Master thesis

A population dynamics model on the social damage by disinformation spread

2025

Yang XIAO

C3IM1019

Department of Computer and Mathematical Sciences
Graduate School of Information Sciences
Tohoku University
Aramaki-Aza-Aoba 6-3-09, Aoba-ku, Sendai, Miyagi 980-8579
JAPAN

Acknowledgements

First and foremost, I extend my deepest gratitude to my supervisor, Professor Hiromi Seno, for his unwavering support and insightful critiques throughout my two-year study journey. His deep commitment to academic excellence and meticulous attention to detail have significantly shaped this dissertation.

My appreciation also goes to the faculty and staff at Tohoku University, whose resources and assistance have been invaluable. I would also like to acknowledge my peers, whose encouragement and the intellectually stimulating environment at the Graduate School of Information Sciences have been vital in completing my research.

I am profoundly grateful to my parents, who nurtured my curiosity, patiently listened to my unconventional ideas, and have supported my academic journey both emotionally and financially since childhood. Their unwavering belief in me has been a constant source of motivation.

Lastly, I extend my heartfelt thanks to everyone who, in one way or another, has contributed to the completion of this dissertation. Your support, whether through academic guidance, thoughtful discussions, or personal encouragement, has made this endeavor possible.

Contents

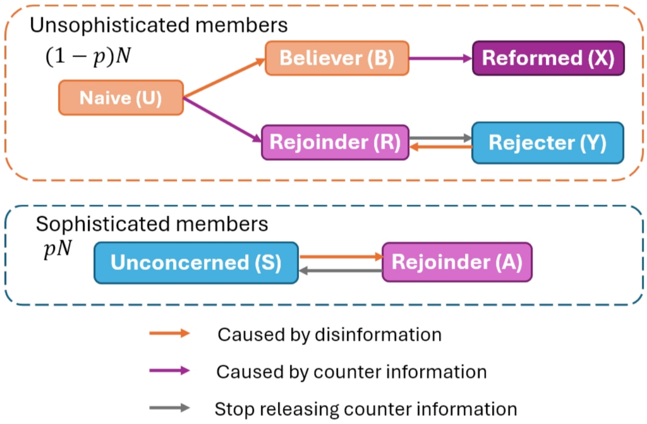

Figure 2.1: Scheme for the state transitions of members in the netizen community with spreading disinformation and its counter information. For detail, see the main text.

1 Introduction

In recent decades, the development of the Internet has made the information exchange much easier and faster than before. Anyone who has a device connected to the Internet can post any piece of information on a platform. Such information can reach another user of the same platform even on the other side of the earth immediately. On the other hand, the spread of a disinformation benefits from the developing Internet facility too. It is not easy for every person to verify the authenticity and accuracy about each piece of information spreading on available platforms. On social occasions like election or in a pandemic situation, people are very easy to encounter various information from unreliable sources. Cuan-Baltazar et al (2020) analysed the contents of 110 websites in the early period of COVID-19 pandemic and found that the information on most of the webpages is not from reliable sources. To make matters worse, some people may deliberately spread some disinformation for their own purposes such as attracting public concern or committing fraud. The spread of disinformation has become one of social issues attracting much concern in recent years. Some governments or societies could release some refutation information to suppress the harmful influence of disinformation. Zhang et al (2022) analysed a mathematical model on the spread of rumors and official rumor-refutation information. Dansu and Seno (2024) studied a population dynamics model on the effect of the timing to release correctional information for the spread of misleading information. Srinivasan and Babu (2021) discussed the model of social immunity to rumors in online social networks.

In this work, we consider a population dynamics model for the reaction between a piece of disinformation and its counter information in a netizen community. We assume some members of the netizen community who can distinguish disinformation by themselves. When they encounter the disinformation, they can identify its falseness and may communicate others about it. We focus on the population structure in the netizen community, which consists of such "sophisticated" members and the others, and try to discuss how the population structure of a netizen community determines the social damage caused by a disinformation spread in it.

2 Assumptions

We consider the spread of information in a netizen community composed of a certain number of members (users). We assume two types of information spreading in the netizen community: a piece of disinformation, and its counter information. We assume the spread of information does not rely on a specific structure of the network. This is applicable for some websites such as forums where the spread of information does not rely on a network of relationships. Every member can post information on it, and each member has a certain chance to contact information from any other member. Members could be regarded as surrounded by the ocean of information.

The members can be influenced by them, and could change their behaviours as follows:

- The total number of members, that is, the total population size of the netizen community does not change during the period that the information spreads in it.
- Members are classified into two classes: “sophisticated” and “unsophisticated”, depending on their ability to identify the falseness of disinformation. Unsophisticated members cannot distinguish the disinformation by themselves. In the time scale to consider the population dynamics related to the disinformation spread, the class is unchanged for every member of the netizen community.
- Sophisticated members never believe the disinformation when they receive it.
- Unsophisticated members who have not received nor accepted either disinformation or its counter information are classified into the “naive” class. Naive members may be influenced by the disinformation and become “believers” who accept and believe it.
- Believers spread the disinformation in the netizen community. That is, they become the source of spreading disinformation.
- Believers may discard their belief on the disinformation if they accept the counter information. Those members who were once believers but no longer believe the disinformation are classified into the “reformed” class.
- Reformed members may release the counter information too. They have experienced the “believer” state so that they could have a reasonable motivation to call members’ attention on the spreading disinformation. The “reformed” members do not transfer to any other state.
- Naive members are susceptible to the counter information too. They may accept the counter information and become “rejoinders”. Any member never believes disinformation once accepting the counter information. Such a member can be regarded as immune to the disinformation.
- Rejoinders release the counter information in the netizen community. Once a rejoinder stops releasing the counter information with the loss of interest about it, such a member transfers to the “rejecter” state.
- Although rejecters are immune to the disinformation and never believe it again, they may return to the rejoinder state and restart releasing the counter information. It could be motivated by the encounter to the spreading disinformation.
- Sophisticated members take one of “rejoinder” and “unconcerned” states. At the rejoinder state, they release the counter information, and at the unconcerned state, they do not act anything with respect to the spreading information. The sophisticated members may change the state between them during the spread of disinformation. Before the emergence of the disinformation in the netizen community, all the sophisticated members are at the unconcerned state.

3 Modelling

3.1 Transition form naive class to believer class

The chance that a naive member becomes a believer could be assumed to have a positive correlation to the concentration of disinformation in the netizen community. The concentration of disinformation is now assumed to be positively correlated to the believer population size in the netizen community since the disinformation is released by believers. In our modelling, we introduce a linear positive correlation between them, that is, the concentration of disinformation in the netizen community is assumed to be proportional to the believer population size.

The probability that a naive member becomes a believer during time interval $[t, t+\Delta t]$ with sufficiently small Δt is now assumed to be given by

$$\beta B(t)\Delta t + \mathrm{o}(\Delta t),$$

where $\mathrm{o}(\Delta t)$ indicates a higher order terms of Δt. Thus, the rate of a naive member to become a believer is given by $\beta B(t)$. Since the naive population size at time t is given by $U(t)$, the net rate, that is, the momental flux of naive members to become believers at time t is given by $\beta B(t)U(t)$ (see Figure 3.1). As β gets larger, the naive members are easier to believe the disinformation and become believers. In other words, the disinformation spreads more easily in the netizen community with larger β.

3.2 State transition by the counter information

Members at three states could contribute to the spread of counter information in the netizen community: unsophisticated rejoinders (R), sophisticated rejoinders (A), and reformed believers (X). We assume that the concentration of counter information is proportional to the population size of members who release it. Thus, by the expression $\kappa_1 R + \kappa_2 A + \kappa_3 X$ we denote the total concentration of counter information in the netizen community in our model. The positive parameters κ_1, κ_2, and κ_3 characterize the differences in their contributions to the release of counter information in the netizen community.

Naive members may accept the counter information and become rejoinders. We assume that the probability with which a naive member becomes a rejoinder is proportional to the concentration of counter information in the netizen community. Thus, the probability during time interval $[t, t+\Delta t]$ with sufficiently small Δt is now given as

$$\sigma\{\kappa_1 R(t) + \kappa_2 A(t) + \kappa_3 X(t)\}\Delta t + \mathrm{o}(\Delta t).$$

Hence the rate at which a naive member becomes a rejoinder at time t is given by $\sigma\{\kappa_1 R(t) + \kappa_2 A(t) + \kappa_3 X(t)\}$. The net rate of the transition of members from the naive state to the rejoinder one is given by $\sigma(\kappa_1 R + \kappa_2 A + \kappa_3 X)U$. We do not take account of the difference in the quality of information from different sources, and we assume that the "strength" or "persuade power" of information is the same for all the counter information independently of the source. As the coefficient σ gets larger, the naive members are easier to accept the counter information and become rejoinders.

The believers may relinquish their belief of disinformation when they accept the counter information. Similarly, the chance that a believer becomes a reformed member is assumed to be proportional to the concentration of counter information in the netizen community. During time interval $[t, t+\Delta t]$ with sufficiently small Δt, the chance is given by

$$\gamma\{\kappa_1 R(t) + \kappa_2 A(t) + \kappa_3 X(t)\}\Delta t + \mathrm{o}(\Delta t).$$

Thus, the per capita rate of such a transition at time t is $\gamma(\kappa_1 R(t) + \kappa_2 A(t) + \kappa_3 X(t))$. Then the momental net flux of members from the believer state to the reformed state is given by $\gamma(\kappa_1 R + \kappa_2 A + \kappa_3 X)B$.

3.3 Transition between rejoinder and rejecter states

Rejoinders can be classified into two: One is of unsophisticated members after accepting the counter information (R), and the other of sophisticated members (A). They release the counter information like volunteers. As we assumed in Section 2, the rejoinders may gradually lose interest in the information and stop releasing the counter information which mean the state transition to other states, that is, the rejecter state or the unconcerned state. We assume that leaving the rejoinder state is random independently of the situation in the netizen community.

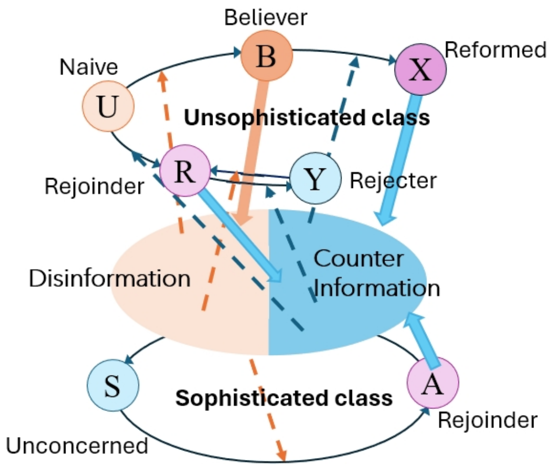

Figure 3.1: Scheme of state transitions by the influence of disinformation and counter information in our model (3.1).

