Robust fake-post detection against real-coloring adversaries

Khushboo Agarwal[1,*], Veeraruna Kavitha

IEOR, IIT Bombay, Powai, Mumbai, 400076, Maharashtra, India

Abstract

The viral propagation of fake posts on online social networks (OSNs) has become an alarming concern. The paper aims to design control mechanisms for fake post detection while negligibly affecting the propagation of real posts. Towards this, a warning mechanism based on crowd-signals was recently proposed, where all users actively declare the post as real or fake. In this paper, we consider a more realistic framework where users exhibit different adversarial or non-cooperative behaviour: (i) they can independently decide whether to provide their response, (ii) they can choose not to consider the warning signal while providing the response, and (iii) they can be real-coloring adversaries who deliberately declare any post as real. To analyze the post-propagation process in this complex system, we propose and study a new branching process, namely total-current population-dependent branching process with multiple death types. At first, we compare and show that the existing warning mechanism significantly under-performs in the presence of adversaries. Then, we design new mechanisms which remarkably perform better than the existing mechanism by cleverly eliminating the influence of the responses of the adversaries. Finally, we propose another enhanced mechanism which assumes minimal knowledge about the user-specific parameters. The theoretical results are validated using Monte-Carlo simulations.

Keywords: Warning Mechanism, Crowd-Signals, Online Social Networks, Branching Processes, Stochastic Approximation, Ordinary Differential Equations

1. Introduction

The prevalence of online social networks (OSNs), like Facebook or Twitter, is unprecedented today. A variety of content is available on the OSNs for users to consume, which can either be for education, entertainment, advertisement or awareness purposes, among many more. Users also read news on such platforms instead of using classical mediums like newspapers.

One of the reasons for such high usage of OSNs is the ease with which users can access or share information. Further, there is no instant check to ensure that the shared post is authentic. On one hand, this freedom allows users to express their views freely, but at the same time, it provides users with the flexibility to post fake content - the one that contains fabricated (mis)information that propagates through OSNs like authentic posts (see [1]). Once a post is shared on the OSN with an initial set of users, called seed users, the post can be further shared repeatedly by the recipients of the post to the extent of getting viral (the copies of the post grow significantly with time), or the post can get extinct in the initial phase ([2, 3, 4, 5]).

Now, there are several reasons for a fake-post to get viral. Authors in [6] theorize that users may share any information obtained from their reliable source, or they can share any exciting post to seek their peers' attention and have a sense of belonging. Also, users share posts that match their beliefs to continue using social media (due to its perceived usefulness). There have been many instances in the past where fake-posts have proven to be fatal, and the most controversial of all is the 2016 US Presidential elections ([7]). Thus, studies on the generation, propagation, detection, and control of fake posts are the need of the hour. In this paper, we focus on the detection aspect of fake-posts.

[*] Corresponding author

Email addresses: agarwal.khushboo@iitb.ac.in (Khushboo Agarwal), vkavitha@iitb.ac.in (Veeraruna Kavitha)

[1] The work of first author is partially supported by Prime Minister's Research Fellowship (PMRF), India.

Machine learning or deep learning is one of the commonly used approaches for fake-post detection (see [8, 9, 10, 11]). However, as argued in [11], such algorithms often face difficulty in obtaining training datasets in certain languages, and it gets difficult to determine the actuality using only the content ([9]). Another approach used for fake-post identification is using crowd-signals. The basic idea is to allow users to declare any post as real or fake, and then leverage user responses to identify the actuality of the post. Such an approach is being used by Facebook[2], where any user can report any post on the OSN. They can also provide specific reasons for reporting the post. When a post is reported, it is reviewed by third-party fact-checking organizations and is removed if it is against their policies. However, until the post is reviewed, the users on the OSN can view it without any warning.

In [12], the authors design a warning-based mechanism to control fake-posts using crowd-signals. The idea is to leverage users' fake/real responses (tags) to the post and generate a warning signal for future recipients. Since the real-time warning signal/status of the post is continuously displayed to the users, this approach of using crowd-signals is different and should be more effective than that of Facebook. The objective is to ensure the maximal correct identification of the fake-post, while maintaining the proportion of fake-tags for the real-post within a given threshold. The paper assumes that each user participates in the tagging process.

In this paper, we consider a more realistic framework. Firstly, we assume that not all users would be willing to tag. Secondly, if a user tags, it can consider the warning signal provided by the OSN; or it can tag without viewing the warning. And lastly, the users can be adversarial- these users always assign the real-tag to any post.

For such a system, we compare and show that the warning mechanism in [12] is insufficient. With just 1% (with 2%) adversaries in the system, while everyone else tagging exactly as in [12], we observed that the performance decreases approximately by 10% (nearly 18.2%). This observation highlights the need for mechanisms which are robust against adversaries. We precisely achieve the same in this paper.

The new warning mechanisms are designed by cleverly eliminating the effect of adversarial users. We derive a one-dimensional ordinary differential equation (ODE) that captures the performance of any such general warning mechanism, and utilizing that ODE, we design the new warning mechanisms as well as illustrate the improved performance guarantees theoretically.

We have also presented Monte-Carlo simulation-based exhaustive numerical study to confirm our theoretical findings. The performance is expressed in two ways: (i) quality of service (QoS) which measures the proportion of fake-tags for the fake-post, and (ii) improved QoS (i-QoS) which represents the proportions only from non-adversarial users. The second metric i-QoS provides better interpretation for the performance of warning mechanisms, as actions of adversarial users can not be controlled. Note that, accordingly the threshold with respect to the real-post also changes, to consider the responses only from non-adversarial users.

According to the parameters in [12], the non-adversarial users are assumed to be smart (i.e., have high intrinsic ability to identify the actuality of the posts). Thus, no warning mechanism can accentuate their ability beyond a limit – we observe minor improvements in QoS of 2.66% and 5.34% with 1% and 2% of adversary respectively; these numbers translate to 98.64% and 98.63% of i-QoS under new mechanisms as compared to 95.8% and 92.53% with the mechanism as in [12].

In another instance, where users are less informed and more likely to wrongly recognize the posts (as is the case in reality), significant improvements are noticed even for a larger fraction of adversaries. Under newly proposed mechanism, the QoS is 52.89% (i-QoS is 80.86%), which is only 45.31% (i-QoS is only 45.31%) under old mechanism, when 32.5% of adversarial users are involved. In fact, this performance is achieved with minimal knowledge about users sensitivity to the warning, and their behavioural type.

The warning dynamics are modelled using a new variant of branching processes (BPs). This paper also contributes towards total-current population-dependent two-type branching processes with population dependent death rates and also considers a variety of unnatural deaths. In particular, we derive all possible limits and limiting behaviours of the population sizes as time progresses.

Related Literature for Branching processes with unnatural deaths: The literature on BPs has previously investigated unnatural deaths in a restricted setting. The BP analyzed in [13] is population-independent, while the authors in [14] consider unnatural deaths due to competition, modelled using a quadratic function of population size. The BP with pairwise interaction in [15] models natural births and deaths, along with additional births and deaths occurring due to cooperation and competition. Further, the birth and death rates in [15] are proportional to current

[2] https://www.facebook.com/help/1753719584844061

population sizes. Our work provides a much more generalised framework where the interactions are not limited to cooperation or competition. Further, the birth and death rate functions can additionally depend on the total and current population-sizes.

2. Problem description

Consider an OSN with a large user base like Facebook or Twitter. Any post, u on the OSN can be either fake ($u = F$) or real ($u = R$). The OSN aims to identify the actuality of the post. In [12], the authors have proposed a warning mechanism where the recipients of the post themselves are guided in such a way that it leads to correct identification. We first study its robustness against adversarial users and then propose improved mechanisms.

We begin by describing the system and the warning mechanism of [12]. The posts are stored in a last-in-show-at-top structure named timeline for each user. The users are provided a warning for each post, and are asked to assign a tag (fake or real) to it. Whenever a user views the post on its timeline, it guesses the actuality of the post, assigns the tag as real or fake accordingly and then forwards the same to its friends. This results in more unread copies of the post tagged as fake or real. The process continues when another user with the post on its timeline visits the OSN. The warning mechanism relies on the tags provided by the users and is updated with each new tag.

We will now introduce a few notations and then discuss the propagation and tagging dynamics of the post. Let the fake and real tagged copies of the u-post be denoted as x-type and y-type, respectively. Further, let $C_x(t)$ and $C_y(t)$ be the number of users who have received the post tagged as fake and real, respectively but have not yet read/shared it; thus, these are the number of unread copies of the post with fake or real tag. The total number of users who have received the post tagged as fake or real are represented by $T_x(t)$ and $T_y(t)$ respectively; these are read plus unread copies of the post. Let $\Phi(t) := (C_x(t), C_y(t), T_x(t), T_y(t))$ be the tuple of number of copies at time t.

Each post contains two pieces of information: first, the sender's tag and second, the warning by the OSN, which is available at the click of a button (see Figure 1). Users can exhibit different behaviours about utilising the provided information. For example, some users may prefer to read the warning before tagging, while others may not. Therefore, motivated by [16], we broadly divide user behaviour into four categories.

2.1. *Warning-ignoring (wi) users*

These are the users who tag the post only based on the sender's tag and their intrinsic ability to judge the post's actuality, not the warning. They prefer to invest less time in the system. Let τ be the time when a wi-user (with an unread copy of the post) reads it. At this time, the user will tag and then share the post with its friends. Let $I_{x,wi}(\Phi(\tau^-))$ and $I_{y,wi}(\Phi(\tau^-))$ be the indicator that the *wi*-user with fake or real-tagged copy of the post tags it as fake.

If the sender has tagged the post as fake, then the recipient tags the post as fake or real with probability (w.p.) $p_x^u \in (0, 1)$ and $1 - p_x^u$ respectively. Similarly, let $p_y^u \in (0, 1)$ be the probability of fake-tagging the post, received with real-tag. Therefore:

$$P(I_{x,wi}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = p_x^u \text{ and } P(I_{y,wi}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = p_y^u, \quad (1)$$

where $\mathcal{G}_t$ is the sigma-algebra generated by $\{\Phi(t'); t' \leq t\}$. Naturally, the users get more suspicious about the post when received with fake-tag. Thus, we assume $p_x^u > p_y^u$ for any $u \in \{R, F\}$.

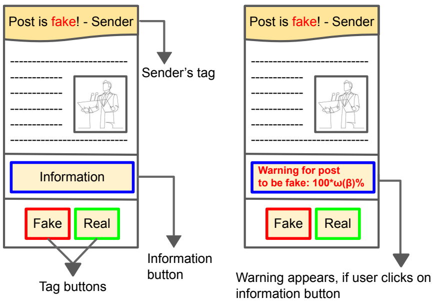

Figure 1: Design of the post

As said before, the user forwards the post to some/all of its friends after tagging. The number of shares depends on how attractive the post is, which we measure by $\eta^u \in (0, 1)$. As argued in [17], the design of fake-posts is deceptive and more appealing; therefore, we assume $\eta^F > \eta^R$.

Let $\mathcal{F}$ be the number of friends of a typical user of the OSN and assume that $\mathcal{F}$ is independent and identically distributed across various users. Let τ^+ and τ^- be the usual limits, e.g., $C_x(\tau^-) := \lim_{t \uparrow \tau} C_x(t)$. When a wi-user receives a post with fake-tag and shares it with fake-tag, it generates $\xi_{xx,wi}$ number of fake-tagged copies. Similarly, when it tags the post as real, it shares to $\xi_{xy,wi}$ friends. Define $\xi_{yx,wi}$ and $\xi_{yy,wi}$ in a similar manner. We assume (k is some constant):

$$\xi_{ix,wi}(\Phi(\tau^-)) = \xi_{iy,wi}(\Phi(\tau^-)) \sim Bin\left(\mathcal{F}, \eta^u + \frac{k}{(Z(\tau^-))^2}\right) \text{ for } i \in \{x, y\}, \tag{2}$$

$Bin(\cdot,\cdot)$ denotes a binomial random variable; many times, users receive the post more than once, however, they may not be interested in it again - thus, the new effective shares in (2) reduces with the total copies/shares of the post generated so far, i.e., $Z(\tau^-) := T_x(\tau^-) + T_y(\tau^-)$, for example as in (2). The distribution considered in (2) is a specific example; however, our analysis can extend to any total-current shares-dependent sharing-distribution that satisfies assumption **C.2** (see Section 3).

2.2. Warning-seeking (ws) users

These users click on the warning button also - they incorporate the sender's tag, their innate capacity and the warning provided by the OSN to decide the tag.

Say a ws-user views the fake-tagged post at time τ. Let ω_τ be the warning at this time. Then, as in [12], we assume that such user tags the post as fake (real) w.p. $\min\{\alpha_x^u \omega_\tau, 1\}$ (respectively, $1 - \min\{\alpha_x^u \omega_\tau, 1\}$) before sharing; here, $\alpha_x^u > 0$ is the sensitivity parameter to the warning when the post is received with fake-tag. Similarly, if the post received by the ws-user has a real-tag, then it tags the post as fake or real w.p. $\min\{\alpha_y^u \omega_\tau, 1\}$ and $1 - \min\{\alpha_y^u \omega_\tau, 1\}$, respectively, where $\alpha_y^u > 0$ is the sensitivity parameter when the post is received with real-tag. Thus, we have:

$$P(I_{x,ws}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = \min\{\alpha_x^u \omega_\tau, 1\} \text{ and } P(I_{y,ws}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = \min\{\alpha_y^u \omega_\tau, 1\}. \tag{3}$$

The *sensitivity parameters are indicative of the user's intrinsic ability to recognize the actuality of the post.* These parameterize *warning-aided identification*, while p_F^u, p_R^u are the probabilities of *un-aided identification*; both are characteristics of the users of the OSN. We thus assume a linear dependence between the two as in [16], i.e., we assume a $\rho \in (0, 1)$ such that

$$p_F^u = \alpha_x^u \rho \text{ and } p_R^u = \alpha_y^u \rho. \tag{4}$$

Now, similar to wi-users, a ws-user also shares the post with its friends. Using notations as in (2), we have (k is some constant):

$$\xi_{ix,ws}(\Phi(\tau^-)) = \xi_{iy,ws}(\Phi(\tau^-)) \sim Bin\left(\mathcal{F}, \eta^u + \frac{k}{(Z(\tau^-))^2}\right) \text{ for } i \in \{x, y\}. \tag{5}$$

2.3. Adversaries (a)

As is usually the case, there can be a small fraction of adversarial users on the OSN. These users aim to harm the efficacy of the system-generated warning by incorrectly tagging the post. Their agenda for doing so can be in self-interest or political. Often, such users do not have prior information about the actuality of the post, but to meet their objective they target to confuse the users about the actuality of the posts. Towards this, we consider that they always tag any post as real. In a way, such users are the ones who wish to color (tag) the posts as real, irrespective of the actuality of the posts.

Let $I_{x,a}(\Phi(\tau^-))$ and $I_{y,a}(\Phi(\tau^-))$ be the indicator that an a-user with a fake or real-tagged copy of the post tags the post as fake, where τ is the time when an a-user views the post. Here, we have:

$$P(I_{x,a}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = P(I_{y,a}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = 0. \tag{6}$$

An adversarial user shares the post with a real-tag to its friends with probability $\eta_a \in (0, 1)$, irrespective of the attractiveness of the post. Therefore, we have (k is some constant):

$$\xi_{ix,a}(\Phi(\tau^-)) \equiv 0 \text{ and } \xi_{iy,a}(\Phi(\tau^-)) \sim Bin\left(\mathcal{F}, \eta_a + \frac{k}{(Z(\tau^-))^2}\right) \text{ for } i \in \{x, y\}. \tag{7}$$

2.4. Non-participants (np)

In [12], it is assumed that all users viewing the post share and tag it. In reality, there can be users named as non-participants who neither participate in the tagging process nor share the post. In other words, when they receive a copy of the post, they do not respond, which we capture as:

$$P(I_{i,np}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = P(I_{i,np}(\Phi(\tau^-)) = 1|\mathcal{G}_\tau) = 0, \tag{8}$$

and shares to none, i.e.,

$$\xi_{ix,np}(\Phi(\tau^-)) = \xi_{iy,np}(\Phi(\tau^-)) \equiv 0, \text{ for } i \in \{x, y\}. \tag{9}$$

Number of shares: Let $\mathcal{U} := \{\text{wi, ws, a, np}\}$ be the set of types of users in the system. Let $\mu_0, \mu_1, \mu_2, \mu_a$ be the respective proportions of np, wi, ws, a-users on the OSN such that $\mu_1 + \mu_2 + \mu_a + \mu_0 = 1$; *we assume that the OSN knows these proportions*. Since our approach is based on crowd-signals, therefore, it is meaningful to assume that $\mu_2 \in (0, 1)$. Any user of the OSN visits it after a random time which is exponentially distributed with parameter 1 (without loss of generality); this is a commonly made assumption in the literature (see, for example, [3, 18, 5]). If required, one can model different users visiting the OSN at different rates, for example, a-users might visit more often; our framework can easily extend to such a case. Any user of j-type, after viewing the post with fake-tag ($i = x$) or real-tag ($i = y$), generates $\Gamma_{ix,j}$ and $\Gamma_{iy,j}$ number of new fake and real-tagged copies of the post respectively, where:

$$\begin{aligned} \Gamma_{ix,j}(\Phi(\tau^-)) &:= I_{i,j}(\Phi(\tau^-))\xi_{ix,j}(\Phi(\tau^-)), \text{ and} \\ \Gamma_{iy,j}(\Phi(\tau^-)) &:= \Big(1 - I_{i,j}(\Phi(\tau^-))\Big)\xi_{iy,j}(\Phi(\tau^-)), \text{ for } i \in \{x, y\} \text{ and } j \in \mathcal{U}. \end{aligned} \tag{10}$$

Next, we discuss some meaningful assumptions (inspired by [12]).

Regime of parameters and assumptions: The probability of a user fake-tagging any u-post is higher when the sender's tag is fake, thus, $\alpha^u_x > \alpha^u_y$, for $u \in \{R, F\}$. We assume that the users are more likely to tag fake-post as fake, as compared to tagging real-post as fake, irrespective of sender's tag, i.e., $\alpha^F_i > \alpha^R_i$, for each $i \in \{x, y\}$. Since the intent of a-users is to share the post rigorously, therefore, we assume $\eta_a > \eta^u$, for each u, only in the numerical experiments; the theoretical results follow even if $\eta_a \leq \eta^u$. Thus, in all, we assume the following:

$$\begin{aligned} &\alpha^u_x > \alpha^u_y > 0, \text{ for each } u \in \{R, F\}, \alpha^F_i > \alpha^R_i \text{ for each } i \in \{x, y\}, \\ &\eta_a > \eta^F > \eta^R > 0, \mu_2 \in (0, 1) \text{ and } \rho \in (0, 1). \end{aligned} \tag{11}$$

For the sake of clarity, we summarize all the notations which will be used consistently throughout the paper:

Sr. No.	Notation	Description
1.	$\mathcal{U} = \{\text{wi, ws, a, np}\}$	types of users: warning-ignoring, warning-seeking, adversarial, non-participating
2.	$\mu_0, \mu_1, \mu_2, \mu_a$	proportion of np, wi, ws and a-users
3.	$u \in \{R, F\}$	actuality of the post as real or fake respectively
4.	η^u, η_a	probability of a user/adversary sharing the post to its friend
5.	x, y	fake or real tag by the sender
6.	α^u_x, α^u_y	sensitivity of a user towards the warning when received with fake or real tag

Table 1: Summary of the notations

2.5. Warning Mechanism (WM) - system-generated warning

In [12], the authors designed a warning mechanism (WM) by leveraging upon the responses of the users. They assumed all users are ws-users and did not consider the adversaries (i.e., $\mu_2 = 1$). The main idea behind the design of the mechanism is to exploit the collective wisdom of the users (via responses of all users), as depicted in Figure 2 (left side). The warning considered in [12] is:

$$\omega_t = \left(\frac{wC_x(t)}{C_x(t) + bC_y(t)} + \gamma \right) = \left(\frac{w\mathrm{B}(t)}{\mathrm{B}(t) + b(1 - \mathrm{B}(t))} + \gamma \right), \text{ where } \mathrm{B}(t) := \frac{C_x(t)}{C_x(t) + C_y(t)} \tag{12}$$

represents the relative fraction of (unread) fake-tagged copies at time t; w and b are the control parameters; $\gamma > 0$ is the parameter which captures the prior knowledge OSN has about the post via some fact-check mechanism. Here, $w \in [0, \overline{w}]$ for $\overline{w} := \frac{1}{\alpha_x^F} - \gamma$. This ensures that a ws-user tags the fake-tagged copy of the post as fake with probability $\min\{\alpha_x^u \omega(\beta), 1\} = \alpha_x^u \omega(\beta)$ for any $\beta \in [0, 1]$, when the warning is as in (12); thus, $\min\{\alpha_y^u \omega(\beta), 1\} = \alpha_y^u \omega(\beta)$ (since $\alpha_y^u < \alpha_x^u$, see (11)). Further, the parameter $b \in [0, \infty)$. The warning in (12) is generated individually for each post.

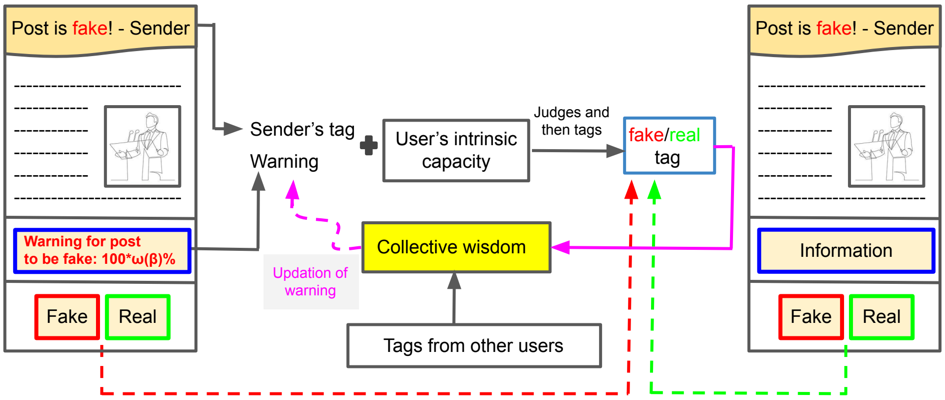

Figure 2: On the left, ws-user tags the post as fake. On the right, a-user tags the post as real, without checking the warning or sender's tag.

In this paper, we are considering a variety of user behavior. Therefore, the warning is now influenced by the responses of users who ignore the warning while tagging or are purposely providing incorrect tags. In Figure 2, we depict that the warning is updated by the response (fake) of the ws-user (left side of the figure) and also by that of a-user (right side of the figure). Similarly, one can visualize how a warning gets updated when a wi-user tags. This suggests that the warning (12) needs to be studied for our complex and more realistic system.

