Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(7): 13113–13132.
DOI: 10.3934/mbe.2023584
Received: 29 March 2023
Revised: 15 May 2023
Accepted: 23 May 2023
Published: 06 June 2023

Research article

Preventing online disinformation propagation: Cost-effective dynamic budget allocation of refutation, media censorship, and social bot detection

Yi Wang[1,*], **Shicheng Zhong**[2] **and Guo Wang**[3]

[1] School of Big Data and Information Industry, Chongqing City Management College, Chongqing 400000, China

[2] Chongqing Longjin Technology Co., Ltd, Chongqing 400000, China

[3] College of Mechanical Engineering, Chongqing Wuyi Polytechinc College, Chongqing 400000, China

* **Correspondence:** Email: wangyi201902@cqc.edu.cn.

Abstract: Disinformation refers to false rumors deliberately fabricated for certain political or economic conspiracies. So far, how to prevent online disinformation propagation is still a severe challenge. Refutation, media censorship, and social bot detection are three popular approaches to stopping disinformation, which aim to clarify facts, intercept the spread of existing disinformation, and quarantine the source of disinformation, respectively. In this paper, we study the collaboration of the above three countermeasures in defending disinformation. Specifically, considering an online social network, we study the most cost-effective dynamic budget allocation (DBA) strategy for the three methods to minimize the proportion of disinformation-supportive accounts on the network with the lowest expenditure. For convenience, we refer to the search for the optimal DBA strategy as the DBA problem. Our contributions are as follows. First, we propose a disinformation propagation model to characterize the effects of different DBA strategies on curbing disinformation. On this basis, we establish a trade-off model for DBA strategies and reduce the DBA problem to an optimal control model. Second, we derive an optimality system for the optimal control model and develop a heuristic numerical algorithm called the DBA algorithm to solve the optimality system. With the DBA algorithm, we can find possible optimal DBA strategies. Third, through numerical experiments, we estimate key model parameters, examine the obtained DBA strategy, and verify the effectiveness of the DBA algorithm. Results show that the DBA algorithm is effective.

Keywords: disinformation propagation; refutation; media censorship; social bot detection; collaborative dynamic budget allocation; optimal control

1. Introduction

1.1. Background

Undoubtedly, today's online social networks (OSNs), such as Weibo and Twitter, have dramatically accelerated information diffusion. However, as a double-edged sword, OSNs also speed up the spread of rumors, posing a potential threat to human society. According to the motivation, false rumors can be classified into two categories: those deliberately fabricated for political or economic conspiracies and those emerging spontaneously without plotters or profound purposes. Generally, the former is referred to as *disinformation* [1], whereas the latter is referred to as *misinformation* [2]. In this paper, we focus on the prevention of online disinformation propagation.

In most cases, disinformation campaigns are premeditatedly launched by resource-sufficient organizations, which can often be sponsored by commercial competitors, terrorists, and even hostile political powers [3, 4]. Launching a disinformation campaign can typically be sketched by the following three steps [5, 6]. First, the plotter buys a large number of social bots (namely, a type of software agent that can autonomously publish online comments as if human beings [7]) from underground darknets. Second, the plotter employs a group of immoral journalists to fabricate convincing fake news by elaborately distorting the facts. Third, the plotter releases fabricated fake news through its own social bots to deceive online users as many as possible and further manipulate public opinions. Once the plotter's conspiracy is successful, there can be serious consequence, ranging from corporation reputation damage to country fragmentation [8]. Hence, it is indispensable to curb disinformation.

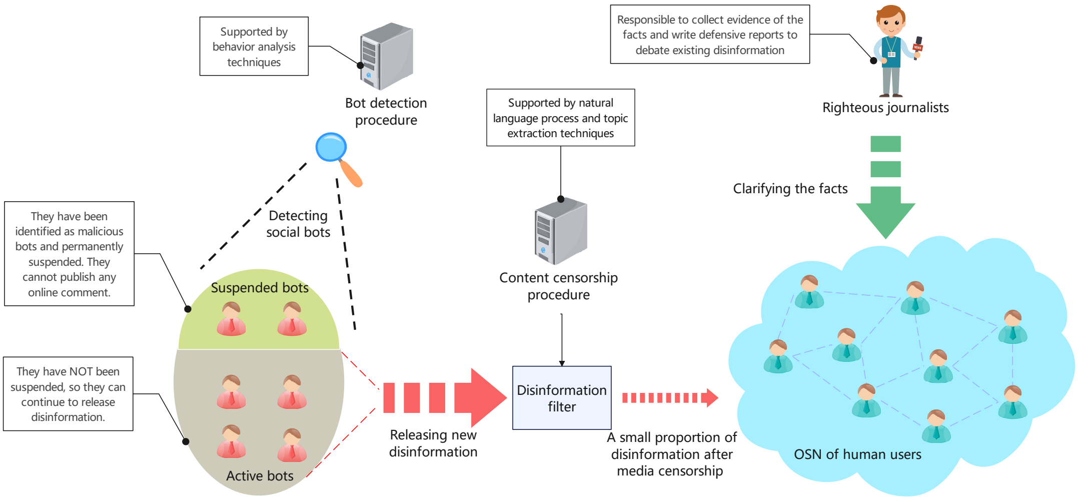

Supported by behavior analysis techniques
Bot detection procedure
Responsible to collect evidence of the facts and write defensive reports to debate existing disinformation
Righteous journalists
Supported by natural language process and topic extraction techniques
Clarifying the facts
They have been identified as malicious bots and permanently suspended. They cannot publish any online comment.
Detecting social bots
Suspended bots
Content censorship procedure
They have NOT been suspended, so they can continue to release disinformation.
Disinformation filter
Releasing new disinformation
A small proportion of disinformation after media censorship
OSN of human users
Active bots

Figure 1. A diagram of refutation, media censorship, and social bot detection in preventing the propagation of online disinformation.

Roughly speaking, OSN carriers can choose three popular approaches to stop disinformation propagation. The first approach is refutation [9, 10]. Refutation means clarifying existing disinformation to inform deceived people of the truth and protect ignorant people from potential deception. OSN carriers may employ righteous journalists to collect evidence of the facts and write defensive materials to debate existing disinformation. The second approach is media censorship [11, 12], which aims to intercept the spread of disinformation by censoring and filtering malicious online comments. Practically, this is realized by computer programs. OSN carriers may need to deploy dedicated cloud servers to run disinformation filtering procedures, which are typically supported by natural language processing (NLP) and topic extraction techniques [13, 14]. The third approach is social bot detection [15, 16]. As the name suggests, this method intends to identify malicious bots and permanently suspend them to make them unavailable—if a bot is suspended, it can no longer publish any comment on the OSN and becomes useless in affecting other users. Because disinformation is mainly released from malicious bots, the bot detection method can weaken the plotter's attacking strength at the root. In addition, bot detection is realized by automatic programs as well. OSN carriers may deploy bot detection procedures on cloud servers to analyze users' behaviors, recognize who have done certain abnormal operators (e.g., sending the same tweet to a large number of other users in a short space of time [17]), and remove them from the OSN. A diagram illustrating the three approaches is given in Figure 1.

Also, the above three countermeasures come at different costs. Recall that refutation requires a group of righteous journalists to write defensive reports. Typically, the more the budget is allocated, the faster such preparatory work can be done. Besides, because media censorship and bot detection are realized by computer algorithms, OSN carriers must buy or rent sufficient computation resources to perform relevant automatic procedures. In many cases, the more the budget is allocated to purchase computation resources, the more the accounts to which media censorship and bot detection can be applied.

1.2. Problem statement

As the three countermeasures have different characteristics and costs, it is natural to ask how to make them collaborate to control disinformation. To this end, a crucial issue is to determine an effective budget allocation scheme for them. To our best knowledge, there is no research on the this topic. In this paper, we intend to fill this research gap by addressing the following problem:

Dynamic Budget Allocation (DBA) problem: Suppose a piece of disinformation or a series of disinformation with the same theme is spreading over an OSN. Consider a finite time horizon. Develop a dynamic budget allocation scheme on this time horizon for the refutation, media censorship, and social bot detection approaches to reduce the proportion of online disinformation-supportive accounts as much as possible with reasonably low expenditure.