Thus, the probability that a rejoinder lose interest of the information during time $[t, t + \Delta t]$ is now given by

$$\delta \Delta t + o(\Delta t).$$

The rate δ is a constant. Hence, the net transition rate of unsophisticated rejoinders to rejecters is given by δR, and that of sophisticated rejoinders to unconcerned members is given by δA. As δ gets larger, the rejoinders are easier to lose interest and stop releasing the counter information. We can regard $1/\delta$ as the expected duration that releases the counter information. In this work, we assume the same transition rate δ for both of sophisticated and unsophisticated rejoinders.

Influenced by the spreading disinformation, the rejecters or unconcerned members may restart releasing the counter information. For a rejecter or an unconcerned sophisticated member, the probability to become rejoinder during time period $[t, t + \Delta t]$ with sufficiently small Δt is given as

$$\alpha B(t) \Delta t + o(\Delta t).$$

The per capita transition rate from rejecter or unconcerned to rejoinder state is αB. The net rate from rejecters (Y) to rejoinders is αBY, and unconcerned sophisticated members (S) to rejoinders is αBS. We assume the same per capita transition rate α for both of rejecters and unconcerned sophisticated members to rejoinders state. As α gets larger, rejecters and unconcerned members are more voluntary to release the counter information for the other members in the netizen community.

3.4 System to govern the population dynamics

$$
\begin{aligned}
\frac{dU}{dt} &= -\beta BU - \sigma(\kappa_1 R + \kappa_2 A + \kappa_3 X)U;\\
\frac{dB}{dt} &= \beta BU - \gamma(\kappa_1 R + \kappa_2 A + \kappa_3 X)B;\\
\frac{dR}{dt} &= \sigma(\kappa_1 R + \kappa_2 A + \kappa_3 X)U - \delta R + \alpha BY;\\
\frac{dX}{dt} &= \gamma(\kappa_1 R + \kappa_2 A + \kappa_3 X)B;\\
\frac{dY}{dt} &= \delta R - \alpha BY;\\
\frac{dS}{dt} &= -\alpha BS + \delta A;\\
\frac{dA}{dt} &= \alpha BS - \delta A.
\end{aligned}
\tag{3.1}
$$

The proportion of sophisticated members in the netizen community is denoted by a constant $p \in (0, 1)$. The total population size is given by N. Thus, the sophisticated population size is pN, and the unsophisticated population size is $(1 - p)N$.

The initial state of the system is defined for the moment when the disinformation emerges in the netizen community. At the initial moment, there are some initial believers B_0 who carry a disinformation in the netizen community, and other members have no contact to the information. The initial condition of system (3.1) is given as

$$
(U, B, R, X, Y, S, A) = (U_0, B_0, 0, 0, 0, pN, 0), \tag{3.2}
$$

where $B_0 > 0$ is the population size of initial believers. $U_0 = (1-p)N - B_0$ is the population size of naive members, and they are unsophisticated members who are unaware of the disinformation at the initial. We assume that the number of initial believers is sufficiently small relative to the total population size of unsophisticated members, that is, $B_0/(1 - p)N \ll 1$.

3.5 Social damage

In this work, we define the social damage caused by the disinformation as how many members in the netizen community ever believed the disinformation. Since the transition from the naive state to the believer state is one-directional (see Figure 3.1), the social damage can be expressed by the sum of the population size of believer and reformed members at the final state, as we will show in the later section that the population sizes converge to equilibrium values. Therefore, in our modelling, the social damage is represented by $B^* + X^*$, where B^* and X^* are the equilibrium population sizes of believer and reformed members at the final state.

We also define the damage risk per unsophisticated member, that corresponds to the probability that an unsophisticated member is cheated by the disinformation and becomes a believer in the course of the population dynamics with the considered spreading disinformation. It is given by

$$
\frac{\text{Population size of members who once believed the disinformation}}{\text{Total population size of unsophisticated members}} = \frac{B^* + X^*}{(1 - p)N}.
$$

4 Netizen community with no voluntary sophisticated members

In this section, we consider a case where there is no voluntary sophisticated member in the netizen community. In this case, all sophisticated members are always unconcerned to the disinformation and never become rejoinders, which means $S = pN$ and $A = 0$. Under this assumption, any sophisticated member does not join the dynamics, and does not release any counter information

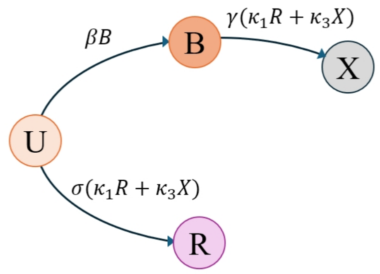

Figure 4.1: Scheme of state transition in the netizen community with no voluntary sophisticated member.

even with contacting the disinformation. Furthermore, we omit the state transition from the unsophisticated rejoinders to rejecters with $\delta = 0$ and $Y = 0$ (see Figure 4.1).

As for the initial condition, we set a positive initial value of unsophisticated rejoinders, that is, $R_0 > 0$. If $R_0 = 0$, we have $R(t) = 0$ for any $t > 0$, and then there will be no counter information in the netizen community. In such a case, the members in the netizen community are unable to have any counter information because the sophisticated members are all unconcerned. The initial condition with $R_0 > 0$ means that the counter information comes from an "external source" on which we do not mention anymore since there would be a variety of possibilities to have it.

4.1 With stiff believer

Here we consider the dynamics where the believers are stiff and will not lose interest in spreading the disinformation. We have $\gamma = 0$ and $X = 0$ with this assumption. As seen from Figure 4.1, the model is described by the following system:

$$\begin{aligned} \frac{dU}{dt} &= -\beta BU - \sigma\kappa_1 RU; \\ \frac{dB}{dt} &= \beta BU; \\ \frac{dR}{dt} &= \sigma\kappa_1 RU, \end{aligned} \tag{4.1}$$

where $U + B + R = (1-p)N$. The initial condition of the system is given as $(U, B, R) = (U_0, B_0, R_0)$. Then, we can get the following result on the convergence of the solution for (4.1) (Appendix A):

Theorem 4.1. *The solution of system (4.1) has the convergent nature as*

$$(U, B, R) \to (0, B^*, R^*) \quad \text{as } t \to \infty,$$

where R^* and X^* are uniquely determined by the root of equations

$$\begin{cases} (1-p)N = R^* + B_0\left(\dfrac{R^*}{R_0}\right)^{\frac{\beta}{\sigma\kappa_1}}; \\ B^* = (1-p)N - R^*. \end{cases} \tag{4.2}$$

Social damage As defined in Section 3.5, we find from Theorem 4.1 for the model (4.1) that the value of B^* represents the population size of those who once believed the disinformation, that is, the social damage. Given initial condition B_0, R_0 and proportion p, the final size (B^*, R^*) depends on the ratio $\rho = \beta/\sigma\kappa_1$. As the ratio ρ gets larger, the disinformation will cause a more serious social damage (see Figure 4.2).

A larger ρ can be caused by larger β. It means that if the naive members are easier to believe the disinformation, there will be more members being cheated by the disinformation, and the

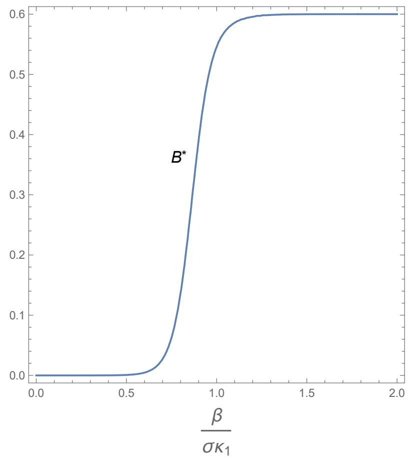

Figure 4.2: Numerical calculation of B^* for (4.2), with $N = 1.0$, $p = 0.4$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 1.0 \times 10^{-8}$.

social damage will be greater. A smaller σ or a smaller κ_1 will also make ρ larger. This can be interpreted as the role of counter information. If the counter information is not easy to be accepted by naive members (small σ), or if the rejoinders are less frequently to release counter information (small κ_1), the social damage of disinformation will be more serious.

4.2 With mild believer

In this section, we consider the dynamics where the believers are mild and they may lose interest in spreading the disinformation. The believers will become reformed members who no longer believe the disinformation (see Figure 4.1). The model is described by the following system:

$$\begin{aligned}
\frac{dU}{dt} &= -\beta BU - \sigma(\kappa_1 R + \kappa_3 X)U; \\
\frac{dB}{dt} &= \beta BU - \gamma(\kappa_1 R + \kappa_3 X)B; \\
\frac{dR}{dt} &= \sigma(\kappa_1 R + \kappa_3 X)U; \\
\frac{dX}{dt} &= \gamma(\kappa_1 R + \kappa_3 X)B,
\end{aligned} \tag{4.3}$$

where $U + B + R + X = (1 - p)N$. The initial condition of the system is given as $(U, B, R, X) = (U_0, B_0, R_0, 0)$. We can get the following result on the convergence of the solution for (4.3) (Appendix B):

Theorem 4.2. *The solution of system (4.3) has the convergent nature as*

$$(U, B, R, X) \to (0, 0, R^*, X^*) \quad \text{as} \quad t \to \infty,$$

where R^* and X^* are positive values, $R^* + X^* = (1 - p)N$.

In this case, we have $B \to 0$ which means there will be no believer and no circulating disinformation in the end. The value X^* is the population size of those who were once been cheated by the disinformation, which is the social damage.

WITH REFORMED MEMBERS RELEASING NO COUNTER INFORMATION

Here we assume that the reformed members just "silently" stop believing the disinformation without releasing counter information, which means $\kappa_3 = 0$. Thus, the counter information is

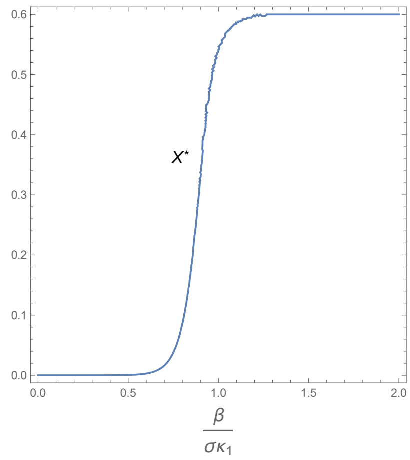

Figure 4.3: Numerical calculated of final size X^* for equation (4.4), with $N = 1.0$, $p = 0.4$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 1.0 \times 10^{-8}$.

only released by rejoinders (R) in the netizen community . We get the following result for the special case with $\gamma = \sigma$ (Appendix C).

