

Harnessing Network Effect for Fake News Mitigation: Selecting Debunkers via Self-Imitation Learning

Xiaofei Xu, Ke Deng, Michael Dann, Xiuzhen Zhang

School of Computing Technologies, RMIT University, Melbourne, Victoria 3000 Australia
xiaofei.xu2@student.rmit.edu.au, {ke.deng, michael.dann, xiuzhen.zhang}@rmit.edu.au

Abstract

This study aims to minimize the influence of fake news on social networks by deploying *debunkers* to propagate true news. This is framed as a reinforcement learning problem, where, at each stage, one user is selected to propagate true news. A challenging issue is *episodic reward* where the “net” effect of selecting individual debunkers cannot be discerned from the interleaving information propagation on social networks, and only the collective effect from mitigation efforts can be observed. Existing Self-Imitation Learning (SIL) methods have shown promise in learning from episodic rewards, but are ill-suited to the real-world application of fake news mitigation because of their poor sample efficiency. To learn a more effective debunker selection policy for fake news mitigation, this study proposes NAGASIL – Negative sampling and state Augmented Generative Adversarial Self-Imitation Learning, which consists of two improvements geared towards fake news mitigation: learning from negative samples, and an augmented state representation to capture the “real” environment state by integrating the current observed state with the previous state-action pairs from the same campaign. Experiments on two social networks show that NAGASIL yields superior performance to standard GASIL and state-of-the-art fake news mitigation models.

1 Introduction

There have been significant efforts to combat the spread of fake news on social networks. Beyond fake news detection, another important strategy is mitigation, whereby debunkers – users who propagate true news – are deployed to counter the spread of fake news. Unfortunately, despite the efforts of “official” debunkers, such as fact-checking services and authoritative organisations (e.g., WHO), fake news still proliferates widely on social media. It is therefore important to unleash the power of crowd debunking from online users for fake news mitigation (Vo and Lee 2018, 2020). Previous work on this problem has explored using reinforcement learning to optimize fake news mitigation campaigns. Some studies focus on optimizing the intensity with which given debunkers spread true news (Farajtabar et al. 2017; Goindani and Neville 2020a,b). Other studies optimize the selection of crowd debunkers from online users over multiple stages with a budget constraint (Xu, Deng, and Zhang 2022).

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

A crucial issue in reinforcement learning-based multi-stage fake news mitigation strategies but one that is largely overlooked in the literature (Farajtabar et al. 2017; Goindani and Neville 2020a,b; Xu, Deng, and Zhang 2022) is the lack of direct, “net” rewards for agent actions. Since news propagation on social networks is a long-lasting process with intensity decay over time, the effects of debunkers on fake news mitigation are interleaved across the network at different stages. The “net” effect of individual debunkers cannot be directly measured at intermediate stages; only the cumulative effect of all debunkers can be observed when the mitigation campaign finishes. In other words, the reward function for the task is episodic.

In this paper, we propose a reinforcement learning approach to multi-stage fake news mitigation that explicitly addresses the issue of episodic reward. In a multi-stage fake news mitigation campaign, one user is selected at each stage to post true news such that the number of users believing in fake news is minimized at the conclusion of the campaign. To address the issue of episodic rewards, we propose to learn the debunker selection policy via self-imitation learning (Oh et al. 2018; Gangwani, Liu, and Peng 2018). Instead of learning directly from immediate rewards, self-imitation learning aims to mimic the agent’s own past behaviour from highly rewarding episodes. Existing self-imitation learning algorithms (Ho and Ermon 2016; Gangwani, Liu, and Peng 2018; Oh et al. 2018) have shown promise in learning from episodic rewards (Gangwani, Liu, and Peng 2018), but have two weaknesses that potentially limit their effectiveness for fake news mitigation on social networks:

First, fake news mitigation campaigns are cost-intensive (Farajtabar et al. 2017). Real-world campaigns often have budget constraints and it is essential to learn mitigation policies from a small number of episodes. It is therefore desirable to exploit all available information about episodes for more efficient sampling and hence more efficient learning. Existing self-imitation learning methods imitate past *good* experiences only, and state-action pairs that appear frequently in past good experiences are typically treated as favourable, but this may not be true if they also appear frequently in past bad experiences. Second, existing methods assume that the full environment state can be observed. However, in fake news mitigation, due to interleaved information propagation and the complexity of social networks,

it is hard to observe users' reactions to news in real time.

In this paper, we argue that leveraging negative samples – past experiences where the agent received a small episodic reward – can yield more efficient sampling of good experiences and hence boost policy learning efficiency. Negative sampling informs the agent about undesirable behaviours and can therefore improve the efficiency of sampling good experiences. In addition, we devise an augmented state representation that better captures the “true” environment state by integrating the current observations with previous state-action pairs in the same campaign.

The contributions of this study are twofold: First, we propose, for the first time to our best knowledge, a multi-stage fake news mitigation approach designed for the realistic setting where rewards are episodic. Second, we propose NAGASIL – Negative sampling and state Augmented Generative Adversarial Self-Imitation Learning – which improves self-imitation learning for fake news mitigation via negative samples and augmented states. In addition to theoretically proving the advantage of negative samples and augmented states, we test NAGASIL on two social networks: one is a large synthetic network based on commonly used social network parameters, while the other is a widely used real-world social network for rumour propagation. Experiments demonstrate NAGASIL's superior performance compared to state-of-the-art fake news mitigation models¹

2 Related Work

There have been many studies on fake news detection on social networks. To reduce the cost and time burden of manual fact-checking, automated fact-checking of fake news and credibility analysis of social media posts have been proposed, using information such as network features (Benamira et al. 2019), multi-modal features (Wang et al. 2018) and combined features (Shu et al. 2017; Shu, Bernard, and Liu 2019). Other studies detect fake news and their spreaders on social networks based on linguistic and personality features (Tian et al. 2020; Tian, Zhang, and Lau 2022; Shrestha, Spezzano, and Joy 2020). For a more complete review of automatic fact-checking systems, please refer to a recent survey (Guo, Schlichtkrull, and Vlachos 2022).

Beyond fake news detection, research on strategies for propagating true news, such as fact-checked content, to mitigate the spread of fake news is attracting more attention. Various solutions have been investigated to select users as debunkers to propagate true news. These studies can be categorised along two lines: One line of research focuses on heuristics for one-off selection of debunkers for mitigation. These studies heuristically select the top- k most influential users as debunkers (Saxena et al. 2020; Saxena, Saxena, and Gera 2020). They assume that users with high social influence will produce wide propagation of true news on social networks. However, research has shown that overall influence on social networks may not translate to wide mitigation propagation and reach users exposed to fake news as expected (Farajtabar et al. 2016).