1.3. Contributions

In this paper, we are devoted to addressing the DBA problem. Specifically, our contributions are as follows.

- From a mathematical modeling perspective, we reduce the DBA problem to an optimal control model. First, we formalize DBA strategies and establish a trade-off model to evaluate different DBA strategies. Then, we propose a disinformation propagation model to estimate the trade-offs of different DBA strategies. On this basis, we formulate an optimal control problem with DBA strategies as decision variables, the trade-off model as the objective functional, and the disinformation propagation model as a constraint.

- By applying Pontraygin Maximum Principle, we derive a set of necessary conditions for the optimality of the formulated optimal control model. Then, we convert the optimal control model to a two-point-boundary-value problem and develop a heuristic numerical algorithm called the DBA algorithm to iteratively solve it. By running the DBA algorithm, we can attain a possible optimal DBA strategy.
- We conduct a series of numerical experiments to verify the DBA algorithm. First, we estimate crucial parameter values for the disinformation propagation model with a commonly used rumor dataset. Second, with the estimated parameters, we examine the possible optimal DBA strategy attained by running the DBA algorithm. Third, we compare the possible optimal DBA strategy with other common heuristic strategies in terms of their trade-offs and effectiveness. Results show that the possible optimal DBA strategy outperforms other heuristic strategies and thus can be considered effective in practice.

The remainder of this paper is structured by the following manner. Section 2 reviews the related work and highlight the novelty of our work. Section 3 formulates a mathematical optimization model. Section 4 discusses the solution to the optimization model. Section 5 shows a series of numerical experiments. Section 6 closes this paper.

2. Related works

This section discusses the related work and highlights the novelty of our work.

Recall that OSNs have greatly accelerated the spread of rumors and posed a potential threat to society. In this context, developing effective strategies to stop rumors has become an urgent task. Optimal control theory is a widely used methodology in this field, using which optimal dynamic anti-rumor strategies can be obtained. Because our work is an application of optimal control theory, in this section, we focus on the related studies that aim to develop anti-rumor strategies with optimal control theory. In addition, as rumors can be divided into misinformation and disinformation according to their motivations, in the following we mainly review the recent contributions that use optimal control theory to contain these two types of rumors.

To date, research on developing anti-misinformation strategies is rich. Misinformation refers to false rumors spontaneously emerging without a plotter or a profound purpose—usually, misinformation is started just for someone's mischief [2]. According to the essential intentions, existing anti-misinformation strategies can be roughly classified into conversion-based [18–20] and isolation-based [21–23]. The former aims to convert deceived people to misinformation-aware people by using all kinds of possible measures, e.g., clarifying the truth to deceived people, whereas the latter intends to isolate deceived people from the OSN by filtering misinformation-supportive online comments or suspending misinformation-supportive accounts to intercept the spread of misinformation. Also, as different countermeasures have different characteristics, recently there has been a trend to develop collaborative anti-misinformation strategies based on multiple countermeasures. Theoretically, multi-countermeasure strategies can outperform single-countermeasure strategies, because the latter can be considered as a particular instance of the former. See [24–28] for some examples.

However, though misinformation and disinformation are both false rumors and have similar characteristics, existing anti-misinformation strategies may not be perfectly applied to the containment of disinformation. Different from misinformation, disinformation is generally fabricated by a certain plotter for illicit political or economic benefits, and is deliberately diffused onto OSNs through malicious social bots controlled by the plotter [1]. In this process, malicious social bots, as the source of disinformation, play a crucial role in disinformation spread. Therefore, the design of anti-disinformation strategies has to emphasize the influences of malicious social bots and take measures to remove them to destroy the source of disinformation. Unfortunately, as far as we know, there exists no research on developing anti-disinformation strategies which specially consider the elimination of malicious social bots.

To fill this research gap, in this paper, we consider the effect of eliminating malicious social bots and propose an anti-disinformation strategy based on multiple countermeasures. Specifically, we develop a collaborative anti-disinformation strategy combining the refutation, media censorship, and social bot detection countermeasures simultaneously. To our best knowledge, this is the first time to make such an attempt, so our work is of novelty.

3. Problem formulation

This section reduces the DBA problem to an optimal control model. First, we formalize the mathematical form of DBA strategies and establish a trade-off model to evaluate different DBA strategies. Second, we propose a disinformation propagation model to estimate the trade-offs of different DBA strategy. Third, we formulate an optimal control problem to represent the DBA problem from a mathematical modeling perspective, with DBA strategies as decision variables, the trade-off model as the objective functional, and the disinformation propagation model as a constraint. After solving the optimal control model, a cost-effective DBA strategy can be obtained.

3.1. DBA strategy

Suppose a piece of disinformation or a series of disinformation with different contents but the same theme is spreading over an OSN. Suppose we intend to control disinformation in the finite time horizon $[0, T]$ by collaboratively using the refutation, media censorship, and social bot detection approaches. At any time $t \in [0, T]$, denote the *expenditure rates* (i.e., the average financial expenditure per unit time) of refutation, media censorship, and social bot detection by $u_1(t)$, $u_2(t)$, and $u_3(t)$, respectively. Then, we refer to the function

$$u(t) = (u_1(t), u_2(t), u_3(t)),\ 0 \le t \le T, \tag{3.1}$$

as the *mathematical form of DBA strategies*.

Assume that DBA strategies are piecewise continuous functions defined on the time horizon $[0, T]$. Denote the space of all 3-dim piecewise continuous functions on $[0, T]$ by Ω. Denote $u_{\max}$ as the total budget per unit time. By definition, the feasible set of DBA strategies is

$$U = \left\{ u \in \Omega \,\middle|\, u_i(t) \ge 0,\ i = 1, 2, 3,\ \sum_{i=1}^{3} u_i(t) \le u_{\max},\ 0 \le t \le T \right\}. \tag{3.2}$$

In order to select the best DBA strategy from the feasible set U, we need to establish a trade-off model as a criterion to evaluate different DBA strategies. For a given DBA strategy, a reasonable trade-off model should account for both the effectiveness of the strategy on controlling disinformation propagation and the cost of conducting the strategy. In this paper, the goal of controlling disinformation is to minimize the proportion of online accounts which support disinformation. In this context, the effect of a DBA strategy can be reflected by the change in the proportion of disinformation-supportive accounts before and after conducting the strategy. Denote $y(t)$ as the proportion of disinformation-supportive accounts at time t. Then, we calculate the effect of the strategy u by

$$E(u) = y(0) - y(T). \tag{3.3}$$

Besides, by definition, the cost of conducting the strategy u is calculated as

$$C(u) = \int_0^T \sum_{i=1}^{3} u_i(t)dt. \tag{3.4}$$

Hence, the trade-off of the strategy u is set to be

$$J(u) = \omega E(u) - C(u) = \omega[y(0) - y(T)] - \int_0^T \sum_{i=1}^{3} u_i(t)dt, \tag{3.5}$$

where ω is a weight coefficient that quantifies the financial gain due to eliminating all disinformation-supportive users on the network.

3.2. *Disinformation propagation model*

Notice that the trade-off model $J(u)$ in (3.5) is dependent on $y(0)$ and $y(T)$ together. Next, we need to estimate these values for any given strategy u.