It is clear from the discussion so far that the end goal of the OSN is to nudge users towards the correct identification of the posts. Let $B^u(t)$ represents the proportion of fake-tags, given that the actuality of the post is $u \in \{R, F\}$. Then, similar to [12], we aim to optimally choose w, b:

- to maximize the proportion of fake-tags for the fake-post, max $\lim_{t\to\infty} \mathrm{B}^F(t)$, and
- to ensure that the proportion of fake-tags for the real-post, $\lim_{t\to\infty} \mathrm{B}^R(t)$, is at most δ, for some $\delta \in (0, 1)$.

The above objective is well defined if the limits in the above exist and are unique almost surely. By Theorem 3 stated in section 4, we prove that the limits indeed exist (but need not be unique) for any general warning mechanism. Hence, define $\mathcal{L}^u := \{\lim_{t\to\infty} \mathrm{B}^u(t)\}$ as the set of all possible limits for u-post, across all sample paths, and consider the following optimization problem:

$$\max_{w,b} \inf(\mathcal{L}^F) \text{ subject to } \sup(\mathcal{L}^R) \leq \delta. \tag{13}$$

Further, we shall investigate the following two questions:

1. How does the optimal WM in (12) perform in the presence of wi-users and a-users?
2. If the performance degrades, can we design improved WMs which are robust against adversaries?

2.6. *Warning dynamics and Branching process*

It is clear that when a user tags the post as fake, the fake number of copies (represented by x) gets updated; otherwise, the real (y) number of copies gets updated. Further, the user who receives the post can be one among the type i, for $i \in \mathcal{U}$, w.p. given by the proportion of the type it belongs to; for example, the recipient can be a wi-user w.p. μ_1. As discussed in (2), (5), (7) and (9), the distribution of the number of shares depends on the type of the user who received the post.

Let τ be the time when a type-i user views the post on its timeline with a fake-tag. Then, the number of fake-tagged and real-tagged copies of the underlying post evolves at time τ as follows:

$$\begin{aligned} &C_x(\tau^+) = C_x(\tau^-) - 1 + \Gamma_{xx,i}(\Phi(\tau^-)), C_y(\tau^+) = C_y(\tau^-) + \Gamma_{xy,i}(\Phi(\tau^-)), \\ &T_x(\tau^+) = T_x(\tau^-) + \Gamma_{xx,i}(\Phi(\tau^-)), \text{ and } T_y(\tau^+) = T_y(\tau^-) + \Gamma_{xy,i}(\Phi(\tau^-)). \end{aligned} \tag{14}$$

We argued before that once a user reads a post, it is seldom interested in the same post again; thus, the current (unread) number of fake-tagged copies decreases by 1. Similarly, when a type-i user who received the post with the real-tag views the post, the system evolves as:

$$\begin{aligned} &C_x(\tau^+) = C_x(\tau^-) + \Gamma_{yx,i}(\Phi(\tau^-)), C_y(\tau^+) = C_y(\tau^-) - 1 + \Gamma_{yy,i}(\Phi(\tau^-)), \\ &T_x(\tau^+) = T_x(\tau^-) + \Gamma_{yx,i}(\Phi(\tau^-)), \text{ and } T_y(\tau^+) = T_y(\tau^-) + \Gamma_{yy,i}(\Phi(\tau^-)). \end{aligned} \tag{15}$$

We shall briefly call the above warning-mechanism aided dynamics as warning dynamics. At this point, it is important to state that the dynamics described above can be modelled as a continuous-time total-current population-dependent branching process (TC-BP) of [19], except for varying death rates. We will discuss how such correspondences can be made in Section 4; in particular, we will see that the viewing of the post can be modelled as a death in an appropriate TC-BP and hence, will have different death-types and rates owing to different types of users. However, we first analyze the TC-BPs with multiple death types in the next section using ODE based stochastic approximation technique, which will be instrumental for our study.

Informal outline for design of WMs: We consider any general warning mechanism $\omega(\beta)$, which depends on the proportion of fake-tags (β) provided by the previous recipients of the post. The limiting behaviour of the warning-guided post-propagation process is analyzed using the ODE derived via the analysis of the underlying BP. In particular, we will show that the analysis of a one-dimensional ODE suffices to study the limits of the underlying process; of course, the limits depend upon the warning mechanism utilized. The main idea is to reverse-engineer: consider the design of the warning mechanism (to achieve the desired output), based on the anticipated attractors of the one-dimensional ODE. We will follow this approach in section 4 and thereafter, where we bring our attention back to the control of fake-post propagation over OSNs.

3. Total-Current population-dependent Branching Process (TC-BP) with multiple death types

Consider two types of populations, namely x and y-types, and let $c_{x,0}$ and $c_{y,0}$ be their respective initial sizes. An individual can either die naturally, or it may die differently due to unnatural circumstances. We refer any death which is not natural as 'unnatural death'[3]. Let $D_i := \{0, 1, \ldots, d_i\}$ be the set of variety of deaths for i-type individual, where $d_i \in [0, \infty)$. Here, $d = 0$ represents the natural death and $d \in D_i - \{0\}$ represents an unnatural death; D_x need not equal D_y as some circumstances may affect only one population. We shall briefly refer to the death of variety d as d-death.

Now, given that the interest of this paper is in controlling the fake post propagation over OSNs, our focus is on the time-asymptotic proportion of the population (fake-tags). Therefore, it is sufficient to study the embedded process (discrete-time chain defined at death instances) of the continuous-time Markov process. In [19], the authors analysed the TC-BP using stochastic approximation based approach, where only natural deaths occur. In this section, we will follow the same approach to incorporate different varieties of deaths. We begin by introducing few notations which are exactly as in [19], however are re-written here for the ease of reading.

[3] In biological systems, unnatural deaths may occur due to exposition to a virus, competition with other species, etc. We discuss unnatural deaths for the application at hand in section 4.

Let τ_n be the time at which n-th individual dies. Consider any $n \geq 1$. Let $\Phi_n := (C_{x,n}, C_{y,n}, T_{x,n}, T_{y,n})$, where $C_{x,n}, C_{y,n}$ represent the *current population* and $T_{x,n}, T_{y,n}$ are the *total population* sizes immediately after τ_n, e.g., $C_{x,n} = C_x(\tau_n^+)$. Let $S_n := C_{x,n} + C_{y,n}$ be the sum current population, again immediately after τ_n. Let $\phi = (c_x, c_y, t_x, t_y)$ be a realisation of the random vector Φ. Any individual can die naturally or unnaturally. We assume that the time till d-death of an i-type individual is exponentially distributed with parameter $\lambda_{i,d} \in (0, \infty)$. An individual in the population will die according to the first death (variety) event that occurs. By memoryless property, after any given instance of time (e.g., τ_n), the death-time of any i-type individual in the population is again exponentially distributed with parameter $\sum_d \lambda_{i,d}$, and hence the first death in the two populations is exponentially distributed with parameter $\left(\sum_d \lambda_{x,d} + \sum_d \lambda_{y,d}\right)$. We further assume that the parameter $\lambda_{i,d}$ depends on the population-size, i.e., $\lambda_{i,d}(\phi_n)$, conditioned on ϕ_n, for each $i \in \{x, y\}$. Observe that we have population-dependency even for the natural deaths, in contrast to the classical models studying only population-independent natural deaths (see, for example, [20, 21, 22]).

The current population can get extinct, and thus let $\nu_e := \inf\{n : S_n = 0\}$ be the extinction epoch, with the usual convention that $\nu_e = \infty$, when $S_n > 0$ for all n. *For the sake of completion, define* $\Phi_n := \Phi_{\nu_e}$ *and* $\tau_n := \tau_{\nu_e}$*, for all* $n \geq \nu_e$*, when* $\nu_e < \infty$. We refer the sample paths in which $\nu_e = \infty$ as the non-extinction paths, and the complementary ones as the extinction paths. *Define* $\mathrm{B}_n := C_{x,n}/S_n$ *as the proportion of x-type population among current population.* Let $\beta = c_x/(c_x + c_y)$ be a realisation of B.

3.1. Evolution of embedded process

In classical BPs, each individual lives for a random time which is exponentially distributed with a common parameter (say) $\lambda > 0$. Thus, an individual to die at n-th epoch is of x-type w.p.[4] β_n, conditioned on $\Phi_n = \phi_n$. In similar lines, with the possibility of unnatural deaths, the probability that an i-type individual d-dies is given by:

$$\begin{aligned}
&P(x\text{-type individual } d\text{-dies}|\phi) = \frac{\lambda_{x,d}(\phi)\beta}{d(\phi)} \text{ and} \\
&P(y\text{-type individual } d\text{-dies}|\phi) = \frac{\lambda_{y,d}(\phi)(1-\beta)}{d(\phi)}, \text{ where } d(\phi) := \sum_{d \in D_x} \lambda_{x,d}(\phi)\beta + \sum_{d \in D_y} \lambda_{y,d}(\phi)(1-\beta).
\end{aligned} \tag{16}$$

In all, the overall probability that an i-type individual is the first to die after previous death instance, τ, is given by:

$$P(x\text{-type individual dies}|\phi) = \frac{\beta \sum_{d \in D_x} \lambda_{x,d}(\phi)}{d(\phi)} =: f_\beta(\phi) \text{ and } P(y\text{-type individual dies}|\phi) = 1 - f_\beta(\phi). \tag{17}$$

Say an individual of i-type dies at n-th epoch. Then, the current size (not the total size) of i-type reduces by 1 due to death. Further, if it d-dies for $d \in D_i$, it produces $\Gamma_{ii,d}(\Phi_{n-1})$ and $\Gamma_{ij,d}(\Phi_{n-1})$ offspring of i-type and j-type ($j \neq i$) respectively, conditioned on the sigma algebra $\sigma\{\Phi_{n-1}\}$, where $\Gamma_{ij,d}(\Phi_{n-1})$ is an integer-valued random variable. Basically, when $\Phi_{n-1} = \phi_{n-1}$, the random offspring are represented by $\Gamma_{ij,d}(\phi_{n-1})$ for each i, j and d. Thus, the embedded process immediately after an i-type individual d-dies at n-th epoch is given by:

$$\begin{aligned}
C_n^i &= C_{n-1}^i + \Gamma_{ii,d}(\Phi_{n-1}) - 1, \quad & A_n^i &= A_{n-1}^i + \Gamma_{ii,d}(\Phi_{n-1}), \\
C_n^j &= C_{n-1}^j + \Gamma_{ij,d}(\Phi_{n-1}), & A_n^j &= A_{n-1}^j + \Gamma_{ij,d}(\Phi_{n-1}), \text{ for } i \neq j.
\end{aligned} \tag{18}$$

Now, conditioned on ϕ, we assume the ϕ-dependent random offspring satisfy the following, which also ensures throughout super-criticality, a notion defined in [19]:

C.1 There exist two integrable random variables $\overline{\Gamma}$ and $\underline{\Gamma}$ which bound the random offspring as: $0 \leq \underline{\Gamma} \leq \Gamma_{ix,d}(\phi) + \Gamma_{iy,d}(\phi) \leq \overline{\Gamma}$ almost surely (a.s.), for each ϕ, for each d. Also, $E[\overline{\Gamma}^2] < \infty$ and $E[\underline{\Gamma}] > 1$. Further, $\Gamma_{ii,d}(\phi) \geq 0$ a.s., for each i, ϕ, d. Furthermore, assume that $\inf_\phi \lambda_{x,d}(\phi) > 0$ for each $d \in D_x$ and $\inf_\phi \lambda_{y,d}(\phi) > 0$ for each $d \in D_y$.

[4] This happens due to the memory-less property of exponential distribution and as minimum of k independent and identically distributed exponentially distributed random variables with parameter λ is exponentially distributed with parameter $k\lambda$.

3.2. Mean matrix

Let $m_{ij,d}(\phi) := E[\Gamma_{ij,d}(\phi)]$ *denote the expectation of the number of j-type offspring, when an i-type parent d-dies, conditioned on* ϕ, *for* $i, j \in \{x, y\}$ *and* $d \in D_i$. Further, define the mean matrix $M(\phi) := [m_{ij}(\phi)]_{i,j \in \{x,y\}}$ as given below:

$$M(\phi) := \begin{bmatrix} \frac{\sum_{d \in D_x} \lambda_{x,d}(\phi) m_{xx,d}(\phi)}{\sum_{d \in D_x} \lambda_{x,d}(\phi)} & \frac{\sum_{d \in D_x} \lambda_{x,d}(\phi) m_{xy,d}(\phi)}{\sum_{d \in D_x} \lambda_{x,d}(\phi)} \\ \\ \frac{\sum_{d \in D_y} \lambda_{y,d}(\phi) m_{yx,d}(\phi)}{\sum_{d \in D_y} \lambda_{y,d}(\phi)} & \frac{\sum_{d \in D_y} \lambda_{y,d}(\phi) m_{yy,d}(\phi)}{\sum_{d \in D_y} \lambda_{y,d}(\phi)} \end{bmatrix}. \tag{19}$$

Then, for $j \in \{x, y\}$, we have (see (16), (17) and (19)):

$$\begin{aligned} E[j\text{-type offspring produced by an } x\text{-type parent}|\phi] &= \sum_{d \in D_x} \frac{\lambda_{x,d}(\phi)\beta}{d(\phi)} m_{xj,d}(\phi) = f_\beta(\phi) m_{xj}(\phi), \\ E[j\text{-type offspring produced by a } y\text{-type parent}|\phi] &= \sum_{d \in D_y} \frac{\lambda_{y,d}(\phi)(1-\beta)}{d(\phi)} m_{yj,d}(\phi) = (1 - f_\beta(\phi)) m_{yj}(\phi). \end{aligned} \tag{20}$$

As in [19, Lemma 2, Appendix A], one can prove the dichotomy for the sum current population of TC-BP with multiple death types, as in the following:

Lemma 1. *Assume* ***C.1*** *and define* $\underline{m} =: E[\underline{\Gamma}]$. *Then, we have:*

$$P\left(\left\{\liminf_n S_n e^{-\underline{\lambda}(\underline{m}-1)\tau_n} > 0\right\} \cup \left\{\lim_{n\to\infty} S_n = 0\right\}\right) = 1,$$

where $\underline{\lambda} := \inf_\phi\{\lambda_{x,0}(\phi), \dots, \lambda^x_{d_x}(\phi), \lambda_{y,0}(\phi), \dots, \lambda^y_{d_y}(\phi)\} > 0.$

Thus, the sum current population either gets extinct or in non-extinction paths, it explodes (i.e., it grows exponentially larger at rate $\underline{\lambda}(\underline{m} - 1)$).

3.3. Main Result

We will now provide the first main result of the paper which determines the limit proportion, $\lim_{t\to\infty} \mathrm{B}^c(t)$ in non-extinction paths and additionally, provides the deterministic approximate trajectories for the underlying BP. The result follows in similar lines to [19, Theorem 1], while accommodating some important changes for multiple deaths. As established in Lemma 1, the underlying BP can explode. In such a case, it is a common practice to scale the process appropriately that enables convergence to a finite limit (see, for example, [19, 12]).

To this end, define the scaled ratios $\Psi^c_n := S_n/n$ and $\Theta^c_n := C_{x,n}/n$. Let $Z_n := T_{x,n} + T_{y,n}$ be the total population size immediately after τ_n, and then analogously, define Ψ^a_n and Θ^a_n for the total population. Let $\boldsymbol{\Upsilon}_n := (\Psi^c_n, \Theta^c_n, \Psi^a_n, \Theta^a_n)$, and let $\boldsymbol{\Upsilon}_0 := (s^c_0, c_{x,0}, s^c_0, c_{x,0})$ denote the initial population, where $s^a_0 = s^c_0 := c_{x,0} + c_{y,0}$. Let $\Upsilon := (\psi^c, \theta^c, \psi^a, \theta^a)$ be a realisation of $\boldsymbol{\Upsilon}$.

In [19, **A.2**], the authors assumed that the total-current population-dependent mean functions converge to proportion dependent mean functions, which can further be discontinuous. Similar to that, we assume that the resultant mean functions ($m_{ij}(\phi)$, and not $m_{ij,d}(\phi)$) converge to proportion-dependent mean functions at a certain rate. However, to accommodate for the variety of deaths, we assume that the lifetime parameters of the populations also become proportion-dependent asymptotically (at the same rate of convergence as that of the mean functions).

C.2 Define $\beta(\Upsilon) := \theta^c/\psi^c = c_x/s$. As sum current population, $s \to \infty$:

$$\begin{aligned} |m_{ij}(\phi) - m^\infty_{ij}(\beta(\Upsilon))| &\le \frac{1}{(s)^\alpha}, \text{ for each } i, j \in \{x, y\} \text{ and} \\ |\lambda_{i,d}(\phi) - \lambda^\infty_{i,d}(\beta(\Upsilon))| &\le \frac{1}{(s)^\alpha}, \text{ for each } d \in D_i \text{ for each } i \in \{x, y\}, \text{ for some } \alpha \ge 1. \end{aligned}$$

Further, under **C.2**, the function $f_\beta(\phi)$ converges to $f_\beta^\infty(\beta)$ as given below (see (17)):

$$
\begin{aligned}
|f_\beta(\phi) - f_\beta^\infty(\beta)| \le \frac{1}{(s)^\alpha}, \text{ where } f_\beta^\infty(\beta) := \frac{\beta \sum_{d\in D_x} \lambda^\infty_{x,d}(\beta)}{d^\infty(\beta)} \text{ with}\\
d^\infty(\beta) := \beta \sum_{d\in D_x} \lambda^\infty_{x,d}(\beta) + (1-\beta) \sum_{d\in D_y} \lambda^\infty_{y,d}(\beta).
\end{aligned}
\tag{21}
$$

In all, under **C.1**-**C.2**, we analyze the ratios $\boldsymbol{\Upsilon}_n$ using the solutions of the following ODE:

$$
\begin{aligned}
\dot{\Upsilon} = \mathbf{g}(\Upsilon) = \mathbf{h}(\beta)1_{\{\psi^c>0\}} - \Upsilon, &\text{ where } \mathbf{h}(\beta) := (h^c_\psi, h^c_\theta, h^a_\psi, h^a_\theta), \text{ with}\\
h^c_\psi(\beta) = f_\beta^\infty(\beta)\Big(m^\infty_{xx}(\beta) + m^\infty_{xy}(\beta)\Big) &+ \Big(1 - f_\beta^\infty(\beta)\Big)\Big(m^\infty_{yy}(\beta) + m^\infty_{yx}(\beta)\Big) - 1,\\
h^c_\theta(\beta) = f_\beta^\infty(\beta)\Big(m^\infty_{xx}(\beta) - 1\Big) &+ \Big(1 - f_\beta^\infty(\beta)\Big) m^\infty_{yx}(\beta),\\
h^a_\psi(\beta) = f_\beta^\infty(\beta)\Big(m^\infty_{xx}(\beta) + m^\infty_{xy}(\beta)\Big) &+ \Big(1 - f_\beta^\infty(\beta)\Big)\Big(m^\infty_{yy}(\beta) + m^\infty_{yx}(\beta)\Big) \text{ and}\\
h^a_\theta(\beta) = f_\beta^\infty(\beta) m^\infty_{xx}(\beta) &+ \Big(1 - f_\beta^\infty(\beta)\Big) m^\infty_{yx}(\beta).
\end{aligned}
\tag{22}
$$

Now, exactly as in [19, **A.3**], we assume the following (see [19, Definition 1] for the definition of extended solution):

C.3 There exists a unique solution $\Upsilon(\cdot)$ for ODE (22) in the extended sense over any bounded interval.

As per [19, Definition 2], let $\mathcal{A}$ be the attractor set and $\mathcal{S}$ be the saddle set with respect to the ODE (22). For systems modelling the BPs, the following subset of the combined domain of attraction of $\mathcal{A}$ and $\mathcal{S}$ is relevant (recall the definition of ratios Υ):

$$
\mathcal{D} := \{\Upsilon \in (\mathbb{R}^+)^4 : \theta^c \le \psi^c \le \psi^a, \theta^a \le \psi^a \text{ and } \Upsilon(t) \to \mathcal{A} \cup \mathcal{S} \text{ as } t \to \infty, \text{ if } \Upsilon(0) = \Upsilon\}. \tag{23}
$$

Therefore, we will be interested in initial conditions $\Upsilon(0) \in \mathcal{D}_I$ for the ODE (22).

In [19, Definition 4], we introduced a new notion of limiting behavior of the stochastic process, named 'hovering around the saddle set' - here, the stochastic trajectory visits every neighborhood of $\mathcal{S}$ infinitely often (i.o.), but also leaves some neighborhood of $\mathcal{S}$ i.o. The main result in [19] states that the random trajectory either converges to the attractor set or it converges to/hovers around a special kind of saddle set. In particular, if any non-zero saddle point, $\Upsilon^* \neq \mathbf{0}$, is attracted exponentially to $\mathcal{S}$ along a particular affine sub-space, $\mathbb{S}(\Upsilon^*) := \{\Upsilon : \beta(\Upsilon) = \beta(\Upsilon^*)\}$ and to $\mathcal{A}$ in the remaining space, then such Υ^* are named as (quasi) q-attractor in [19]. We have a similar result for the case with multiple deaths.

Similar to [19], under above definition, we finally assume the following:

C.4 Let $\mathcal{A} \cap \mathcal{D}_I$ be the attractor set and each $\Upsilon \in \mathcal{S} \cap \mathcal{D}_I$ be the q-attractor. Consider $\mathcal{D}$ as in (23) and let $\mathcal{D}_b := \mathcal{D} \cap \{\psi^a \le b\}$, for some $b > 0$, be a compact subset of combined domain of attraction. Assume $p_b := P(\mathcal{V}) > 0$, where $\mathcal{V} := \{\omega : \boldsymbol{\Upsilon}_n(\omega) \in \mathcal{D}_b \text{ i.o.}\}$.

We have the following result:

Theorem 1. *Under **C.1**-**C.3**, we have:*

(i) For every $T > 0$, a.s. there exists a sub-sequence (n_l) such that:

$$
\sup_{k: t_k \in [t_{n_l}, t_{n_l}+T]} d(\boldsymbol{\Upsilon}_k, \Upsilon(t_k - t_{n_l})) \to 0 \textit{ as } l \to \infty, \textit{ where } t_n := \sum_{k=1}^n \frac{1}{k} \textit{ and}
$$

$\Upsilon(\cdot)$ *is the extended solution of ODE* (22) *which starts at* $\Upsilon(0) = \lim_{n_l \to \infty} \boldsymbol{\Upsilon}_{n_l}$.