Another line of research, which most strongly relates to ours, leverages reinforcement learning to optimize the cumulative effect across multiple stages of fake news mitigation (Farajtabar et al. 2017; Goindani and Neville 2020a,b; Xu, Deng, and Zhang 2022). In some studies (Farajtabar et al. 2017; Goindani and Neville 2020a,b), a set of debunkers is given in advance, and each stage focuses on optimizing the intensity with which the debunkers post true news. Because the debunkers are fixed, even if they post with high intensity, the true news they propagate may not reach all users exposed to fake news, given the unknown origin and dynamic propagation of fake news on social networks (Xu, Deng, and Zhang 2022). Not assuming fixed debunkers, Xu, Deng, and Zhang (2022) propose to select debunkers in a cost-effective way for multi-stage fake news mitigation. However, all of these studies assume that the mitigation effect of each stage can be immediately observed before the next stage, and overlook the issue of episodic reward that we described in the introduction, which can lead to less than optimal mitigation policies. In this paper, we address the critical issue of episodic reward, and our solution can be generally applied to other mitigation settings of selecting debunkers (Xu, Deng, and Zhang 2022).

In addition to network-level mitigation, there are also studies on individual-level fake news mitigation (Wang et al. 2022; He, Ahamad, and Kumar 2023). Wang et al. (2022) propose a personalised true news recommender system to counteract fake news by modelling both user interests and news veracity for individuals. With the recent advance of generative models (Ouyang et al. 2022), a recent study focuses on generating personalized counter-misinformation posts (He, Ahamad, and Kumar 2023).

3 Preliminaries and Problem Statement

Information Propagation Models. Various information propagation models have been applied in research on fake news mitigation on social networks, including Hawkes processes (Farajtabar et al. 2017; Lacombe 2018; Shu, Bernard, and Liu 2019; Goindani and Neville 2020a,b; Murayama et al. 2021), Linear Threshold / Independent Cascade models (Pham et al. 2019; Saxena, Saxena, and Gera 2020), Information Aggregation Games (Aymanns et al. 2020), and epidemic models (Zhao et al. 2013; Wen et al. 2014b,a; Tan et al. 2019). Our proposed NAGASIL is independent of the propagation model. Without loss of generality, the epidemic model is employed as the environment for reinforcement learning in this study.

To meet the specific requirements of different applications, several variants of the epidemic model have been proposed (Brauer 2008; Gumel et al. 2004). In this study, we adopt SEIR (Susceptible-Exposed-Infected-Recovered). Let $X_i(t)$ be the *epidemic state* (or *e-state* for simplicity) of user i at time t . $X_i(t)$ is always in one of four e-states: *Susceptible*, *Exposed*, *Recovered*, or *Infected*. $X_i(t)$ can transition from one e-state to another. If the user has not received any news yet, $X_i(t)$ is *Susceptible*; if they have received fake/true news, $X_i(t)$ is *Exposed*. Let P_i^I (P_i^R) be the probability of user i transitioning from some other e-state to *Infected* (*Recovered*).

¹The source code for our experiments is available at <https://github.com/xxfwin/NAGASIL>

$$P_i^I = \begin{cases} L_i(N_i^F - N_i^M) & \text{if } N_i^F > N_i^M \\ 0 & \text{else} \end{cases} \quad (1)$$

$$P_i^R = \begin{cases} L_i(N_i^M - N_i^F) & \text{if } N_i^M > N_i^F \\ 0 & \text{else} \end{cases} \quad (2)$$

where N_i^F and N_i^M are the number of fake and true news items received by user i respectively, and $L_i(x)$ is defined as:

$$L_i(x) = \frac{1}{1 + e^{-\delta(x-x_i)}} \quad (3)$$

where x_i is the logistic function's midpoint, which is determined by the number of followers user i has on the social network, δ is the logistic growth rate and $\delta = 1$ by default. If a user has received more true news than fake news (i.e., $N_i^M > N_i^F$) and is not in e-state *Recovered*, they will have a non-zero probability of transitioning to e-state *Recovered*; otherwise, if they have received more fake news than true news (i.e., $N_i^M < N_i^F$) and are not in e-state *Infected*, they will have a non-zero probability of transitioning to e-state *Infected*. In both cases, the probability is lower if user i has more followers. The rationale is that users with more followers tend to be harder to convince by news received (Smit et al. 2022).

Once a user transitions to e-state *Infected* (or *Recovered*), they have an initial intensity (i.e., probability) for spreading fake news (or true news). The intensity decays over time. Specifically, the intensity of user i at time t is:

$$\iota_i(t) = \xi_i e^{-\omega(t-t_c)} \quad (4)$$

where ξ_i is the initial intensity, ω controls the intensity decay rate ($\omega = 1$ by default), and t_c is the time when the e-state of user i changed to *Infected* (or *Recovered*). The more time elapsed, the less likely the user is to spread fake/true news.

Note that the users in the information propagation model continuously spread fake/true news based on their intensities. This leads to two consequences: First, the probability of each user transitioning to e-state *Infected* (or *Recovered*) changes over time, since it is determined by the number of fake/true news items received (Eq. 1 and 2). Second, the intensity of users spreading fake/true news changes continuously so it is difficult to observe intrinsic properties that drive users to interact with the environment (e.g., initial intensity).

Problem Statement. A social network is modelled as a directed graph $G(U, E)$ where U and E denote the social network users and the directed links between users, respectively. The information propagation on social networks is modelled by SEIR. Initially, the environment state is s_0 , where the e-state of each user in U is *Susceptible*, *Exposed* or *Infected*, and the users in e-state *Infected* have different intensities to spread fake news. B is a budget and each user i has a cost c_i . If user i is selected as a debunker, the budget is reduced by c_i . Given a social network $G(U, E)$ with initial state s_0 , the problem of selecting debunkers under budget B for optimal mitigation can be mapped to a reinforcement learning problem, where the goal is to design a debunker selection policy such that, each w time steps (i.e., a stage), the

agent selects one user from U as a new debunker based on the environment state at that time (s_0 for the first stage), and sets the user's e-state to *Recovered* with the initial intensity to spread true news on G , until the remaining budget is less than the cost of any user in U . The optimization objective is to minimize the number of users in e-state *Infected* in U at time t_f ($\gg w$) after the multi-stage fake news mitigation campaign concludes.