Suppose there are another two possible attitudes to disinformation besides the supportive attitude: *reserved* and *denying*. Denote $R(t)$, $S(t)$, and $D(t)$ as the number of human users who hold the reserved, supportive, and denying attitudes at time t, respectively. Besides, denote $B(t)$ as the number of *unsuspended social bots* at time t. Denote the total number of *active* (i.e., unsuspended) accounts on the OSN (including human users and unsuspended bots) as $N(t) = R(t) + S(t) + D(t) + B(t)$. Let

$$r(t) = \frac{R(t)}{N(t)},\ s(t) = \frac{S(t)}{N(t)},\ d(t) = \frac{D(t)}{N(t)},\ b(t) = \frac{B(t)}{N(t)},\ 0 \le t \le T. \tag{3.6}$$

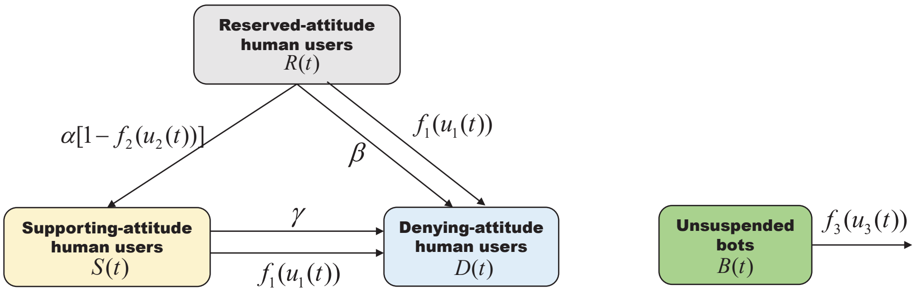
Because $r(t) + s(t) + d(t) + b(t) \equiv 1$ for all t, we refer to the function

$$x(t) = (s(t), d(t), b(t)),\ 0 \le t \le T, \tag{3.7}$$

as the *network state trajectory* for convenience.

In addition, suppose a user can contact (e.g., browsing the latest published tweets) averagely k other users per unit time. Let α' be the average rate at which a user transitions from a reserved attitude to a supportive attitude because of contacting a user who holds a supportive attitude. Let β' be the average rate at which a user transitions from a reserved attitude to a denying attitude because of contacting a user who holds a denying attitude. Let γ' be the average rate at which a user transitions from a supportive attitude to a denying attitude because of contacting a user who holds a denying attitude.

For convenience, let $\alpha = \alpha' k$, $\beta = \beta' k$, and $\gamma = \gamma' k$. Also, denote $f_1(u_1)$ as the average rate at which anti-disinformation stories are released on the OSN when the budget for refutation is u_1. Denote $f_2(u_2)$ as the proportion of disinformation-supportive tweets filtered due to media censorship when the budget for media censorship is u_2. Denote $f_3(u_3)$ as the average rate at which the proportion of unsuspended bots decreases due to bot detection when the budget for bot detection is u_3. A diagram illustrating the changes in the number of various types of accounts is shown in Figure 2.



Reserved-attitude human users $R(t)$
$\alpha[1-f_2(u_2(t))]$
β
$f_1(u_1(t))$
Supporting-attitude human users $S(t)$
γ
$f_1(u_1(t))$
Denying-attitude human users $D(t)$
Unsuspended bots $B(t)$
$f_3(u_3(t))$

Figure 2. A diagram of the changes in the number of various types of accounts.

Remark 1. *In Figure 2, it is seen that all the attitude transitions are considered to be unidirectional, namely, the trajectory of attitude transitions is either "reserved-supporting-denying" or "reserved-denying", and the attitude will ultimately reach "denying". This is because the term "disinformation" in this paper refers to a single piece of tweet or a series of tweets with different contents but the same theme. In this context, we think the power of refuting information far exceeds that of disinformation, such that online users will not be deceived by disinformation again once they have been informed of the truth (as if they have gained long-term immunity to disinformation). In fact, this assumption derives from some existing rumor propagation models (see [29–31]), from which we can find similar thoughts of mathematical modeling. In addition, as there are essential similarities between disinformation propagation and virus spreading, unidirectional attitude transitions can be analogically explained by a classical epidemic SIR model introduced in [32]. Nonetheless, it is worth mentioning that if disinformation refers to tweet variants with different themes such that users cannot gain long-term immunity from refuting information, user attitude transitions can be bidirectional. An example for this topic can be found in the rumor propagation model proposed in [33].*

Then, by applying [32] directly, we propose a disinformation propagation model as follows.

Theorem 1. *Given the initial network state $x(0) = x_0 = (s_0, d_0, b_0)$, the network state trajectory $x(t)$ is determined by the dynamic system (3.8).*

$$\begin{cases} \dfrac{ds}{dt}(t) = \alpha[1 - f_2(u_2(t))][1 - s(t) - b(t) - d(t)][s(t) + b(t)] - \gamma s(t)d(t) - f_1(u_1(t))s(t), \ \ 0 \le t \le T, \\ \dfrac{dd}{dt}(t) = \beta d(t)[1 - s(t) - b(t) - d(t)] + \gamma s(t)d(t) + f_1(u_1(t))[1 - d(t) - b(t)], \ \ 0 \le t \le T, \\ \dfrac{db}{dt}(t) = -f_3(u_3(t))b(t), \ \ 0 \le t \le T. \end{cases} \tag{3.8}$$

Proof. See Appendix. □

Through the dynamic system (3.8), we can predict the network state $x(t)$ for any time t. By definition, the proportion of accounts which support disinformation at time t can be obtained by

$$y(t) = s(t) + b(t),\ 0 \leq t \leq T. \tag{3.9}$$

3.3. Optimal control problem

With the feasible set (3.2), the trade-off model (3.5), the disinformation propagation model (3.8), and the relationship (3.9), we formulate the following optimal control model to represent the DBA problem from a mathematical modeling perspective.

$$\begin{aligned} &\max_{u\in U}\ J(u) = \omega[y(0) - y(T)] - \int_0^T \sum_{i=1}^{3} u_i(t)dt \\ &s.t \begin{cases} s(t),\ d(t),\ b(t) \text{ satisfy the disinformation spread model (3.8)}, \\ x(0) = x_0, \\ y(t) = s(t) + b(t),\ 0 \leq t \leq T. \end{cases} \end{aligned} \tag{3.10}$$

After solving it with appropriate optimization methods, the optimal DBA strategy can be attained.

4. Solution

In the previous section, we reduced the DBA problem to the optimal control model (3.10). In this section, we discuss the solution to the optimal control model (3.10). First, we derive a set of necessary conditions for the optimality of the optimal control problem by applying Pontraygin Maximum Principle (PMP) [34], and then convert the optimal control problem to a two-point boundary value (TPBV) problem [35]. Second, we develop a heuristic algorithm for solving the TPBV problem iteratively.

According to PMP, we construct the following Hamiltonian function for the optimal control model (3.10).

$$\begin{aligned} H(u, x, \lambda) &= -\sum_{i=1}^{3} u_i(t) + \lambda^s \frac{ds}{dt} + \lambda^d \frac{dd}{dt} + \lambda^b \frac{db}{dt} + \lambda^z \frac{dz}{dt} \\ &= -\sum_{i=1}^{3} u_i(t) + \lambda^s[\alpha(1 - f_2(u_2))(1 - s - b - d)(s + b) - \gamma sd - f_1(u_1)s] \\ &\quad + \lambda^d[\beta d(1 - s - b - d) + \gamma sd + f_1(u_1)(1 - d - b)] - \lambda^b f_3(u_3)b, \end{aligned} \tag{4.1}$$

where $\lambda = (\lambda^s, \lambda^d, \lambda^b)$ is a co-state vector. Then, we derive the following necessary conditions for the optimality of the optimal control problem (3.10).