*(ii) Further, assume **C.4**. Then, $P(C_1 \cup C_2) \ge p_b$, where*

$$
C_1 := \{\boldsymbol{\Upsilon}_n \to (\mathcal{A} \cup \mathcal{S}) \cap \mathcal{D}_I \textit{ as } n \to \infty\}, \textit{ and } C_2 := \{\boldsymbol{\Upsilon}_n \textit{ hovers around } \mathcal{S}\}. \quad \square
$$

The proof of the above Theorem and all forthcoming results will be provided in Appendix A.

3.4. Derivation of attractor and saddle sets

It is evident from Theorem 1 that the limit proportion, $\lim_{n\to\infty} \mathrm{B}_n$ can be deduced if one derives the attractor and saddle (specifically, q-attractor) sets. In [19], the authors proposed a procedure to derive these sets for the ODE (22), when only natural deaths occur. The main idea was to exploit the dependence of limit mean functions on β as in **C.2** and finally, it is showed that the analysis of $\beta(\Upsilon)$-ODE suffices. We extend the same approach for the new process with both natural and unnatural deaths. Towards this, one can derive the following limit β-ODE, using (22):

$$\begin{aligned}\dot{\beta} &= \frac{1}{\psi^c} g_\beta(\beta) 1_{\{\psi^c>0\}}, \text{ where}\\ g_\beta(\beta) &:= -f_\beta^\infty(\beta) m_{xy}^\infty(\beta) + (1-f_\beta^\infty(\beta)) m_{yx}^\infty(\beta) + \beta - f_\beta^\infty(\beta)\\ &\qquad + (1-\beta) f_\beta^\infty(\beta)(m_{xx}^\infty(\beta) + m_{xy}^\infty(\beta)) - \beta(1-f_\beta^\infty(\beta))(m_{yy}^\infty(\beta) + m_{yx}^\infty(\beta)).\end{aligned} \tag{24}$$

Similar to [19], we will also show that *the asymptotic analysis of β is independent of other components of* Υ. In particular, the result stated below shows that the analysis of the following one-dimensional ODE suffices:

$$\dot{\beta} = g_\beta(\beta). \tag{25}$$

Theorem 2. *Consider the interval* $[0,1]$ *such that* $g_\beta(0) \geq 0$ *and* $g_\beta(1) \leq 0$. *Define* $\mathcal{I} := \{x^* : g_\beta(x^*) = 0\}$ *and say* $\mathcal{I} = \{x_i^* : 1 \leq i \leq n\}$, *for some* $1 \leq n < \infty$. *For each i, let there exist an open/closed/half-open non-empty interval around* $x_i^* \in \mathcal{I}$, *say* $\mathcal{N}_i^*$, *such that* $\cup_{1\leq i\leq n} \mathcal{N}_i^* = [0,1]$ *and* $\mathcal{N}_i^* \cap \mathcal{N}_j^* = \emptyset$ *for* $i \neq j$. *Define* $\mathcal{N}_i^- := \mathcal{N}_i^* \cap [0, x_i^*)$ *and* $\mathcal{N}_i^+ := \mathcal{N}_i^* \cap (x_i^*, 1]$. *Let* $g_\beta(x)$ *be Lipschitz continuous on* $\mathcal{N}_i^-$ *and* $\mathcal{N}_i^+$ *for each i:*

(i) if $g_\beta(x) > 0$ *for all* $x \in \mathcal{N}_i^-$, $g_\beta(x) < 0$ *for all* $x \in \mathcal{N}_i^+$, *then,* x_i^* *is an attractor for ODE* (25)*;*

(ii) if $g_\beta(x) < 0$ *for all* $x \in \mathcal{N}_i^-$ *and* $g_\beta(x) > 0$ *for all* $x \in \mathcal{N}_i^+$, *then,* x_i^* *is a repeller for ODE* (25)*;*

(iii) else if $g_\beta(x) > 0$ *(or* $g_\beta(x) < 0$*) for all* $x \in \mathcal{N}_i^-$ *and* $g_\beta(x) > 0$ *(or* $g_\beta(x) < 0$ *respectively) for all* $x \in \mathcal{N}_i^+$, *then,* x_i^* *is a saddle point for ODE* (25).

Further, ODE (22) *satisfies* ***C.3***. *Furthermore, the attractor and saddle sets in* $\mathcal{D}_I$ *are respectively given by:*

$$\begin{aligned}\mathcal{A} &:= \{\mathbf{h}(x^*) : x^* \in \mathcal{I} \text{ is an attractor for the ODE } (25)\},\\ \mathcal{S} &:= \{\mathbf{h}(x^*) : x^* \in \mathcal{I} \text{ is a repeller or saddle point for the ODE } (25)\} \cup \{\mathbf{0}\}, \text{ and}\end{aligned}$$

entire $\mathcal{D}_I$ *is the combined domain of attraction for* (22). □

The above result provides the limiting behaviour of a one-dimensional ODE with possibly discontinuous right hand sides, that typically arises while studying our type of application. The condition $g_\beta(0) \geq 0$ and $g_\beta(1) \leq 0$ ensures that the interval $[0,1]$ is positive invariant for the ODE (25). It is important to note that in [19, Theorem 2], the authors consider the function g_β such that its zeroes could be either attractors or repellers for the ODE (25). The above result is an extension of the former as here the zeroes of the function g_β can be either attractors or repellers or saddle points for the ODE (25). Such an extended result is required for the application at hand as we will see in the coming sections.

4. Modelling of warning dynamics using TC-BP with multiple deaths

We begin this section by demonstrating how the warning dynamics can be modelled using TC-BP with multiple deaths discussed in the previous section. Towards this, we model the copies with fake and real tags as the x and y-type populations respectively. The time instance when a user views, tags and shares the post corresponds to the time of death of an individual in the BP. As seen in section 2, in (1)-(10), the distribution of shares, types of shares, etc., depends on the type-d of the user that reads the post with $d \in \mathcal{U}$. Thus, one can correspond each d-type user to a d-death because of the following details. When a d-type user reads and shares the post, the said post becomes a read copy, resulting in a d-death. Further, clearly $D_x = D_y = \mathcal{U}$. At any given time, the proportions of the users of any type are given by μ_0, μ_1, μ_2 and μ_a, which also correspondingly represent the proportions of unread copies with np, wi, ws and a-users. Thus, one can easily infer that a type-d user reads the post first among the existing unread copies, or in other words, d-type death occurs first with probability $\mu_d/(\mu_0 + \mu_1 + \mu_2 + \mu_a) = \mu_d$. Therefore, one can set the parameter of d-death as:

$$\lambda_{z,d}(\phi) := \mu_d \text{ for all } \phi, \text{ for each } d \in \mathcal{U} \text{ and } z \in \{x, y\}. \tag{26}$$

Now, after viewing the post, if a ws-user with fake-tagged copy shares the post with fake-tag, then we say that the number of shares, $\Gamma_{xx,ws}$, corresponds to the number of x-type offspring produced by an x-type parent, when ws-death occurs. In general, the number of shares with fake and real-tag correspond to offspring of x and y-type respectively, see (10); the number of shares (offspring) also depend upon $\Phi(\cdot)$.

The underlying TC-BP with multiple death-types that models the warning dynamics (10) is exactly like the well-known irreducible BP, except for the inclusion of multiple death-types (see [23]). In irreducible BPs, the extinction occurs only when both the population-types die; individual extinction of a population-type is not possible. The same is the case with our model. For example, say there are no unread copies with fake-tag, i.e., $C_x(t) = 0$ at some time $t > 0$, while the system still has real-tagged unread copies ($C_y(t) > 0$). Then, if at some time $t' > t$, a wi-user or ws-user reads and shares the post, then, with non-zero probability, it can tag the post as fake (see (1), (3)). If so happens, then it will lead to new unread copies with fake-tag, i.e., $C_x(t') > 0$. Thus, *the number of fake-tagged copies can be regenerated even after they are not present on the OSN, as long as there are some unread copies of the post on the OSN.*

Next, we provide the general framework for analyzing the warning dynamics with respect to any warning mechanism (ω). Observe that when any real/fake post gets extinct, then it's effect is harmless. Therefore, our focus shall only be on the non-extinction paths.

4.1. Analysis of warning dynamics for general WM

Consider a general warning mechanism defined using a continuous-function $\omega : [0, 1] \to \mathbb{R}^+$ which depends only on the proportion of fake-tags β. Further, consider any post with actuality, $u \in \{R, F\}$. Then, for each u, it is clear from the previous section that the analysis of the TC-BP with multiple-death types, and hence the warning dynamics, is driven by the limit mean matrix (see (19) and **C.2**). Thus, we first construct the limit mean matrix, $M^{\infty,u}(\beta) := [m^{\infty,u}_{ij}(\beta)]_{\{i,j\in\{x,y\}\}}$, as follows (see (1)-(10)):

$$M^{\infty,u}(\beta) = \begin{bmatrix} \Big(\mu_1\rho\alpha^u_x + \mu_2 \min\{\omega(\beta)\alpha^u_x, 1\}\Big)m_f\eta^u & \Big(\mu_1(1-\alpha^u_x\rho) + \mu_2(1-\min\{\omega(\beta)\alpha^u_x, 1\})\Big)m_f\eta^u + \mu_a m_f \eta_a \\ \Big(\mu_1\rho\alpha^u_y + \mu_2 \min\{\omega(\beta)\alpha^u_y, 1\}\Big)m_f\eta^u & \Big(\mu_1(1-\alpha^u_y\rho) + \mu_2(1-\min\{\omega(\beta)\alpha^u_y, 1\})\Big)m_f\eta^u + \mu_a m_f \eta_a \end{bmatrix}. \tag{27}$$

Next, we will identify the attractor, repeller and saddle sets for the ODE (25) which will lead to the limits for the stochastic trajectory corresponding to the warning dynamics by using Theorem 2 and Theorem 1. Towards this, observe that $d(\phi) = d^\infty(\beta) = 1$, as $\sum_d \lambda_{z,d}(\phi) = 1$ for any ϕ and $z \in \{x, y\}$. This implies, $f^\infty_\beta(\beta) = \beta$ (see (21)). Thus, by (24), the function g^u_β and the corresponding ODE (25) for the warning dynamics for both types of posts, $u \in \{R, F\}$, is given by:

$$\dot{\beta}^u = g^u_\beta(\beta) \text{ where} \tag{28}$$

$$g^u_\beta(\beta) := \Big(-\beta\mu_2 - \beta\mu_1(1-\alpha^u_x\rho) + (1-\beta)\mu_1\rho\alpha^u_y + \mu_2\Big(\beta\min\{\omega(\beta)\alpha^u_x, 1\} + (1-\beta)\min\{\omega(\beta)\alpha^u_y, 1\}\Big)\Big)m_f\eta^u - \beta\mu_a m_f\eta_a.$$

Define $\mathcal{A}^u_\beta$ as the set of attractors in $[0, 1]$ and $\mathcal{S}^u_\beta$ as the combined set of repellers and saddle points in $[0, 1]$ for the above ODE. Then, we have the following result:

Theorem 3. *Consider the warning dynamics as in* (14) *and* (15)*. Let the distribution of number of friends, $\mathcal{F} \geq 0$ be such that $E[\mathcal{F}]\eta_R > 1$ and $E[\mathcal{F}^2] < \infty$. Then, the following statements are true for each u, the actuality of post:*

*(i) the assumptions **C.2** and **C.3** hold for the ODE* (22)*; hence Theorem 1(i) is true,*

(ii) the set $\mathcal{A}^u_\beta \neq \emptyset$, and then $\boldsymbol{\Upsilon}_n$ converges to $\mathcal{A}^u \cup \mathcal{S}^u$, as $n \to \infty$ or hovers around $\mathcal{S}^u$ w.p. 1*, where $\mathcal{A}^u = \{\mathbf{h}(\beta) : \beta \in \mathcal{A}^u_\beta\}$ and $\mathcal{S}^u = \{\mathbf{0}\} \cup \{\mathbf{h}(\beta) : \beta \in \mathcal{S}^u_\beta\}$.*

(iii) *Further, any potential limit proportion corresponding to the warning dynamics, i.e., $\beta \in \mathcal{A}^u_\beta \cup \mathcal{S}^u_\beta$, can be bounded as below:*

$$0 < \frac{\mu_1 \rho \alpha^u_y \eta^u}{q^u} =: \underline{\beta}^u < \beta \leq \overline{\beta}^u := \frac{(\mu_2 + \mu_1 \rho \alpha^u_y)\eta^u}{q^u} \leq 1, \text{ where} \tag{29}$$

the constant $q^u := \Big(\mu_2 + \mu_1(1 - (\alpha^u_x - \alpha^u_y)\rho)\Big)\eta^u + \mu_a \eta_a.$ □

At first, observe that any warning mechanism ω only affects the likelihood of tagging the post as real or fake by a ws-user (see (3)). It does not affect the probability of a post getting viral or extinct as extinction depends on the sum current number of unread copies (i.e., sum current population in the BP). Now, given that our interest lies in non-extinction paths, the above Theorem gives a generalised result which holds for any warning dynamics. It is important to note that viral paths are possible only when the probability of non-extinction is non-zero; this is possible if $E[\mathcal{F}]\eta_R > 1$, as then the TC-BP with multiple deaths can be in throughout super-critical regime (see Lemma 1).

Theorem 3(i) implies that the warning dynamics can be approximated by the solution of the ODE (28) over any finite-time window, where the limit mean functions are given by (27). The more important result for our context is the second part of the Theorem which states that the stochastic trajectory $\boldsymbol{\Upsilon}_n$ either converges to $\mathcal{A}^u \cup \mathcal{S}^u$ or hovers around $\mathcal{S}^u$. The set $\mathcal{S}^u$ contains $\mathbf{0}$ which represents the limiting behavior of the stochastic trajectory in the extinction paths. *Thus, all the results henceforth will focus on deriving the limits which are not equal to* $\mathbf{0}$*, which in turn provide the limit proportion of fake-tags for the warning dynamics in non-extinction paths.*

Further, Theorem 3 provides the above limits using the zeroes $\{\beta^{\infty,u}\}$ of g^u_β (see (28)). Now, observe that the function g^u_β and therefore the zeroes $\{\beta^{\infty,u}\}$ depends on χ, where $\chi := \{\mu_1, \mu_2, \mu_a, b, w\}$ is the set of parameters. For some warning mechanisms, the function g^u_β can have multiple zeroes, and the warning dynamics can converge to one of them. Thus, one would want to ensure that the maximum limit proportion of fake-tags for the real-post is within a given limit and optimise the minimum proportion of fake-tags for the fake-post. This aspect is considered in the optimization problem (33) given in the next section. In this context, we define the following Quality of Service (QoS) for any warning mechanism:

$$Q := \inf\{\beta : \beta \in \mathcal{A}^F_\beta \cup \mathcal{S}^F_\beta\}. \tag{30}$$

Observe here that $Q = \inf(\mathcal{L}^F)$, the objective function of (13) and is the almost sure lower bound on the limit proportion of fake-tags when the underlying post is fake. It measures the minimal extent to which a fake-post is identified by the users. From (29) of Theorem 3, $Q \in (\underline{\beta}^F, \overline{\beta}^F]$. We would see in the coming sections how (optimal) Q varies with different warning mechanisms.

Next, in Theorem 4, we will derive some properties of $\{\beta^{\infty,u}\}$ with respect to each parameter in χ, when g^u_β has a unique zero. This result will be instrumental in deriving important results in the coming sections. To keep it simple, we shall write $\beta^{\infty,u}(\kappa)$ and $g^u_\beta(\beta;\kappa)$ to show the dependency on the required parameter κ, an element of the tuple χ. Towards this, we require the following difference term (note that $g^u_\beta(\beta^\infty(\kappa);\kappa) = 0$):

$$\nabla^u(\kappa, \partial\kappa) := g^u_\beta(\beta^\infty(\kappa); \kappa + \partial\kappa) - g^u_\beta(\beta^\infty(\kappa);\kappa) = g^u_\beta(\beta^\infty(\kappa); \kappa + \partial\kappa). \tag{31}$$

Theorem 4. *Consider any warning mechanism, $\omega(\beta)$. Let κ be any parameter. Let $g^u_\beta(\beta;\kappa)$ be either a convex or concave or linear function of β with a unique zero, $\beta^{\infty,u}(\kappa) \in (0,1)$, for each $u \in \{R, F\}$. Keeping all parameters in χ fixed, other than κ, if difference term $\nabla^u(\kappa, \partial\kappa) > 0$ for some $\partial\kappa > 0$, then $\beta^{\infty,u}(\kappa + \partial\kappa) > \beta^{\infty,u}(\kappa)$. Else if $\nabla^u(\kappa, \delta\kappa) < 0$, then $\beta^{\infty,u}(\kappa + \partial\kappa) < \beta^{\infty,u}(\kappa)$. Else, $\beta^{\infty,u}(\kappa + \partial\kappa) = \beta^{\infty,u}(\kappa)$.* □

Hereon, we will analyse the warning dynamics for some specific mechanisms.

5. Analysis of Extended Original WM (eo-WM)

In this section, we will analyse the warning dynamics when the OSN provides the warning as in (12), which is originally proposed in [12]. Recall that in [12], the system has only ws-users who interact with the warning mechanism. Since we study the original mechanism (12) under the influence of a variety of user behaviour, we refer to ω as extended original warning mechanism (eo-WM) in our context.

Consider any post with actuality $u \in \{R, F\}$. Recall that we have $w \leq \overline{w} := \frac{1}{\alpha_x^F} - \gamma$, thus leading to $\alpha_j^u \omega(\beta) < 1$ for each $j \in \{x, y\}$ and for any $\beta \in [0, 1]$. We begin the analysis by analyzing the ODE (28) for the eo-WM. The g_β^u defined in (28) for the eo-WM, henceforth denoted as $g_\beta^{o,u}$, is as given below:

$$g_\beta^{o,u}(\beta) = -\beta\mu_2 m_f \eta^u - \beta\mu_1(1 - \alpha_x^u\rho)m_f\eta^u + (1-\beta)\mu_1\rho\alpha_y^u m_f\eta^u + \mu_2\omega(\beta)\Big(\beta\alpha_x^u + (1-\beta)\alpha_y^u\Big)m_f\eta^u - \beta\mu_a m_f\eta_a. \tag{32}$$

Let $\mathcal{A}_\beta^{o,u} \subset [0, 1]$ be the corresponding attractor set and $\mathcal{S}_\beta^{o,u} \subset [0, 1]$ be the union of the corresponding repeller and saddle sets, i.e., with respect to ODE $\dot{\beta}^u = g_\beta^{o,u}(\beta)$. We study these sets in the following.

Corollary 1. *There exists a unique zero, $\beta^{o,\infty,u}$, of $g_\beta^{o,u}$ in* $[0, 1]$. *Further, $\beta^{o,\infty,u} \in (0, 1)$, $\mathcal{A}_\beta^{o,u} = \{\beta^{o,\infty,u}\}$ and $\mathcal{S}_\beta^{o,u} = \emptyset$.* □

Thus, there is a unique attractor, $\beta^{o,\infty,u}$, of ODE (28). By Theorem 3, the stochastic trajectory $\boldsymbol{\Upsilon}_n$ under eo-WM either converges to $\mathbf{h}(\beta^{o,\infty,u})$ or $\mathbf{0}$, or hovers around $\mathbf{0}$ almost surely. We re-iterate that our focus is on the non-extinction paths, and thus, the relevant proportion of fake-tags is unique and equals $\beta^{o,\infty,u}$. Further, by Theorem 3, for the given choice of w, b and given $\mu_1, \mu_2, \mu_a \in \chi$, $\beta^{o,\infty,u} \in (\underline{\beta}^u, \overline{\beta}^u]$.

We now consider the following robust optimization problem for the OSN discussed before:

$$\sup_{w\in[0,\overline{w}],b\in[0,\infty)} \quad \beta^{o,\infty,F}(w, b) \text{ subject to } \beta^{o,\infty,R}(w, b) \leq \delta, \text{ for some } \delta \in (0, 1). \tag{33}$$

By uniqueness of the attractors in the non-extinction paths, the above constrained optimization problem optimizes the QoS defined in (30), $Q = \beta^{o,\infty,F}$ under eo-WM by choosing w, b, while ensuring that the unique zero for the real-post, $\beta^{o,\infty,R} \leq \delta$. The problem in (33) is exactly the same as in [12], but for the inclusion of different user behaviour in our model. Thus, we need to extend the solution of [12] to the case that includes wi, ws, a, and np-users. Observe that δ is a design parameter for the OSN.

Before we solve the above problem, we observe the following qualitative behaviour which is true by the virtue of Theorem 4 – this behavior is important for further analysis:

Corollary 2. *For each $u \in \{R, F\}$, the limit $\beta^{o,\infty,u}(w, b)$ strictly increases with w and strictly decreases with b.* □

The above Corollary intuitively indicates to choose the largest w, i.e., $\overline{w}$ and the smallest b, i.e., 0. However, since the optimal w, b needs to satisfy the constraint for the real-post as in (33), therefore, formal analysis is required.

Theorem 5. [Optimal eo-warning design] *The following statements hold for the optimization problem* (33)*:*

(i) If $\beta^{o,\infty,R}(\overline{w}, 0) > \delta$, then the optimizer (w^*, b^*) of (33) *is as below and satisfies $\beta^{o,\infty,R}(w^*, b^*) = \delta$:*

$$w^* = \overline{w} \text{ and } b^* = \left(\frac{\delta}{1-\delta}\right)\left(\frac{w^*\eta^R\mu_2(\delta\alpha_x^R + (1-\delta)\alpha_y^R)}{\delta((\mu_1+\mu_2)\eta^R + \mu_a\eta_a) - \eta^R(\mu_1\rho + \mu_2\gamma)(\delta\alpha_x^R + (1-\delta)\alpha_y^R)} - 1\right). \tag{34}$$

(ii) Else, if $\beta^{o,\infty,R}(\overline{w}, 0) \leq \delta$, then $(w^*, b^*) = (\overline{w}, 0)$ and satisfies $\beta^{o,\infty,R}(w^*, b^*) \leq \delta$. □

Thus, as anticipated, $w^* = \overline{w}$. Interestingly, contrary to the expectation, b^* is not always 0. If $\beta^{o,\infty,R}(\overline{w}, 0) > \delta$, then the optimal choice for b is given by $b^* > 0$. Such b^* is achieved by solving for $\beta^{o,\infty,R}(w^*, b) = \delta$, i.e., relaxing the constraint for the real-post to the maximum δ-level in a bid to achieve the maximum $\beta^{o,\infty,F}$ for fake-post at optimality. In view of Corollary 2, it is then easy to see that, $\beta^{o,\infty,F}(w^*, b^*) < \beta^{o,\infty,F}(w^*, 0)$, when $b^* > 0$. For simpler notations, henceforth we will refer to $\beta^{o,\infty,F}(w^*, b^*)$ as β^o and $\beta^{o,\infty,R}(w^*, b^*)$ as $\beta^{o,R}$.