Since it is impractical to observe the full environment state in a realistic setting, we mask certain state features. The rationale is that we cannot observe the intrinsic properties that drive users to interact with the environment. The full environment state s_E and the observed environment state s are defined as follows:

$$s_E = [\mathbf{P}^I; \mathbf{r}^I; \mathbf{d}^I; \mathbf{P}^R; \mathbf{r}^R; \mathbf{d}^R; \boldsymbol{\iota}; \mathbf{e}], \quad (5)$$

$$s = [\mathbf{r}^I; \mathbf{d}^I; \mathbf{r}^R; \mathbf{d}^R; \mathbf{e}].$$

where $\mathbf{P}^I \in [0, 1]^n$ ($\mathbf{P}^R \in [0, 1]^n$) is a vector where each element indicates the probability of a user changing e-state to *Infected* (*Recovered*), $\mathbf{r}^I \in \mathbb{Z}_2^n$ ($\mathbf{r}^R \in \mathbb{Z}_2^n$) is a vector indicating for each of the $n = |U|$ users whether they are in e-state *Infected* (*Recovered*) or not, $\mathbf{d}^I \in \mathbb{R}^n$ ($\mathbf{d}^R \in \mathbb{R}^n$) is a vector indicating the number of times each user had propagated fake news (true news) since the beginning of fake news mitigation, $\boldsymbol{\iota} \in \mathbb{R}^n$ is a vector indicating the intensity of each user, and $\mathbf{e} \in \mathbb{R}^n$ is a vector indicating the number of followers of each user. The action, a , to be chosen at each stage is which user $i \in U$ is to be selected as the debunker.

To learn an optimal fake news mitigation policy under this reinforcement learning framework, we face a key challenge that the reward for selecting debunkers is not immediately available. Due to the continual and interleaving nature of information propagation on social networks, the net effect of selecting a debunker to post true news on the network is not directly and immediately observable; instead only the *network effect* can be seen at the end of a campaign. Thus, we define the episodic reward of an episode τ as:

$$V(\tau) = -\log\left(\frac{\|\mathbf{r}_\tau^I\|_1}{n}\right). \quad (6)$$

where $\mathbf{r}_\tau^I \in \mathbb{Z}_2^n$ is a vector indicating whether each of the n users is in e-state *Infected* after the fake news mitigation finishes. There are two possible methods to decide whether a user is *Infected*: First, if a user shares a piece of fake news, the user is *Infected*. Second, a user is *Infected* based on the probability provided by Eq 1. In this paper, we use the second method. The episodic reward is measured at time t_f after the multi-stage fake news mitigation ends.

4 Methodology

Under the reinforcement learning framework, multi-stage fake news mitigation campaigns have episodic rewards. One approach to addressing episodic rewards is self-imitation learning (Oh et al. 2018; Guo et al. 2018). These methods imitate the past good behaviours of the learner itself, using the signal provided by the episodic reward, and can learn strong policies.

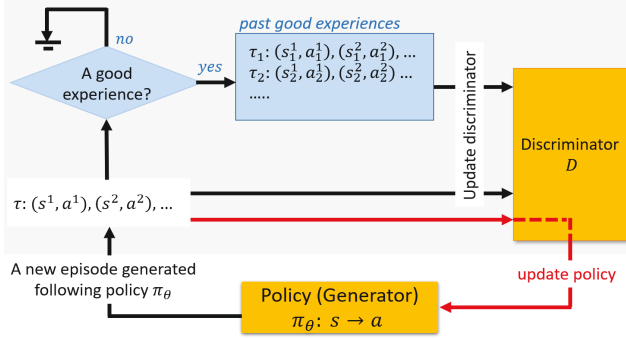


Figure 1: Generative Adversarial Self-Imitation Learning.

In this section, we first introduce the self-imitation learning framework and then describe our approach, NAGASIL, namely Negative sampling and state Augmented Generative Adversarial Self-Imitation Learning.

Self Imitation Learning

Originating from applications such as self-driving cars and drone manipulation, the imitation learning framework learns a policy to produce episodes similar to those of a human demonstrator, rather than learning from feedback on agent actions at intermediate steps (Bojarski et al. 2016; Ross et al. 2013; Hester et al. 2017; Vecerík et al. 2017). The same framework is used to derive Generative Adversarial Imitation Learning (GAIL), which is motivated by minimizing the divergence between the agent’s rollouts and expert demonstrations (Ho and Ermon 2016). GAIL has been extended to GASIL (Generative Adversarial Self-Imitation Learning) (Guo et al. 2018) by replacing expert data with the past good experiences of the learner itself (i.e., episodes with large episodic reward), which is a form of self-imitation learning. Given a policy π , occupancy measure ρ_π is the distribution of state-action pairs that an agent encounters when navigating the environment under this policy. GAIL finds a policy π whose occupancy measure ρ_π minimizes the Jensen-Shannon divergence to the distribution of state-action pairs in past good experiences, i.e., it minimizes $D_{JS}(\rho_\pi, \rho_{\pi_E})$ where π_E is the mixture policy represented by past good experiences. The causal entropy $H(\pi) \triangleq \mathbb{E}_\pi[-\log \pi(a|s)]$ is included as a policy regularizer to guard against collapse to a deterministic policy. To improve exploration, an ensemble of self-imitating agents is explicitly encouraged to visit different, non-overlapping regions of the state-action space, i.e., to simultaneously learn multiple diverse policies that explore different regions of the task (Gangwani, Liu, and Peng 2018).

Fig. 1 illustrates GASIL. Episodes are generated following policy π_θ parameterised by θ . In fake news mitigation, each episode is a sequence of (s, a) pairs where a is the debunker selected at state s . The episodes with the highest reward are considered as past good experiences. Higher reward means that the number of users in e-state *Infected* after the fake news mitigation finishes, is smaller. Other experiences will be discarded during the learning process. Dis-

criminator $D_\phi(s, a) : S \times A \rightarrow [0, 1]$ is parameterized by ϕ . The policy is trained to select action a for state s in a similar way to the mixture policy π_E represented by past good experiences. The discriminator is trained to evaluate the discrepancy between the distribution of (s, a) generated following policy π_θ and the distribution of (s, a) in past good experiences. The policy and discriminator are improved alternately. To summarise, the aim is to find a policy π_θ whose occupancy measure ρ_{π_θ} minimizes the Jensen-Shannon divergence to the distribution of state-action pairs in past good experiences, i.e., to minimize $D_{JS}(\rho_\pi, \rho_{\pi_E})$ (Ho and Ermon 2016; Guo et al. 2018; Gangwani, Liu, and Peng 2018).

NAGASIL

We next describe our algorithm, NAGASIL, for learning a policy for multi-stage fake news mitigation on social networks in a sample-efficient manner.