Theorem 2. *Let $u(\cdot)$ denote the optimal solution to the optimal control problem (3.10), $x(\cdot)$ denote the network state trajectory with respect to $u(\cdot)$, and $\lambda(\cdot)$ denote the co-state function with respect to*

$u(\cdot)$ and $x(\cdot)$. Then, $x(\cdot)$ must satisfy the disinformation propagation model (3.8), $\lambda(\cdot)$ must satisfy the dynamic system

$$\begin{cases} \frac{d\lambda^s}{dt}(t) = -\lambda^s(t)\{\alpha[1-f_2(u_2(t))][1-2s(t)-2b(t)-d(t)] - \gamma d(t) - f_1(u_1(t))\} - (\gamma-\beta)\lambda^d(t)d(t), \\ \qquad 0 \le t \le T, \\ \frac{d\lambda^d}{dt}(t) = \lambda^s(t)\{\alpha[1-f_2(u_2(t))][s(t)+b(t)] + \gamma s(t)\} + \lambda^d(t)[\beta d(t) - \gamma s(t) + f_1(u_1(t))],\ 0 \le t \le T, \\ \frac{d\lambda^b}{dt}(t) = -\alpha\lambda^s(t)[1-f_2(u_2(t))][1-2s(t)-2b(t)-d(t)] + \lambda^d(t)[\beta d(t) + f_1(u_1(t))] + \lambda^b(t)f_3(u_3(t)), \\ \qquad 0 \le t \le T. \end{cases} \tag{4.2}$$

with $\lambda(T) = (-\omega, 0, -\omega)$, and $u(\cdot)$ must satisfy the condition

$$u(t) \in \arg\max_{u' \in U'} H(u', x(t), \lambda(t)),\ 0 \le t \le T, \tag{4.3}$$

where

$$U' = \left\{ u' \in R^3 \,\middle|\, u'_i \ge 0,\ i = 1,2,3,\ \sum_{i=1}^{3} u'_i \le u_{\max} \right\}. \tag{4.4}$$

Proof. See Appendix. □

The results of Theorem 2, including the condition (4.3) and the dynamic systems (3.8) and (4.2), are referred to as the *optimality system* of the optimal control problem (3.10). Normally, it is difficult to directly find the optimal solution to an optimal control problem. A more practical approach is to find the solutions that satisfy all the known necessary conditions for optimality and then eliminate the obtained solutions by examining their effectiveness. Thus, the optimality system plays an important role in solving the optimal DBA strategy because it helps us search the optimal DBA strategy indirectly. Any solution that satisfies the optimality system is called a possible optimal DBA strategy.

Solving a possible optimal DBA strategy from the optimality system is essentially solving a TPBV problem, which is complex as well. A widely used approach called the *indirect shooting method* [36] suggests iteratively updating the network state trajectory and the co-state function from an initial guess until they satisfy the optimality system. In this process, a core issue is to determine a criterion for updating the network state trajectory and the co-state function in each iteration. Based on the main idea of the indirect shooting method, we develop a heuristic algorithm as shown in Algorithm 1, which is called the DBA algorithm. As we have difficulty in proving the convergence of the DBA algorithm from a theoretical perspective, we will examine the convergence explicitly in our numerical experiments conducted in the next section.

5. Numerical experiments

This section shows a series of numerical experiments to verify the DBA algorithm developed in the previous section. First, we estimate some key parameters involved in the disinformation propagation model (3.8) with a widely used Twitter rumor dataset. Second, with the estimated parameters, we examine the possible optimal DBA strategy yielded from the DBA algorithm. Third, we introduce

Algorithm 1 DBA

Input: An initial guess of the optimal DBA strategy $u^{(0)}(\cdot)$, convergence error ϵ, and the update step length θ for each iteration.

Output: A possible optimal DBA strategy $u^{*}(\cdot)$.

```
$k \leftarrow 0$;
repeat
    //Calculate the $x(\cdot)$ and $\lambda(\cdot)$ with respect to $u^{(k)}(\cdot)$
    Calculate $x(\cdot)$ from the disinformation propagation model (3.8) with $u(\cdot) = u^{(k)}(\cdot)$;
    $x^{(k)}(\cdot) \leftarrow x(\cdot)$;
    Calculate $\lambda(\cdot)$ from the dynamic system (4.2) with $u(\cdot) = u^{(k)}(\cdot)$ and $x(\cdot) = x^{(k)}(\cdot)$;
    $\lambda^{(k)}(\cdot) \leftarrow \lambda(\cdot)$;
    // Calculate the solution $u(\cdot)$ for $x^{(k)}(\cdot)$ and $\lambda^{(k)}(\cdot)$
    Calculate $u(\cdot)$ from (4.3) with $x(\cdot) = x^{(k)}(\cdot)$ and $\lambda(\cdot) = \lambda^{(k)}(\cdot)$;
    // If $u^{(k)}(\cdot)$ is the optimal solution, $u^{(k)}(\cdot)$ should be equal to $u(\cdot)$
    $\Delta \leftarrow \sum_{i=1}^{3} \int_0^T |u_i^{(k)}(t) - u_i(t)| dt$;
    if $\Delta < \epsilon$ then
        return $u^{(k)}(\cdot)$; // Break the loop and return
    else
        // $u^{(k)}(\cdot)$ is not the optimal solution, so update the solution $u^{(k)}(\cdot)$ with the step length $\theta$
        $u^{(k+1)}(\cdot) \leftarrow u^{(k)}(\cdot) + \theta[u(\cdot) - u^{(k)}(\cdot)]$;
    end if
    $k \leftarrow k + 1$;
until True
```

several heuristic DBA strategies and compare them with the possible optimal DBA strategy to verify the effectiveness of the DBA algorithm.

5.1. Estimation of model parameters

The disinformation propagation model (3.8) involves three objectively determined parameters: the one indicating the average rate at which ignorant people are deceived due to disinformation propagation, called α, the one indicating the average rate at which ignorant people are informed of the truth due to the spread of true stories, called β, and the one indicating the average rate at which deceived people come to reason again due to the spread of true stories, called γ.

To estimate the actual values for the above three parameters, we introduce a widely-used dataset called NERT (Newly Emerged Rumors in Twitter) [37], which results from an empirical study on the spreading process of newly emerged rumors in Twitter. The NERT dataset is structured as a table—each row represents an online comment relevant to the concerned rumor, and each column explains an attribute of the comment, such as the user ID related to this comment, the published date and time of this comment, the attitude of this comment to the concerned rumor, and so on.

From the NERT dataset, we can extract the actual curves of the changes in the proportion of users of different attitudes to the concerned rumor, as discussed in [38]. The optimal parameter estimation should make the results of the disinformation propagation model (3.8) match the actual curves extracted from the dataset as much as possible. Denote the actual proportion of disinformation-supportive human users, disinformation-denying human users, and social bots at time t by $s(t)$, $d(t)$, and $b(t)$, respectively. Besides, denote the proportion predicted from the disinformation propagation model (3.8) by $\hat{s}(t)$, $\hat{d}(t)$ and $\hat{b}(t)$. Then, given the feasible sets $\alpha \in \Omega_\alpha, \beta \in \Omega_\beta$, and $\gamma \in \Omega_\gamma$, the optimal-estimated values of the parameters α, β, and γ can be obtained by solving

$$(\alpha^*, \beta^*, \gamma^*) = \arg \min_{\alpha \in \Omega_\alpha, \beta \in \Omega_\beta, \gamma \in \Omega_\gamma} \int_0^T \{[s(t) - \hat{s}(t)]^2 + [d(t) - \hat{d}(t)]^2 + [b(t) - \hat{b}(t)]^2\} dt. \tag{5.1}$$

Let one unit time be 6 hours. From the NERT dataset, we extract the curves of $s(t)$, $d(t)$, and $b(t)$ for a 18-hour time duration from 04:00 on Oct. 27th to 22:00 on Oct. 27th, 2018. It is worth mentioning that, however, human users and social bots have not been distinguish in the NERT dataset. Thus, in our experiments, we suppose the accounts that support the concerned rumor at the initial time are all malicious social bots and the number of bots keeps constant because there is no bot detection available in the NERT dataset. After calculating the parameters from the spaces $\Omega_\alpha = \Omega_\beta = \Omega_\gamma = \{0.000, 0.001, \ldots, 1.000\}$ with the initial network state $x_0 = (0, 0.280901, 0.311545)$ given in the NERT dataset, we attain the optimal-estimated parameters $\alpha^* = 0.351, \beta^* = 0.288$, and $\gamma^* = 0.000$. Figure 3 compares the actual and estimated proportion curves, which shows that our disinformation propagation model can well match the actual situation.