Corollary 4.1. *Under the conditions that $\kappa_3 = 0$ and $\gamma = \sigma$, R^* and X^* at the final state of system (4.3) are uniquely determined by the root of equations:*

$$\begin{cases} (1-p)N = R_0 + \sum_{n=0}^{\infty} \frac{1}{\frac{n\beta}{\sigma\kappa_1}+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n \Big\{R^*\Big(\frac{R^*}{R_0}\Big)^{\frac{n\beta}{\sigma\kappa_1}} - R_0\Big\}; \\ X^* = (1-p)N - R^*. \end{cases} \tag{4.4}$$

Social damage Given B_0, R_0 and p, the root R^* of equation (4.4) monotonically decreases in terms of the ratio $\rho = \beta/\sigma\kappa_1$. It means a larger ρ will make R^* smaller, and make social damage $X^* = (1-p)N - R^*$ larger. A numerical calculated X^* value is shown in Figure 4.3. We can see that there exists a certain range of ρ that the slope of the curve is much more steep (ρ between 0.8 to 1.0). This implies that within this range, a small change of ratio ρ may result in a large difference on the social damage X^*.

WITH REFORMED MEMBERS RELEASING SOME COUNTER INFORMATION

Here we assume that the reformed members release counter information at a certain frequency ($\kappa_3 > 0$). We get the following result for the special case with $\gamma = \sigma$, $\beta = \sigma\kappa_1$ and $\kappa_1 = \kappa_3$ (Appendix D).

Corollary 4.2. *Under the condition that $\kappa_1 = \kappa_3$, $\gamma = \sigma$ and $\beta = \sigma\kappa_1$, The value X^* at the final state of system (4.3) is uniquely determined by the root of equation*

$$(1-p)N = X^* + R_0 + R_0\frac{U_0}{B_0}\ln\frac{\frac{(1-p)N}{R_0} + \frac{U_0}{B_0}}{1+\frac{U_0}{B_0}} \tag{4.5}$$

in $(0, (1-p)N - R_0)$.

4.3 Social damage

In this section, we compare the social damage in different netizen community which is determined by (4.2), (4.4) and (4.5) respectively (Appendices E and F).

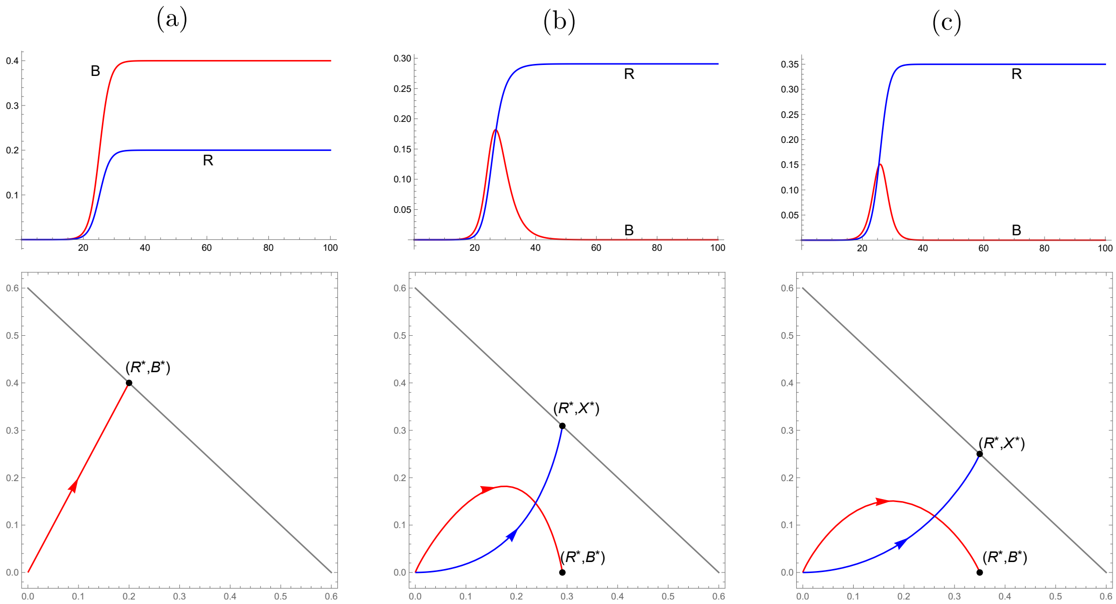

Figure 4.4: Numerical calculation for the systems: (a) (4.1); (b) $\kappa_3 = 0$ for (4.3) ; (c) $\kappa_3 = \kappa_1$ for (4.3), commonly with $N = 1.0$, $p = 0.4$, $\beta = 1.0$, $\sigma\kappa_1 = 1.0$, $\gamma\kappa_1 = 1.0$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 5.0 \times 10^{-8}$.

Corollary 4.3. *Given same parameters β, σ, κ_1, same initial state (U_0, B_0, R_0) and the condition $\gamma = \sigma$, when B_0 and R_0 are relatively small, the social damage B^* determined by (4.2) is greater than the social damage X^* from (4.4).*

This means that given the same conditions, the netizen community with the stiff believer (4.4) suffers more social damage compared to a netizen community with mild believers (4.2). The stiffness of the believer will make the influence of the disinformation more serious.

Corollary 4.4. *Given same parameters β, σ, κ_1, same initial state (U_0, B_0, R_0) and the conditions $\kappa_1 = \kappa_3$, $\gamma = \sigma$ and $\beta = \sigma\kappa_1$, X^* determined by (4.4) is greater than the X^* from (4.5).*

It means that if the reformed believers do not release any counter information (4.4), the social damage will be more serious compared to netizen community with reformed believers release some counter information (4.5). In Figure 4.4, numerical calculation matches this result: a netizen community with stiff believers (Figure 4.4(a)) receives more social damage by the disinformation, compared to the netizen communities with mild believers; the netizen community with reformed members releasing no counter information (Figure 4.4(b)) receives less social damage; the netizen community with reformed members releasing some counter information (Figure 4.4(c)) receives the least social damage among the three netizen communities.

5 Netizen community with naive members indifferent to counter information

In this section, we assume that the sophisticated members may release counter information in this netizen community. However, the naive members are not affected by the counter information and never become rejoinders. This means that the naive members never gain the immunity to the disinformation from the counter information. This assumption corresponds to $\sigma = 0$ in the

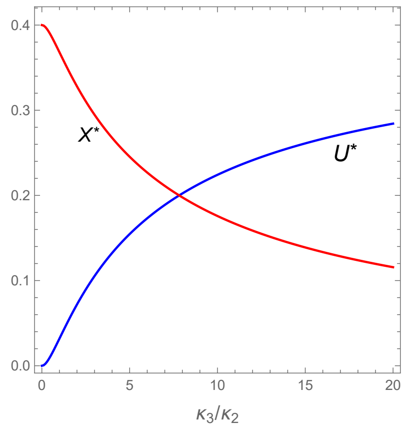

Figure 5.1: Numerical calculation of final sizes U^* and X^* for the system (5.1) with $N = 1.0$, $p = 0.6$, $\beta = 1.0$, $\gamma\kappa_2 = 0.8$, $\alpha = 5.0$, $\delta = 6.0$, $B_0 = 1.0 \times 10^{-7}$.

system, and we have $R \equiv 0$ and $Y \equiv 0$ for any $t \geq 0$. The model of the system is given as

$$\begin{aligned}\frac{dU}{dt} &= -\beta BU;\\ \frac{dB}{dt} &= \beta BU - \gamma(\kappa_2 A + \kappa_3 X)B;\\ \frac{dX}{dt} &= \gamma(\kappa_2 A + \kappa_3 X)B;\\ \frac{dS}{dt} &= -\alpha BS + \delta A;\\ \frac{dA}{dt} &= \alpha BS - \delta A,\end{aligned} \tag{5.1}$$

with initial condition

$$(U, B, X, S, A) = (U_0, B_0, 0, pN, 0). \tag{5.2}$$

Theorem 5.1. *The solution of system (5.1) has the convergent nature as*

$$(U, B, X, S, A) \to (U^*, 0, X^*, pN, 0) \quad \text{as } t \to \infty,$$

where the value of U^* follows

$$\begin{cases} U^* = 0, & \text{if } \kappa_3 = 0;\\ U^* > 0, & \text{if } \kappa_3 > 0.\end{cases}$$

In Appendix G, we provide proof of the theorem. In the final state, X^* is the size of reformed members, and U^* is the size of the naive members. A numerical calculation (see Figure 5.1) shows that a greater κ_3 makes the social damage X^* smaller.

When $\kappa_3 = 0$, the reformed members do not release counter information, it results in $U^* = 0$. In this case, it is not possible to have anyone remained in the naive state U. The size of reformed members in the final state is $X^* = (1 - p)N$, which means all the unsophisticated members will be at the reformed class in the end. Since reformed members come from believers, every member who is unsophisticated will experience a certain period of being a believer.

When $\kappa_3 > 0$, reformed members release the counter information at some frequency. Then there will be some unsophisticated members who stay in the naive state U without believing the disinformation. In this case, the reformed members release counter information and their efforts will save some unsophisticated members from the disinformation.

This indicates the important role of unsophisticated members who release counter information. When counter information is only released by sophisticated members ($\kappa_3 = 0$), their voluntary behaviour is not enough to prevent naive members from believing the disinformation.

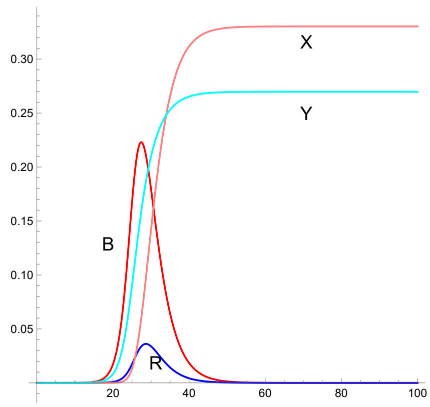

Figure 6.1: Numerical calculation of temporal variation for the system (3.1), with $N = 1.0$, $p = 0.4$, $\beta = 1.0$, $\kappa_1 = \kappa_2 = \kappa_3 = \kappa$, $\gamma\kappa = 0.8$, $\sigma\kappa = 1.5$, $\alpha = 5.0$, $\delta = 6.0$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 0.0$.

Everyone who is unsophisticated will experience being cheated by the disinformation. In other words, if we want to protect unsophisticated members from disinformation, it is necessary to have some counter information accepted and spread by the unsophisticated members themselves.

6 Contribution of sophisticated members to moderate the social damage

In this section, we will analyse how the proportion of sophisticated members in the netizen community (p) contributes in the netizen community to suppress the disinformation in the generic model (3.1). We have the following property of the system (Appendix H):

Theorem 6.1. *The system (3.1) with $p > 0$ and the initial condition (3.2) has the convergent nature as*

$$(U, B, R, X, Y, S, A) \to (0, 0, 0, X^*, Y^*, pN, 0) \quad \text{as} \quad t \to \infty.$$

For the final state, U, B, R and A converges to 0. There will be no believer in the end and there will be no disinformation spreading in the community. There will also be no naive members remained. All the unsophisticated embers will be either reformed members or rejecters in the final state. It implies that each unsophisticated member has once accepted either the disinformation or the counter information. A numerical calculation is shown in Figure 6.1.