In [12], the optimization problem (33) is solved partially. Firstly, only the case with the hypothesis of Theorem 5(i) is analyzed. It is shown that the optimal value is achieved for the value of b which satisfies $\beta^{o,\infty,R} = \delta$. However, the optimal choice of w is not derived; rather a projected gradient descent algorithm is suggested to evaluate w^*. Furthermore, [12] considers $w \in [0, 1]$, while one can allow w to be as large as $\overline{w}$, which can be larger than 1. As we have proved that $w^* = \overline{w}$, therefore, *our optimal eo-WM should perform better than the optimal WM designed in [12].* We numerically show this aspect in the next sub-section.

5.1. *QoS under eo-WM*

It is clear from Corollary 1 and Theorem 5 that the QoS under optimal eo-WM, say Q^o equals β^o. Now, fix any configuration,

$$C := \left\{ \{\alpha_i^u\}_{\{i \in \{x,y\}, u \in \{F,R\}\}}, \eta_a, \{\eta^u\}_{\{u \in \{F,R\}\}}, \rho, \gamma, m_f, w, \mu_1, \mu_2 \right\},$$

and let μ_a vary. Then, we want to investigate how Q^o changes with μ_a. Towards this, define:

$$\beta^o_{\text{na}} := \beta^o(\mu_a = 0) = Q^o(\mu_a = 0), \tag{35}$$

as the proportion of fake-tags for the fake-post at optimality when there is no adversary. Recall that a-users deliberately tag any post as real. Therefore, one can anticipate that the OSN achieves the maximum QoS when there is no adversary, i.e., $\beta^o(\mu_a) = Q^o(\mu_a) < \beta^o_{\text{na}}$, when $\mu_a > 0$. We precisely prove this in the next result for an appropriate range of δ.

Corollary 3. *For given configuration C, there exists a $\bar{\delta} > 0$ such that $\beta^o(\mu_a) < \beta^o_{na}$ for all $\delta \leq \bar{\delta}$, for each $\mu_a \in (0, 1 - \mu_1 - \mu_2]$.* □

Thus, the above corollary confirms our anticipation that the performance degrades with introduction of the a-users in the system, however for a smaller range of δ; observe that the OSN is interested in keeping δ as small as possible, therefore, such choices of δ are indeed meaningful. *Henceforth, we consider such δ, i.e., $\delta \leq \bar{\delta}$.* In the next subsection, we will validate this result numerically and reinforce the requirement to design better WMs in the presence of adversaries.

5.2. *Numerical analysis for eo-WM*

At first, we would like to compare eo-WM with the mechanism in [12] with just a-users added – in the first example, any user on the OSN can either be a ws-user or an a-user ($\mu_2 + \mu_a = 1$). Thus, there is no wi-user and everyone participates. Further, let the parameters be as in [12]:

$$m_f = 28, \eta^F = 0.08, \eta^R = 0.05, \gamma = 0.1, \eta_a = 0.55, \delta = 0.02, \alpha_x^F = 0.85, \alpha_y^F = 0.6375, \alpha_x^R = 0.3 \text{ and } \alpha_y^R = 0.09. \tag{36}$$

For such parameters, we perform Monte-Carlo (MC) simulation, and also evaluate the zeroes of $g_\beta^{o,u}$ for each $u \in \{R, F\}$. In Figure 3, we plot the outputs of MC simulations and the theoretical limits against μ_a, which can be seen to be close to each other. The constraint for the real-post is satisfied. In fact, the proportion of tags (for fake post) decreases with μ_a, which is intuitive as a-users deliberately real-tag the posts.

Under the optimal eo-WM, 99.981% of users can identify the fake-post as fake; this optimal value is higher than the reported 90% in [12], as we use $w^* = 1.076$, while algorithm in [12] uses $w^* = 1$. Now, it is interesting to note that with just 1% and 2% of a-users on the OSN, the performance of the eo-WM decreases to 89.798% and 81.74% respectively (in fact, there is degradation with respect to the new QoS defined in (39) which focuses only on non-adversarial users; 99.981% decreases to 95.8% and 92.53% respectively with 1% and 2% of a-users). This depicts that the original WM is not sufficient to control the fake-post propagation in the presence of adversaries.

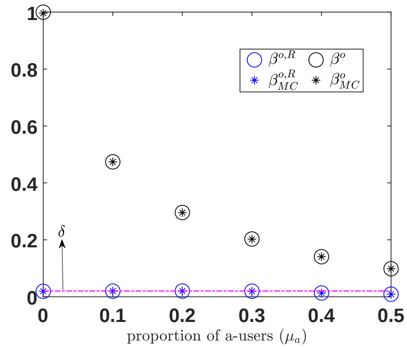

Figure 3: Limits of warning dynamics under eo-WM

Next, we consider a second example with parameters almost as in (36), but with proportion of ws-users (μ_2) fixed and with μ_a, the proportion of a-users varying. We set $\mu_2 = 0.5$, $\mu_1 = 0$ and let the fraction of non-participants equal $0.5 - \mu_a$. For ease of reference, the users of this example are referred to as 'smart users', as here $\alpha_x^F - \alpha_x^R = 0.55$ and $\alpha_y^F - \alpha_y^R = 0.5475$ indicating that the users are capable of distinguishing the fake posts from real posts to a good extent, even without external aid and irrespective of sender tag.

We compare smart users with users in another example scenario where $\alpha_x^F - \alpha_x^R = 0.18$ and $\alpha_y^F - \alpha_y^R = 0.135$. As the differences between the distinguishing parameters are small, these users are referred to as 'naive users'. For this example, the remaining parameters are fixed as below (for diversity, we also consider more attractive posts):

$$\begin{aligned} \rho = 0.9, m_f = 30, \eta^F = 0.52, \eta^R = 0.4, \gamma = 0.1, \eta_a = 0.55, \delta = 0.05, \\ \alpha_x^F = 0.3, \alpha_y^F = 0.225, \alpha_x^R = 0.12, \alpha_y^R = 0.09, \mu_1 = 0.15 \text{ and } \mu_2 = 0.5. \end{aligned} \tag{37}$$

Typically, the users may be naive – may not possess sufficient intrinsic ability to distinguish between the posts to the level that smart users can. Interestingly as seen below, the proposed mechanism is effective to guide even naive users.

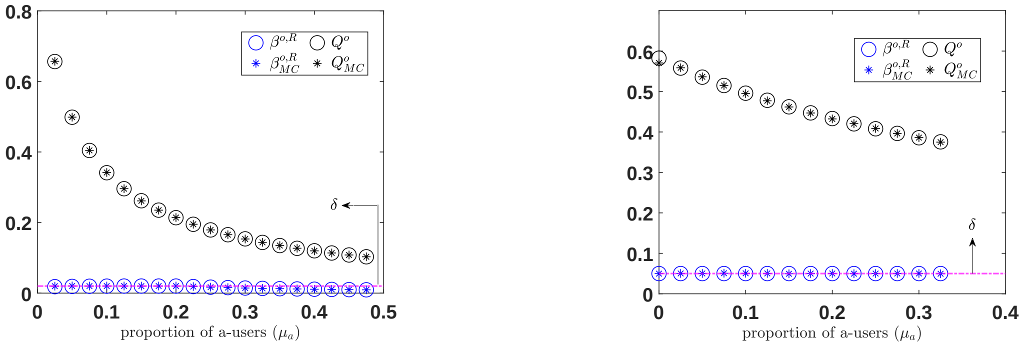

Figure 4: Limits of warning dynamics under eo-WM with smart (left) and naive (right) users respectively

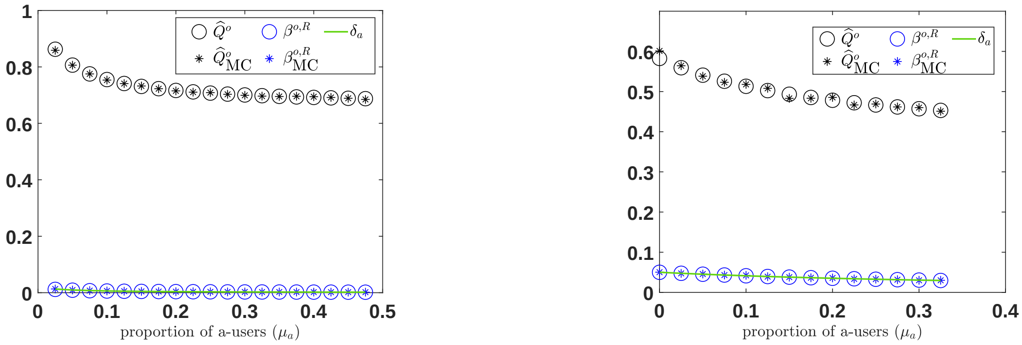

Figure 5: i-QoS under eo-WM with smart (left) and naive (right) users respectively

In Figure 4, we illustrate the QoS (Q defined in 30) and the proportion of fake-tags for the real-post for examples with smart and naive users in left and right sub-figures respectively. Many of the observations are similar to that in the first example: QoS decreases with an increase in μ_a, and the proportion of fake-tags for real-post is at most δ. The QoS in the left sub-figure with smart users is also less than that for first example provided in Figure 3, which also considers smart users – however, for the example in Figure 4(left), the proportion of ws-users (μ_2) is lesser than that in Figure 3 and the number of np-users is non-zero. Furthermore, as one may anticipate, the QoS with naive users is even smaller.

5.3. *Improved QoS – QoS among non-adversaries*[5]

It is important to note that the OSN can control/guide the fake-tags only from non-adversarial users. In fact, the aim is also confined to the correct identification of the actuality of the posts by such users. Hence, it is more appropriate

[5] We would like to thank the reviewers for their feedback which motivated us to design a new performance metric (named improved QoS (i-QoS)) that helps in explaining the numerical results in a more meaningful way.

to consider a metric/QoS focused on the proportion of fake-tags only from ws and wi-users. We precisely aim to capture the same in this sub-section, and define the appropriate optimization problem. Towards this, let X_1^u, X_2^u, X_a^u be the respective proportion of wi, ws and a-users at limit who fake-tag the u-post; observe $X_a^u = 0$ and recall, np-users do not participate. Similarly, define Y_1^u, Y_2^u, Y_a^u as the corresponding proportion of users who real-tag; observe $Y_a^u = 1$. The limit approaches when the number of users that read the post, $t \uparrow \infty$, and consider a large enough t. Then, the number of fake-tags by ws-users after t-th user reads the post can be approximated by $tX_2^u m_f \eta^u$ (one can anticipate this by law of large numbers and because of **C.2**). The other numbers can be approximated in a similar way and as a result, one can re-write the overall proportion of fake-tags as:

$$\beta^u \approx \frac{(X_1^u + X_2^u) m_f \eta^u}{(X_1^u + X_2^u + Y_1^u + Y_2^u) m_f \eta^u + Y_a^u m_f \eta_a}.$$

In a similar manner, the proportion of fake-tags from non a-users represented by β_a^u, the quantity of actual interest, can be approximated as below:

$$\beta_a^u \approx \frac{(X_1^u + X_2^u) m_f \eta^u}{(X_1^u + X_2^u + Y_1^u + Y_2^u) m_f \eta^u} = \frac{X_1^u + X_2^u}{X_1^u + X_2^u + Y_1^u + Y_2^u}.$$

Thus, one can relate the two QoS metrics as follows:

$$\beta_a^u = \left(\frac{(\mu_1 + \mu_2)\eta_u + \mu_a \eta_a}{(\mu_1 + \mu_2)\eta^u} \right) \beta^u. \tag{38}$$

The above discussion motivates us to define an 'improved quality of service (i-QoS)' with respect to any warning-mechanism:

$$\widehat{Q} := \inf \left\{ \left(\frac{(\mu_1 + \mu_2)\eta^u + \mu_a \eta_a}{(\mu_1 + \mu_2)\eta^u} \right) \beta : \beta \in \mathcal{A}_\beta^F \cup \mathcal{S}_\beta^F \right\} = \left(\frac{(\mu_1 + \mu_2)\eta^u + \mu_a \eta_a}{(\mu_1 + \mu_2)\eta^u} \right) Q. \tag{39}$$

One can interpret $\widehat{Q}$ as the almost sure lower bound on the limit proportion of fake-tags for fake-post from non a-users. Henceforth, we also consider comparison of various warning mechanisms using this more relevant metric, i-QoS. Further, we illustrate a lot more improvement when optimization problem (33) is instead designed using i-QoS. Observe that i-QoS is simply a constant multiple of QoS, and hence by Corollary 1 and by (39), the i-QoS for eo-WM (represented by $\widehat{Q}^o$) is unique. Thus, the original optimization problem (33) changes to the following, for some $\delta \in (0, 1)$:

$$\sup_{w \in [0, \overline{w}], b \in [0, \infty)} \widehat{Q}^o(w, b) \text{ subject to } \beta^{o,\infty,R}(w, b) \leq \delta_a := \frac{\delta((\mu_1 + \mu_2)\eta^R)}{((\mu_1 + \mu_2)\eta^R + \mu_a \eta_a)}. \tag{40}$$

Observe that the above optimization problem has exactly the same structure as in (33), except that δ is replaced by δ_a; hence, w^*, b^* can be derived by Theorem 5 directly. The optimal value of the above problem represents the fraction of non a-users (wi and ws-users) who correctly identify the fake-post as fake. When $\mu_a > 0$ and is sufficiently large, then QoS is sufficiently small (lesser than $1 - \mu_a$), as it includes the effects of a-users real-tagging. But this does not imply that the WM failed; in fact on the contrary, at the extreme end, WM is completely successful in eliminating the effect of adversaries if optimal $\widehat{Q}^o = 1$ (indicating that all the non a-users correctly identify the fake-post).

In Figure 5, we continue with the two examples of Figure 4, where we plot i-QoS and its MC estimates, and the corresponding quantities for the real-post; the left sub-figure has smart users and right sub-figure has naive users. It is clear that the proportion of fake-tags for the real-post ($\beta^{o,R}$, see blue curves) are within δ_a-threshold for both cases. More interestingly, the results of the said figure for the fake-post indicate that the results of Figure 4 are mis-leading; the latter figure shows extremely high level of degradation in QoS with μ_a, while the same is not the case in the former; this is obviously because the latter also counts the (intentional) real-tags from a-users. For example, when $\mu_a = 0.3$, the QoS is 15.38% in Figure 4(left), while the actual fraction of fake tags among the smart non a-users is around 70.06%. Thus, the degradation with μ_a may not be as large as depicted in Figure 4, nonetheless there is sufficient degradation as μ_a increases (for example, from 99.981% at $\mu_a = 0$ to 70.06% for $\mu_a = 0.3$).

The above illustrations motivate us to design better warning mechanisms, which achieve higher performance. In fact, the underlying theme of the entire paper is to optimize/increase the proportion of fake-tags for the fake-post, while still ensuring that the constraint in (40) for the real-post is satisfied. In this section, we optimized the performance of the eo-WM for the fake-post, and achieved exactly δ-threshold for the real-post. In the coming sections, we will attempt at designing WMs which increase the performance, without compromising over the real-post. As mentioned before, this goal is achieved by designing appropriate WMs such that the resultant g_β of (28) has zeroes with desirable properties, which in turn dictate the limiting behaviour of WM as confirmed by Theorem 3. To this end, the first idea is to eliminate the effect of adversaries, which we consider next.

6. Eliminating Adversarial Effect WM (ea-WM)

The OSN may not know the exact set of adversarial users, but it knows the proportion of adversarial users (μ_a). We aim to use this knowledge to design a new improved WM which performs better even when μ_a is large. *The idea is to construct a WM specific to any given $\mu_a > 0$, namely $\omega^a(\beta)$, such that g_β^F under the new WM exactly matches that corresponding to $g_\beta^{o,F}$ with $\mu_a = 0$, at optimality (see* (32)*).* In other words, using the knowledge of μ_a, we are creating a hypothetical situation with no adversaries, and hence we name ω^a as eliminating adversarial effect WM (ea-WM). If that is possible, then one can anticipate that the performance will improve for the fake-post under ea-WM; we will identify such conditions below. Further, one still needs to ensure that the performance of real-post is not degraded beyond δ as in (33) (beyond δ_a as in (40) when i-QoS is considered); this is ensured by the WM proposed in this section (and for coming WMs as well). Towards this, we simply define ω^a as:

$$\omega^a(\beta) = \omega(\beta) + \frac{\beta\mu_a m_f \eta_a}{\mu_2 m_f \eta^F\Big(\beta\alpha_x^F + (1-\beta)\alpha_y^F\Big)}. \tag{41}$$

Consider w, b and β such that $\min\{\alpha_x^u\omega^a(\beta), 1\} = \alpha_x^u\omega^a(\beta)$. Then g_β^F under ea-WM, henceforth denoted as $g_\beta^{a,F}$, matches with $g_\beta^{o,F}(\beta; \mu_a = 0)$, because (see (28)):

$$\begin{aligned} g_\beta^{a,F}(\beta;\mu_a > 0) &= -\beta\mu_2 m_f\eta^F - \beta\mu_1(1-\alpha_x^F\rho)m_f\eta^F + (1-\beta)\mu_1\rho\alpha_y^F m_f\eta^u + \mu_2\omega^a(\beta)\Big(\beta\alpha_x^F + (1-\beta)\alpha_y^F\Big)m_f\eta^F \\ &= g_\beta^F(\beta;\mu_a = 0). \end{aligned} \tag{42}$$

Thus, if $\min\{\alpha_x^u\omega^a(\beta), 1\} = \alpha_x^u\omega^a(\beta)$ is satisfied in a neighborhood of β_{na}^o, then one can design the required ea-WM, if further the performance of real-post is within δ-threshold (or δ_a-threshold). In view of Theorem 5, we set w, b as follows for the new ea-WM (similarly, with δ_a):

$$w = \overline{w} \text{ and } b = \begin{cases} b^*|_{\mu_a=0} = \left(\frac{\delta}{1-\delta}\right)\left(\frac{\overline{w}\mu_2(\delta\alpha_x^R+(1-\delta)\alpha_y^R)}{\delta(\mu_1+\mu_2)-(\mu_1\rho+\mu_2\gamma)(\delta\alpha_x^R+(1-\delta)\alpha_y^R)} - 1\right), & \text{if } \beta^{a,\infty,R}(\overline{w}, 0) > \delta, \\ 0, & \text{otherwise.} \end{cases}$$

Now, similar to eo-WM, for each $u \in \{R, F\}$, we will first identify the set of attractors ($\mathcal{A}_\beta^{a,u}$) and the combined set of repellers and saddle points ($\mathcal{S}_\beta^{a,u}$) for the ODE (28) under ea-WM, i.e., $\dot\beta^u = g_\beta^{a,u}(\beta)$.

Theorem 6. *Define*

$$\Delta_a := \mu_2\eta^F\left(\frac{1}{\alpha_x^F} - \omega(\beta_{na}^o)\right)\left(\frac{\beta_{na}^o\alpha_x^F + (1-\beta_{na}^o)\alpha_y^F}{\beta_{na}^o\eta_a}\right). \tag{43}$$

Then, the following statements are true for the fake-post:

(i) If $0 < \mu_a \le \min\{1-\mu_1-\mu_2, \Delta_a\}$, then $\beta^a \ge \beta_{na}^o$ for all $\beta^a \in \mathcal{A}_\beta^{a,F} \cup \mathcal{S}_\beta^{a,F}$.

(ii) Else, i.e., if $\Delta_a < \mu_a < 1-\mu_1-\mu_2$, then $\beta^a \in (\beta^o, \beta_{na}^o)$ for all $\beta^a \in \mathcal{A}_\beta^{a,F} \cup \mathcal{S}_\beta^{a,F}$.

For the real-post, $\beta^{a,R} < \delta$ for all $\beta^{a,R} \in \mathcal{A}^{a,R}_{\beta} \cup \mathcal{S}^{a,R}_{\beta}$. □

In view of the above and Theorem 3, we get that the stochastic iterates Υ_n under ea-WM for the u-post either converge to $\{\mathbf{h}(\beta) : \beta \in \mathcal{A}^{a,u}_{\beta} \cup \mathcal{S}^{a,u}_{\beta}\} \cup \{\mathbf{0}\}$, or hover around $\{\mathbf{h}(\beta) : \beta \in \mathcal{S}^{a,u}_{\beta}\} \cup \{\mathbf{0}\}$. Unlike eo-WM, above Theorem does not guarantee unique limit for the warning dynamics under ea-WM in the non-extinction paths, but Theorem 3(ii) ensures that there exists at least one attractor of the ODE (28), as $\mathcal{A}^{a,u}_{\beta} \neq \emptyset$.

Now, note that ea-WM provides higher warning in comparison to the eo-WM, even for the real-post. Even with such a WM, it is proved above that the proportion of the real-post is maintained[6] within δ-threshold. Further, due to higher warning, we expect a higher QoS under ea-WM; next we discuss the same. Let the QoS (30) under ea-WM be represented by Q^a. In view of Theorem 6, we claim that $Q^a > Q^o$ for the following reasons:

(i) when μ_a is small, i.e., when $\mu_a \leq \Delta_a$, we have $Q^a \geq \beta^o_{\text{na}} > Q^o$ (by Theorem 6(i) and Corollary 3). Thus, ea-WM with adversaries achieves higher QoS than the original eo-WM with no adversary. Then, one can say that the former eliminated the effect of adversaries completely.

(ii) when μ_a is larger, i.e., when $\mu_a > \Delta_a$, ea-WM still improves over eo-WM as $Q^a > Q^o$ by Theorem 6(ii). However, in this case, the QoS under ea-WM is lesser than the QoS under eo-WM with no adversary as $Q^a < \beta^o_{\text{na}}$. Thus, in this case, the effect of adversaries is not completely eliminated by ea-WM.

Similar design and observations follow when one attempts to design ea-WM with i-QoS, i.e., by replacing δ with δ_a. Recall again that with i-QoS, we consider a more relevant problem that focuses only on the responses from non a-users.

6.1. Numerical analysis for ea-WM

In this sub-section, we will numerically quantify the improvement achieved by ea-WM, in comparison to eo-WM; we consider only i-QoS based problems and results. In Figure 6, we continue with the two examples considered in Figure 3 (i.e., with smart and naive users) for ea-WM. We plot the i-QoS with respect to ea-WM (denoted as $\widehat{Q^a}$) evaluated via the exact zeroes of $g^{a,u}_{\beta}$ and the corresponding MC estimates for the ea-WM. We once again observe a close match between the theoretical expressions and the corresponding MC estimates.