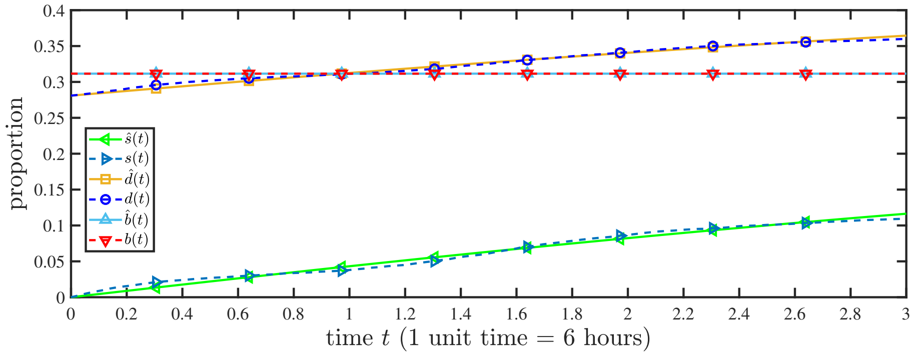


Figure 3. Comparison of the actual and estimated proportion curves.

5.2. *Possible optimal DBA strategy*

Next, we examine the possible optimal DBA strategy yielded from the DBA algorithm. We have to mention that in addition to the three parameters discussed above (namely, α, β, and γ), the remaining parameters (e.g., the weight coefficient ω) are also needed to be estimated by actual situations. However, due to the lack of real-world datasets on these parameters, in our numerical experiments we can only set their values by experience. The details are as follows.

First, let us set the weight coefficient ω by experience, which reflects the financial gain brought by eliminating all disinformation-supportive users from an OSN. To estimate it, let us consider a notorious disinformation event in 2013, in which Barack Obama was claimed to get injured in an explosion. As [39] reports, related disinformation in this event has swept out the whole Twitter and finally wiped out about 130 billion dollars in the stock market. So, if we can prevent all Twitter users from being deceived in this event, we can achieve the financial gain of 130 billion dollars. Hence, we set $\omega = 1.3 \times 10^{11}$ (dollars).

Second, let us define the function f_1 by experience. Recall that the value of $f_1(u_1)$ means the average rate at which rebuttal information is released on the OSN when the expenditure rate of refutation is u_1 dollars per unit time. Particularly, we suppose the cost of producing rebuttal reports mainly comes from employing righteous journalists to collect evidence, write reports, and so on. In this context, we roughly assume that it can take 6 hours on average for one journalist to independently accomplish all the preparatory work. As reports in [40], in 2013 journalists could earn 127.98 dollars on average per 6 hours. Recall that one unit time is defined as 6 hours. So, if the collaboration of multiple journalists can be assumed as linearly accumulative, the function f_1 can be considered to be $f_1(u_1) = \frac{1}{127.98}u_1$ (per unit time).

Next, let us define the function f_2 by experience. Recall that the value of $f_2(u_2)$ means the proportion of disinformation filtered by media censorship procedures when the expenditure rate of renting computation resources for censorship procedures is u_2 dollars per unit time. From related references, we can learn that in 2013 there were 125 million tweets produced by users per 6 hours [41], a filtering procedure can process about 864 million tweets on average per 6 hours [42], and running one censorship procedure on cloud servers can cost 2.608 dollars per 6 hours [43]. So, if the collaboration of multiple filtering procedures can be considered linearly accumulative, the function f_2 can be defined as $f_2(u_2) = \frac{864}{2.608\times125}u_2$ (per unit time), where the condition $u_2 \leq \frac{2.608\times125}{864}$ must be satisfied to guarantee $f(u_2) \leq 1$.

Finally, let us define the function f_3 by experience. Recall that the value of $f_3(u_3)$ means the average rate at which online accounts are examined by bot detection procedures when the expenditure rate of renting computation resources for detection procedures is u_3 dollars per unit time. From related references, we can learn that a bot detection procedure needs 0.71 seconds on average to examine one online account [44] and running one bot detection procedure on cloud servers can cost 2.608 dollars per 6 hours [43]. So, if the collaboration of multiple detection procedures can be considered linearly accumulative, the function f_3 can be defined as $f_3(u_3) = \frac{1}{0.71\times60\times60\times6\times2.608}u_3 = \frac{1}{6666.048}u_3$ (per unit time).

Table 1. A summary for experiment settings.

Parameter	Meaning	Value	Unit	References
ω	the expected financial gain of reducing all disinformation-supportive accounts from an OSN.	1.3×10^{11}	dollar	[39]
$f_1(u_1)$	the average rate at which rebuttal reports are published on an OSN when the expenditure rate is u_1.	$\frac{1}{127.98}u_1$	per unit time	[40]
$f_2(u_2)$	the proportion of which disinformation is filtered by censorship procedures when the expenditure rate is u_2.	$\frac{864}{2.608\times125}u_2$	per unit time	[41–43]
$f_3(u_3)$	the average rate at which online accounts are examined by bot detection procedures when the expenditure rate is u_3.	$\frac{1}{6666.048}u_3$	per unit time	[43,44]

According to the above discussions, a summary for experiment settings is given in Table 1. Then, we conduct the following experiment:

Experiment 1. *Consider the parameters displayed in Table 1. Besides, suppose OSN carriers intend to control disinformation in 3 hours with a maximum budget rate of 10,000 dollars per 6 hours (i.e.,* $T = 0.5$, $u_{\max} = 10,000$*). Then, run the DBA algorithm with the convergence error* $\epsilon = 0.001$ *and step length* $\theta = 0.1$*. The results are shown in Figures 4 and 5.*

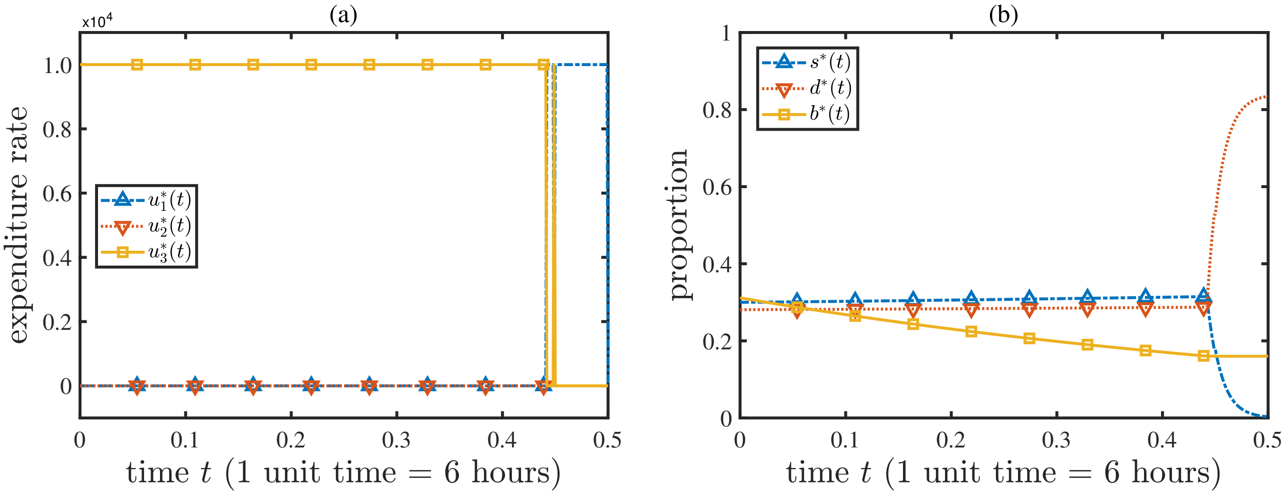


Figure 4. Results of Experiment 1: (a) a possible optimal DBA strategy u^*; (b) the network state trajectory x^* with respect to u^*.

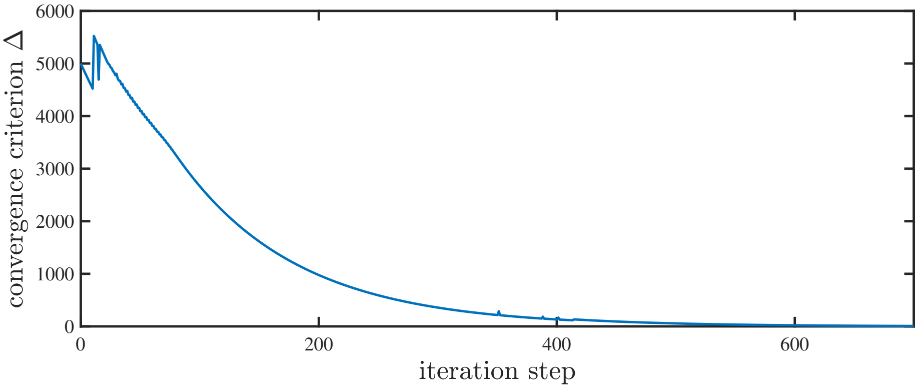


Figure 5. The convergence curve of the DBA algorithm with respect to Experiment 1.