Negative Samples. Fake news mitigation is cost-intensive, meaning that we aim to learn a policy with a limited number of episodes given some budget constraints. To this end, the strategy of negative sampling is explored. With past bad experiences only, we train a machine learning model \mathcal{M} to predict the probability of actions being selected in state s . The output is a vector denoted as $\mathcal{M}(A|s)$ where each element corresponds to one action. Let $\pi_\theta(A|s)$ be the vector indicating the probability of each action following policy π_θ . If $\mathcal{M}(A|s)$ indicates an action with a probability higher than that of the same action in $\pi_\theta(A|s)$, it implies we should avoid that action. We define the negative samples regularizer:

$$\mathcal{N}(\pi_\theta) \triangleq \|\pi_\theta(A|s) - \mathcal{F}(\pi_\theta(A|s) - \mathcal{M}(A|s))\|_2^2. \quad (7)$$

where $\mathcal{F}(V)$ returns 0 for each negative element and returns the value in $\pi_\theta(A|s)$ for each positive element. That is, the regularizer punishes the negative elements but ignores the positive elements. The effectiveness of negative samples is analyzed in Proposition 1, where it is shown that the policy learned from negative samples on top of good experiences is guaranteed to be at least as good as the policy learned with only good experiences.

Proposition 1.

Provided that low Q -value state-action pairs appear more often in past bad experiences, $\mathbb{E}[\pi_\theta(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)]$ where π_θ is learnt with the negative samples while π_{θ_1} is learnt without the negative samples. Proof can be found in Appendix A.

Augmented State. As the information propagation on social networks is interleaved, it is impossible to observe all the features needed for action selection. To address this issue, we propose to augment the state s , derived from the observable information, with all previous state-action pairs from the same episode. The rationale is to include as much historical information in the input as possible so that the agent can infer missing features in the observable state based on historical data. The augmented state s' is defined as follows:

$$s'_{i+1} = \frac{1}{i} \sum_{m=1}^i \psi^{i-m} [s_m; a_m]. \quad (8)$$

Algorithm 1 Selecting Debunkers via NAGASIL

```

1: Initialize discriminator  $D$  with random parameters  $\phi$ ,
   policy  $\pi$  with random parameters  $\theta$ 
2: Initialize past good experience memory  $M_E$ , past bad
   experience memory  $M_B$ 
3: for iterations do
4:   Generate episode  $\tau_{\pi_\theta}$  following policy  $\pi_\theta$ 
5:   Update past good and bad experience memory  $M_E$ 
     and  $M_B$  using  $\tau_{\pi_\theta}$ 
6:   Sample minibatch  $\tau_E$  from  $M_E$ 
7:   Sample minibatch  $\tau_B$  from  $M_B$ 
8:   Update machine learning model  $\mathcal{M}$  with  $\tau_B$ 
9:   Update  $D$  by ascending stochastic gradient  $\nabla_\phi \mathcal{L}$ 
     with  $\tau_E$  and  $\tau_{\pi_\theta}$ 
10:  Update  $\pi$  by ascending stochastic gradient  $\nabla_\theta \mathcal{L}$ 
     with  $\tau_{\pi_\theta}$ 
11: end for

```

Here $\psi \in [0, 1]$ is the discount rate. Instead of finding policy $\pi_\theta(a|s)$, we aim to find policy $\pi_\theta(a|s, s')$.

Proposition 2 provides a theoretical justification of the augmented state, showing that the policy trained with the augmented state will be at least as strong as the policy trained otherwise.

Proposition 2.

$\mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_2}(a|s)Q(s, a)]$ where π_{θ_1} is trained with the augmented state while π_{θ_2} is trained without the augmented state. Both policies are learnt via Eq. 9 without negative samples. $Q(s, a)$ is the action-value function. Proof can be found in the Appendix A.

We integrate the negative samples and augmented state with GASIL, such that the objective of the discriminator and generator is defined as:

$$\begin{aligned} \arg \min_{\theta} \arg \max_{\phi} \mathcal{L} = & \mathbb{E}_{\pi_\theta} [\log D_\phi(s, s', a)] \\ & + \mathbb{E}_{\pi_E} [\log(1 - D_\phi(s, s', a))] - \lambda \mathcal{H}(\pi_\theta) + \lambda_1 \mathcal{N}(\pi_\theta). \end{aligned} \quad (9)$$

where $\mathcal{H}(\pi) = \mathbb{E}[-\log \pi(a|s)]$ is causal entropy (Ho and Ermon 2016). The optimized θ and ϕ minimize $D_{JS}(\rho_\pi, \rho_{\pi_E})$ which is:

$$\begin{aligned} D_{JS}(\rho_{\pi_\theta}, \rho_{\pi_E}) = & \max_{\phi} \mathbb{E}_{(s, s', a) \sim \rho_{\pi_\theta}} [\log D_\phi(s, s', a)] \\ & + \mathbb{E}_{(s, s', a) \sim \rho_{\pi_E}} [\log(1 - D_\phi(s, s', a))]. \end{aligned} \quad (10)$$

Selecting Debunkers via NAGASIL. Our algorithm for fake news mitigation is presented in Alg. 1. Following policy π_θ , the generated episode τ_{π_θ} is evaluated; if the episodic reward is one of the highest, τ_{π_θ} is inserted into M_E ; if it is one of the lowest, τ_{π_θ} is inserted into M_B (line 4-5). The parameters ϕ of discriminator D are updated (line 9) via stochastic gradient ascent:

$$\begin{aligned} \nabla_\phi \mathcal{L} = & \mathbb{E}_{\tau_{\pi_\theta} \sim \pi_\theta} [\nabla_\phi \log D_\phi(s, s', a)] \\ & + \mathbb{E}_{\tau_E \sim \pi_E} [\nabla_\phi \log(1 - D_\phi(s, s', a))]. \end{aligned} \quad (11)$$

The parameters θ of policy π are updated (line 10) via stochastic gradient ascent:

$$\begin{aligned} \nabla_\theta \mathcal{L} = & \mathbb{E}_{\tau_{\pi_\theta} \sim \pi_\theta} [\nabla_\theta \log D_\phi(s, s', a)] \\ & - \lambda \nabla_\theta \mathcal{H}(\pi_\theta) + \lambda_1 \nabla_\theta \mathcal{N}(\pi_\theta). \end{aligned} \quad (12)$$

The algorithm continuously updates the policy, which takes the observed environment state as input, and outputs a selected user as a debunker for fake news mitigation.