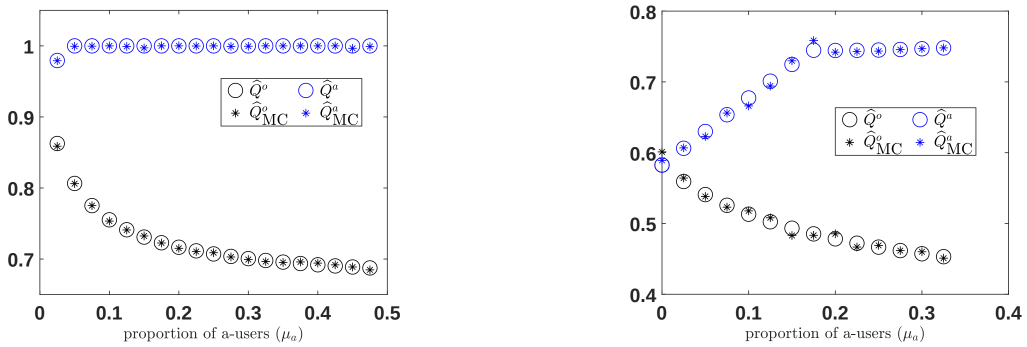

Figure 6: Comparison of i-QoS under eo-WM and ea-WM, with smart (left) and naive (right) users respectively

Further as seen from the figure (6), in all the case studies, the i-QoS improves; nonetheless this way of improvement does not degrade the performance of the real-post, as confirmed by Theorem 6) and also as observed in Figure 7 which plots the performance for the real-post. More interestingly, the i-QoS and the improvement (with respect to eo-WM) both increase sharply with μ_a. Thus, even in the presence of a larger fraction of a-users confusing the WM, ea-WM is able to nudge the non a-users to correctly identify the fake-post as fake. In view of Theorem 6, this may be true as ea-WM provides increasingly high warning levels with increase in μ_a (see (41)). One probably can design

[6] Some equilibrium points can be saddle points and according to Theorem 3, the warning dynamics can hover around such points. But then the warning dynamics moves arbitrarily close to such points and we assume the equilibrium points to be the representative of the limiting behaviour. This leads to a small level of inaccuracy in the sense that the warning dynamics can go above or below the point, in case of hovering around.

a better WM that provides higher warning levels even with smaller value of μ_a (and which again ensures the required real performance) and the quest further is precisely for the same.

From Figure 6(left), for the case study with smart users, observe that $\widehat{Q}^a = 1$, the maximum possible i-QoS, for $\mu_a \geq 0.05$. However, ea-WM fails to achieve such high i-QoS with naive users — i-QoS is less than 0.8 in right sub-figure of Figure 6. The quest again is for a better WM which works well even for naive users, and this is considered in the immediate next.

7. Enhanced WM (eh-WM)

In this section, we design an improved version of ea-WM. The idea is to design a warning ω^h such that $\omega^a(\beta) < \omega^h(\beta)$ for all $\beta \in [0, 1]$. In lines of Theorem 4, such monotonicity of the WM will ensure that the zeroes of the function g^F_β (see (28)) corresponding to the new WM are larger than that of $g^{a,F}_\beta$. However, the design should be such that the performance of the new WM for the real-post is not compromised. Towards this, we design an enhanced warning mechanism (eh-WM) as follows:

$$\omega^h(\beta) = \phi\omega^a(\beta), \text{ for an appropriate choice of } \phi > 1, \text{ with } w, b \text{ as in ea-WM.} \tag{44}$$

For given ϕ, denote the g^u_β of (28) corresponding to the eh-WM as $g^{h,u}_{\beta,\phi}$. Further, define β^h_ϕ as a zero of $g^{h,F}_{\beta,\phi}$ in $[0, 1]$ and $\beta^{h,R}_\phi$ as a zero of $g^{h,R}_{\beta,\phi}$ in $[0, 1]$. Observe that:

$$\begin{aligned} g^{h,F}_{\beta,\phi}(\beta) &= g^{a,F}_\beta(\beta) + \mu_2 m_f \eta^F \Big\{\beta\Big(\min\{1, \phi\omega^a(\beta)\alpha^F_x\} - \min\{1, \omega^a(\beta)\alpha^F_x\}\Big) + (1-\beta)\Big(\min\{1, \phi\omega^a(\beta)\alpha^F_y\} - \min\{1, \omega^a(\beta)\alpha^F_y\}\Big)\Big\} \\ &\geq g^{a,F}_\beta(\beta), \end{aligned}$$

with equality only if $\alpha^F_j\omega^a(\beta) > 1$ for each $j \in \{x, y\}$. This implies that any zero of $g^{h,F}_{\beta,\phi}$ is larger or equal to the smallest zero of $g^{a,F}_\beta$. Thus, it clear that $\beta^h_\phi \geq Q^a$ for any $\beta^h_\phi \in \mathcal{A}^{h,F}_{\beta,\phi} \cup \mathcal{S}^{h,F}_{\beta,\phi}$. Therefore, we have:

$$\inf\{\beta : \beta \in \mathcal{A}^{h,F}_{\beta,\phi} \cup \mathcal{S}^{h,F}_{\beta,\phi}\} =: Q^h_\phi \geq Q^a.$$

That is, the QoS under eh-WM (for any ϕ) is higher or at par with the QoS corresponding to ea-WM.

Now, one can anticipate that higher the warning level is, the more cautiously users tag the posts. Thus, as ϕ increases, the proportion of fake-tags must increase. However, one can not choose an arbitrarily large ϕ as then the performance for the real-post is degraded. Thus, we consider the following problem to optimally choose $\phi = \phi^*$ such that Q^h_ϕ is maximized, while satisfying constraint in (33):

$$\max_\phi \; Q^h_\phi \text{ subject to } \beta \leq \delta \text{ for each } \beta \in \mathcal{A}^{h,R}_{\beta,\phi} \cup \mathcal{S}^{h,R}_{\beta,\phi}. \tag{45}$$

We have the following optimal design for the eh-WM (proof is in appendix):

Theorem 7. *Define the constant*

$$\overline{\phi} := \frac{\delta\Big(\mu_2\eta^R + \mu_1(1 - \alpha^R_x\rho)\eta_R + \mu_a\eta_a\Big) - (1-\delta)\mu_1\rho\alpha^R_y\eta_R}{\mu_2\omega^a(\delta)\Big(\delta\alpha^R_x + (1-\delta)\alpha^R_y\Big)\eta^R}. \tag{46}$$

The ϕ^* defined below is greater than 1 *and is the optimizer of the problem* (45)*:*

$$\phi^* := \begin{cases} \overline{\phi}, & \textit{if } \; \overline{\phi} < \frac{1}{\alpha^R_y\omega^a(\delta)}, \; \textit{or if } \; \overline{\phi} \geq \frac{1}{\alpha^R_y\omega^a(\delta)}, \; \underline{\beta}^F = 0 \textit{ and } b = 0, \\ \frac{1}{\omega^a(\underline{\beta}^F)\alpha^F_y}, & \textit{else.} \end{cases} \qquad \square \tag{47}$$

Thus, the choice of ϕ which gives the maximum proportion of fake-tags for the fake-post is given by ϕ^*. Such a ϕ^* also ensures that the performance of eh-WM for the real-post is not degraded beyond δ-level. The problem (45) can also be designed and solved in terms of the better metric i-QoS and by replacing δ by δ_a analogously. Henceforth, when we refer eh-WM, it corresponds to the case with $\phi = \phi^*$ and when $\delta = \delta_a$. We present the numerical results with respect to eh-WM directly in terms of i-QoS and the correspondingly modified δ_a-threshold.

7.1. Numerical analysis for eh-WM

We now (MC) simulate the warning dynamics under eh-WM for the two examples with smart and naive users and the MC-estimates again well match the theoretical values, as seen from Figure 7 (for real-post) and Figure 8 (for fake-post). Next, we discuss the qualitative analysis. To begin with, the Figure 7 re-affirms the results of Theorem 7 with regard to the real-post — the proportion of fake-tags for the real-post is at most δ_a.

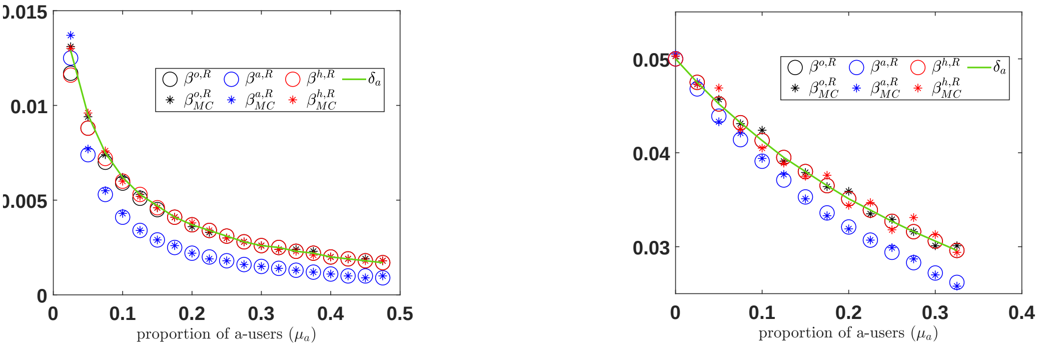

Figure 7: Limits of warning dynamics for real-post under three WMs with smart (left) and naive (right) users respectively

In Figure 8, we plot the i-QoS under eh-WM $\widehat{Q}^h$ (i.e., with ϕ^*), along with that corresponding to the previous two WMs. For the example with smart users, eh-WM performs at par with ea-WM; recall, ea-WM almost achieved $\widehat{Q}^a = 1$. However, for the case with naive users, $\widehat{Q}^h \gg \widehat{Q}^a$; thus, eh-WM is more robust against adversaries than ea-WM. Therefore, one can say that eh-WM is able to guide the naive non a-users about the actuality of fake-posts better than ea-WM.

As an example, when 10% of a-users are trying to harm the system, the eh-WM ensures that 76.29% of naive non a-users correctly identify the fake-post, while this fraction is only 67.73% under ea-WM (observe, $\widehat{Q}^h - \widehat{Q}^a$ is as large as 0.0856, for $\mu_a = 0.1$).

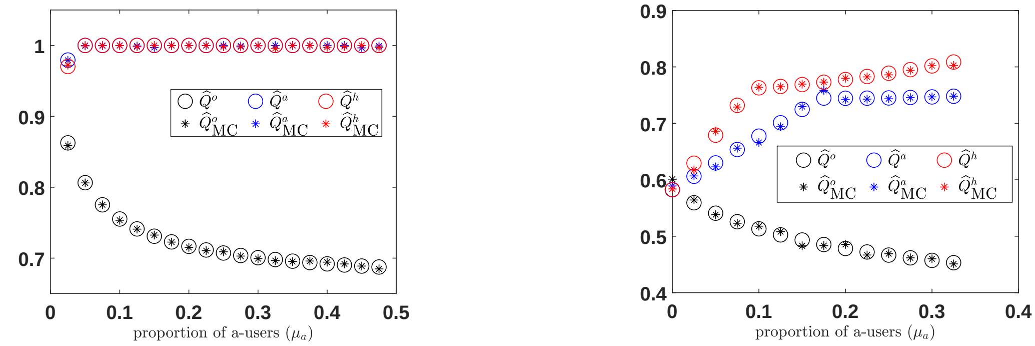

Figure 8: Comparison of i-QoS under three WMs with smart (left) and naive (right) users respectively

As seen from the example with naive users, eh-WM (red curve) performs significantly better than ea-WM (blue curve). Even then, the i-QoS under eh-WM is much better with higher values of μ_a. This probably calls for a very different design of WM, which can generate high warning levels even for smaller values of μ_a. This is attempted in the immediate next.

8. Enhanced-2 WM (eh2-WM) and learning[7]

It is intuitive that as warning increases, the users are alarmed rigorously about the actuality of the posts; this should lead to more users correctly identifying the posts, and thus higher QoS; in fact, Theorem 4 precisely captures this intuition. If one can control the warning such that it does not harm the performance of the real-post beyond δ-threshold, providing higher warning should be effective. We designed ea-WM and eh-WM along these intuitions with higher warning than eo-WM (recall, there is an additive term in (41) and multiplicative term in (44)), and still managed to ensure the performance of the real-post is within the desired level (see Theorem 6 and Theorem 7). Further motivated by this, in this section, we aim to design another improved version of eo-WM, named enhanced-2 WM (eh2-WM) and denoted by ω^{h2}, which provides higher warning signals to the users (in fact, even for the cases with smaller μ_a); this mechanism also facilitates learning the required parameters b and w.

To achieve the same, we again utilize the eo-WM but now with a bigger w, and ensure that there is a unique limit proportion for the real-post which satisfies the δ-threshold. From (12), a bigger w results in higher warning levels, hence, we simply set $w = w^{h2} := {}^1/_{\alpha_x^R} - \gamma$ and choose a corresponding b as in Theorem 5. This value of w ensures that $\alpha_j^R \omega^{h2}(\beta) \leq 1$ for all $j \in \{x, y\}$ and all $\beta \in [0, 1]$ for real post (i.e., when $u = R$) and hence using exactly the same logic as in Corollary 1, we have a unique zero/attractor for the real-post; further the choice of b as in Theorem 5 ensures the said unique attractor $\beta^{R,h2}$ corresponding to the real-post is within the required threshold δ. However, unlike eo-WM, with larger w we may not have a unique limit proportion for the fake-post under eh2-WM. Nonetheless, the resultant QoS (and hence i-QoS) is bigger than that with eo-WM by Theorem 4, as with bigger w, $\omega^{h2}(\beta) > \omega(\beta)$ for all β.

It is important to observe here that the new enhanced WM (eh2-WM) generates high levels of warning signals, and its design does not depends on parameters like μ_a. Thus, one can anticipate that it will enhance the performance even for the smaller values of μ_a. To illustrate the same, we tabulate the i-QoS, $\widehat{Q}^{h2}(w^{h2}, b(w^{h2}))$, achieved under eh2-WM for the case with naive users: Clearly, the i-QoS under eh2-WM is consistently higher than that with eh-WM (see Figure 8, where the red curve is below 0.8 for all μ_a). More importantly, the i-QoS under eh2-WM is almost the same for all values of μ_a.

	$\mu_a = 0$	$\mu_a = 0.1$	$\mu_a = 0.2$	$\mu_a = 0.3$
$\widehat{Q}^{h2}(w^{h2}, b(w^{h2}))$	0.8289	0.8270	0.8257	0.8246

Table 2: i-QoS under perfect knowledge of user sensitive parameters

Learning the parameters: At this point, it is important to note that all the discussion so far assumed that the user sensitive parameters (ρ and (α_i^u) for each $i \in \{x, y\}$ and $u \in \{R, F\}$) and proportions of users of different types (μ_1, μ_2 and μ_a) are known to the OSN. However, such information is not easily accessible to the OSN and the purpose now is to design a WM without such knowledge. Towards this, *we propose an algorithm which directly learns the parameters of the WM, b and w*. We only require that there is a non-zero proportion of ws-users[8], i.e., $\mu_2 > 0$ and the knowledge of ratio α_x^R/α_y^R (details are given below). The design would only utilize various random quantities that are observed during the post propagation process.

The main idea is to consider a real-post which is known to the OSN and train the parameters w and b using the responses of the users.

Basically, we add a SA-based step which tunes b such that the corresponding $\beta^{o,R}$ eventually approaches δ - recall, the constraint in optimization problem (33) requires that $\beta^{o,R} \leq \delta$. Further, w is tuned such that $\alpha_y^R \omega^{h2}(1)$ approaches $1 - \kappa$, where constant $\kappa \geq 1 - \alpha_y^R/\alpha_x^R$. From (12), $\omega^{h2}(1; w, b) = w + \gamma$, and hence such a tuning ensures that w approaches ${}^{(1-\kappa)}/_{\alpha_y^R} - \gamma$ (and by choice of κ, eventually $w \leq {}^{(1-\kappa)}/_{\alpha_y^R} - \gamma$) — thus, eventually $\alpha_j^R \omega^{h2}(1) \leq 1$ for each $j \in \{x, y\}$, as planned for the real-post. Here, we would like to stress that the tuning of w is done with respect to α_y^R, instead of α_x^R, as there may not be sufficient estimates corresponding to fake-tags for the real-posts (recall, δ is typically a small value). Thus, the algorithm requires some idea on the ratio α_x^R/α_y^R. In all, if such a tuning (of both w and b) is possible, then it would ensure a unique attractor below δ-threshold for the real-post.

[7] We would also like to thank the reviewers of the paper, as their feedback motivated us to design a better warning mechanism (discussed in this section) that does not require the knowledge of system parameters and the proportions.

[8] it can be checked by noticing the users who click on the information button (see Figure 1)

Algorithm 1: Design of learning WM

```
(i) Consider a real-post.
(ii) Initialize $C_x(\tau_0)$ and $C_y(\tau_0)$; calculate $B_0^{h2,R}$. Fix a large enough $\mathbb{S} < \infty$.
(iii) Initialize $b_0$ and $\eta_0$ sufficiently small, and choose a $w_0 > 1$.
(iv) At $k$-th epoch, $\tau_k$, when $k$-th user reads the post, for $k \in \{1, 2, \ldots, \mathbb{S}\}$:
    • set the $w$-update flag, $J_{ws} = 0$
    • if the reader is a ws-user, then provide warning, $\omega^{h2}$, which is set as below:
        – toss a biased coin such that $P(\text{head appears}) = \eta_{k-1} > 0$, let $\eta_{k-1} \to 0$
        – if head appeared and if the said user received with post with real-tag,
            ∗ set warning corresponding to $\beta = 1$, i.e., set $\omega^{h2}(B_{k-1}^{h2,R}) := w_{k-1} + \gamma$
            ∗ set the indicator $J_{ws} = 1$
        – else, set warning as per WM, i.e., set $\omega^{h2}(B_{k-1}^{h2,R})$ as in (50)
    • observe the tag $I_k$ and the number of shares by the said user; accordingly, update proportion of fake tags,
      $B_k^{h2,R} = \frac{C_x(\tau_{k-1}^+)}{C_x(\tau_{k-1}^+)+C_y(\tau_{k-1}^+)}$
    • update the parameters, using the new estimate $B_k^{h2,R}$ and $I_k$
        – if $w$-update flag, $J_{ws} = 1$, then update $w_k$ as in (48)
        – update $b_k$ as in (49)
```

The above tuning for w requires warning levels $\omega^{h2}(1)$, corresponding to $\beta = 1$; however, in the eo-WM, the warning levels were generated according to the then estimates of β, the proportion of fake-tags. To minimally disrupt the normal functioning of the WM, we propose some special epochs at which such special warning is provided – at time epoch k, if a ws-user who received the post with real-tag clicks on the information button, the OSN generates such a warning with probability η_k, where $\eta_k \downarrow 0$, as $k \to \infty$. Only such special epochs are used to learn w. To summarize, the updates for w at epoch k are as follows: if a ws-user that received the post with real-tag reads the post, then we have:

$$w_k \leftarrow \max\{1, w_{k-1} - \epsilon_k (I_k - (1-\kappa))\}, \text{ with probability } \eta_k, \tag{48}$$

where I_k is the indicator that the user tags the post as fake and $\epsilon_k := c_1(\frac{1}{k+1})^{c_2}$ with some appropriate $c_1 > 0$ and $c_2 \in (0.5, 1]$. In all other cases, we set $w_k = w_{k-1}$.

Next, we discuss the updates for b. For each $k \geq 1$, update b_k as below:

$$b_k \leftarrow \max\left\{0, b_{k-1} + \epsilon_k(B_k^{h2,R} - \delta)\right\}, \text{ where as before } B_k^{h2,R} := \frac{C_x(\tau_k^-)}{C_x(\tau_k^-) + C_y(\tau_k^-)}, \tag{49}$$

and the post-propagation process updates as in (14) and (15) — the warning shown to the k-th user reading the post would have been generated using (w_k, b_k) as below:

$$\omega^{h2}(B_k^{h2,R}) := \omega(B_k^{h2,R}) = \frac{w_k B_k^{h2,R}}{B_k^{h2,R} + b_k(1 - B_k^{h2,R})} + \gamma, \tag{50}$$

at the normal epochs (when w_k is not updated); for the special epochs, the warning $\omega^{h2}(B_k^{h2,R}) := \omega(1) = w_k + \gamma$ is generated.

The brief idea behind such a design is that as is usually the case with SA algorithms, the SA iterates b_k and w_k converge so as to ensure the expected values of the respective update-terms $B_k^{h2,R} - \delta$ in (49) and $I_k - (1-\kappa)$ in (48) converge to 0 as $k \to \infty$. That is, $\beta_k^{h2,R} = E[B_k^{h2,R}]$ approaches δ and w_k approaches[9] $(1-\kappa)/\alpha_y^R - \gamma$. As already mentioned, such a limit of w ensures that the unique limit for the real-post $\beta^{h2,R}$ is near δ; thus, the constraint in (33) is satisfied and the discussion in the beginning of this section also ensures that the QoS is strictly improved in comparison to the eo-WM.

The learning algorithm is summarized in Algorithm 1. The analysis of the above learning algorithm would require rigorous two-time scale (projected) SA-based tools - observe w_k is updated minimally and further probability $\eta_k \downarrow 0$. We skip the analysis here, but validate and illustrate the improved performance of the learning WM (referred to as l-eh2-WM) via numerical examples in the next sub-section.

8.1. Numerical analysis for l-eh2-WM

In Table 3, we continue with the example with naive users to test the learning algorithm. Towards this, we fix $\kappa = 1 - \alpha_y^R/\alpha_x^R + 10^{-3}$, $\eta_k = 1.5(1/k)^{0.8}$, $\eta_0 = 0.008$, $w_0 = 6$ and $b_0 = 10^{-4}$. The choice of ϵ_k for learning b and w is $2.2(1/k)^{0.7}$. We initialize the system such that a real-post is shared by the content provider to 20 users with the real-tag.

For a given sample-size (number of samples available for learning and represented by $\mathbb{S}$), we consider 150 sample paths for the post-propagation of the real-post under l-eh2-WM; the idea is to measure the efficacy of l-eh2-WM algorithm via the fraction of times it achieves an i-QoS within ±0.05 of that corresponding to the case with perfect information (i.e., $\widehat{Q}^{h2}(w^{h2}, b(w^{h2})))$. We consider different sample-sizes $\mathbb{S}$ in the range 10^4 to 10^5. In each sample-path, l-eh2-WM algorithm is used to update the estimates of $\{(w_k, b_k)\}$ for $\mathbb{S}$ number of epochs and then the i-QoS $\widehat{Q}_k^o(w_{\mathbb{S}}, b_{\mathbb{S}})$ (fake-post) corresponding to eo-WM using the last estimate $(w_{\mathbb{S}}, b_{\mathbb{S}})$ is computed.