First, let us describe the obtained experiment results. From Experiment 1, we can attain a possible optimal DBA strategy u^* shown in Figure 4(a). It is seen that under the possible optimal DBA strategy, the expenditure rate of refutation (i.e., $u_1^*(t)$) first stays at zero and then increases sharply to nearly the maximum after time $t = 0.45$, the expenditure rate of media censorship (i.e., $u_2^*(t)$) stays at zero during the whole time horizon, and the expenditure rate of bot detection (i.e., $u_3^*(t)$) first stays at the maximum and then keenly drops to zero after time $t = 0.45$. Besides, Figure 4(b) shows the corresponding network state trajectory x^*. It is seen that the proportion of social bots decreases from 0.3 to nearly 0.2, the proportion of disinformation-supportive users increases at a very low speed from 0.3 to 0.35 during the time duration $0 \leq t \leq 0.45$ and then decreases quickly to zero after the time $t = 0.45$, and the proportion of disinformation-denying users first keeps stable and then increases rapidly from about 0.3 to 0.85 after the time $t = 0.45$.

Second, let us analyze the obtained experiment results. From the above results, we can acquire some conclusions for the case of Experiment 1: (a) media censorship is the least important way to control disinformation, so there is no need to allocate any budget for it; (b) in the first 2.7 hours (i.e., from $t = 0$ to $t = 0.45$), the most important thing is to reduce the number of malicious social bots to destroy the source of disinformation, and thus bot detection is the most effective way to control disinformation; (c) in the last 0.3 hours (i.e., from $t = 0.45$ to $t = 0.5$), because half of social bots have been suspended, the most important thing is to clarify the truth to reduce the number of disinformation-supportive human users, and thus refutation becomes the most effective way to control disinformation; (d) as Figure 3 shows the network state trajectory for the case of no anti-disinformation countermeasure, we can learn from Figures 3 and 4 that the possible optimal DBA strategy can dramatically increase the number of disinformation-denying accounts and meanwhile decrease the number of disinformation-supportive accounts.

In addition, let us examine the convergence of the DBA algorithm. Figure 5 shows the change in the convergence criterion Δ of the DBA algorithm over iteration steps for the case in Experiment 1. It is seen that the DBA algorithm converges within 600 iteration steps because the convergence criterion Δ have approximately reduced to zero at the 600-th iteration step.

5.3. *Effectiveness of the DBA algorithm*

Next, we verify if the possible optimal DBA strategy yielded from the DBA algorithm is effective. First, we introduce several commonly used heuristic DBA strategies as follows.

- No-Countermeasure (NC) strategy: Do not take any countermeasure, i.e., $u_(t) = 0$, $0 \leq t \leq T$.
- All-Refutation (AR) strategy: Allocate all the budget to refutation, i.e., $u_1(t) = u_{\max}$, $u_2(t) = 0$, $u_3(t) = 0$, $0 \leq t \leq T$.
- All-Censorship (AC) strategy: Allocate all the budget to media censorship, i.e., $u_1(t) = 0$, $u_2(t) = u_{\max}$, $u_3(t) = 0$, $0 \leq t \leq T$.
- All-Detection (AD) strategy: Allocate all the budget to bot detection, i.e., $u_1(t) = 0$, $u_2(t) = 0$, $u_3(t) = u_{\max}$, $0 \leq t \leq T$.
- Average (Avg) strategy: Allocate the total budget to refutation, censorship, and detection equally, i.e., $u_1(t) = u_2(t) = u_3(t) = \frac{1}{3}u_{\max}$, $0 \leq t \leq T$.

Then, we perform the following experiment to calculate the trade-offs of the above five heuristic strategies.

Experiment 2. *Consider the case described in Experiment 1. Calculate the trade-offs of the NC, AR, AC, AD, and Avg strategies. The result is shown in Figure 6.*

Let $\Delta y = y(0) - y(T)$, where $y(t) = s(t) + b(t)$ is defined in (3.9). The variable Δy, which means the change in the proportion of disinformation-supportive accounts on the network, indicates the effectiveness of conducting a DBA strategy on curbing disinformation. Then, Figure 6 compares the possible optimal DBA strategy and the five heuristic strategies in terms of their effectiveness and ultimate trade-offs. From Figure 6, it is seen that the possible optimal strategy outperforms the five heuristic strategies because the trade-off of the possible optimal strategy is much higher than those of the heuristic strategies. Besides, by comparing the Δy of these strategies, it is seen that the possible optimal strategy can dramatically reduce the proportion of disinformation-supportive accounts. Hence, the possible optimal strategy obtained from the DBA algorithm is effective.

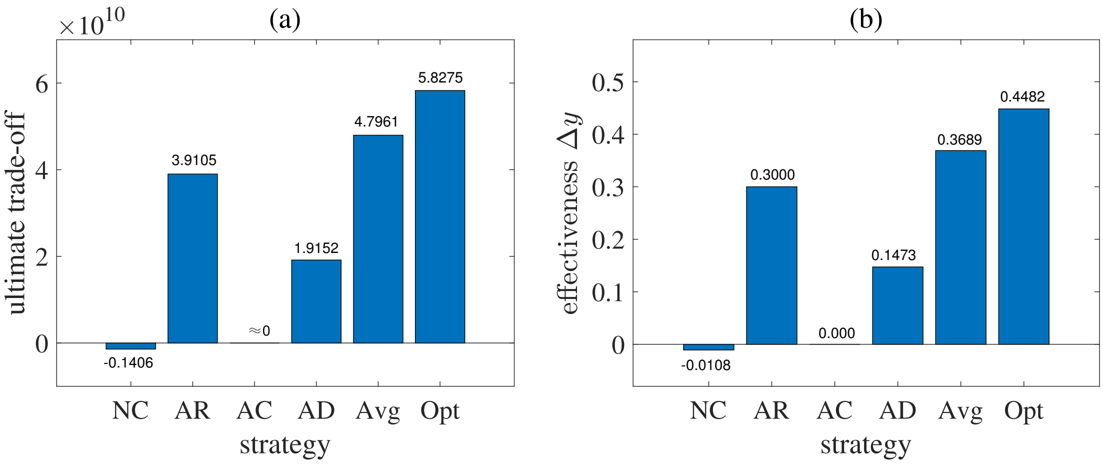


Figure 6. Comparison between the possible optimal DBA strategy and five proposed heuristic strategies in terms of their effectiveness and ultimate trade-offs.

6. Conclusions

In this paper, we have addressed the DBA problem. First, we have proposed a disinformation propagation model to characterize the influences of different DBA strategies on curbing disinformation, and then, established a trade-off model to evaluate DBA strategies. On this basis, we have reduced the DBA problem to an optimal control problem. Second, we have derived a set of necessary conditions called the optimality system for the optimal DBA strategy and developed an iterative heuristic algorithm called the DBA algorithm to numerically solve the optimality system. Third, we have conducted massive numerical experiments to estimate key model parameters, examine the obtained optimal DBA strategy, and verify the effectiveness of the DBA algorithm.