Social damage As shown in Figure 6.2(a), a larger p makes X^* smaller. The rejecters in final state Y^* can be regarded as how many unsophisticated members have been saved by the counter information. It has a peak which means that with a certain p value, the netizen community will have the largest amount of unsophisticated members saved by counter information. However, this does not means that the p value at the peak is "optimal". A larger p also makes the total sophisticated population size $(1 - p)N$ smaller, that is the why a large p may make Y^* small.

Dependence of the risk In this system, the risk of unsophisticated members can be denoted by

$$\frac{X^*}{(1 - p)N},$$

which can be regarded as the probability for an unsophisticated member to be cheated by disinformation. As shown in Figure 6.2(b), the risk is monotonically decreasing in terms of p. As p value gets larger, the risk of an unsophisticated member will decrease. However, an infimum of the risk can be found. Even if p is sufficiently large and close to 1, which means that

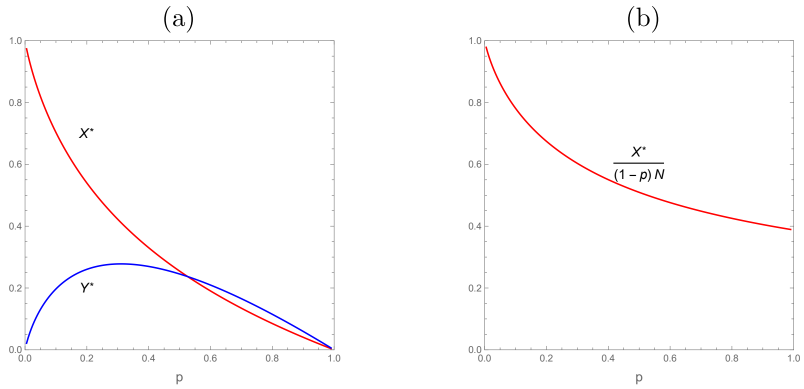

Figure 6.2: Parameter dependence of final size and risk in terms of p for the system (3.1), numerical calculation with $N = 1.0$, $\beta = 1.0$, $\kappa_1 = \kappa_2 = \kappa_3 = \kappa$, $\gamma\kappa = 0.8$, $\sigma\kappa = 1.5$, $\alpha = 5.0$, $\delta = 6.0$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 0.0$.

almost everyone in the netizen community is sophisticated, the non-zero risk still exists for the unsophisticated members.

In Appendix I, we apply fast process assumption for the limit of risk as $p \to 1$. Under the fast process assumption, we find the analytical expression for the case that $\kappa_3 = 0$:

$$\frac{X^*}{(1-p)N} \to \frac{1}{1 + \kappa_2 \alpha \sigma N / \beta \delta}, \quad \text{as} \quad p \to 1.$$

In this fast process assumption, we assume that the state transition between A-S (sophisticated rejoinder and unconcerned unsophisticated members) and R-Y (unsophisticated rejoinder and rejecter) are much faster than transition between other states. In this special consideration, the value $1/(1 + \kappa_2 \alpha \sigma N / \beta \delta)$ can be regarded as the limit of the risk for the case that reformed members release no counter information.

This result is under a rather special condition, but it can be regarded as an implication for the existence of the infimum of risk. In a more general case, from Figure 6.2(b), we know that the risk cannot be suppressed to 0. Although sophisticated members can reduce the risk of unsophisticated members, there exists a lower boundary of the risk. Even if the proportion of sophisticated members is sufficiently large ($p \to 1$), the risk cannot be reduced towards 0, unsophisticated members are not safe.

Social damage in netizen community with no voluntary sophisticated members In Section 4, we conducted analysis on the netizen community with no voluntary sophisticated members by setting a fixed proportion of sophisticated members (p). Since a larger p means less unsophisticated members $(1-p)N$, it can be inferred that a larger p also makes social damage X^* smaller. As for the risk of unsophisticated members, we will show a numerical calculated example.

From Figure 6.3, the p value has some influence on the risk of unsophisticated members. This influence is not significant unless p is sufficiently large. It is understandable, since the contribution of p value is to change the population size of unsophisticated members. Hence, the p may not have a strong relation to the risk as the sophisticated members do not involved in the dynamics. It seems that when p value large, the p dependence will be stronger. This is because in the calculation, we manually set a positive initial value of B_0 and R_0 that does not change by p. If p becomes large, then the unsophisticated population size $(1-p)N$ will be small. It has the similar result as "to amplify the initial value of B_0 and R_0". For the same reason, it is also possible that the p-dependence of the risk is not decreasing, but increasing. As shown in Figure 6.3(b), it is possible to have some situation that a larger p brings higher risk to the unsophisticated members.

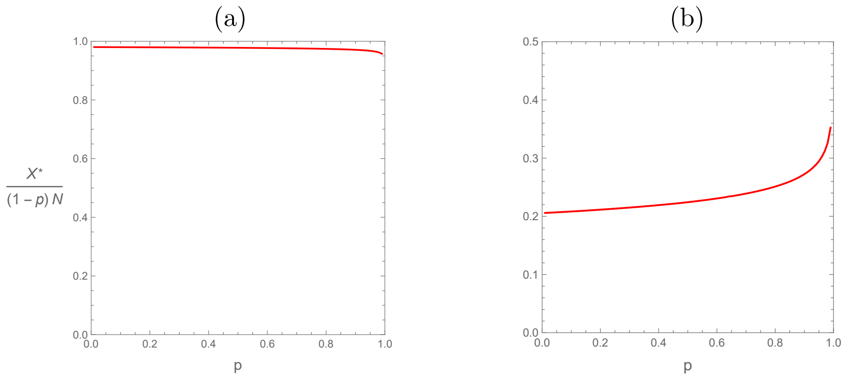

Figure 6.3: An example of numerical calculated risk in terms of p for the system (4.3), with $N = 1.0$, $\beta = 1.0$, $\kappa_1 = \kappa_3 = \kappa$, $\gamma\kappa = 0.8$, $\alpha = 5.0$, $\delta = 6.0$, $B_0 = 1.0 \times 10^{-7}$, $R_0 = 1.0 \times 10^{-8}$, (a) $\sigma\kappa = 0.8$, (b) $\sigma\kappa = 1.2$.

7 Concluding remarks

In this work, we have analysed the social damage by the disinformation of different communities. For the netizen community without voluntary sophisticated members, there needs to have some "external" source of counter information, such as some initial spreader of counter information. In these cases, the social damage depends on the members' release and acceptance of disinformation and counter information. Generally speaking, if members in the netizen community have a stronger tendency to release or accept the disinformation, the social damage will be greater. If the netizens are more willing to release or accept the counter information, the social damage will be released. We also compare the analytical results of social damage between different cases. The netizen community with stiff believers will suffer greater social damage. For the case with naive members indifferent to the counter information, it is important to have some unsophisticated members who release counter information (reformed members).

Then we turn to the communities with sophisticated members releasing counter information. These communities do not necessarily need an "external" source of counter information. The sophisticated members can identify disinformation and generate counter information themselves. From the numerical calculation, we find that a community with a larger proportion of sophisticated members can better defend the disinformation. As for the unsophisticated members, their risk also becomes less in a community with more sophisticated members. However, the risk cannot be sufficiently small, there exists an infimum of the risk. This type of condition can be considered when most members can distinguish the disinformation. For example, some scams do not sound real, most members will not believe them. But the targets of scams are those members who cannot distinguish them. Even though the proportion of these unsophisticated members is relatively small in the community, they still have a positive chance of being cheated by the disinformation.

Bibliography

[1] Cuan-Baltazar, J.Y., Muñoz-Perez, M.J., Robledo-Vega, C., Pérez-Zepeda, M.F., and Soto-Vega, E. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health Surveill*, 6(2): e18444, 2020. https://doi.org/10.2196/18444https://doi.org/10.2196/18444.

[2] Zhang, Y., Xu ,J., Nekovee, M., and Li, Z. The impact of official rumor-refutation information on the dynamics of rumor spread. *Physica A*, 607: 128096, 2022. https://doi.org/10.1016/j.physa.2022.128096https://doi.org/10.1016/j.physa.2022.128096.

[3] Srinivasan, S. and Babu, L.D.D. A social immunity based approach to suppress rumors in online social networks. *International Journal of Machine Learning and Cybernetics*, 12: 1281-1296, 2021. https://doi.org/10.1007/s13042-020-01233-0https://doi.org/10.1007/s13042-020-01233-0.

[4] Dansu, E.J. and Seno, H. A rejoinder model for the population dynamics of the spread of two interacting pieces of information. *Afrika Matematika*, 35(2), 2024. https://doi.org/10.1007/s13370-023-01134-9https://doi.org/10.1007/s13370-023-01134-9.

Appendix A Proof of Theorem 4.1

Convergence of final state

From the model (4.1), we can see that

$$\frac{dU}{dt} + \frac{dB}{dt} + \frac{dR}{dt} = 0.$$

It means the sum $U+B+R$ is constant for any $t > 0$, and is equal to the initial sum $U_0+B_0+R_0 = (1-p)N$.
We can factorize dU/dt by U by following equation:

$$\frac{dU}{dt} = \{-\beta B - \sigma\kappa_1 R\}U,$$

where the term $-\beta B - \sigma\kappa_1 R$ is linear of B and R. Therefore, if the initial condition follows $U_0 > 0$, $B_0 > 0$, $R_0 > 0$, we have

$$U > 0, B > 0, R > 0 \quad \text{for any } t > 0.$$

Since $U + B + R(1-p)N$, we have $B < (1-p)N$, which means B is bounded above. From the equation, $dB/dt > 0$. Therefore, B is monotonically increasing and bounded above. Then,

$$B \to B^* \quad \text{as} \quad t \to \infty.$$

For the same argument,

$$R \to R^* \quad \text{as} \quad t \to \infty.$$

As for U, we have

$$\frac{dB}{dt} = \beta BU.$$

The final size satisfies $dB/dt \to 0$ as $t \to \infty$. We already know $B \to B^*$, therefore,

$$U \to 0 \quad \text{as} \quad t \to \infty.$$

Hence, the final state of is $(U, B, R) \to (0, B^*, R^*)$.

Derivation of equation (4.2)

Conserved quantity For convenience of expression, we denote $\rho := \frac{\beta}{\sigma\kappa_1}$. From equation (4.3), we have the differential equation

$$\frac{dB}{dR} = \frac{\rho B}{R}.$$

Then we can derive the solution

$$\frac{B}{B_0} = \left(\frac{R}{R_0}\right)^{\rho}. \tag{A1}$$

This equation holds for any $t \geq 0$.