In Table 3, for different values of $\mathbb{S}$, we tabulate $f_{\mathbb{S}}$, the fraction of sample paths for which $|\widehat{Q}^{h2}(w^{h2}, b(w^{h2})) - \widehat{Q}_{\mathbb{S}}^o(b_{\mathbb{S}}, w_{\mathbb{S}})| \leq 0.05$.

	$\mathbb{S}$				
	10^4	$2.5 * 10^4$	$5 * 10^4$	$7.5 * 10^4$	10^5
$\mu_a = 0$	0.73	0.89	0.91	0.95	0.93
$\mu_a = 0.1$	0.41	0.57	0.76	0.84	0.91
$\mu_a = 0.2$	0.19	0.44	0.64	0.74	0.79

Table 3: Fraction of sample paths that learnt the parameters (b, w) sufficiently well and achieved the desired level of i-QoS under l-eh2-WM

It can be seen from the table that the fraction of sample paths with the desired property ($f_{\mathbb{S}}$) increases with $\mathbb{S}$, thus depicting that the l-eh2-WM is progressively able to achieve the performance close to the case with perfect knowledge. One may anticipate that more iterations/shares should be required to achieve i-QoS of eh2-WM (i.e., with perfect knowledge) as μ_a increases; the same is evident from the table; for example, when $\mathbb{S} = 10^5$, $f_{\mathbb{S}}$ is as large as 0.91 for $\mu_a = 0.1$, but with $\mu_a = 0.2$, it is much smaller and equals 0.79. Thus, this example illustrates that l-eh2-WM has managed to learn and tune the WM sufficiently well, when number of samples $\geq 7.5 * 10^4$ for proportion of a-users up to 0.2.

The performance of the l-eh2-WM is sensitive to the initial conditions and the parameters of the two-timescale algorithm (like, ϵ_k), as is the usual case with SA-based algorithms. Using trial-and-error method, we picked a good enough set of values, while extensive study on better choice of these parameters is outside the scope of this work.

Next, in Figure 9, we continue with the two examples considered in Figure 4. In the left and right sub-figures, we consider the instance with smart and naive users respectively and present the results directly in terms of i-QoS. The learning algorithm is again initialized and tuned appropriately, and now with a large sample-size, $\mathbb{S} = 10^6$.

From the figure, it can be seen that for all values of μ_a, the i-QoS under l-eh2-WM (marked in diamond) is higher than the eo-WM; in fact, it performs superior to all the previous WMs. Of course, the i-QoS can not be further improved for smart users — even l-eh2-WM achieves i-QoS close to 1, as eh-WM. The superior performance of eh2-WM (actually that of l-eh2-WM with large $\mathbb{S}$) is clearly depicted in the case with naive users. From Table 2 and Figure 8, it is clear that the eh2-WM outperforms eh-WM and further has similar performance for all values of μ_a. The l-eh2-WM with large $\mathbb{S}$ has exactly similar performance traits, as can be seen from Figure 9. Furthermore, the proportion of fake-tags for the real-post is also within the δ_a-threshold, thus satisfying the constraint in (40).

[9] Observe that the the conditional expected value conditioned that the user is a ws-user who received the post with real-tag, $E[I_k] = \alpha_y^R \omega(1) = \alpha_y^R(w_k + \gamma)$.

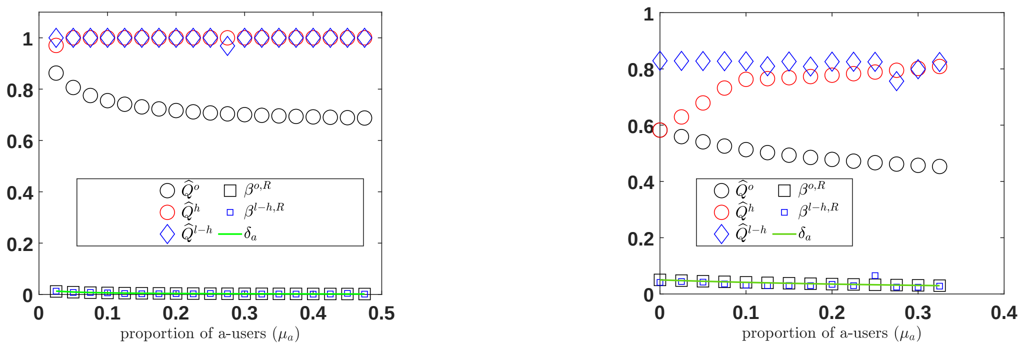

Figure 9: Comparison of limits of warning dynamics under eo, eh and l-eh2-WM with smart (left) and naive (right) users respectively

9. Conclusions and Future Work

There is a huge requirement to identify fake posts on ever active OSNs. Further, any algorithm attempting to identify fake posts faces challenges from adversarial users and users unwilling to participate. Our first aim in this paper is to derive the performance of a promising recently proposed algorithm in the presence of adversaries who always real-tag any post. A severe degradation in performance is observed with just 1% of adversaries.

The algorithm collects binary signals (fake/real tags) from all previous users, generates a warning based on the fraction of fake tags and compels further users to judge and consume the post cautiously based on the warning level provided. Using new results in branching processes (also derived in this paper), we obtain a one-dimensional ordinary differential equation (ODE) that analyses any generic iterative warning mechanism depending upon the fraction of fake tags. This ODE is instrumental in deriving robust adaptations of the previous mechanism – in particular, we use concepts like eliminating the effects of adversaries, the inherent monotone characteristics of relevant performance on certain parameters, etc. The new mechanisms illustrate significant performance improvement both in the presence and the absence of adversaries compared to the previous method. An algorithm which provides improvement over the existing method, without relying on the usually inaccessible users-specific information, is also proposed.

This paper also contributes towards total-current population-dependent two-type branching processes with population dependent death rates and also considers a variety of unnatural deaths. In particular, we derive all possible limits and limiting behaviours of the population sizes as time progresses.

In future, one can think of several new directions. The one-dimensional ODE can also be utilized to study other types of adversaries, like always fake-tagging adversaries or more informed adversaries that mis-tag both posts (fake-tag authentic post and real-tag the fake post). One can again derive improved algorithms, as we illustrated with real-tagging adversaries. One can also study the influence of users that share but refuse to tag or other important behavioural characteristics. Further, we designed two types of enhanced warning mechanisms, both of which improved over the existing mechanism. However, the two new mechanisms are not comparable, as one can perform better than the other in some instances. In future, one can attempt to design a combination of the two, which outperforms all of them, and also design the corresponding learning mechanism.

Appendix A.

Proof of Theorem 1. The proof follows exactly as in [19], except for some changes due to unnatural deaths. Here, we directly mention the SA based scheme for the new process, and necessary details where ever required.

From (18), the embedded process immediately after n-th death, when say an x-type individual d-dies, is given by:

$$\begin{aligned} C_{x,n} &= C_{x,n-1} + \Gamma_{xx,d,n}(\Phi_{n-1}) - 1, \quad T_{x,n} = T_{x,n-1} + \Gamma_{xx,d,n}(\Phi_{n-1}), \\ C_{y,n} &= C_{y,n-1} + \Gamma_{xy,d,n}(\Phi_{n-1}), \quad A_{y,n} = A_{y,n-1} + \Gamma_{xy,d,n}(\Phi_{n-1}). \end{aligned} \tag{A.1}$$

The ratios in $\boldsymbol{\Upsilon}_n$ can be re-written as (with $\epsilon_{n-1} := 1/n$):

$$
\begin{aligned}
&\boldsymbol{\Upsilon}_n = \boldsymbol{\Upsilon}_{n-1} + \epsilon_{n-1}\mathbf{L}_{n-1}, \text{ where } \mathbf{L}_{n-1} := (L^{\psi,c}_{n-1}, L^{\theta,c}_{n-1}, L^{\psi,a}_{n-1}, L^{\theta,a}_{n-1})^t, \text{ with}\\
&L^{\psi,c}_{n-1} := \left\{ \sum_{d\in[d_x]} \left(H^x_{n,d}(\Gamma_{x,d,n}(\Phi_{n-1}) - 1)\right) + \sum_{d\in[d_y]} \left(H^y_{n,d}(\Gamma_{y,d,n}(\Phi_{n-1}) - 1)\right) \right\} 1_{\Psi^c_{n-1}>0} - \Psi^c_{n-1},\\
&L^{\theta,c}_{n-1} := \left\{ \sum_{d\in[d_x]} \left(H^x_{n,d}(\Gamma_{xx,d,n}(\Phi_{n-1}) - 1)\right) + \sum_{d\in[d_y]} \left(H^y_{n,d}\Gamma_{yx,d,n}(\Phi_{n-1})\right) \right\} 1_{\Psi^c_{n-1}>0} - \Theta^c_{n-1},\\
&L^{\psi,a}_{n-1} := \left\{ \sum_{d\in[d_x]} \left(H^x_{n,d}\Gamma_{x,d,n}(\Phi_{n-1})\right) + \sum_{d\in[d_y]} \left(H^y_{n,d}\Gamma_{y,d,n}(\Phi_{n-1})\right) \right\} 1_{\Psi^c_{n-1}>0} - \Psi^a_{n-1}, \text{ and}\\
&L^{\theta,a}_{n-1} := \left\{ \sum_{d\in[d_x]} \left(H^x_{n,d}\Gamma_{xx,d,n}(\Phi_{n-1})\right) + \sum_{d\in[d_y]} \left(H^y_{n,d}\Gamma_{yx,d,n}(\Phi_{n-1})\right) \right\} 1_{\Psi^c_{n-1}>0} - \Theta^a_{n-1}, \text{ where}\\
&\Gamma_{x,d,k} := \Gamma_{xx,d,k} + \Gamma_{xy,d,k}, \;\; \Gamma_{y,d,k} := \Gamma_{yy,d,k} + \Gamma_{yx,d,k},
\end{aligned}
\tag{A.2}
$$

$H^x_{k,d} \in \{0,1\}$ indicates that an x-type individual d-dies at k-th epoch such that $\sum_{d\in D_x} H^x_{k,d} \in \{0,1\}$ and $\sum_{d\in D_y} H^y_{k,d} := 1 - \sum_{d\in D_x} H^x_{k,d}$.

Henceforth, the proof of part (i) has two major steps: (a) to construct a sequence of piece-wise constant interpolated trajectories for almost all sample-paths; (b) to prove that the designed trajectories are equicontinuous in extended sense. We will provide the proof in terms of θ^c-component of the vector Υ, when the proof for the remaining components goes through in exactly similar manner.

Define $\varrho = (\rho^c_\psi, \rho^c_\theta, \rho^a_\psi, \rho^a_\theta)$ as the conditional expectation, $E[\mathbf{L}_n|\mathcal{F}_n] =: \varrho(\boldsymbol{\Upsilon}_n, t_n)$, with respect to the sigma algebra, $\mathcal{F}_n := \sigma\{\Phi_k : 1 \le k < n\}$ (see [19, (16)]). Let $\boldsymbol{\Upsilon}^n(\cdot) := (\Psi^{n,c}(\cdot), \Theta^{n,c}(\cdot), \Psi^{n,a}(\cdot), \Theta^{n,a}(\cdot))$ be the constant piece-wise interpolated trajectory defined as below (see (A.2), and recall $t_n = \sum_{i=1}^n \epsilon_{i-1}$):

$$
\begin{aligned}
\Theta^{n,c}(t) &:= \Theta^c_n + \int_0^t g^c_\theta(\boldsymbol{\Upsilon}^n(s))ds + \sum_{i=n}^{\eta(t_n+t)-1} \epsilon_i L^{\theta,c}_i - \int_0^t g^c_\theta(\boldsymbol{\Upsilon}^n(s))ds\\
&= \Theta^c_n + \int_0^t g^c_\theta(\boldsymbol{\Upsilon}^n)ds + M^{n,\theta,c}(t) + R^{n,\theta,c}(t) + D^{n,\theta,c}(t), \text{ where}\\
M^{n,\theta,c}(t) &:= \sum_{i=n}^{\eta(t_n+t)-1} \epsilon_i \left(L^{\theta,c}_i - \rho^c_\theta(\boldsymbol{\Upsilon}_i, t_i)\right),\\
R^{n,\theta,c}(t) &:= \sum_{i=n}^{\eta(t_n+t)-1} \epsilon_i g^c_\theta(\boldsymbol{\Upsilon}_i) - \int_0^t g^c_\theta(\boldsymbol{\Upsilon}^n)ds,\\
D^{n,\theta,c}(t) &:= \sum_{i=n}^{\eta(t_n+t)-1} \epsilon_i D^{\theta,c}_i, \text{ where } D^{\theta,c}_i := \rho^c_\theta(\boldsymbol{\Upsilon}_i, t_i) - g^c_\theta(\boldsymbol{\Upsilon}_i),
\end{aligned}
\tag{A.3}
$$

$\Psi^{n,c}(t), \Psi^{n,a}(t)$ and $\Theta^{n,a}(t)$ are defined analogously. As in [19], the extended equicontinuity can be proved for $M^{n,\theta,c}(\cdot)$, $R^{n,\theta,c}(\cdot)$. For, $D^{n,\theta,c}(\cdot)$ the procedure again follows as in [19] when $S_n \to 0$; however for sample paths where $S_n \nrightarrow 0$, the arguments for proving the equicontinuity for $D^{n,\theta,c}(\cdot)$ slightly changes as below:

$$
\begin{aligned}
|D^{\theta,c}_i| &\le |f_\beta(\Phi_i)(m_{xx}(\Phi_i) - 1) - f^\infty_\beta(\mathrm{B}^c_i)(m^\infty_{xx}(\mathrm{B}^c_i) - 1)| + |(1 - f_\beta(\Phi_i))m_{yx}(\Phi_i) - (1 - f^\infty_\beta(\mathrm{B}^c_i))m^\infty_{yx}(\mathrm{B}^c_i)|\\
&\le |f_\beta(\Phi_i)m_{xx}(\Phi_i) - f^\infty_\beta(\mathrm{B}^c_i)m^\infty_{xx}(\mathrm{B}^c_i)| + |f_\beta(\Phi_i) - f^\infty_\beta(\mathrm{B}^c_i)|\\
&\qquad + |m_{yx}(\Phi_i) - m^\infty_{yx}(\mathrm{B}^c_i)| + |f_\beta(\Phi_i)m_{yx}(\Phi_i) - f^\infty_\beta(\mathrm{B}^c_i)m^\infty_{yx}(\mathrm{B}^c_i)|
\end{aligned}
\tag{A.4}
$$

In the above, under **C.2**, the third term is bounded above by $1/(S_i)^\alpha$. The second term can be bounded above as follows:

$$
\begin{aligned}
|f_\beta(\Phi_i) - f_\beta^\infty(\mathrm{B}_i^c)| &= \mathrm{B}_i^c \left| \frac{\sum_{d\in D_x} \lambda_{x,d}(\Phi_i)}{d(\Phi_i)} - \frac{\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)}{d^\infty(\mathrm{B}_i^c)} \right| \\
&= \mathrm{B}_i^c \left| \frac{\sum_{d\in D_x} \lambda_{x,d}(\Phi_i)}{d(\Phi_i)} - \frac{\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)}{d(\Phi_i)} + \frac{\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)}{d(\Phi_i)} - \frac{\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)}{d^\infty(\mathrm{B}_i^c)} \right| \\
&\le \frac{\mathrm{B}_i^c}{d(\Phi_i)} \sum_{d\in D_x} |\lambda_{x,d}(\Phi_i) - \lambda_{x,d}^\infty(\mathrm{B}_i^c)| + \mathrm{B}_i^c \left| \sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c) \left(\frac{1}{d(\Phi_i)} - \frac{1}{d^\infty(\mathrm{B}_i^c)} \right) \right| \\
&\le \frac{\mathrm{B}_i^c}{d(\Phi_i)} \left(\frac{|D_x|}{(S_i^c)^\alpha} + \frac{\left|\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)\right|}{d^\infty(\mathrm{B}_i^c)} \left| d^\infty(\mathrm{B}_i^c) - d(\Phi_i) \right| \right) \\
&\le \frac{\mathrm{B}_i^c}{d(\Phi_i)} \left(\frac{|D_x|}{(S_i^c)^\alpha} + \frac{\left|\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)\right|}{d^\infty(\mathrm{B}_i^c)} \frac{(|D_x| + |D_y|)}{(S_i^c)^\alpha} \right) \\
&\le \frac{\mathrm{B}_i^c}{d(\Phi_i)} \frac{|D_x| + |D_y|}{(S_i^c)^\alpha} \left(1 + \frac{\left|\sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c)\right|}{d^\infty(\mathrm{B}_i^c)} \right) \\
&\le \frac{\mathrm{B}_i^c}{d(\Phi_i)} \frac{|D_x| + |D_y|}{(S_i^c)^\alpha} \left(1 + \frac{1}{B_i^c} \right) \qquad \left(\text{since } d^\infty(\mathrm{B}_i^c) \ge B_c^i \sum_{d\in D_x} \lambda_{x,d}^\infty(\mathrm{B}_i^c) \right) \\
&= \frac{(|D_x| + |D_y|)(B_i^c + 1)}{d(\Phi_i)(S_i^c)^\alpha} \le \frac{2(|D_x| + |D_y|)}{d(\Phi_i)(S_i^c)^\alpha}
\end{aligned}
$$

Define $\Delta_1 := \min\left\{\inf_\Phi \lambda_{x,d}(\Phi), \inf_\Phi \lambda_{y,d}(\Phi)\right\} > 0$, by **C.1**. Then, $d(\Phi_i) \ge \mathrm{B}_i^c \inf_\Phi \lambda_{x,d}(\Phi) + (1 - \mathrm{B}_i^c) \inf_\Phi \lambda_{y,d}(\Phi) \ge \Delta_1$. Thus, we have:

$$
|f_\beta(\Phi_i) - f_\beta^\infty(\mathrm{B}_i^c)| \le \frac{2(|D_x| + |D_y|)}{(S_i^c)^\alpha} \frac{1}{\Delta_1} \tag{A.5}
$$

The first term in (A.4) can be bounded as follows under **C.2** and (A.5):

$$
\begin{aligned}
|f_\beta(\Phi_i) m_{xx}(\Phi_i) - f_\beta^\infty(\mathrm{B}_i^c) m_{xx}^\infty(\mathrm{B}_i^c)| &\le |f_\beta(\Phi_i)| |m_{xx}(\Phi_i) - m_{xx}^\infty(\mathrm{B}_i^c)| + |m_{xx}^\infty(\mathrm{B}_i^c)| |f_\beta(\Phi_i) - f_\beta^\infty(\mathrm{B}_i^c)| \\
&\le \frac{1}{(S_{i})^\alpha} + \frac{2(|D_x| + |D_y|)}{(S_i^c)^\alpha} \frac{1}{\Delta_1} \left(E[\bar{\Gamma}] + \frac{1}{(S_{i})^\alpha} \right).
\end{aligned}
$$

Similarly, the fourth term in (A.4) can be upper bounded as follows:

$$
|f_\beta(\Phi_i) m_{yx}(\Phi_i) - f_\beta^\infty(\mathrm{B}_i^c) m_{yx}^\infty(\mathrm{B}_i^c)| \le \frac{1}{(S_{i})^\alpha} + \frac{2(|D_x| + |D_y|)}{(S_i^c)^\alpha} \frac{1}{\Delta_1} \left(E[\bar{\Gamma}] + \frac{1}{(S_{i})^\alpha} \right).
$$

Thus, $D_i^{\theta,c}$ can be upper bounded as follows for some $K < \infty$ (recall, $\alpha \ge 1$):

$$
\begin{aligned}
D_i^{\theta,c} &\le 2\left(\frac{1}{(S_{i})^\alpha} + \frac{2(|D_x| + |D_y|)}{(S_i^c)^\alpha} \frac{1}{\Delta_1} \left(E[\bar{\Gamma}] + \frac{1}{(S_{i})^\alpha} \right) \right) + \frac{1}{(S_i^c)^\alpha} + \frac{2(|D_x| + |D_y|)}{(S_i^c)^\alpha} \frac{1}{\Delta_1} \\
&\le \frac{K}{(S_i^c)^\alpha} \le \frac{K}{S_i^c} = \frac{K}{\Psi_i^c \eta(t_i)} \le \frac{K}{\Delta i}.
\end{aligned}
$$

This implies that, (recall $\epsilon_i = 1/(i+1)$ and $\alpha \ge 1$)

$$
|D^{n,\theta,c}(t)| = \left| \sum_{i=n}^{\eta(t_n+t)-1} \epsilon_i D_i^{\theta,c} \right| \le \sum_{i=n}^{\eta(t_n+t)-1} \frac{K}{\Delta i(i+1)} \le \sum_{i=n}^{\infty} \frac{K}{\Delta i(i+1)}, \text{ for any } t.
$$

Thus, $D^{n,\theta,c}(t)$ uniformly converges to 0 as $n \to \infty$. In all, $(\Theta^{n,c}(\cdot))$ is equicontinuous in the extended sense.

The proof of part (ii) follows exactly as in [19]. □

Proof of Theorem 2. Observe that each point $x_i^* \in \mathcal{I}$ can either be a point of dis-continuity or continuity for g_β. In the former case, when x_i^* is either an attractor or repeller of the ODE (25), the result can be proved exactly as in [19, Theorem 2]. In fact, when x_i^* is a saddle point of the ODE (25), the analysis can be easily extended similar to the case when x_i^* is a repeller.

Now consider $x_i^* \in \mathcal{I}$ such that g_β is continuous at x_i^*. Let $\Upsilon(0) \in \mathcal{D}_I$ with $\psi^c(0) > 0$. By [19, Lemma 5.], $\psi^c(t) > 0$ for all $t \geq 0$, thus ODE (22) simplifies to $\dot{\Upsilon} = \mathbf{h}(\beta(\Upsilon)) - \Upsilon$. Now, we will prove the claim for different possibilities of x^* as in the hypothesis separately. Firstly for all cases global solution exists because of Lipschtiz continuity.

Part (i) Without loss of generality, let $\beta(0) \in \mathcal{N}_i^-$. Then, by [19, Lemma 4(a)(i)], $\beta(t)$ increases to x_i^* for all $t < \tau := \inf\{t : \beta(t) = x_i^*\}$. If $t < \infty$, then $\beta(t) = x_i^*$ for all $t \geq \tau$ (as x_i^* is an equilibrium point). Then, clearly, $\beta(t) \to x_i^*$ and $\Upsilon(t) \to \mathbf{h}(x_i^*)$ as $t \to \infty$, as above.