Still, there are some open problems. First, in our numerical experiments, we have only estimated a proportion of parameters for the proposed disinformation propagation model. So in the future extensions, it is urgent to estimate the remaining parameters, such as the effect functions of different countermeasures, to perform more practical experiments. In fact, many related studies (such as [33,45,46]) also meet this challenge. Second, in our disinformation propagation model, the rates at which a user transitions its current attitude from one to another are assumed to be constant for simplicity. However, these rates can be dynamic in some cases as disinformation campaigns are generally driven by large organizations that can change these rates over time [47]. Hence, extending this work by considering a dynamic-rate disinformation propagation model is valuable, though it may introduce more complexity in system modeling and problem solving.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. D. Fallis, What is disinformation?, *Library Trends*, **63** (2015), 401–426. https://doi.org/10.1353/lib.2015.0014
2. J. D. West, C. T. Bergstrom, Misinformation in and about science, *Proc. Natl. Acad. Sci.*, **118** (2021), e1912444117. https://doi.org/10.1073/pnas.1912444117
3. T. Lin, M. Chang, C. Chang, Y. Chou, Government-sponsored disinformation and the severity of respiratory infection epidemics including COVID-19: A global analysis, 2001–2020. *Soc. Sci. Med.*, **296** (2022), 114744. https://doi.org/10.1016/j.socscimed.2022.114744
4. S. Bradshaw, P. N. Howard, The global organization of social media disinformation campaigns, *J. Int. Aff.*, **71** (2018), 23–32.
5. A. Bessi, E. Ferrara, Social bots distort the 2016 US Presidential election online discussion, *First Monday*, **21** (2016). https://doi.org/10.5210/FM.V21I11.7090
6. T. R. Keller, U. Klinger, Social bots in election campaigns: Theoretical, empirical, and methodological implications, *Political Commun.*, **36** (2019), 171–189. https://doi.org/10.1080/10584609.2018.1526238
7. E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, *Commun. ACM*, **59** (2016), 96–104. https://doi.org/10.1145/2818717
8. N. J. Cull, V. Gatov, P. Pomerantsev, A. Applebaum, A. Shawcross, Soviet subversion, disinformation and propaganda: How the West fought against it, *London LSE Consult.*, **68** (2017), 1–77.
9. Z. Li, Q. Zhang, X. Du, Y. Ma, S. Wang, Social media rumor refutation effectiveness: Evaluation, modelling and enhancement, *Inform. Proc. Manage.*, **58** (2021), 102420. https://doi.org/10.1016/j.ipm.2020.102420
10. P. Ozturk, H. Li, Y. Sakamoto, Combating rumor spread on social media: The effectiveness of refutation and warning, in *2015 48th Hawaii international conference on system sciences*, IEEE, (2015), 2406–2414. https://dx.doi.org/10.2139/ssrn.2564249
11. G. Simons, D. Strovsky, Censorship in contemporary Russian journalism in the age of the war against terrorism: A historical perspective, *Eur. J. Commun.*, **21** (2006), 189–211. https://doi.org/10.1177/0267323105064
12. M. Eid, The new era of media and terrorism, *Stud. Conflict Terrorism*, **36** (2013), 609–615. https://doi.org/10.1080/1057610X.2013.793638

13. S. M. Alzanin, A. M. Azmi, Detecting rumors in social media: A survey, *Proc. Comput. Sci.*, **142** (2018), 294–300. https://doi.org/10.1016/j.procs.2018.10.495

14. F. Xu, V. S. Sheng, M. Wang, Near real-time topic-driven rumor detection in source microblogs, *Knowl. Based Syst.*, **207** (2020), 106391. https://doi.org/10.1016/j.knosys.2020.106391

15. E. Alothali, N. Zaki, E. A. Mohamed, H. Alashwal, Detecting social bots on twitter: a literature review, in *2018 International conference on innovations in information technology (IIT)*, SAGA, (2018), 175–180. https://doi.org/10.1109/INNOVATIONS.2018.8605995

16. N. Hajli, U. Saeed, M. Tajvidi, F. Shirazi, Social bots and the spread of disinformation in social media: the challenges of artificial intelligence, *Br. J. Manage.*, **33** (2022), 1238–1253. https://doi.org/10.1111/1467-8551.12554

17. C. Cai, L. Li, D. Zengi, Behavior enhanced deep bot detection in social media, in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, IEEE, (2017), 128–130. https://doi.org/10.1109/ISI.2017.8004887

18. J. Li, H. Jiang, X. Mei, C. Hu, G. Zhang, Dynamical analysis of rumor spreading model in multi-lingual environment and heterogeneous complex networks, *Inform. Sci.*, **536** (2020), 391–408. https://doi.org/10.1016/j.ins.2020.05.037

19. J. Chen, C. Chen, Q. Song, Y. Zhao, L. Deng, R. Xie, et al., Spread mechanism and control strategies of rumor propagation model considering rumor refutation and information feedback in emergency management, *Symmetry*, **13** (2021), 1694. https://doi.org/10.3390/sym13091694

20. L. Zhu, F. Yang, G. Guan, Z. Zhang, Modeling the dynamics of rumor diffusion over complex networks, *Inform. Sci.*, **562** (2021), 240–258. https://doi.org/10.1016/j.ins.2020.12.071

21. S. Yu, Z. Yu, H. Jiang, Stability, hopf bifurcation and optimal control of multilingual rumor-spreading model with isolation mechanism, *Mathematics*, **10** (2022), 4556. https://doi.org/10.3390/math10234556

22. T. Li, Y. Guo, Nonlinear dynamical analysis and optimal control strategies for a new rumor spreading model with comprehensive interventions, *Qualitative theory of dynamical systems*, **20** (2021), 1–24. https://doi.org/10.1007/s12346-021-00520-7

23. Z. Liu, T. Qin, Q. Sun, S. Li, H. H. Song, Z. Chen, SIRQU: Dynamic quarantine defense model for online rumor propagation control, *IEEE Trans. Comput. Soc. Syst.*, **9** (2022), 1703–1714. https://doi.org/10.1109/TCSS.2022.3161252

24. X. Wang, X. Wang, F. Hao, G. Min, L. Wang, Efficient coupling diffusion of positive and negative information in online social networks, *IEEE Trans. Network Serv. Manage.*, **16** (2019), 1226–1239. https://doi.org/10.1109/TNSM.2019.2917512

25. J. Zhao, L. Yang, X. Zhong, X. Yang, Y. Wu, Y. Y. Tang, Minimizing the impact of a rumor via isolation and conversion, *Phys. A Stat. Mech. Appl.*, **526** (2019), 120867. https://doi.org/10.1016/j.physa.2019.04.103

26. Y. Lin, X. Wang, F. Hao, Y. Jiang, Y. Wu, G. Min, et al., Dynamic control of fraud information spreading in mobile social networks, *IEEE Trans. Syst. Man Cybernetics Syst.*, **51** (2019), 3725–3738. https://doi.org/10.1109/TSMC.2019.2930908

27. Y. Cheng, L. Zhao, Dynamical behaviors and control measures of rumor-spreading model in consideration of the infected media and time delay, *Inform. Sci.*, **564** (2021), 237–253. https://doi.org/10.1016/j.ins.2021.02.047

28. J. B. Bak-Coleman, I. Kennedy, M. Wack, A. Beers, J. S. Schafer, E. S. Spiro, et al., Combining interventions to reduce the spread of viral misinformation, *Nat. Hum. Behav.*, **6** (2022), 1372–1380. https://doi.org/10.1038/s41562-022-01388-6

29. Z. Zhao, Y. Liu, K. Wang, An analysis of rumor propagation based on propagation force, *Phys. A Stat. Mech. Appl.*, **443** (2016), 263–271. https://doi.org/10.1016/j.physa.2015.09.060

30. A. Yang, X. Huang, X. Cai, X. Zhu, L. Lu, ILSR rumor spreading model with degree in complex network, *Phys. A Stat. Mech. Appl.*, **531** (2019), 121807. https://doi.org/10.1016/j.physa.2019.121807

31. Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, Y. Li, Cost-efficient strategies for restraining rumor spreading in mobile social networks, *IEEE Trans. Veh. Technol.*, **66** (2016), 2789–2800. https://doi.org/10.1109/TVT.2016.2585591

32. L. Zino, M. Cao, Analysis, prediction, and control of epidemics: A survey from scalar to dynamic network models, *IEEE Circuits Syst. Mag.*, **21** (2021), 4–23. https://doi.org/10.1109/MCAS.2021.3118100

33. J. Chen, L. Yang, X. Yang, Y. Y. Tang, Cost-effective anti-rumor message-pushing schemes, *Phys. A Stat. Mech. Appl.*, **540** (2020), 123085. https://doi.org/10.1016/j.physa.2019.123085