5 Experiments

We evaluated NAGASIL on both real-world rumour datasets and social networks, as well as large synthetic networks, and benchmarked it against baselines from both the fake news mitigation and self-imitation learning literature. Experiments were conducted on a cluster where each node has 64 cores, 2.0GHz CPUs and 256G RAM. All deep neural networks are implemented using Tensorflow (Abadi et al. 2016) (distributed with the Apache License 2.0), all social networks are implemented using NetworkX (Hagberg, Swart, and S Chult 2008) (distributed with the 3-clause BSD License) and the epidemic model SEIR is implemented based on EoN (Miller and Ting 2020) (distributed with the MIT License).

Data. We use PHEME (Zubiaga et al. 2016), a widely used dataset for modeling rumour propagation on social networks. PHEME includes the source and the propagation path of messages, i.e., who spreads the news first and how the news propagates on social networks. From PHEME, the probability distribution that users propagate received news is extracted and used in our experiments. As the public PHEME dataset does not include the underlying social network, we utilize the Facebook social network (Leskovec and McAuley 2012) from SNAP (Leskovec and Krevl 2014) and also generate several synthetic Twitter social networks according to Twitter network settings from the literature (Myers et al. 2014).

Baselines and Evaluation Metrics. We compare our proposed NAGASIL against baselines from both the fake news mitigation and reinforcement learning literature. Each method selects a debunker for every stage – w time steps – until a given budget is used up. For each method, performance is evaluated based on the episodic reward of the episodes generated following the policy in the testing stage (i.e., the last 100 episodes). The episodic reward is defined by Eq. 6. Three baselines are state-of-the-art models from the fake news mitigation literature:

- *Deep Q-Network with Future State Predictor* (DQN-FSP) (Xu, Deng, and Zhang 2022) is a state-of-the-art RL-based method for fake news mitigation. When training DQN-FSP with single debunker selection, a reward will be provided following Eq. 6.
- *Maximum Influence* (MAX-INF) (Saxena, Saxena, and Gera 2020) is a widely used heuristic approach in the fake news mitigation literature. It selects the user with the highest number of followers (i.e., the most influential users) as the debunker for each stage.
- *Maximum Defense* (MAX-DEF) (Saxena, Saxena, and Gera 2020) is another heuristic approach for fake news

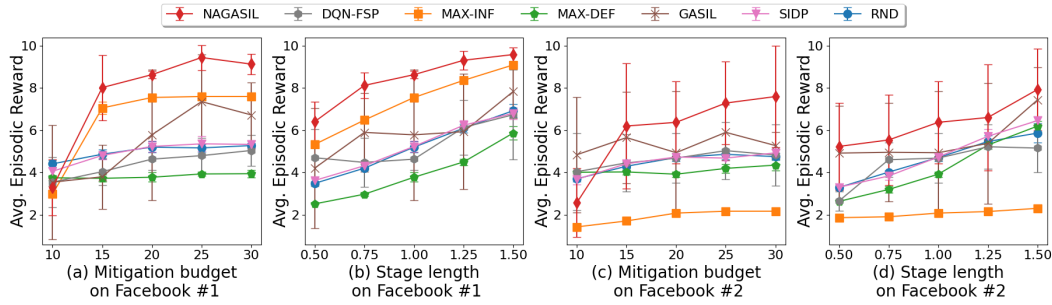


Figure 2: Performance for rumour mitigation on a real-world Facebook social network.

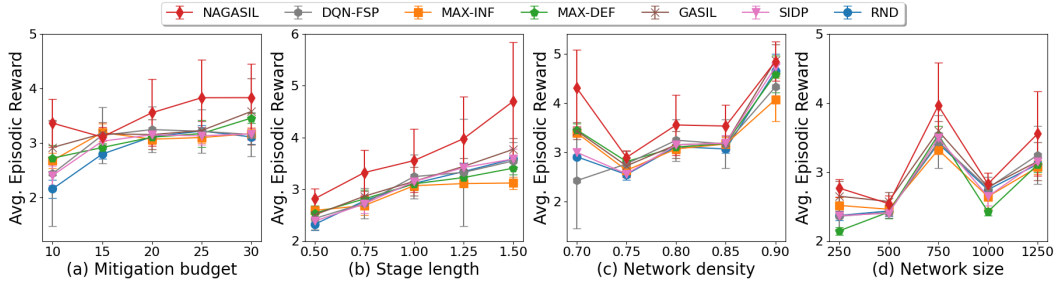


Figure 3: Performance for rumour mitigation on synthetic Twitter networks with various settings.

mitigation. It selects the most active user propagating fake news as the debunker for each stage.

Two more baselines can be formed by applying existing self-imitation learning models for fake news mitigation:

- *Generative Adversarial Self-Imitation Learning* (GASIL) is a popular self-imitation learning method proposed in (Guo et al. 2018) (source code distributed with MIT License).
- *Self-Imitating Diverse Policies* (SIDP) is a self-imitation learning method, improved to achieve better exploration of the environment (Gangwani, Liu, and Peng 2018).

As a sanity check, we include as a baseline the basic policy of randomly selecting one user as the debunker at each stage (RND). Note that we have experimented with other general RL-based methods, such as PPO and DQN, but they perform consistently worse than RND and thus do not pass our sanity check. A possible reason for this is that these methods are not designed for episodic reward settings. Results for these methods, therefore, are not reported. The details of experiment settings are presented in Appendix B.

Performance w.r.t. Mitigation Settings

Figure 2 and 3 report the average episodic reward of five runs using different random seeds and their standard deviation for PHEME propagation on a real-world Facebook social network and (synthetic) Twitter networks with different settings. From the Facebook data, we randomly pick two ego networks, named Facebook #1 and #2. Figure 2(a) and (c) show the performance with respect to the budget

B , where the budget increases from 10 to 30, on Facebook networks. Given more budget, more mitigation effect is expected, since the fake news mitigation campaign will continue for more stages. The experiment results show that NAGASIL performs better on all budget settings except on extremely low budgets. This might be caused by the extremely limited number of actions, where the problem effectively reduces to one-off selections. Figure 2(b) and (d) show the performance with respect to stage length w on Facebook networks. The greater stage length allows the agent to observe more mitigation effects from previous actions; however, the greater stage length means it takes a longer time to select the next debunker. To investigate the impact caused by stage length, we increase the stage length from 0.5 to 1.5. We can see that NAGASIL has outstanding performance in all settings against all baselines. MAX-INF demonstrates much better performance in Facebook #1 than in Facebook #2. It suggests that the performance of MAX-INF is unstable, i.e., it is heavily influenced by the underlying social network structure.

Figure 3(a) shows the performance with respect to the budget B on a Twitter network of 1250 nodes (see Appendix B). We also vary the budget B from 10 to 30. The experiment results demonstrate that NAGASIL performs consistently better under most budget settings. Figure 3(b) shows the performance with respect to stage length w , varying from 0.5 to 1.5 on the Twitter network. We can see that NAGASIL has consistently outstanding performance in all settings against all baselines.