Else say $\tau = \infty$; then for every $\delta > 0$, there exists a $T_\delta < \infty$ (guaranteed as before by [19, Lemma 4(a)(i)] because by continuity the RHS of ODE can be uniformly bounded by non-zero values) such that:

$$x_i^* - \delta \leq \beta(t) \leq x_i^* + \delta \text{ for all } t \geq T_\delta.$$

Thus, $\beta(t) \to x_i^*$ as $t \to \infty$. This also implies that:

$$\underline{\mathbf{h}}_\delta(x_i^*) - \Upsilon \leq \dot{\Upsilon} \leq \overline{\mathbf{h}}_\delta(x_i^*) - \Upsilon \text{ for all } t \geq T_\delta, \text{ for } \overline{\mathbf{h}}_\delta(x_i^*) := \sup_{x \in \overline{\mathcal{N}}_\delta(x_i^*)} \mathbf{h}(x) \text{ and } \underline{\mathbf{h}}_\delta(x_i^*) := \inf_{x \in \overline{\mathcal{N}}_\delta(x_i^*)} \mathbf{h}(x).$$

By Comparison Theorem in [24] for ODEs having Lipschitz continuous right hand sides and using classical methods to derive the upper and lower bounds, we get:

$$\underline{\mathbf{h}}_\delta(x_i^*) + e^{-t+T_\delta}(\Upsilon(T_\delta) - \underline{\mathbf{h}}_\delta(x_i^*)) \leq \Upsilon(t) \leq \overline{\mathbf{h}}_\delta(x_i^*) + e^{-t+T_\delta}(\Upsilon(T_\delta) - \overline{\mathbf{h}}_\delta(x_i^*)) \text{ for all } t \geq T_\delta.$$

Then clearly by considering limits $t \to \infty$ we have :

$$\underline{\mathbf{h}}_\delta(x_i^*) \leq \liminf_{t\to\infty} \Upsilon(t) \leq \limsup_{t\to\infty} \Upsilon(t) \leq \overline{\mathbf{h}}_\delta(x_i^*), \text{ and now letting } \delta \to 0, \mathbf{h}(x_i^*) \leq \liminf_{t\to\infty} \Upsilon(t) \leq \limsup_{t\to\infty} \Upsilon(t) \leq \mathbf{h}(x_i^*).$$

Hence, $\Upsilon(t) \to \mathbf{h}(x_i^*)$ as $t \to \infty$.

Part (ii) If $\beta(0) = x_i^*$, then clearly $\beta(t) = x_i^*$ for all $t \geq 0$ and $\Upsilon(t) \to \mathbf{h}(x_i^*)$ as $t \to \infty$. However if $\beta(0) \in \mathcal{N}_i^-$, then it can be shown as above that $\beta(t) \to y^* := \max\{y \in \mathcal{I} : y < x_i^*\}$. Similarly, if $\beta(0) \in \mathcal{N}_i^+$, then $\beta(t) \to y^* := \min\{y \in \mathcal{I} : y > x_i^*\}$. Thus, x_i^* is a repeller for ODE (25) and $\mathbf{h}(x_i^*)$ is a saddle point for ODE (22).

Part (iii) If $\beta(0) = x_i^*$, then clearly $\beta(t) = x_i^*$ for all $t \geq 0$ and $\Upsilon(t) \to \mathbf{h}(x_i^*)$ as $t \to \infty$. Say $g(x) > 0$ for all $x \in \mathcal{N}_i^-$ and $g(x) > 0$ for all $x \in \mathcal{N}_i^+$. Then, if $\beta(0) \in \mathcal{N}_i^-$, $\beta(t) \to x_i^*$, as shown for part 1. While if $\beta(0) \in \mathcal{N}_i^+$, then $\beta(t) \to y^* := \min\{y \in \mathcal{I} : y > x_i^*\}$, as shown for part 2. Thus, x_i^* is a saddle point for ODE (25) and $\mathbf{h}(x_i^*)$ is a saddle point for ODE (22).

Lastly, consider the initial condition $\Upsilon(0) \in \mathcal{D}_I$ with $\psi^c(0) = 0$, then ODE (22) simplifies to $\dot{\Upsilon} = -\Upsilon$, which clearly has unique solution and $\Upsilon(t) \to \mathbf{0}$ as $t \to \infty$. We have shown above that whenever $\psi^c(0) > 0$, $\Upsilon(t) \nrightarrow \mathbf{0}$. Therefore, $\mathbf{0} \in \mathcal{S}$. □

Proof of Theorem 3. At first, observe that in view of the hypothesis regarding $\mathcal{F}$ and (26) with $\sum_i \mu_i = 1$, the assumption **C.1** holds. Further, it is clear from (1)-(10), (19) and (27) that the assumption **C.2** holds.

We will now prove that $\mathcal{A}_\beta^u \neq \emptyset$, which will then imply that the assumption **C.3** holds, by Theorem 2; this would complete Theorem 1(i). Towards proving the claim, note that (recall $\mu_2 > 0$):

$$\begin{aligned}
&\underline{g}_\beta^u(\beta) < g_\beta^u(\beta) \leq \overline{g}_\beta^u(\beta) \text{ for all } \beta \in [0,1] \text{ where} \\
&\underline{g}_\beta^u(\beta) := \left(-\beta\mu_2 - \beta\mu_1(1-\alpha_x^u\rho) + (1-\beta)\mu_1\rho\alpha_y^u\right)m_f\eta^u - \beta\mu_a m_f\eta_a, \text{ and} \\
&\overline{g}_\beta^u(\beta) := \left(-\beta\mu_1(1-\alpha_x^u\rho) + (1-\beta)\mu_1\rho\alpha_y^u + \mu_2(1-\beta)\right)m_f\eta^u - \beta\mu_a m_f\eta_a.
\end{aligned} \tag{A.6}$$

Now, $\underline{g}_\beta^u(0) = \mu_1\rho\alpha_y^u m_f\eta^u \geq 0$; thus, $g_\beta^u(0) > 0$. Further, $\overline{g}_\beta^u(1) = -\mu_1(1-\alpha_x^u\rho)m_f\eta^u - \mu_a m_f\eta_a \leq 0$; thus, $g_\beta^u(1) \leq 0$. Since $g_\beta^u(\beta)$ is a continuous function of β, therefore there exists at least one zero of g_β^u, say $\beta^{u,\infty}$ such that $g_\beta^u > 0$ in $\mathcal{N}_\epsilon^-(\beta^{u,\infty})$ and $g_\beta^u < 0$ in $\mathcal{N}_\epsilon^+(\beta^{u,\infty})$; $\mathcal{N}_\epsilon^+(\beta^{u,\infty}) = \emptyset$ if $\beta^{u,\infty} = 1$. Then, by Theorem 2, $\beta^{u,\infty} \in \mathcal{A}_\beta^u$; thus, $\mathcal{A}_\beta^u \neq \emptyset$.

Since $\overline{g}^u_\beta(\beta)$ is a linear function such that $\overline{g}^u_\beta(0) > 0$ and (recall) $\overline{g}^u_\beta(1) \leq 0$, therefore, $\overline{\beta}^u \in (0, 1]$, given in (29), is the unique zero of $\overline{g}^u_\beta(\beta)$. Further, since $g^u_\beta \leq \overline{g}^u_\beta$ and $\overline{g}^u_\beta(\beta) < 0$ for all $\beta \in (\overline{\beta}^u, 1]$ when $\overline{\beta}^u < 1$, therefore, there exists no zero of g^u_β in $(\overline{\beta}^u, 1]$; if $\overline{\beta}^u = 1$, then also, any zero of g^u_β is atmost 1. Thus, if at all, there is any zero of g^u_β, which can be an attractor or repeller or saddle point of (28), it is lesser than or equals to $\overline{\beta}^u$. Next, notice that there is a unique zero of the function $\underline{g}^u_\beta$, namely $\underline{\beta}^u \in (0, 1)$, as given in (29). Again using similar arguments as before, we get that $\beta^{u,\infty} > \underline{\beta}^u$. This proves (29).

Now, by Theorem 2, the attractor and saddle sets are as in the hypothesis with subset of the combined domain of attraction as $\mathcal{D}_I$.

We will now identify the compact sub-domain of $\mathcal{D}_I$ for completing the proof using Theorem 1. From **C.1** for our case, one can bound Ψ^a_n:

$$0 \leq \Psi^a_n \leq \overline{\Psi}^a_n := \frac{1}{n}\left(\sum_{k=1}^{\min\{\nu_e, n\}} 2\mathcal{F} 1_{\{\Psi^c_k > 0\}} + s^c_0\right).$$

By strong law of large numbers, $\overline{\Psi}^a_n \to 2E[\mathcal{F}]$ a.s. in survival paths and $\overline{\Psi}^a_n \to 0$ in extinction paths, as $n \to \infty$. Thus, $\mathcal{D}_b := \mathcal{D}_I \cap \{\Upsilon : \psi^a \in [0, 2E(\mathcal{F})]\}$ is the compact subset of $\mathcal{D}_I$ and $p_b := P(\boldsymbol{\Upsilon}_n \text{ visits } \mathcal{D}_b \text{ i.o.}) = 1$. Hence, by Theorem 1(ii), the claim holds. □

Proof of Theorem 4. Let all parameters except κ be fixed. Consider the case when $\nabla^u(\kappa, \kappa+\partial\kappa) = g^u_\beta(\beta^{\infty,u}(\kappa); \kappa+\partial\kappa) > 0$ for some $\partial\kappa > 0$. Since $g^u_\beta(\beta; \kappa + \partial\kappa)$ is either a convex or concave or linear function of β with a unique zero in $(0, 1)$, therefore, there exists a $\beta^{\infty,u}(\kappa + \partial\kappa) > \beta^{\infty,u}(\kappa)$ such that $g^u_\beta(\beta^{\infty,u}(\kappa + \partial\kappa); \kappa + \partial\kappa) = 0$. One can prove the claim similarly when $\nabla^u(\kappa, \kappa + \partial\kappa) < 0$. Lastly if $\nabla^u(\kappa, \kappa + \partial\kappa) = 0$, then again due to uniqueness, $\beta^{\infty,u}(\kappa + \partial\kappa) = \beta^{\infty,u}(\kappa)$. □

Proof of Corollary 1. We will first show that the function $g^{o,u}_\beta$ is either convex or concave or linear depending upon warning-specific and user-specific parameters. Towards this, note that for each u:

$$\begin{aligned}\frac{dg^{o,u}_\beta(\beta)}{d\beta} &= -(\mu_1 + \mu_2)m_f\eta^u + (\alpha^u_x - \alpha^u_y)(\mu_1\rho + \mu_2\omega(\beta))m_f\eta^u + (\beta\alpha^u_x + (1 - \beta)\alpha^u_y)\frac{bw\mu_2 m_f\eta^u}{(\beta + b(1 - \beta))^2} - \mu_a m_f\eta_a \\ \implies \frac{d^2 g^{o,u}_\beta(\beta)}{d\beta^2} &= \frac{2m_f\eta^u bw\mu_2}{(\beta + b(1 - \beta))^3}\left(b\alpha^u_x - \alpha^u_y\right).\end{aligned} \tag{A.7}$$

Thus, if $bw\mu_2(b\alpha^u_x - \alpha^u_y) = 0$ or < 0 or > 0, then $g^{o,u}_\beta$ is a linear, concave or convex function respectively. From (32):

$$g^{o,u}_\beta(0) = (\mu_1\rho + \mu_2\gamma)\,\alpha^u_y m_f\eta^u > 0, \text{ and } g^{o,u}_\beta(1) = -\Big(\mu_1 m_f\eta^u\,(1 - \alpha^u_x\rho) + \mu_2(1 - \alpha^u_x(w + \gamma))m_f\eta^u + \mu_a m_f\eta_a\Big) < 0;$$

the last inequality in above holds as $\alpha^u_x(w + \gamma) \leq 1$ for each u and $\alpha^u_x\rho < \alpha^u_x < 1$. Therefore, there exists a unique $\beta^{o,\infty,u} \in (0, 1)$ such that $g^u_\beta(\beta^{o,\infty,u}) = 0$, $g^u_\beta(\beta) > 0$ for all $\beta \in [0, \beta^{o,\infty,u})$ and $g^u_\beta(\beta) < 0$ for all $\beta \in (\beta^{o,\infty,u}, 1]$. This implies that for the ODE (28), $t \mapsto \beta^u(t)$ is strictly increasing if $\beta^u(0) \in [0, \beta^{o,\infty,u})$ and strictly decreasing if $\beta^u(0) \in (\beta^{o,\infty,u}, 1]$. Thus, $\mathcal{A}^{o,u}_\beta = \{\beta^{o,\infty,u}\}$ with the domain of attraction as $[0, 1]$. Lastly, observe that $g^{o,u}_\beta(\beta) \leq \overline{g}^u_\beta(\beta)$ for each $\beta \in [0, 1]$, therefore, $\beta^{o,\infty,u} \leq \overline{\beta}^u$, as these two zeroes are unique zeroes of their respective functions (see (A.6)).

Proof of Corollary 2. Recall from Corollary 1, $g^{o,u}_\beta$ has a unique attractor, $\beta^{o,\infty,u} \in (0, 1)$, for each $u \in \{R, F\}$. Observe further that $g^{o,u}_\beta(\beta^{o,\infty,u}(w); w) = 0$ and $g^u_\beta(\beta^{o,\infty,u}(b); b) = 0$. Henceforth, the corollary will be proved using

Theorem 4. For any $\partial w > 0$ and $\partial b > 0$, we get:

$$\begin{aligned}\nabla^u(w, w+\partial w) &= g_\beta^{o,u}(\beta^{o,\infty,u}(w); w+\partial w)\\ &= g_\beta^{o,u}(\beta^{o,\infty,u}(w); w) + m_f\eta^u\mu_2\Big(\alpha_x^u\beta^{o,\infty,u}(w) + \alpha_y^u(1-\beta^{o,\infty,u}(w))\Big)\left(\frac{\partial w\beta^{o,\infty,u}(w)}{\beta^{o,\infty,u}(w) + (1-\beta^{o,\infty,u}(w))b}\right) > 0 \text{ and}\\ \nabla^u(b, b+\partial b) &= g_\beta^{o,u}(\beta^{o,\infty,u}(b); b+\partial b)\\ &= g_\beta^{o,u}(\beta^{o,\infty,u}(b); b) - \partial b m_f\eta^u\mu_2\frac{w\beta^{o,\infty,u}(b)\Big(\alpha_x^u\beta^{o,\infty,u}(b) + \alpha_y^u(1-\beta^{o,\infty,u}(b)\Big)}{\Big(\beta^{o,\infty,u}(b) + (1-\beta^{o,\infty,u}(b))(b+\partial b)\Big)\Big(\beta^{o,\infty,u}(b) + (1-\beta^{o,\infty,u}(b))b\Big)} < 0.\end{aligned}$$

Thus, by Theorem 4, $\beta^{o,\infty,u}(w, b)$ strictly increases with w and strictly decreases with b for any $u \in \{R, F\}$. □

Proof of Theorem 5. In this proof, we explicitly show the dependency of zeros of (32) on design parameters (w, b). Part (i) Consider a $\delta > 0$ such that $\beta^{o,\infty,R}(\overline{w}, 0) > \delta$. Then, $w \in [0, \overline{w}] = W_1 \cup W_2$, where $W_1 := \{w : \beta^{o,\infty,R}(w, 0) > \delta\}$ and $W_2 := \{w : \beta^{o,\infty,R}(w, 0) \leq \delta\}$. If $W_2 \neq \emptyset$, by Corollary 1, there exists a $\widetilde{w} > 0$ such that $\beta^{o,\infty,R}(\widetilde{w}, 0) = \delta$, $W_1 = \{w : w > \widetilde{w}\}$, and $W_2 := \{w : w \leq \widetilde{w}\}$. The proof for case with $W_2 = \emptyset$ is trivially true once the other case is proved. Hence, consider $W_2 \neq \emptyset$.

Consider $w \in W_1$. Then, by Corollary 2, there exists a unique $b(w;\delta) > 0$ such that $\beta^{o,\infty,R}(w, b(w;\delta)) = \delta$ (i.e., the zero of $g_\beta^{o,F}$ equals δ) and hence:

$$b(w;\delta) := \left(\frac{\delta}{1-\delta}\right)(wp(\delta) - 1), \text{ where } p(\delta) := \frac{\eta^R\mu_2(\delta\alpha_x^R + (1-\delta)\alpha_y^R)}{\delta((\mu_1+\mu_2)\eta^R + \mu_a\eta_a) - \eta^R(\mu_1\rho + \mu_2\gamma)(\delta\alpha_x^R + (1-\delta)\alpha_y^R)}. \tag{A.8}$$

Thus, again by Corollary 2 and because $[0, \overline{w}] \cap W_1 = (\widetilde{w}, \overline{w}]$ (as said before):

$$\sup_{w\in[0,\overline{w}]\cap W_1; b\in[0,\infty); \beta^{o,\infty,R}(w,b)\leq\delta} \beta^{o,\infty,F}(w, b) = \sup_{w\in(\widetilde{w},\overline{w}]} \beta^{o,\infty,F}(w, b(w;\delta)). \tag{A.9}$$

By Lemma 2, $\beta^{o,\infty,F}(w, b(w;\delta))$ strictly increases with w, for every $\delta > 0$. Then, the optimal value for the problem in (A.9) is given by:

$$\sup_{w\in(\widetilde{w},\overline{w}]} \beta^{o,\infty,F}(w, b(w;\delta)) = \beta^{o,\infty,F}(\overline{w}, b(\overline{w};\delta)). \tag{A.10}$$

Now, consider $w \in W_2$. Then, $\beta^{o,\infty,R}(w, 0) \leq \delta$. Further by Corollary 2, for any $w < \widetilde{w}$ and $b > 0$, we have:

$$\beta^{o,\infty,F}(\widetilde{w}, 0) > \beta^{o,\infty,F}(w, 0) > \beta^{o,\infty,F}(w, b), \text{ and } \beta^{o,\infty,F}(\widetilde{w}, 0) > \beta^{o,\infty,F}(\widetilde{w}, b).$$

Thus, we have:

$$\sup_{w\in[0,\overline{w}]\cap W_2; b\in[0,\infty); \beta^{o,\infty,R}(w,b)\leq\delta} \beta^{o,\infty,F}(w, b) = \beta^{o,\infty,F}(\widetilde{w}, 0). \tag{A.11}$$

In all, by (A.10), (A.11), we have:

$$\sup_{w\in[0,\overline{w}]; b\in[0,\infty); \beta^{o,\infty,R}(w,b)\leq\delta} \beta^{o,\infty,F}(w, b) = \max\left\{\beta^{o,\infty,F}(\overline{w}, b(\overline{w};\delta)), \beta^{o,\infty,F}(\widetilde{w}, 0)\right\}. \tag{A.12}$$

Let us now consider a sequence of $w \downarrow \widetilde{w}$ and observe $\frac{\partial b(w;\delta)}{\partial w} = \left(b(w;\delta) + \frac{\delta}{1-\delta}\right)\frac{1}{w} > 0$. Thus, $b(w;\delta)$ decreases as w decreases. We claim that $\lim_{w\downarrow\widetilde{w}} b(w;\delta) = 0$. Let us suppose on the contrary that the limit is positive; note that the limit can not be negative as $b(w;\delta) > 0$. By continuity of $b(w;\delta)$ with respect to w (see (A.8)), there exists a $w' < \widetilde{w}$ such that $b(w';\delta) > 0$, and further $\beta^{o,\infty,R}(w', b(w';\delta)) = \delta$, by definition of $b(w';\delta)$. However, since $w' \in W_2$, we also have $\beta^{o,\infty,R}(w', 0) \leq \delta$, leading to a contradiction. Thus, the limit is 0.