34. R. E. Kopp, Pontryagin maximum principle, *Math. Sci. Eng.*, (1962), 255–279. https://doi.org/10.1016/S0076-5392(08)62095-0

35. S. N. Ha, A nonlinear shooting method for two-point boundary value problems, *Comput. Math. Appl.*, **42** (2001), 1411–1420. https://doi.org/10.1016/S0898-1221(01)00250-4

36. A. V. Rao, A survey of numerical methods for optimal control, *Adv. Astronaut. Sci.*, **135** (2009), 497–528.

37. A. Bodaghi, J. Oliveira, The characteristics of rumor spreaders on Twitter: A quantitative analysis on real data, *Comput. Commun.*, **160** (2020), 674–687. https://doi.org/10.1016/j.comcom.2020.07.017

38. Z. Yu, S. Lu, D. Wang, Z. Li, Modeling and analysis of rumor propagation in social networks, *Inform. Sci.*, **580** (2021), 857–873. https://doi.org/10.1016/j.ins.2021.09.012

39. M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, B. On, Fake news stance detection using deep learning architecture (CNN-LSTM), *IEEE Access*, **8** (2020), 156695–156706. https://doi.org/10.1109/ACCESS.2020.3019735

40. M. Yglesias, This is the real truth about journalists' pay, *Vox*, 2015.

41. Twitter Usage Statistics. Available from: https://www.internetlivestats.com/twitter-statistics/.

42. S. Antoniadis, I. Litou, V. Kalogeraki, A model for identifying misinformation in online social networks, in *On the Move to Meaningful Internet Systems: OTM 2015 Conferences: Confederated International Conferences*, Springer, (2015), 473–482. https://doi.org/10.1007/978-3-319-26148-5_32

43. How Much Does a Cloud Server Cost for a Small Business. Available from: https://siriusofficesolutions.com/cloud-server-price/.

44. Y. Feng, J. Li, L. Jiao, X. Wu, Towards learning-based, content-agnostic detection of social bot traffic, *IEEE Trans. Dependable Secure Comput.*, **18** (2020), 2149–2163. https://doi.org/10.1109/TDSC.2020.3047399

45. D. Huang, L. Yang, P. Li, X. Yang, Y. Y. Tang, Developing cost-effective rumor-refuting strategy through game-theoretic approach, *IEEE Syst. J.*, **15** (2020), 5034–5045. https://doi.org/10.1109/JSYST.2020.3020078

46. D. Huang, L. Yang, X. Yang, Y. Y. Tang, J. Bi, Defending against online social network rumors through optimal control approach, *Discrete Dyn. Nat. Soc.*, **2020** (2020), 1–13. https://doi.org/10.1155/2020/6263748

47. S. Asur, B. A. Huberman, G. Szabo, C. Wang, Trends in social media: Persistence and decay, in *Proceedings of the International AAAI Conference on Web and Social Media*, (2011), 434–437. https://doi.org/10.1609/icwsm.v5i1.14167

Appendix

Proof of Theorem 1

Proof. Let Δt be a small time interval. According to the methodology summarized in [32], the following hold true for any time $t \in [0, T]$:

- At any time t, the average rate at which a user transitions from a reserved attitude to a supportive attitude because of contacting a user who holds a supportive attitude is $\alpha'' = \alpha'[1 - f_2(u_2(t))]$.
- During the time horizon $[t, t + \Delta t]$, the expected number of human users who transition from a reserved attitude to a supportive attitude due to disinformation spread is $\Delta_{RS}(t) = \alpha'' k \Delta t R(t) \frac{S(t)+B(t)}{N(t)}$.
- During the time horizon $[t, t + \Delta t]$, the expected number of human users who transition from a reserved attitude to a denying attitude due to refutation is $\Delta_{RD1}(t) = f_1(u_1(t))\Delta t R(t)$.
- During the time horizon $[t, t + \Delta t]$, the expected number of human users who transition from a reserved attitude to a denying attitude due to the spread of facts is $\Delta_{RD2}(t) = \beta' k \Delta t R(t) \frac{D(t)}{N(t)}$.
- During the time horizon $[t, t + \Delta t]$, the expected number of human users who transition from a supportive attitude to a denying attitude due to refutation is $\Delta_{SD1} = f_1(u_1(t))\Delta t S(t)$.
- During the time horizon $[t, t + \Delta t]$, the expected number of human users who transition from a supportive attitude to a denying attitude due to the spread of facts is $\Delta_{SD2}(t) = \gamma' k \Delta t S(t) \frac{D(t)}{N(t)}$.
- During the time horizon $[t, t + \Delta t]$, the expected number of newly suspended bots due to bot detection is $\Delta_B(t) = f_3(u_3(t))\Delta t B(t)$.

Then, the changes in the number of various types of accounts during the time horizon $[t, t + \Delta t]$ are calculated by

$$\begin{cases} S(t + \Delta t) - S(t) = \Delta_{RS}(t) - \Delta_{SD1}(t) - \Delta_{SD2}(t), \ 0 \le t \le T, \\ D(t + \Delta t) - D(t) = \Delta_{RD1}(t) + \Delta_{RD2}(t) + \Delta_{SD1}(t) + \Delta_{SD2}(t), \ 0 \le t \le T, \\ B(t + \Delta t) - B(t) = -\Delta_D(t), \ 0 \le t \le T. \end{cases} \tag{6.1}$$

Because

$$\begin{cases} \dfrac{ds}{dt}(t) = \lim\limits_{\Delta t \to 0} \dfrac{S(t+\Delta t) - S(t)}{N(t)\Delta t}, \ 0 \le t \le T, \\ \dfrac{dd}{dt}(t) = \lim\limits_{\Delta t \to 0} \dfrac{D(t+\Delta t) - D(t)}{N(t)\Delta t}, \ 0 \le t \le T, \\ \dfrac{db}{dt}(t) = \lim\limits_{\Delta t \to 0} \dfrac{B(t+\Delta t) - B(t)}{N(t)\Delta t}, \ 0 \le t \le T, \end{cases} \tag{6.2}$$

the dynamic system (3.8) is obtained by calculation. The proof is complete. □

Proof of Theorem 2

Proof. According to PMP, the optimal network state trajectory $x(\cdot)$ satisfies

$$\frac{ds}{dt}(t) = \frac{\partial H}{\partial \lambda^s}(u(t), x(t), \lambda(t)), \ \frac{dd}{dt}(t) = \frac{\partial H}{\partial \lambda^d}(u(t), x(t), \lambda(t)), \ \frac{db}{dt}(t) = \frac{\partial H}{\partial \lambda^b}(u(t), x(t), \lambda(t)), \tag{6.3}$$

for all $0 \le t \le T$. Through calculation, the disinformation propagation model (3.8) holds true exactly. Besides, the optimal co-state function $\lambda(\cdot)$ satisfies

$$\frac{d\lambda^s}{dt}(t) = -\frac{\partial H}{\partial s}(u(t), x(t), \lambda(t)), \ \frac{d\lambda^d}{dt}(t) = -\frac{\partial H}{\partial d}(u(t), x(t), \lambda(t)), \ \frac{d\lambda^b}{dt}(t) = -\frac{\partial H}{\partial b}(u(t), x(t), \lambda(t)), \tag{6.4}$$

for all $0 \le t \le T$. Through calculation, the dynamic system (4.2) is attained. Also, as the terminal time T is fixed whereas the terminal state $x(T)$ is free, there are transversality conditions

$$\lambda^s(T) = \frac{\partial h}{\partial s}(s(T), d(T), b(T)), \ \lambda^d(T) = \frac{\partial h}{\partial d}(s(T), d(T), b(T)), \ \lambda^b(T) = \frac{\partial h}{\partial b}(s(T), d(T), b(T)), \tag{6.5}$$

where $h(s, d, b) = \omega(s_0 + b_0 - s - b)$. So, we can attain $\lambda(T) = (-\omega, 0, -\omega)$. Finally, the condition (4.3) is attained from PMP directly. The proof is complete. □

AIMS Press