Equation of final state Since $B \to B^*, R \to R^*$, and $B^* + R^* = (1-p)N$, we may substitute them in the equation (A1), then we can get the equation determines the final size:

$$(1-p)N = R^* + B_0\Big(\frac{R^*}{R_0}\Big)^{\rho}. \tag{A2}$$

Existence and uniqueness of root in equation (4.2)

Firstly, we may define a function

$$f_1(z) = z + B_0\Big(\frac{z}{R_0}\Big)^{\rho} - (1-p)N,$$

with the domain of $f_1(z)$ to be $z \in [R_0, (1-p)N]$. The root R^* of Equation (4.2) satisfies

$$f_1(R^*) = 0.$$

We want to show the root R^* is existed and unique.

Let us prove the root R^* exists in $(R_0, (1-p)N)$. This can be done by considering value of $f(R_0)$ and $f((1-p)N)$:

$$f_1(R_0) = R_0 + B_0\Big(\frac{R_0}{R_0}\Big)^{\rho} - (1-p)N = R_0 + B_0 - (1-p)N < 0.$$

From the meaning, $R_0 + B_0$ is the initial believer and rejecter. Their sum must be less than the total population $(1-p)N$.

$$f_1((1-p)N) = (1-p)N + B_0\Big(\frac{(1-p)N}{R_0}\Big)^{\rho} - (1-p)N = \Big(\frac{(1-p)N}{R_0}\Big)^{\rho} > 0.$$

We get $f(R_0) < 0$ and $f((1-p)N) > 0$. Since the function $f_1(z)$ is continuous in $[R_0, (1-p)N]$, there must exist $R^* \in (R_0, (1-p)N)$ such that $f_1(R^*) = 0$.
Then, we will prove the uniqueness of the root. Since

$$\frac{df_1(z)}{dz} = 1 + \frac{B_0}{R_0^{\rho}}\rho z^{\rho-1} > 0,$$

$f_1(z)$ is monotonically increasing in terms of z. Therefore, the root of $f_1(z) = 0$ is unique.

Appendix B Proof of Theorem 4.2

By using similar argument as Appendix A, we have the positivity for each state in finite time:

$$U > 0, B > 0, R > 0, X > 0 \quad \text{for any } t > 0.$$

We can find (omitted here) from the equations that both R and X are monotonically increasing and bounded above. Hence, we have the positivity of the value R^* and X^*:

$$R \to R^* > 0,\ X \to X^* > 0 \quad \text{as} \quad t \to \infty. \tag{B1}$$

As for U, we will show the convergence using proof by contradiction. Since we have

$$\frac{dU}{dt} = -\beta BU - \sigma(\kappa_1 R + \kappa_3 X)U \le -\sigma\kappa_1 RU,$$

if $U \to U^* > 0$, the term $\sigma\kappa_1 R^* U^* < 0$. It will cause the negativity for the final state:

$$\lim_{t\to\infty} \frac{dU}{dt} \le -\sigma\kappa_1 R^* U^* < 0.$$

This violates the definition of the final state that

$$\lim_{t\to\infty} \frac{dU}{dt} = 0.$$

Hence, it is impossible to have $U \to U^* > 0$, the value of U must converge to 0:

$$U \to 0 \quad \text{as} \quad t \to \infty. \tag{B2}$$

As for the final state of B, we have

$$\frac{dB}{dt} = \{\beta U - \gamma(\kappa_1 R + \kappa_3 X)\}B.$$

Since $U \to 0$ and $R \to R^*$, there exists a finite time T such that

$$\beta U - \gamma\kappa_1 R < -\frac{1}{2}\gamma\kappa_1 R, \quad \text{for any } t > T.$$

Then, we have

$$\frac{dB}{dt} = \{\beta U - \gamma(\kappa_1 R + \kappa_3 X)\}B < -\frac{1}{2}\gamma\kappa_1 RB, \quad \text{for any } t > T.$$

If $B \to B^* > 0$, we will have

$$\lim_{t\to\infty} \frac{dB}{dt} < -\frac{1}{2}\gamma\kappa_1 R^* B^* < 0.$$

It violates the definition of the final state. Hence, the value of B must converge towards 0:

$$B \to 0 \quad \text{as} \quad t \to \infty. \tag{B3}$$

From (B1), (B2) and (B3), we have shown The final state of system (4.3) is

$$(U, B, R, X) \to (0, 0, R^*, X^*), \quad \text{as } t \to \infty.$$

Appendix C Proof of Corollary 4.1

Derivation of equation (4.4)

Conserved quantity with $\kappa_3 = 0$, $\gamma = \sigma$ For calculation convenience, let us define $\varphi := B/U$ which is the ratio of believer over naive members. We can get the differential equation

$$\frac{dX}{dR} = \frac{\gamma}{\sigma}\frac{B}{U} = \varphi. \tag{C1}$$

As for the variable φ, we could get

$$\frac{d\varphi}{dR} = \varphi\frac{d\ln\varphi}{dR} = \varphi\frac{d(\ln B - \ln U)}{dR}$$

From the system (4.3) and the further assumption $\kappa_3 = 0$, $\gamma = \sigma$, we have:

$$\begin{aligned}
\frac{d\ln B}{dt} &= \beta U - \gamma(\kappa_1 R + \kappa_3 X) = \beta U - \sigma\kappa_1 R; \\
\frac{d\ln U}{dt} &= -\beta B - \sigma(\kappa_1 R + \kappa_3 X) = -\beta B - \sigma\kappa_1 R.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}\frac{d\varphi}{dR} &= \varphi\frac{d(\ln B - \ln U)}{dR}\\ &= \varphi\frac{\beta U - \sigma\kappa_1 R - (-\beta B - \sigma\kappa_1 R)}{\sigma\kappa_1 RU}\\ &= \varphi\frac{\beta(U+B)}{\sigma\kappa_1 RU}\\ &= \varphi\frac{\beta(U+\varphi U)}{\sigma\kappa_1 RU}\\ &= \frac{\beta\varphi(1+\varphi)}{\sigma\kappa_1 R}\end{aligned}$$

For convenience of calculation, we may define

$$\rho := \frac{\beta}{\sigma\kappa_1}.$$

We can solve the ODE by following steps:

$$\begin{aligned}\frac{d\varphi}{dR} &= \rho\frac{\varphi(1+\varphi)}{R}\\ \frac{d\varphi}{\varphi(1+\varphi)} &= \rho\frac{dR}{R}\\ \Big(\frac{d\varphi}{\varphi} - \frac{d\varphi}{1+\varphi}\Big) &= \rho\frac{dR}{R}\\ d\{\ln\varphi - \ln(1+\varphi)\} &= \rho\, d\ln R\\ d\Big(\ln\frac{\varphi}{1+\varphi}\Big) &= \rho\, d\ln R\\ \ln\frac{\varphi}{1+\varphi} - \ln\frac{\varphi_0}{1+\varphi_0} &= \rho(\ln R - \ln R_0)\\ \frac{\varphi}{1+\varphi} &= \frac{\varphi_0}{1+\varphi_0}\Big(\frac{R}{R_0}\Big)^{\rho}\end{aligned}$$

Consider the initial condition $\varphi_0 = B_0/U_0$ when $t = 0$, we have

$$\frac{\varphi}{1+\varphi} = \frac{B_0/U_0}{1+B_0/U_0}\Big(\frac{R}{R_0}\Big)^{\rho} = \frac{B_0}{U_0+B_0}\Big(\frac{R}{R_0}\Big)^{\rho} = c_1 R^{\rho}, \tag{C2}$$

where

$$c_1 = \frac{B_0}{(U_0+B_0)R_0^{\rho}}; \qquad \rho = \frac{\beta}{\sigma\kappa_1}.$$

From (C1), $\frac{dX}{dR} = \varphi$, X can be expressed by

$$X = \int_{R_0}^{R}\varphi dR.$$

From (C2), the equality $\varphi = c_1 R^{\rho}(1+\varphi)$ holds. We can use them to obtain an equation that includes X and R. We notice that

$$\begin{aligned}\int_{R_0}^{R}\varphi dR &= \int_{R_0}^{R} c_1 R^{\rho}(1+\varphi)dR\\ &= c_1\int_{R_0}^{R} R^{\rho}dR + c_1\int_{R_0}^{R} R^{\rho}\varphi dR\\ &= c_1\frac{1}{\rho+1}\left(r^{\rho+1} - R_0^{\rho+1}\right) + c_1\int_{R_0}^{R} R^{\rho}\varphi dR.\end{aligned}$$

We notice that there exists a recurrence relation for any $k \geq 1$:

$$
\begin{aligned}
\int_{R_0}^{R} \varphi R^{(k-1)\rho} dR &= \int_{R_0}^{R} c_1 R^{k\rho}(1+\varphi) dR \\
&= c_1 \int_{R_0}^{R} R^{k\rho} dR + c_1 \int_{R_0}^{R} R^{k\rho} \varphi dR \\
&= c_1 \frac{1}{\rho+1} \left(r^{\rho+1} - R_0^{\rho+1} \right) + c_1 \int_{R_0}^{R} R^{k\rho} \varphi dR
\end{aligned}
$$

Hence, we have

$$
X = \int_{R_0}^{R} \varphi dR = \sum_{n=1}^{M} \frac{1}{n\rho+1} c_1^n \left(r^{n\rho+1} - R_0^{n\rho+1} \right) + c_1^M \int_{R_0}^{R} R^{M\rho} \varphi dR.
$$

We will examine if the residual term $c_1^M \int_{r_0}^{r} r^{M\rho} \varphi dr$ converge to 0 as $M \to \infty$. Since $dR/dt > 0$, $\varphi > 0$ and

$$
\frac{d\varphi}{dR} = \frac{\beta \varphi (1+\varphi)}{\sigma \kappa_1 R},
$$

it is clear that $d\varphi/dt > 0$. From (C2), the relation

$$
\frac{\varphi}{1+\varphi} = c_1 R^{\rho}
$$

holds. Since $R \to R^*$, Hence there exists a value φ^* which satisfies

$$
\frac{\varphi^*}{1+\varphi^*} = c_1 R^{*\rho} \tag{C3}
$$

and is finite. Since $d\varphi/dt > 0$, the value φ^* can be regarded as an upper bound of φ.