Consider function $L(\beta; w) := \left(g_\beta^{o,F}(\beta(w, b(w))\right)^2$. Clearly this function is jointly continuous and has a unique minimum at $\beta^{o,\infty,F}(w, b(w))$ for each w (as it is the unique zero of $g_\beta^o(\cdot)$). Hence by Maximum Theorem:

$$\beta^{o,\infty,F}(w, b(w)) \to \beta^{o,\infty,F}(\widetilde{w}, 0), \text{ as } w \downarrow \widetilde{w} \text{ and further by Lemma 2, } \beta^{o,\infty,F}(w, b(w)) \downarrow \beta^{o,\infty,F}(\widetilde{w}, 0).$$

Thus, $\beta^{o,\infty,F}(\widetilde{w}, 0) \leq \beta^{o,\infty,F}(w, b(w;\delta)) < \beta^{o,\infty,F}(\overline{w}, b(\overline{w};\delta))$, where the last inequality is again due to Lemma 2. Conclusively, by (A.12), we get that

$$\sup_{w\in[0,\overline{w}];b\in[0,\infty);\beta^{o,\infty,R}(w,b)\leq\delta} \beta^{o,\infty,F}(w, b) = \beta^{o,\infty,F}(\overline{w}, b(\overline{w};\delta)).$$

<u>Part (ii)</u> Consider $\delta > 0$ such that $\beta^{o,\infty,R}(\overline{w}, 0) \leq \delta$. Again, by Corollary 2, for all $w \in [0, \overline{w}]$ and $b > 0$:

$$\beta^{o,\infty,F}(\overline{w}, 0) > \beta^{o,\infty,F}(w, 0) > \beta^{o,\infty,F}(w, b), \text{ and } \beta^{o,\infty,F}(\overline{w}, 0) > \beta^{o,\infty,F}(\overline{w}, b).$$

Thus, the optimal value is achieved at $b = 0$ and $w = \overline{w}$, with $\beta^{o,\infty,R}(\overline{w}, 0) \leq \delta$. □

Lemma 2. *The function $\beta^{o,\infty,F}(w, b(w;\delta))$ strictly increases with w, when $w < \overline{w}$, for every $\delta > 0$.*

Proof. Fix w and $\partial w > 0$, we have (for simplicity, denote $\beta^{o,\infty,F}(w, b(w;\delta))$ by $\beta_\delta(w)$):

$$\begin{aligned}
\nabla^F(w, w + \partial w; b) &= g_\beta^{o,F}(\beta_\delta(w); w + \partial w) - g_\beta^{o,F}(\beta_\delta(w); w) \\
&= m_f\eta^F\mu_2\left(\alpha_x^F\beta_\delta(w) + \alpha_y^F(1 - \beta_\delta(w))\right)\left(\left(\frac{(w + \partial w)\beta_\delta(w)}{\beta_\delta(w) + (1 - \beta_\delta(w))b(w + \partial w;\delta)}\right) - \left(\frac{w\beta_\delta(w)}{\beta_\delta(w) + (1 - \beta_\delta(w))b(w;\delta)}\right)\right) \\
&= \frac{m_f\eta^F\mu_2\beta_\delta(w)\left(\alpha_x^F\beta_\delta(w) + \alpha_y^F(1 - \beta_\delta(w))\right)}{\left(\beta_\delta(w) + (1 - \beta_\delta(w))b(w + \partial w;\delta)\right)\left(\beta_\delta(w) + (1 - \beta_\delta(w))b(w;\delta)\right)}\left(\partial w\left((1 - \beta_\delta(w))b(\partial w;\delta) + \beta_\delta(w)\right)\right), \text{ where}
\end{aligned} \tag{A.13}$$

the last equality follows by simple algebra after substituting for $b(\cdot;\delta)$ from (A.8). Since $b(\cdot,\delta) > 0$ and by Theorem 3, $\beta_\delta(w) \in (\underline{\beta}^F, \overline{\beta}^F] \subset [0, 1]$, therefore, $\nabla^F(w, w + \partial w;\delta) > 0$ for any $\delta > 0$. Thus, the proof follows by Theorem 4. □

Proof of Corollary 3. Consider any $\delta > 0$, $\mu_a \in (0, 1 - \mu_1 - \mu_2]$ and let b^*, w^* be as in Theorem 5.
<u>Case 1: when $b^* > 0$:</u> Let $b^*(\mu_a = 0) =: b_0^*$. From (34), observe that b^* is a strictly decreasing function of μ_a, therefore, $b_0^* > b^* > 0$. Further, from (32), we have:

$$\begin{aligned}
g_\beta^{o,F}(\beta_{\text{na}}^o) &= g_\beta^{o,F}(\beta_{\text{na}}^o; \mu_a = 0) - \beta_{\text{na}}^o\mu_a m_f\eta_a \\
&\qquad + \mu_2 m_f\eta^F\left(\beta_{\text{na}}^o\alpha_x^F + (1 - \beta_{\text{na}}^o)\alpha_y^F\right)\left(\frac{w^*\beta_{\text{na}}^o}{\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b^*} - \frac{w^*\beta_{\text{na}}^o}{\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b_0^*}\right) \\
&= 0 + \beta_{\text{na}}^o\mu_a m_f\eta_a\left(\frac{\mu_2\eta^F w^*(1 - \beta_{\text{na}}^o)(\beta_{\text{na}}^o\alpha_x^F + (1 - \beta_{\text{na}}^o)\alpha_y^F)}{\left(\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b_0^*\right)\left(\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b^*\right)}\left(\frac{b_0^* - b^*}{\eta_a\mu_a}\right) - 1\right).
\end{aligned} \tag{A.14}$$

Define $p(\mu_a) := \delta((\mu_1 + \mu_2)\eta^R + \mu_a\eta_a) - \eta^R(\mu_1\rho + \mu_2\gamma)(\delta\alpha_x^R + (1 - \delta)\alpha_y^R)$. Then, by (34), we have:

$$\frac{b_0^* - b^*}{\eta_a\mu_a} = \left(\frac{\delta^2}{1 - \delta}\right)\left(\frac{w^*\eta^R\mu_2(\delta\alpha_x^R + (1 - \delta)\alpha_y^R)}{p(\mu_a)p(0)}\right).$$

Substitute the above term in (A.14) and consider the following limit to analyse (A.14):

$$\lim_{\delta\to0}\left(\frac{\mu_2\eta^F w^*(1 - \beta_{\text{na}}^o)(\beta_{\text{na}}^o\alpha_x^F + (1 - \beta_{\text{na}}^o)\alpha_y^F)}{\left(\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b_0^*\right)\left(\beta_{\text{na}}^o + (1 - \beta_{\text{na}}^o)b^*\right)}\right)\lim_{\delta\to0}\left(\frac{\delta^2}{1 - \delta}\right)\lim_{\delta\to0}\left(\frac{w^*\eta^R\mu_2(\delta\alpha_x^R + (1 - \delta)\alpha_y^R)}{p(\mu_a)p(0)}\right) - 1.$$

In the above, the second limit is clearly 0 and the rate of convergence is independent of other factors. The first and third limits are finite, and the respective terms can be upper-bounded independent of μ_a and other factors. Thus, the product of three limits is 0, and the rate of convergence is uniform in μ_a and b^*, i.e., there exists a $\overline{\delta} > 0$ such that (for example):

$$\left(\frac{\mu_2 \eta^F w^*(1-\beta^o_{\text{na}})(\beta^o_{\text{na}}\alpha^F_x + (1-\beta^o_{\text{na}})\alpha^F_y)}{\Big(\beta^o_{\text{na}} + (1-\beta^o_{\text{na}})b^*_0\Big)\Big(\beta^o_{\text{na}} + (1-\beta^o_{\text{na}})b^*\Big)}\right)\left(\frac{\delta^2}{1-\delta}\right)\left(\frac{w^*\eta^R\mu_2(\delta\alpha^R_x + (1-\delta)\alpha^R_y)}{p(\mu_a)p(0)}\right) - 1 < -\frac{1}{2} \text{ for all } \delta \le \overline{\delta} \text{ and } \mu_a > 0.$$

Thus, from (A.14), $g^{o,F}_\beta(\beta^o_{\text{na}}) < -\beta^o_{\text{na}}\mu_a m_f \eta_a/2 < 0$ for any $\mu_a > 0$ and all $\delta \le \overline{\delta}$.

Recall from the proof of Corollary 2 that $g^{o,F}_\beta(\cdot)$ is either convex/concave/linear with a unique zero in $(0,1)$. Therefore, the unique zero of $g^{o,F}_\beta(\cdot;\mu_a)$, namely $\beta^o(\mu_a) < \beta^o_{\text{na}}$ for all $\delta \le \overline{\delta}$ and for all $\mu_a \in (0, 1-\mu_1-\mu_2]$.
Case 2: when $b^* = 0$: Here, again $b^*(\mu_a = 0) =: b^*_0 > b^* = 0$. Then, similar to (A.14), using (34):

$$g^{o,F}_\beta(\beta^o_{\text{na}}) = \beta^o_{\text{na}}\mu_a m_f \eta_a \left(\frac{\mu_2 \eta^F w^*(1-\beta^o_{\text{na}})(\beta^o_{\text{na}}\alpha^F_x + (1-\beta^o_{\text{na}})\alpha^F_y)b^*_0}{\eta_a\mu_a\beta^o_{\text{na}}\Big(\beta^o_{\text{na}} + (1-\beta^o_{\text{na}})b^*_0\Big)}\left(\left(\frac{\delta}{1-\delta}\right)\left(\frac{w^*\eta^R\mu_2(\delta\alpha^R_x + (1-\delta)\alpha^R_y)}{p(0)} - 1\right)\right) - 1\right).$$

Hereafter, the proof follows as in Case 1. □

Proof of Theorem 6. We begin the proof for the fake-post.
Part (i) Consider $0 < \mu_a \le \min\{1-\mu_1-\mu_2, \Delta_a\}$. Then, by the definition of upper-bound Δ_a and (41), $\alpha^F_x\omega^a(\beta^o_{\text{na}}) \le 1$. Note from (41) that $\omega^a(\beta)$ is a strictly increasing function of β. Therefore, $\alpha^F_x\omega^a(\beta) \le 1$ for all $\beta \le \beta^o_{\text{na}}$, for given μ_a.

This implies that for $\beta \le \beta^o_{\text{na}}$, we have $g^{a,F}_\beta(\beta) = g^{o,F}_\beta(\beta;\mu_a = 0)$ (see (42)). Further, β^o_{na} is a zero of $g^{a,F}_\beta$, as $g^{a,F}_\beta(\beta^o_{\text{na}}) = g^{o,F}_\beta(\beta^o_{\text{na}};\mu_a = 0) = 0$. Furthermore, by uniqueness given in Corollary 1, β^o_{na} is the unique zero of $g^{a,F}_\beta$ in $[0,\beta^o_{\text{na}}]$. Therefore, any $\beta^a \in \mathcal{A}^{a,F}_\beta \cup \mathcal{S}^{a,F}_\beta$ is in $[\beta^o_{\text{na}}, 1]$.
Part (ii) Consider $\mu_a > \Delta_a$. Then, the corresponding $\alpha^F_x\omega^a(\beta^o_{\text{na}}) > 1$. Define the function $h(\beta) := \alpha^F_x\omega^a(\beta) - 1$. It is easy to see that $h(0) < 0$, $h(1) > 0$ and $h(\cdot)$ is a strictly increasing function. Thus, there exists a unique zero of h, denoted by $\widetilde{\beta} \in (0,1)$, i.e., $\alpha^F_x\omega^a(\widetilde{\beta}) = 1$. As $\beta \mapsto \omega^a(\beta)$ is strictly increasing, we further have $\alpha^F_x\omega^a(\beta) < 1$ for all $\beta < \widetilde{\beta}$; furthermore $\widetilde{\beta} < \beta^o_{\text{na}}$ as $\alpha^F_x\omega^a(\beta^o_{\text{na}}) > 1$.

From (42), we have:

$$g^{a,F}_\beta(\beta) = g^{o,F}_\beta(\beta;\mu_a = 0) + \mu_2 m_f \eta^F \left\{\beta\Big(\min\{1,\alpha^F_x\omega^a(\beta)\} - \alpha^F_x\omega^a(\beta)\Big) + (1-\beta)\Big(\min\{1,\alpha^F_y\omega^a(\beta)\} - \alpha^F_y\omega^a(\beta)\Big)\right\}. \tag{A.15}$$

Thus, $g^{a,F}_\beta(\beta) < g^{o,F}_\beta(\beta;\mu_a = 0)$ if $1 < \alpha^F_j\omega^a(\beta)$ for some $j \in \{x,y\}$, and $g^{a,F}_\beta(\beta) = g^{o,F}_\beta(\beta;\mu_a = 0)$ if $\alpha^F_j\omega^a(\beta) \le 1$ for each $j \in \{x,y\}$. As a result, we have:

(a) for $\beta \in [0,\widetilde{\beta}]$, $g^{a,F}_\beta(\beta) = g^{o,F}_\beta(\beta;\mu_a = 0) > 0$, and

(b) for $\beta \in [\beta^o_{\text{na}}, 1]$, $g^{a,F}_\beta(\beta) < g^{o,F}_\beta(\beta;\mu_a = 0) \le 0$.

By Theorem 3, there exists at least one zero of $g^{a,F}_\beta$, say β^a and by above arguments, $\beta^a \in (\widetilde{\beta},\beta^o_{\text{na}})$. We will now claim and show that $\beta^a > \beta^o$, but first observe that $\beta^o < \beta^o_{\text{na}}$ by Corollary 3. Towards this, note that for $\beta \in (\widetilde{\beta},\beta^o_{\text{na}})$, we have:

$$\begin{aligned} g^{a,F}_\beta(\beta) &= g^{o,F}_\beta(\beta) + \mu_2 m_f \eta^F \left\{\beta\Big(\min\{1,\alpha^F_x\omega^a(\beta)\} - \alpha^F_x\omega(\beta)\Big) + (1-\beta)\Big(\min\{1,\alpha^F_y\omega^a(\beta)\} - \alpha^F_y\omega(\beta)\Big)\right\} \\ &= g^{o,F}_\beta(\beta) + \mu_2 m_f \eta^F \left\{\beta\Big(1 - \alpha^F_x\omega(\beta)\Big) + (1-\beta)\Big(\min\{1,\alpha^F_y\omega^a(\beta)\} - \alpha^F_y\omega(\beta)\Big)\right\}. \end{aligned} \tag{A.16}$$

In the above, if $1 > \alpha_y^F \omega^a(\beta)$, then:

$$\begin{aligned} g_\beta^{a,F}(\beta) &= g_\beta^{o,F}(\beta) + \mu_2 m_f \eta^F \left\{\beta\Big(1 - \alpha_x^F \omega(\beta)\Big) + (1-\beta)\alpha_y^F\Big(\omega^a(\beta) - \omega(\beta)\Big)\right\} \\ &= g_\beta^{o,F}(\beta) + \mu_2 m_f \eta^F \left\{\beta\Big(1 - \alpha_x^F \omega(\beta)\Big) + (1-\beta)\alpha_y^F \left(\frac{\beta\mu_a m_f \eta_a}{\mu_2 m_f \eta^F \left(\beta\alpha_x^F + (1-\beta)\alpha_y^F\right)}\right)\right\} > g_\beta^{o,F}(\beta), \end{aligned}$$

as $\omega(\beta)\alpha_x^F < 1$ for all $\beta \in [0,1)$. Further, $\omega(\beta)\alpha_y^F < 1$ for all $\beta \in [0,1)$, hence even with $1 \le \alpha_y^F \omega^a(\beta)$, we have:

$$g_\beta^{a,F}(\beta) = g_\beta^{o,F}(\beta) + \mu_2 m_f \eta^F \left\{\beta\Big(1 - \alpha_x^F \omega(\beta)\Big) + (1-\beta)\Big(1 - \alpha_y^F \omega(\beta)\Big)\right\} > g_\beta^{o,F}(\beta).$$

Now, for $\beta \in (\widetilde{\beta}, \beta^o]$, $g_\beta^{o,F}(\beta) \ge 0$, and thus, $g_\beta^{a,F}(\beta) > 0$. This completes the proof of the claim.

Now, consider the real-post. By Theorem 3, $\mathcal{A}_\beta^{a,R} \neq \emptyset$, therefore, there exists at least one zero of $g_\beta^{a,R}$, say $\beta^{a,R} \in (0,1)$. Now, using arguments as above:

$$\begin{aligned} g_\beta^{a,R}(\beta) = g_\beta^{o,R}(\beta; \mu_a = 0) &+ \mu_2 m_f \eta^R \left\{\beta\Big(\min\{1, \alpha_x^R \omega^a(\beta)\} - \alpha_x^R \omega^a(\beta)\Big) + (1-\beta)\Big(\min\{1, \alpha_y^R \omega^a(\beta)\} - \alpha_y^R \omega^a(\beta)\Big)\right\} \\ &+ \beta\mu_a m_f \eta_a \left(\frac{\eta^R}{\eta^F}\left(\frac{\beta\alpha_x^R + (1-\beta)\alpha_y^R}{\beta\alpha_x^F + (1-\beta)\alpha_y^F}\right) - 1\right) < g_\beta^{o,R}(\beta; \mu_a = 0). \end{aligned}$$

Thus, any zero of $g_\beta^{a,R}(\beta)$ is strictly less than the unique zero of $g_\beta^{o,R}(\beta; \mu_a = 0)$, i.e., $\beta^{a,R} < \beta^{o,R}(\mu_a = 0) \le \delta$, for any $\beta^{a,R} \in \mathcal{A}_\beta^{a,R} \cup \mathcal{S}_\beta^{a,R}$ (see Theorem 5). □

Proof of Theorem 7. We divide the proof in two cases.

Case 1: If $\overline{\phi} < \frac{1}{\alpha_y^R \omega^a(\delta)}$. Then $\overline{\phi}$ is the unique zero of $g_{\beta,\phi}^{h,R}(\delta) = 0$. Further, for any $\phi' \in \left(\overline{\phi}, \frac{1}{\alpha_y^R \omega^a(\delta)}\right)$, $g_{\beta,\phi'}^{h,R}(\delta) > 0$. By (A.6), $g_{\beta,\phi'}^{h,R}(1) < 0$. Thus, there exists at least one zero of $g_{\beta,\phi'}^{h,R}$ greater than δ. Now, consider any $\phi' \ge \frac{1}{\alpha_y^R \omega^a(\delta)}$. Since the function $\phi \mapsto g_{\beta,\phi}^{h,R}(\delta)$ is continuous, therefore, $g_{\beta,\phi'}^{h,R}(\delta) > 0$ for such ϕ. Thus, again as before, there exists at least one zero of $g_{\beta,\phi'}^{h,R}$ greater than δ. Hence, any ϕ satisfying the constraint in (45) is less than or equals to $\overline{\phi}$. Thus, the optimizer of (45) is $\phi^* = \overline{\phi}$.

Case 2: If $\overline{\phi} \ge \frac{1}{\alpha_y^R \omega^a(\delta)}$. Then by monotonicity, for any $\phi \ge \overline{\phi}$:

$$g_{\beta,\phi}^{h,R}(\beta) \le q_\phi(\beta) := \Big(-\beta\mu_2 - \beta\mu_1(1 - \alpha_x^R \rho) + (1-\beta)\mu_1 \rho \alpha_y^R + \mu_2 \phi \omega^a(\beta)\Big(\beta\alpha_x^R + (1-\beta)\alpha_y^R\Big)\Big) m_f \eta^R - \beta\mu_a m_f \eta_a.$$

Thus for all such ϕ, $q_\phi(\delta)$ is a strictly increasing function of ϕ with $q_1(\delta) < 0$ (by Theorem 6) and $q_{\overline{\phi}}(\delta) = 0$. Thus, $g_{\beta,\phi}^{h,R}(\delta) \le q_\phi(\delta) \le 0$.

Further, by strict monotonicity of $\omega^a(\cdot)$ in β, we have $\phi\alpha_y^R \omega^a(\beta) \ge 1$ for all $\beta > \delta$ whenever $\phi \ge \overline{\phi}$. Thus, $g_{\beta,\phi}^{h,R}(\beta)$ is linearly (strictly) decreasing in β, when $\beta > \delta$. As already proved $g_{\beta,\phi}^{h,R}(\delta) < 0$, and hence $g_{\beta,\phi}^{h,R}(\beta) < 0$ for all $\beta > \delta$. Hence, the feasibility condition of (45) is satisfied for any $\phi \ge \overline{\phi}$.

By definition of ϕ^* in this case (the second row), we have:

$$\phi^* \omega^a(\beta)\alpha_y^F = 1 \text{ for all } \beta \ge \underline{\beta}^F.$$

Further, $\min\{1, \phi^* \omega^a(\beta)\alpha_y^F\} = 1$ for all $\beta > \underline{\beta}^F$, when $\phi \ge \phi^*$. Thus, the functions $g_{\beta,\phi}^{h,F}(\beta) = g_{\beta,\phi^*}^{h,F}(\beta)$ for all $\beta \ge \underline{\beta}^F$. Also, by Theorem 3, any zero of $g_{\beta,\phi}^{h,F}$ is larger than $\underline{\beta}^F$. Thus, $\left\{\beta : \beta \in \mathcal{A}_\beta^{h,\phi} \cup \mathcal{S}_\beta^{h,\phi}\right\} = \left\{\beta : \beta \in \mathcal{A}_\beta^{h,\phi^*} \cup \mathcal{S}_\beta^{h,\phi^*}\right\}$. Now, given any β, observe that $\phi \mapsto g_{\beta,\phi}^{h,F}(\beta)$ is an increasing (actually non-decreasing) function. Thus, $\inf\left\{\beta : \beta \in \mathcal{A}_\beta^{h,\phi} \cup \mathcal{S}_\beta^{h,\phi}\right\}$ increases with ϕ. Conclusively, we get that ϕ^* is an optimizer of (45). □

References

[1] Lazer, David MJ, et al. ”The science of fake news.” Science 359.6380 (2018): 1094-1096.

[2] Dhounchak, Ranbir, and Veeraruna Kavitha. ”Decomposable Branching Processes and Viral Marketing.” arXiv preprint arXiv:1907.00160 (2019).

[3] Agarwal, Khushboo, and Veeraruna Kavitha. ”Co-virality of competing content over osns?.” 2021 IFIP Networking Conference (IFIP Networking). IEEE, 2021.

[4] Agarwal, Khushboo, and Veeraruna Kavitha. ”Saturated total-population dependent branching process and viral markets.” 2022 IEEE 61st Conference on Decision and Control (CDC). IEEE, 2022.

[5] Van der Lans, Ralf, et al. ”A viral branching model for predicting the spread of electronic word of mouth.” Marketing science 29.2 (2010): 348-365.

[6] Talwar, Shalini, et al. ”Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior.” Journal of Retailing and Consumer Services 51 (2019): 72-82.

[7] Allcott, Hunt, and Matthew Gentzkow. ”Social media and fake news in the 2016 election.” Journal of economic perspectives 31.2 (2017): 211-236.

[8] Feng, Yuran. ”Misreporting and Fake News Detection Techniques on the Social Media Platform.” Highlights in Science, Engineering and Technology 12 (2022): 142-152.

[9] Sharma, Karishma, et al. ”Combating fake news: A survey on identification and mitigation techniques.” ACM Transactions on Intelligent Systems and Technology (TIST) 10.3 (2019): 1-42.

[10] Ruchansky, Natali, Sungyong Seo, and Yan Liu. ”Csi: A hybrid deep model for fake news detection.” Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.

[11] Ahmed, Alim Al Ayub, et al. ”Detecting fake news using machine learning: A systematic literature review.” arXiv preprint arXiv:2102.04458 (2021).

[12] Kapsikar, Suyog, et al. ”Controlling fake news by collective tagging: A branching process analysis.” IEEE Control Systems Letters 5.6 (2020): 2108-2113.

[13] Fittipaldi, Maria Clara, and Sandra Palau. ”On multitype Branching Processes with Interaction.” arXiv preprint arXiv:2203.09701 (2022).

[14] Etheridge, Alison, Shidong Wang, and Feng Yu. ”Conditioning the logistic branching process on non-extinction.” arXiv preprint arXiv:1310.5766 (2013).

[15] Ojeda, Gabriel Berzunza, and Juan Carlos Pardo. ”Branching processes with pairwise interactions.” arXiv preprint arXiv:2009.11820 (2020).

[16] Agarwal, Khushboo, and Veeraruna Kavitha. ”Single-out fake posts: participation game and its design.” arXiv preprint arXiv:2303.08484 (2023). (pre- sented at American Control Conference (ACC 2023))

[17] Sui, Mingxiao, Ian Hawkins, and Rui Wang. ”When falsehood wins? Varied effects of sensational elements on users’ engagement with real and fake posts.” Computers in Human Behavior 142 (2023): 107654.

[18] Dhounchak, Ranbir, Veeraruna Kavitha, and Eitan Altman. ”Viral marketing branching processes.” Computer Communications 198 (2023): 140-156.

[19] Agarwal, Khushboo, and Veeraruna Kavitha. ”New results in Branching processes using Stochastic Approximation.” arXiv preprint arXiv:2111.14527 (2021).

[20] Klebaner, Fima C. ”Population-dependent branching processes with a threshold.” Stochastic processes and their applications 46.1 (1993): 115-127.

[21] Athreya, Krishna B., and Peter Jagers, eds. Classical and modern branching processes. Vol. 84. Springer Science & Business Media, 2012.

[22] Jagers, Peter. ”The proportions of individuals of different kinds in two-type populations. A branching process problem arising in biology.” Journal of Applied Probability 6.2 (1969): 249-260.

[23] Athreya, Krishna B., Peter E. Ney, and P. E. Ney. Branching processes. Courier Corporation, 2004.

[24] Piccinini, Livio C., Guido Stampacchia, and Giovanni Vidossich. Ordinary differential equations in $\mathbb{R}^n$. Applied Mathematical Sciences 39, 1984.