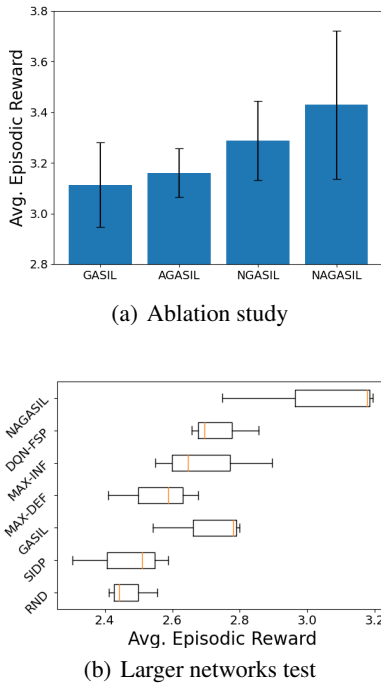


Figure 4: NAGASIL - ablation study and performance on larger networks.

Performance w.r.t. Network Settings

Figure 3(c) shows the performance with respect to Twitter networks of different densities. In the network generation model (Bollobás et al. 2003), the density was controlled by β (by default, $\beta = 0.8$). In the experiments, β is increased from 0.7 to 0.9 to simulate different levels of density. The sum of α , β and γ must be 1. We keep $\gamma = 3\alpha$ to ensure that the social networks maintain the property that the out-bound degree is lower than the in-bound degree (Myers et al. 2014). We can observe the significant advantage of NAGASIL against all baselines at most density levels. Figure 3(d) shows the performance with respect to the size of the Twitter networks where the number of users increases from 250 to 1250. The density of Twitter networks is the same by setting $\beta = 0.8$. Clearly, NAGASIL outperforms all baselines consistently.

Ablation Study

To verify the effectiveness of negative samples and augmented state, we compare the performance of NAGASIL against three ablated versions: NGASIL, AGASIL, and GASIL. NGASIL is a variant of NAGASIL where the augmented state is removed (i.e., s' is removed in Eq. 9 when learning the policy). AGASIL is a variant of NAGASIL where the negative samples are removed (i.e., λ_1 is set to 0 in Eq. 9 when learning the policy). The default experiment settings on Twitter networks (see Appendix B) are applied except the evaluation metric is calculated over all 1000 episodes to demonstrate the performance of different baselines during the whole training process.

The results of the ablation study are reported in Figure 4(a), where the mean and standard deviation of 5 runs are plotted for each method. Firstly, note that NGASIL has significantly better performance than GASIL, indicating that negative samples can clearly boost performance for fake news mitigation. Secondly, AGASIL has better performance compared to GASIL. Even though s' provides additional information, the augmented state increases the state space. A larger state space necessitates more training data to fully realise the benefits brought by additional information. Thirdly, NAGASIL has a significant performance advantage compared to NGASIL, AGASIL and GASIL. The performance gain is attributed to the combination of negative samples and the augmented state; that is, the negative samples provide extra training data to help improve the effectiveness of the augmented state.

Performance on Larger Twitter Networks

To simulate real-world deployment of fake news mitigation campaigns, we further evaluate NAGASIL on a larger Twitter network with 2500 users. The average performance of 3 runs of NAGASIL (default settings) and baselines is reported as a box plot in Figure 4(b). The experiment results clearly show that NAGASIL outperforms all baselines on the larger Twitter network. Note that following real-world fake news mitigation reported in the literature (Farajtabar et al. 2017), our setting simulates that mitigation policy is applied to a group of users who have interactions with the news to be mitigated, which often are in the order of thousands rather than millions.

6 Social Impact and Limitations

Misinformation and fake news on social media can swing public opinions, damage social cohesion, and foster misperception, and are recognised threats to human societies (Weeks 2015; Roozenbeek et al. 2020). Our proposed solution potentially offers direct benefit to social media companies and fact-checking services in their efforts to counteract the spread of misinformation at scale via autonomous crowd debunking.

Our NAGASIL framework provides a data-driven platform for media researchers to conduct large-scale studies and collect data to analyse and discover factors affecting the interaction between misinformation and truthful information propagation on social networks. For example, controlling the settings of NAGASIL for different mitigation stages and stage lengths can be used to simulate and design media responses for future misinformation campaigns.

As for limitations, similar to existing studies, e.g., (Farajtabar et al. 2017; Xu, Deng, and Zhang 2022), our proposed fake news mitigation policy assumes that the truth value – true or fake – of social media news posts is established and fed to the mitigation process. Therefore, the mitigation model must be applied in conjunction with a fake news detection model.

7 Conclusion

This study identified and addressed the issue of *episodic reward*, an essential but overlooked issue for learning fake news mitigation policies via reinforcement learning. Specifically, our solution proposed the *negative samples* to learn the policy for debunker selection by imitating past good experiences and avoiding past bad experiences. We further proposed the *augmented state* to enrich the state of reinforcement learning with features extracted from past state-action pairs in the current episode. The superiority of the proposed solution has been verified against various baselines in different settings. In a broader context, this study improves fake news mitigation, for the first time to our best knowledge, by giving due consideration to the complexity of information propagation on social networks. While NAGASIL is compatible with different information propagation models, a single propagation model cannot fully characterise the complex information propagation on social networks. In future work, we will further investigate NAGASIL with alternative information propagation models to further evaluate its robustness.

A Proof of Propositions

Proposition 1.

$\mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_2}(a|s)Q(s, a)]$ where π_{θ_1} is trained with the augmented state while π_{θ_2} is trained without the augmented state. Both policies are learnt via Eq. 9 without negative samples. $Q(s, a)$ is the action-value function.

proof: $\rho_E(s, a)$ can be calculated by summing the joint probability distribution over all values of s' , that is, $\rho_E(s, a) = \sum_{s'} \rho_E(s, s', a)$. By definition, $\rho_E(s, a) = P(s)\pi_E(a|s)$ and $\rho_E(s, s', a) = P(s) \sum_{s'} \pi_E(a|s, s')$ where $P(s)$ is probability of s in past good experiences. Since $\pi_E(a|s) = \sum_{s'} \pi_E(a|s, s')$, we have $\pi_E(a|s) \geq \pi_E(a|s, s')$ which indicates a is equally or more preferable if considering s than that if considering (s, s') . As a result, $\rho_E(s, a) \geq \rho_E(s, s', a)$ where $\rho_E(s, a)$ includes the probability of (s, s', a) inside, and outside, past good experiences. According to (Ho and Ermon 2016), $\rho_{\pi_{\theta_1}} = \rho_E(s, s', a)$ and $\rho_{\pi_{\theta_2}} = \rho_E(s, a)$ by learning policy via Eq. 9 without $\mathcal{N}(\pi_{\theta})$. The episodes generated following policy π_{θ_1} are equally or more likely to have high rewards compared with those generated following π_{θ_2} . That is, $\mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_2}(a|s)Q(s, a)]$. \square

Proposition 2.