As for the residual term, we have

$$
\begin{aligned}
c_1^M \int_{R_0}^{R} R^{M\rho} \varphi dR &= \int_{R_0}^{R} (c_1 R^{\rho})^M \varphi dR \\
&= \int_{R_0}^{R} \left(\frac{\varphi}{1+\varphi} \right)^M \varphi dR \\
&< \int_{R_0}^{R^*} \left(\frac{\varphi^*}{1+\varphi^*} \right)^M \varphi^* dR \\
&= \left(\frac{\varphi^*}{1+\varphi^*} \right)^M \varphi^* (R^* - R_0).
\end{aligned}
$$

Since both φ^* and $(R^* - R_0)$ are finite, we have

$$
\lim_{M \to \infty} \left(\frac{\varphi^*}{1+\varphi^*} \right)^M \varphi^* (R^* - R_0) = 0.
$$

The residual term also goes to 0 as $M \to \infty$.

$$
\lim_{M \to \infty} c_1^M \int_{R_0}^{R} R^{M\rho} \varphi dR \leq \lim_{M \to \infty} \left(\frac{\varphi^*}{1+\varphi^*} \right)^M \varphi^* (R^* - R_0) = 0.
$$

Therefore, we can use an infinite series to express X and R:

$$
X = \sum_{n=1}^{\infty} \frac{1}{n\rho+1} c_1^n \left(R^{n\rho+1} - R_0^{n\rho+1} \right). \tag{C4}
$$

Value of the final state The value of the final state (R^*, X^*) satisfies the equation (C4), and their sum is the total unsophisticated population $R^* + X^* = (1-p)N$. We can substitute R with R^*, X with $(1-p)N - R^*$ to get the following equation about the final state that only includes R^*:

$$(1-p)N - R^* = \sum_{n=1}^{\infty} \frac{1}{n\rho+1} c_1^n \Big({R^*}^{n\rho+1} - R_0^{n\rho+1} \Big).$$

Note that in the infinite series, if we let $n = 0$, it will be $(R^* - R_0)$, we can get the equation

$$(1-p)N = R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1} c_1^n \Big({R^*}^{n\rho+1} - R_0^{n\rho+1} \Big).$$

By substituting the parameters $c_1 = \frac{B_0}{(U_0+B_0)R_0^\rho}$ and $\rho = \frac{\beta}{\sigma\kappa_1}$, we have:

$$\begin{aligned}
(1-p)N &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1} \Big\{ \frac{B_0}{(U_0+B_0)R_0^\rho} \Big\}^n \Big({R^*}^{n\rho+1} - R_0^{n\rho+1} \Big) \\
&= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1} \Big(\frac{B_0}{U_0+B_0} \Big)^n \Big\{ R^* \Big(\frac{R^*}{R_0} \Big)^{n\rho} - R_0 \Big\} \\
&= R_0 + \sum_{n=0}^{\infty} \frac{1}{\frac{n\beta}{\sigma\kappa_1}+1} \Big(\frac{B_0}{U_0+B_0} \Big)^n \Big\{ R^* \Big(\frac{R^*}{R_0} \Big)^{\frac{n\beta}{\sigma\kappa_1}} - R_0 \Big\},
\end{aligned}$$

which is shown in equation (4.4).

Existence and uniqueness of the root for equation (4.4)

Possible range of R^* We may use the similar method to the proof in Appendix A. Firstly, we need to figure out the possible range of R^*. From (C4), we have

$$c_1 {R^*}^\rho < 1.$$

This gives an upper bound of R^*:

$$R^* < \Big(\frac{1}{c_1} \Big)^{1/\rho} = R_0 \Big(\frac{U_0+B_0}{B_0} \Big)^{1/\rho}.$$

From the positiveness of X^* and $X^* + R^* = (1-p)N$, we have

$$R^* < (1-p)N.$$

Both the conditions need to be satisfied. We can find the condition of initial value B_0, R_0 that makes $(1/c_1)^{1/\rho} < (1-p)N$:

$$\Big(\frac{1}{c_1} \Big)^{1/\rho} = R_0 \Big(\frac{U_0+B_0}{B_0} \Big)^{1/\rho} < (1-p)N, \quad \text{if } B_0 > \frac{r_0^\rho \{(1-p)N - R_0\}}{\{(1-p)N\}^\rho}.$$

The calculation is shown in the following steps:

$$\begin{aligned}
R_0 \Big(\frac{U_0+B_0}{B_0} \Big)^{1/\rho} &< (1-p)N \\
R_0 \Big\{ \frac{(1-p)N - R_0)}{B_0} \Big\}^{1/\rho} &< (1-p)N \\
R_0 \{(1-p)N - R_0\}^{1/\rho} &< (1-p)N B_0^{1/\rho} \\
B_0^{1/\rho} &> \frac{R_0 \{(1-p)N - R_0\}^{1/\rho}}{(1-p)N} \\
B_0 &> \frac{r_0^\rho \{(1-p)N - R_0\}}{\{(1-p)N\}^\rho}.
\end{aligned}$$

Similarly, we can also get the condition for $(1/c_1)^{1/\rho} > (1-p)N$. Then, we can write down the possible range of R^*.

$$\begin{cases} R^* \in [R_0, (1/c_1)^{1/\rho}] & \text{if } B_0 > \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}; \\ R^* \in [R_0, (1-p)N] & \text{if } B_0 \leq \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}. \end{cases}$$

where

$$c_1 = \frac{B_0}{(U_0+B_0)R_0^\rho}; \qquad \rho = \frac{\beta}{\sigma\kappa_1}.$$

Existence of root We can define a function

$$f_2(z) = R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{z\Big(\frac{z}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N,$$

with the domain of $f_2(z)$

$$\begin{cases} z \in [R_0, (1/c_1)^{1/\rho}] & \text{if } B_0 > \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}; \\ z \in [R_0, (1-p)N] & \text{if } B_0 \leq \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}. \end{cases}$$

We want to find the positivity of the function $f_2(z)$ in the lower bound and upper bound of z. For $z = R_0$,

$$\begin{aligned} f_2(R_0) &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{R_0\Big(\frac{R_0}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N \\ &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{R_0 - R_0\Big\} - (1-p)N \\ &= R_0 - (1-p)N \\ &< 0. \end{aligned}$$

As for upper bound of z, firstly let us consider that case $z* \in [R_0, (1/c_1)^{1/\rho}]$, when the initial condition satisfies $B_0 > \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}$.

$$\begin{aligned} f_2((1/c_1)^{1/\rho}) &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{\Big(\frac{1}{c_1}\Big)^{1/\rho}\Big(\frac{(1/c_1)^{1/\rho}}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N \\ &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{\Big(\frac{1}{c_1}\Big)^{1/\rho}\Big(\frac{1}{c_1}\Big)^n\Big(\frac{1}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N \\ &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{R_0\Big(\frac{U_0+B_0}{B_0}\Big)^{1/\rho}\Big(\frac{U_0+B_0}{B_0}\Big)^n - R_0\Big\} - (1-p)N \\ &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n\rho+1}R_0\Big(\frac{U_0+B_0}{B_0}\Big)^{1/\rho} - \sum_{n=0}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n R_0 - (1-p)N \end{aligned}$$

The value diverges to $+\infty$. This is because the sum of infinite series

$$\sum_{n=0}^{\infty} \frac{1}{n\rho+1}R_0\Big(\frac{U_0+B_0}{B_0}\Big)^{1/\rho} = R_0\Big(\frac{U_0+B_0}{B_0}\Big)^{1/\rho}\sum_{n=0}^{\infty} \frac{1}{n\rho+1} = +\infty$$

is infinite, since

$$\sum_{n=0}^{\infty} \frac{1}{n\rho+1} = 1 + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} \geq 1 + \sum_{n=1}^{\infty} \frac{1}{n\rho+n} = 1 + \frac{1}{1+\rho}\sum_{n=1}^{\infty} \frac{1}{n} = +\infty.$$

Meanwhile, the sum of another infinite series converge to finite value.

$$\sum_{n=0}^{\infty}\frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n R_0 < \sum_{n=0}^{\infty}\Big(\frac{B_0}{U_0+B_0}\Big)^n R_0 = \frac{1}{1-\frac{B_0}{U_0+B_0}}R_0 = \frac{U_0+B_0}{U_0}R_0$$

There is no singular point of function $f_2(z)$ in range $[R_0, (1/c_1)^{1/\rho}]$, it is continuous in $(R_0, (1/c_1)^{1/\rho})$. We already have $f_2(R_0) < 0$ and $f_2((1/c_1)^{1/\rho}) > 0$, therefore, it has a root $R^* \in (R_0, (1/c_1)^{1/\rho})$ that makes $f_2(R^*) = 0$.

In another case with $B_0 \le \frac{r_0^\rho\{(1-p)N-R_0\}}{\{(1-p)N\}^\rho}$, we may consider the value of $f_2(z)$ at the upper bound of z as $(1-p)N$. If we consider the term that $n = 0$ of the infinite series in $f_2(z)$, we can get:

$$\frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{z\Big(\frac{z}{R_0}\Big)^{n\rho} - R_0\Big\} = z - R_0 \quad \text{when } n = 0.$$

Then, we can consider the positiveness of $f_2((1-p)N)$.

$$\begin{aligned} f_2((1-p)N) &= R_0 + \sum_{n=0}^{\infty}\frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{(1-p)N\Big(\frac{(1-p)N}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N \\ &= R_0 + \{(1-p)N - R_0\} + \sum_{n=1}^{\infty}\frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{(1-p)N\Big(\frac{(1-p)N}{R_0}\Big)^{n\rho} - R_0\Big\} - (1-p)N \\ &= \sum_{n=1}^{\infty}\frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big\{(1-p)N\Big(\frac{(1-p)N}{R_0}\Big)^{n\rho} - R_0\Big\} \\ &> 0. \end{aligned}$$

We also showed that $f_2(R_0) < 0$ and $f_2((1-p)N) > 0$, therefore, it has a root $R^* \in (R_0, (1-p)N)$ that makes $f_2(R^*) = 0$.

Uniqueness of root The uniqueness of root of $f_2(z)$ can be proved since

$$\frac{\mathrm{d}f_2(z)}{\mathrm{d}z} = \sum_{n=0}^{\infty}\Big(\frac{B_0}{U_0+B_0}\Big)^n\Big(\frac{z}{R_0}\Big)^{n\rho} > 0.$$

This means that $f_2(z)$ monotonically increases in terms of z in its domain. It is impossible to have more than one root.

