. CEUR-WS, p 1–15
89. Tsang A, Wilder B, Rice E, Tambe M, Zick Y(2019) Group-fairness in influence maximization. IJCAI International Joint Conference on Artificial Intelligence. p 5997–6005
90. Tsioutsiouliklis S, Pitoura E, Tsaparas P, Kleftakis I, Mamoulis N (2021) Fairness-aware pagerank. Proceedings of the Web Conference 2021:3815–3826
91. Vosoughi S, Mohsenvand M, Roy D (2017) Rumor gauge: predicting the veracity of rumors on twitter. ACM Trans Knowl Discov 11(4):50
92. Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151
93. Wang X, Deng K, Li J, Yu JX, Jensen CS, Yang X (2018) Targeted influence minimization in social networks., vol. 23. Springer International Publishing
94. Wu P, Pan L (2017) Scalable influence blocking maximization in social networks under competitive independent cascade models. Comput Netw 123:38–50
95. Yao C, Zhang Y, Zhang X, Bian K, Song L(2019) Competitive influence blocking in online social networks: A case study on WeChat. 2018 24th Asia-Pacific Conference on Communications, APCC 2018 pp. 251–256. https://doi.org/10.1109/APCC.2018.8633553
96. Zafarani R, Abbasi MA, Liu H (2014) Social media mining: an introduction. Cambridge University Press
97. Zareie A, Sakellariou R (2022) Rumour spread minimization in social networks: A source-ignorant approach. Online Soc Netw Media 29:100206
98. Zhang P, Bao Z, Niu Y, Zhang Y, Mo S, Geng F, Peng Z (2019) Proactive rumor control in online networks. World Wide Web 22:1799–1818
99. Zhang Y, Xu J, Nekovee M, Li Z (2022) The impact of official rumor-refutation information on the dynamics of rumor spread. Physica A: Stat Mech Appl 607:128096
100. Zhu T, Wang B, Wu B, Zhu C (2014) Maximizing the spread of influence ranking in social networks. Inf Sci 278:535–544

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.