Provided that low Q-value state-action pairs appear more often in past bad experiences, $\mathbb{E}[\pi_{\theta}(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)]$ where π_{θ} is learnt with the negative samples while π_{θ_1} is learnt without the negative samples.

proof: $\pi_{\theta}(a|s, s')$ and $\pi_{\theta_1}(a|s, s')$ are updated in the same way when $\rho_{\hat{E}}(s, s', a)$ changes. Let $\rho_{\hat{E}}$ be the occupancy measure of past bad experiences. If $\rho_{\hat{E}}(s, s', a) = 0$, $\pi_{\theta}(a|s, s')$ is same as $\pi_{\theta_1}(a|s, s')$. If $\rho_{\hat{E}}(s, s', a) > 0$, when it becomes relatively greater

(or less) than $\rho_E(s, s', a)$, the Q-value of corresponding state-action pair $Q(s, s', a)$ will be less (or greater). Accordingly, $\pi_{\theta}(a|s, s')$ will be reduced (or increased); but, $\pi_{\theta_1}(a|s, s')$ will not change since it is unaware relative change between $\rho_{\hat{E}}(s, s', a)$ and $\rho_E(s, s', a)$. Thus, $\mathbb{E}[\pi_{\theta}(a|s, s')Q(s, s', a)] \geq \mathbb{E}[\pi_{\theta_1}(a|s, s')Q(s, s', a)]$. \square

The assumption that low Q-value state-action pairs appear more often in bad experiences is likely to hold much of the time since past bad experiences only include worst-performing episodes.

B Experiment Settings

In this study, unless stated otherwise, all methods run for 1000 episodes, where the last 100 episodes are in the testing stage. By default, the experiments on real-world networks will randomly pick two ego networks, named Facebook #1 and #2, (with radius = 2, anonymized networks released with code) from the Facebook dataset (Leskovec and McAuley 2012). The experiments on synthetic Twitter social networks will have a directed scale-free network with 1250 users generated using the method in (Bollobás et al. 2003) (no self-loop and multiple edges). The parameters of network generation model are $\alpha = 0.05$, $\beta = 0.8$ and $\gamma = 0.15$. Each user i has a cost $(e_i / \max_{j \in U}(e_j)) * 9 + 1$ which is in $[1, 10]$. Users with a higher number of followers will have higher costs. 20 users are randomly picked up as fake news spreaders who are in e-state *Infected* with intensity $\xi_i \sim \mathcal{U}[0.5, 1.5]$ to spread fake news at time 0 and the intensity decays along with time following Eq. 4; other users are in e-state *Susceptible*. At time 5, fake news mitigation starts. For every time period $w = 1$ (i.e., a stage), one user from U is selected as a new debunker until budget $B = 20$ is used up. If user i is selected as a debunker, their e-state is changed to *Recovered* with intensity $\xi_i \sim \mathcal{U}[0.5, 1.5]$ to spread true news immediately. After the last stage, the news spreading continues for a time period of 5 and then the episodic reward is gauged. For self-imitation learning baselines, the number of past good experiences is set to 20, for NAGASIL, we additionally set the number of past bad experiences to 10% of total past experiences.

When a user i receives fake news (true news), the probability of changing from another e-state to *Infected* (*Recovered*) is updated according to Eq. 1 (Eq. 2) where the logistic function's midpoint x_i in Eq. 3 is defined as $(e_i / \max_{j \in U}(e_j)) * 2 + 1$, i.e., a value in $[1, 3]$. If user i has more followers, x_i is higher and the user is less likely to change e-state. Once user i is successfully changed to e-state *Infected* (*Recovered*), the intensity of the user to spread fake news (true news) is initialized to be $\xi_i \sim \mathcal{U}[0.5, 1.5]$ and decays along with time following Eq. 4.

Acknowledgements

This research is supported in part by the ARC Discovery Projects DP200101441, DP210100743 and ARC Linkage Project LP180100750.