Appendix D Proof of Corollary 4.2

Derivation of equation (4.5)

Conserved quantity with $\kappa_1 = \kappa_3$, $\gamma = \sigma$ and $\beta = \sigma\kappa_1$. We denote $\varphi = B/U$ and the system (4.3) has

$$\frac{dX}{dR} = \frac{\gamma}{\sigma}\frac{B}{U} = \varphi.$$

By using similar method to Appendix C, we can get:

$$\frac{d\varphi}{dt} = \varphi\frac{d\ln\varphi}{dt} = \varphi\frac{d(\ln B - \ln U)}{dt} = \varphi\beta(U+B).$$

$$\frac{d\varphi}{d(R+X)} = \frac{\varphi\beta(U+B)}{\sigma\kappa_1(R+X)(U+B)} = \frac{\beta\varphi}{\sigma\kappa_1(R+X)} = \rho\frac{\varphi}{R+X},$$

where $\rho = \beta/\sigma\kappa_1$. Then, we can get a conserved quantity which includes φ and $R+X$.

$$\begin{aligned}
\frac{d\varphi}{d(R+X)} &= \rho\frac{\varphi}{R+X} \\
\frac{d\varphi}{\varphi} &= \rho\frac{d(R+X)}{R+X} \\
\mathrm{d}\ln\varphi &= \rho\,\mathrm{d}\ln(R+X) \\
\ln\varphi - \ln\varphi_0 &= \rho\{\ln(R+X) - \ln(R_0+X_0)\} \\
\ln\frac{\varphi}{\varphi_0} &= \rho\ln\frac{R+X}{R_0+X_0} \\
\frac{\varphi}{\varphi_0} &= \Big(\frac{R+X}{R_0+X_0}\Big)^{\rho} \\
\varphi &= \varphi_0\Big(\frac{R+X}{R_0+X_0}\Big)^{\rho}
\end{aligned}$$

Considering the initial condition $\varphi_0 = B_0/U_0$ and $X_0 = 0$, we have

$$\varphi = \frac{B_0}{U_0}\Big(\frac{R+X}{R_0}\Big)^{\rho} = c_2(R+X)^{\rho}, \tag{D1}$$

where

$$c_2 := \frac{B_0}{U_0 R_0^{\rho}}.$$

We can then substitute the equation $X = \int_{R_0}^{R}\varphi\,dR$ by $\varphi = c_2(R+X)^{\rho}$. In this work, we consider the parameters under constraint $\beta = \sigma\kappa_1$ which means $\rho = 1$. Hence, we can use $\varphi = c_2(R+X)$. We can derive the expression with R and X as below:

$$\begin{aligned}
X &= \int_{R_0}^{R}\varphi dR \\
&= \int_{R_0}^{R} c_2(R+X)dR \\
&= c_2\int_{R_0}^{R} RdR + c_2\int_{R_0}^{R} XdR \\
&= \frac{1}{2}c_2\Big(R^2 - R_0^2\Big) + c_2\int_{R_0}^{R} XdR \\
&= \frac{1}{2}c_2\Big(R^2 - R_0^2\Big) + c_2\Big([XR]_{R_0}^{R} - \int_{X(R_0)}^{X(R)} RdX\Big) \\
&= \frac{1}{2}c_2\Big(R^2 - R_0^2\Big) + c_2XR - c_2\int_{X(R_0)}^{X(R)} RdX, \\
&= \frac{1}{2}c_2\Big(R^2 - R_0^2\Big) + c_2XR - c_2\int_{R_0}^{R} \varphi RdR.
\end{aligned}$$

We notice that the last term $-c_2\int_{R_0}^{R}\varphi RdR$ can further be expanded in similar method. We can find the recurrence relation as follow:

$$\begin{aligned}\int_{R_0}^{R} \varphi R^{n-1} dR &= \int_{R_0}^{R} c_2(R+X)R^{n-1} dR \\ &= c_2 \int_{R_0}^{R} R^n dR + c_2 \int_{R_0}^{R} X R^{n-1} dR \\ &= \frac{1}{n+1} c_2 \Big(R^{n+1} - R_0^{n+1}\Big) + \frac{1}{n} c_2 \int_{R_0}^{R} X dR^n \\ &= \frac{1}{n+1} c_2 \Big(R^{n+1} - R_0^{n+1}\Big) + \frac{1}{n} c_2 \Big([XR^n]_{R_0}^{R} - \int_{X(R_0)}^{X(R)} R^n dX \Big) \\ &= \frac{1}{n+1} c_2 \Big(R^{n+1} - R_0^{n+1}\Big) + \frac{1}{n} c_2 X R^n - \frac{1}{n} c_2 \int_{R_0}^{R} \varphi R^n dR.\end{aligned}$$

If we let the first integration term $\int_{R_0}^{R} \varphi dR$ is $n = 1$, the general form of the integration term can be expressed by

$$(-1)^{n+1} \frac{1}{(n-1)!} c_2^{n-1} \int_{R_0}^{R} \varphi R^{n-1} dR.$$

We can write the expression of X by the sum up to M terms:

$$\begin{aligned}X &= \sum_{n=1}^{M} (-1)^{n+1} \frac{1}{(n-1)!} c_2^{n-1} \Big\{ \frac{1}{n+1} c_2 \Big(R^{n+1} - R_0^{n+1}\Big) + \frac{1}{n} c_2 X R^n \Big\} + (-1)^{M+1} \frac{1}{M!} c_2^M \int_{R_0}^{R} \varphi R^M dR \\ &= \sum_{n=1}^{M} (-1)^{n+1} \frac{c_2^n}{n!} \Big\{ \frac{n}{n+1} (R^{n+1} - R_0^{n+1}) + R^n X \Big\} + (-1)^{M+1} \frac{1}{M!} c_2^M \int_{R_0}^{R} \varphi R^M dR\end{aligned}$$

We need to verify if the the residual term converges to 0 as $M \to \infty$. It can be seen that

$$\frac{1}{M!} c_2^M \int_{R_0}^{R} \varphi R^M dR < \frac{1}{M!} c_2^M \int_{R_0}^{(1-p)N} \varphi^* \{(1-p)N\}^M dR = \frac{1}{M!} c_2^M \varphi^* \{(1-p)N - R_0\} \{(1-p)N\}^M,$$

since $R < (1-p)N$ and $\varphi \leq \varphi^*$. We can use the ratio test to verify that

$$\lim_{M \to \infty} \frac{\frac{1}{(M+1)!} c_2^{M+1} \varphi^* \{(1-p)N - R_0\} \{(1-p)N\}^{M+1}}{\frac{1}{M!} c_2^M \varphi^* \{(1-p)N - R_0\} \{(1-p)N\}^M} = \lim_{M \to \infty} \frac{c_2 (1-p) N}{M+1} = 0.$$

which means

$$\lim_{M \to \infty} \frac{1}{M!} c_2^M \varphi^* \{(1-p)N - R_0\} \{(1-p)N\}^M = 0.$$

Therefore, we have

$$\lim_{M \to \infty} (-1)^{M+1} \frac{1}{M!} c_2^M \int_{R_0}^{R} \varphi R^M dR = 0.$$

Hence, we can write the conserved quantity of X and R.

$$X = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{c_2^n}{n!} \Big\{ \frac{n}{n+1} (R^{n+1} - R_0^{n+1}) + R^n X \Big\}. \tag{D2}$$

Final state As for the final size, we can use substitute the value of final state R^* and X^* in equation (D2) and $X^* = (1-p)N - R^*$.

$$(1-p)N = R^* + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{c_2^n}{n!} \Big\{ \frac{n}{n+1} ({R^*}^{n+1} - R_0^{n+1}) + {R^*}^n \{(1-p)N - R^*\} \Big\}. \tag{D3}$$

We can be rewrite the equation by substituting $c_2 = \frac{B_0}{U_0 R_0}$.

$$(1-p)N = R^* + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n!} \Big(\frac{B_0}{U_0}\Big)^n \Big\{ (1-p)N \Big(\frac{R^*}{R_0}\Big)^n - \frac{1}{n+1} R^* \Big(\frac{R^*}{R_0}\Big)^n - \frac{n}{n+1} R_0 \Big\}. \tag{D4}$$

Expression of X^* We can rewrite the equation (D4) as:

$$
\begin{aligned}
(1-p)N &= R^* + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0}{U_0}\Big)^n\left\{(1-p)N\Big(\frac{R^*}{R_0}\Big)^n - \frac{1}{n+1}R^*\Big(\frac{R^*}{R_0}\Big)^n - \frac{n}{n+1}R_0\right\}\\
&= R^* + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n(1-p)N - \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n\frac{1}{n+1}R^*\\
&\quad - \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0}{U_0}\Big)^n\frac{n}{n+1}R_0.
\end{aligned}
$$

The infinite series of the right hand side can be applied with Taylor's expansion:

$$
\mathrm{e}^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}.
$$

We can get the sum of infinite series from $n=1$ and $n=2$:

$$
\sum_{n=1}^{\infty}\frac{x^n}{n!} = \mathrm{e}^x - 1, \quad \sum_{n=2}^{\infty}\frac{x^n}{n!} = \mathrm{e}^x - 1 - x,
$$

which will be used in the following step. Then, We transform the three terms with infinite series one by one.
The first infinite series term:

$$
\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n(1-p)N = -(1-p)N\sum_{n=1}^{\infty}(-1)^n\frac{1}{n!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n = (1-p)N\Big\{1-\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big)\Big\}
$$

The second term:

$$
\begin{aligned}
-\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n\frac{1}{n+1}R^* &= -R^*\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{(n+1)!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^n\\
&= -R^*\frac{U_0R_0}{B_0R^*}\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{(n+1)!}\Big(\frac{B_0R^*}{U_0R_0}\Big)^{n+1}\\
&= -R_0\frac{U_0}{B_0}\sum_{n=2}^{\infty}\frac{1}{n!}\Big(-\frac{B_0R^*}{U_0R_0}\Big)^n\\
&= -R_0\frac{U_0}{B_0}\Big\{\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) - 1 - \Big(-\frac{B_0R^*}{U_0R_0}\Big)\Big\}\\
&= -R_0\frac{U_0}{B_0}\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) + R_0\frac{U_0}{B_0} - R^*.
\end{aligned}
$$