References

- Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283.
- Aymanns, C.; Weber, M.; Georg, C.-P.; and Foerster, J. 2020. Modeling Fake News in Social Networks with Deep Multi-Agent Reinforcement Learning.
- Benamira, A.; Devillers, B.; Lesot, E.; Ray, A. K.; Saadi, M.; and Malliaros, F. D. 2019. Semi-supervised learning and graph neural networks for fake news detection. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 568–569. IEEE.
- Bojarski, M.; Testa, D. D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L. D.; Monfort, M.; Muller, U.; Zhang, J.; Zhang, X.; Zhao, J.; and Zieba, K. 2016. End to End Learning for Self-Driving Cars. *CoRR*, abs/1604.07316.
- Bollobás, B.; Borgs, C.; Chayes, J. T.; and Riordan, O. 2003. Directed scale-free graphs. In *SODA*, volume 3, 132–139.
- Brauer, F. 2008. Compartmental models in epidemiology. In *Mathematical epidemiology*, 19–79. Springer.
- Farajtabar, M.; Yang, J.; Ye, X.; Xu, H.; Trivedi, R.; Khalil, E.; Li, S.; Song, L.; and Zha, H. 2017. Fake news mitigation via point process based intervention. In *International Conference on Machine Learning*, 1097–1106. PMLR.
- Farajtabar, M.; Ye, X.; Harati, S.; Song, L.; and Zha, H. 2016. Multistage campaigning in social networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4725–4733.
- Gangwani, T.; Liu, Q.; and Peng, J. 2018. Learning self-imitating diverse policies. *arXiv preprint arXiv:1805.10309*.
- Goindani, M.; and Neville, J. 2020a. Cluster-Based Social Reinforcement Learning. *arXiv preprint arXiv:2003.00627*.
- Goindani, M.; and Neville, J. 2020b. Social reinforcement learning to combat fake news spread. In *Uncertainty in Artificial Intelligence*, 1006–1016. PMLR.
- Gumel, A. B.; Ruan, S.; Day, T.; Watmough, J.; Brauer, F.; Van den Driessche, P.; Gabrielson, D.; Bowman, C.; Alexander, M. E.; Ardal, S.; et al. 2004. Modelling strategies for controlling SARS outbreaks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1554): 2223–2232.
- Guo, Y.; Oh, J.; Singh, S.; and Lee, H. 2018. Generative adversarial self-imitation learning. *arXiv preprint arXiv:1812.00950*.
- Guo, Z.; Schlichtkrull, M.; and Vlachos, A. 2022. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10: 178–206.
- Hagberg, A.; Swart, P.; and S Chult, D. 2008. Exploring network structure, dynamics, and function using NetworkX. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- He, B.; Ahamad, M.; and Kumar, S. 2023. Reinforcement learning-based counter-misinformation response generation: a case study of COVID-19 vaccine misinformation. In *Proceedings of the ACM Web Conference 2023*, 2698–2709.
- Hester, T.; Vecerik, M.; Pietquin, O.; Lanctot, M.; Schaul, T.; Piot, B.; Sendonaris, A.; Dulac-Arnold, G.; Osband, I.; Agapiou, J. P.; Leibo, J. Z.; and Gruslys, A. 2017. Learning from Demonstrations for Real World Reinforcement Learning. *CoRR*, abs/1704.03732.
- Ho, J.; and Ermon, S. 2016. Generative adversarial imitation learning. *Advances in neural information processing systems*, 29: 4565–4573.
- Lacombe, R. 2018. Fake News Mitigation in Social Networks.
- Leskovec, J.; and Krevl, A. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>.
- Leskovec, J.; and McAuley, J. 2012. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25.
- Miller, J. C.; and Ting, T. 2020. Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *arXiv preprint arXiv:2001.02436*.
- Murayama, T.; Wakamiya, S.; Aramaki, E.; and Kobayashi, R. 2021. Modeling the spread of fake news on Twitter. *Plos one*, 16(4): e0250419.
- Myers, S. A.; Sharma, A.; Gupta, P.; and Lin, J. 2014. Information network or social network? The structure of the Twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web*, 493–498.
- Oh, J.; Guo, Y.; Singh, S.; and Lee, H. 2018. Self-imitation learning. In *International Conference on Machine Learning*, 3878–3887. PMLR.
- Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35: 27730–27744.
- Pham, C. V.; Phu, Q. V.; Hoang, H. X.; Pei, J.; and Thai, M. T. 2019. Minimum budget for misinformation blocking in online social networks. *Journal of Combinatorial Optimization*, 38(4): 1101–1127.
- Roozenbeek, J.; Schneider, C. R.; Dryhurst, S.; Kerr, J.; Freeman, A. L.; Recchia, G.; Van Der Bles, A. M.; and Van Der Linden, S. 2020. Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, 7(10): 201199.
- Ross, S.; Melik-Barkhudarov, N.; Shankar, K. S.; Wendel, A.; Dey, D.; Bagnell, J. A.; and Hebert, M. 2013. Learning monocular reactive uav control in cluttered natural environments. In *2013 IEEE international conference on robotics and automation*, 1765–1772. IEEE.
- Saxena, A.; Hsu, W.; Lee, M. L.; Leong Chieu, H.; Ng, L.; and Teow, L. N. 2020. Mitigating misinformation in online social network with top-k debunkers and evolving user opinions. In *Companion Proceedings of the Web Conference 2020*, 363–370.

- Saxena, A.; Saxena, H.; and Gera, R. 2020. k-TruthScore: Fake News Mitigation in the Presence of Strong User Bias. In *International Conference on Computational Data and Social Networks*, 113–126. Springer.
- Shrestha, A.; Spezzano, F.; and Joy, A. 2020. Detecting Fake News Spreaders in Social Networks via Linguistic and Personality Features. In *CLEF*.
- Shu, K.; Bernard, H. R.; and Liu, H. 2019. Studying fake news via network analysis: detection and mitigation. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, 43–65. Springer.
- Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1): 22–36.
- Smit, C. R.; Bevelander, K. E.; de Leeuw, R. N.; and Buijzen, M. 2022. Motivating social influencers to engage in health behavior interventions. *Frontiers in Psychology*, 13: 885688.
- Tan, Z.; Wu, D.; Gao, T.; You, I.; and Sharma, V. 2019. AIM: Activation increment minimization strategy for preventing bad information diffusion in OSNs. *Future Generation Computer Systems*, 94: 293–301.
- Tian, L.; Zhang, X.; and Lau, J. H. 2022. DUCK: Rumour Detection by Modelling User and Comment Propagation Networks. In *NAACL*.
- Tian, L.; Zhang, X.; Wang, Y.; and Liu, H. 2020. Early detection of rumours on twitter via stance transfer learning. In *European Conference on Information Retrieval*, 575–588. Springer.
- Vecerík, M.; Hester, T.; Scholz, J.; Wang, F.; Pietquin, O.; Piot, B.; Heess, N.; Rothörl, T.; Lampe, T.; and Riedmiller, M. A. 2017. Leveraging Demonstrations for Deep Reinforcement Learning on Robotics Problems with Sparse Rewards. *CoRR*, abs/1707.08817.
- Vo, N.; and Lee, K. 2018. The rise of guardians: Fact-checking url recommendation to combat fake news. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, 275–284.
- Vo, N.; and Lee, K. 2020. Where Are the Facts? Searching for Fact-checked Information to Alleviate the Spread of Fake News. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7717–7731.
- Wang, S.; Xu, X.; Zhang, X.; Wang, Y.; and Song, W. 2022. Veracity-aware and event-driven personalized news recommendation for fake news mitigation. In *Proceedings of the ACM Web Conference 2022*, 3673–3684.
- Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 849–857.
- Weeks, B. E. 2015. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of communication*, 65(4): 699–719.
- Wen, S.; Haghighi, M. S.; Chen, C.; Xiang, Y.; Zhou, W.; and Jia, W. 2014a. A sword with two edges: Propagation studies on both positive and negative information in online social networks. *IEEE Transactions on Computers*, 64(3): 640–653.
- Wen, S.; Jiang, J.; Xiang, Y.; Yu, S.; Zhou, W.; and Jia, W. 2014b. To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems*, 25(12): 3306–3316.
- Xu, X.; Deng, K.; and Zhang, X. 2022. Identifying Cost-effective Debunkers for Multi-stage Fake News Mitigation Campaigns. In *International Conference on Web Search and Data Mining*.
- Zhao, L.; Cui, H.; Qiu, X.; Wang, X.; and Wang, J. 2013. SIR rumor spreading model in the new media age. *Physica A: Statistical Mechanics and its Applications*, 392(4): 995–1003.
- Zubiaga, A.; Liakata, M.; Procter, R.; Wong Sak Hoi, G.; and Tolmie, P. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3): e0150989.