The third term:

$$
\begin{aligned}
-\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n!}\Big(\frac{B_0}{U_0}\Big)^n\frac{n}{n+1}R_0 &= \sum_{n=1}^{\infty}(-1)^{n}\frac{1}{n!}\Big(\frac{B_0}{U_0}\Big)^n\Big(1-\frac{1}{n+1}\Big)R_0 \\
&= R_0\sum_{n=1}^{\infty}\frac{1}{n!}\Big(-\frac{B_0}{U_0}\Big)^n - R_0\sum_{n=1}^{\infty}\frac{1}{(n+1)!}\Big(-\frac{B_0}{U_0}\Big)^n \\
&= R_0\sum_{n=1}^{\infty}\frac{1}{n!}\Big(-\frac{B_0}{U_0}\Big)^n + R_0\frac{U_0}{B_0}\sum_{n=1}^{\infty}\frac{1}{(n+1)!}\Big(-\frac{B_0}{U_0}\Big)^{n+1} \\
&= R_0\sum_{n=1}^{\infty}\frac{1}{n!}\Big(-\frac{B_0}{U_0}\Big)^n + R_0\frac{U_0}{B_0}\sum_{n=2}^{\infty}\frac{1}{n!}\Big(-\frac{B_0}{U_0}\Big)^{n} \\
&= R_0\Big\{\exp\Big(-\frac{B_0}{U_0}\Big)-1\Big\} + R_0\frac{U_0}{B_0}\Big\{\exp\Big(-\frac{B_0}{U_0}\Big)-1-\Big(-\frac{B_0}{U_0}\Big)\Big\} \\
&= R_0\exp\Big(-\frac{B_0}{U_0}\Big) - R_0 + R_0\frac{U_0}{B_0}\exp\Big(-\frac{B_0}{U_0}\Big) - R_0\frac{U_0}{B_0} + R_0 \\
&= R_0\exp\Big(-\frac{B_0}{U_0}\Big) + R_0\frac{U_0}{B_0}\exp\Big(-\frac{B_0}{U_0}\Big) - R_0\frac{U_0}{B_0}
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
(1-p)N = {} & R^* + (1-p)N\Big\{1-\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big)\Big\} - R_0\frac{U_0}{B_0}\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) + R_0\frac{U_0}{B_0} - R^* + R_0\exp\Big(-\frac{B_0}{U_0}\Big) \\
& + R_0\frac{U_0}{B_0}\exp\Big(-\frac{B_0}{U_0}\Big) - R_0\frac{U_0}{B_0} \\
= {} & (1-p)N - \Big\{(1-p)N + R_0\frac{U_0}{B_0}\Big\}\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) + \Big(R_0 + R_0\frac{U_0}{B_0}\Big)\exp\Big(-\frac{B_0}{U_0}\Big) \\
0 = {} & -\Big\{(1-p)N + R_0\frac{U_0}{B_0}\Big\}\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) + \Big(R_0 + R_0\frac{U_0}{B_0}\Big)\exp\Big(-\frac{B_0}{U_0}\Big).
\end{aligned}
$$

Then, we can get an expression of R^* in terms of R_0:

$$
\begin{aligned}
\Big\{(1-p)N + R_0\frac{U_0}{B_0}\Big\}\exp\Big(-\frac{B_0R^*}{U_0R_0}\Big) &= \Big(R_0 + R_0\frac{U_0}{B_0}\Big)\exp\Big(-\frac{B_0}{U_0}\Big) \\
\frac{\exp\big(-\frac{B_0R^*}{U_0R_0}\big)}{\exp\big(-\frac{B_0}{U_0}\big)} &= \frac{R_0 + R_0\frac{U_0}{B_0}}{(1-p)N + R_0\frac{U_0}{B_0}} \\
\mathrm{e}^{\big(-\frac{B_0}{U_0}\big)\big(\frac{R^*}{R_0}-1\big)} &= \frac{R_0 + R_0\frac{U_0}{B_0}}{(1-p)N + R_0\frac{U_0}{B_0}} \\
\Big(-\frac{B_0}{U_0}\Big)\Big(\frac{R^*}{R_0}-1\Big) &= \ln\frac{R_0 + R_0\frac{U_0}{B_0}}{(1-p)N + R_0\frac{U_0}{B_0}} \\
\frac{R^*}{R_0} &= 1 - \frac{U_0}{B_0}\ln\frac{R_0 + R_0\frac{U_0}{B_0}}{(1-p)N + R_0\frac{U_0}{B_0}}
\end{aligned}
$$

Since

$$
R^* = R_0 + R_0\frac{U_0}{B_0}\ln\frac{\frac{(1-p)N}{R_0} + \frac{U_0}{B_0}}{1 + \frac{U_0}{B_0}} \tag{D5}
$$

and $(1-p)N = R^* + X^*$ are satisfied for the final state, we have

$$
(1-p)N = X^* + R_0 + R_0\frac{U_0}{B_0}\ln\frac{\frac{(1-p)N}{R_0} + \frac{U_0}{B_0}}{1 + \frac{U_0}{B_0}}
$$

as shown by equation (4.5).

Range of the root for equation (4.5)

We want to verify that the range of X^* in equation (4.5) satisfied $X^* \in (0, (1-p)N - R_0)$. From (D4), it is easily to see that $X^* < (1-p)N - R_0$. We will show that $X^* > 0$ in the following parts. For simplicity, let us denote

$$u = \frac{U_0}{B_0}; \quad v = \frac{(1-p)N}{R_0}; \quad x^* = \frac{X^*}{R_0} = \frac{(1-p)N - R^*}{R_0}.$$

We have $u > 0$ and $v > 1$ which represent the initial state. Then, the equation (4.5) becomes

$$v = x^* + 1 + u \ln \frac{v+u}{1+u}.$$

We want to show that $x^* > 0$ which means $X^* > 0$. Since

$$\ln \frac{v+u}{1+u} = \ln \frac{(v-1)+(1+u)}{1+u} = \ln \Big(1 + \frac{v-1}{1+u}\Big) < \frac{v-1}{1+u} < \frac{v-1}{u},$$

we have

$$\begin{aligned} u \ln \frac{v+u}{1+u} &< v - 1 \\ x^* = v - 1 - u \ln \frac{v+u}{1+u} &> 0. \end{aligned}$$

Therefore, $X^* \in (0, (1-p)N - R_0)$.

Appendix E Proof of Corollary 4.3

We denote the root R^* of equations (4.2) is R_1^*, the root of equations (4.4) is R_2^*. We want to show that given same parameters, the relation

$$R_1^* < R_2^*$$

always holds.

Proof For convenience, let's define $\rho = \beta/\sigma\kappa$. We also define the two equations correspond to (4.2) and (4.4):

$$\begin{aligned} g_1(y) &= y + B_0 \Big(\frac{y}{R_0}\Big)^{\rho}; \\ g_2(y) &= y + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} c_1^n \Big(y^{n\rho+1} - R_0^{n\rho+1}\Big). \end{aligned}$$

From (4.2), $g_1(R_1^*) = (1-p)N$; from (4.4), $g_2(R_2^*) = (1-p)N$.

We know that $g_1'(y) > 0, g_2'(y) > 0$ in their domain. Hence, if we find $g_2(R_1^*) < (1-p)N$, it means that $g_2(R_1^*) < g_2(R_2^*)$, which implies $R_2^* > R_1^*$.

$$\begin{aligned} g_2(R_1^*) &= R_1^* + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} c_1^n \Big({R_1^*}^{n\rho+1} - R_0^{n\rho+1}\Big) \\ &= R_1^* + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} \Big(\frac{B_0}{(U_0+B_0)R_0^{\rho}}\Big)^n \Big({R_1^*}^{n\rho+1} - R_0^{n\rho+1}\Big) \\ &= R_1^* + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} \Big(\frac{B_0}{U_0+B_0}\Big)^n \Big\{\Big(\frac{R_1^*}{R_0}\Big)^{n\rho} R_1^* - R_0\Big\} \\ &= R_1^* + \sum_{n=1}^{\infty} \frac{1}{n\rho+1} \Big(\frac{B_0}{U_0+B_0}\Big)^n \Big\{\Big(\frac{(1-p)N - R_1^*}{B_0}\Big)^n R_1^* - R_0\Big\}. \end{aligned}$$

Since R_1^* satisfies $g_1(R_1^*) = (1-p)N$, which means

$$(1-p)N = R_1^* + B_0\Big(\frac{R_1^*}{R_0}\Big)^\rho,$$
$$\Big(\frac{R_1^*}{R_0}\Big)^\rho = \frac{(1-p)N - R_1^*}{B_0}.$$

We have

$$\begin{aligned}
g_2(R_1^*) &= R_1^* + \sum_{n=1}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n \Big\{\Big(\frac{(1-p)N-R_1^*}{B_0}\Big)^n R_1^* - R_0\Big\} \\
&< R_1^* + R_1^* \sum_{n=1}^{\infty} \frac{1}{n\rho+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n \Big(\frac{(1-p)N-R_1^*}{B_0}\Big)^n \\
&= R_1^* + R_1^* \sum_{n=1}^{\infty} \frac{1}{n\rho+1}\Big(\frac{(1-p)N-R_1^*}{U_0+B_0}\Big)^n \\
&< R_1^* + R_1^* \sum_{n=1}^{\infty} \frac{1}{\rho+1}\Big(\frac{(1-p)N-R_1^*}{U_0+B_0}\Big)^n \\
&= R_1^* + R_1^* \sum_{n=1}^{\infty} \frac{1}{\rho+1}\Big(\frac{(1-p)N-R_1^*}{(1-p)N-R_0}\Big)^n \\
&= R_1^* + R_1^* \frac{1}{\rho+1} \frac{(1-p)N-R_1^*}{R_1^*-R_0} \\
&\le R_1^* + (1-p)N - R_1^*, \quad \text{if } \frac{R_1^*}{R_0} \ge 1 + \frac{1}{\rho}, \\
&= (1-p)N.
\end{aligned}$$

The condition $\frac{R_1^*}{R_0} \ge 1 + \frac{1}{\rho}$ is satisfied for our initial value B_0, R_0 sufficiently small. Because if $\frac{R_1^*}{R_0} < 1 + \frac{1}{\rho}$, according to the equation (4.2), we have

$$\frac{B_1^*}{B_0} = \Big(\frac{R_1^*}{R_0}\Big)^\rho < \Big(1+\frac{1}{\rho}\Big)^\rho < \mathrm{e},$$

where B_1^* is the final size of believer in terms of R_1^* by equation (4.2). Then, we can get:

$$\frac{B_0}{B_1^*} > \frac{1}{\mathrm{e}} \quad \text{and} \quad \frac{R_0}{R_1^*} > \frac{\rho}{\rho+1}.$$

Since $B_1^* + R_1^* = (1-p)N$,

$$\frac{B_0+R_0}{(1-p)N} > \min\Big\{\frac{1}{\mathrm{e}}, \frac{\rho}{\rho+1}\Big\}.$$

This implies that at least one of the initial values B_0, R_0 must be relatively large, which conflicts with our assumption that they are sufficiently small. Hence, the condition $\frac{R_1^*}{R_0} \ge 1+\frac{1}{\rho}$ is satisfied. We can safely come to the result that $g_2(R_1^*) < (1-p)N$.

Since R_2^* satisfies $g_2(R_2^*) = (1-p)N$ and the function g_2 is monotonically increasing in its domain, we could obtain the result:

$$R_1^* < R_2^*.$$

Then, the social damage B_1^* from equation (4.2) and X_2^* form (4.4) have:

$$B_1^* > X_2^*.$$

Appendix F Proof of Corollary 4.4

We denote the root of equation (4.4) with $\beta = \sigma\kappa_1$ is R_{21}^*, the final state expressed by equation (4.5) is R_3^*. We want to show that given same parameters, the relation

$$R_{21}^* < R_3^*$$

holds.

Value of R_{21}^* We can rewrite the equations (4.4) with $\beta = \sigma\kappa_1$ to the form without infinite series:

$$(1-p)N = R_0 + \sum_{n=0}^{\infty} \frac{1}{n+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n \Big\{R^*\Big(\frac{R^*}{R_0}\Big)^n - R_0\Big\}. \tag{F1}$$

Notice that the infinite series in the right hand side can be substituted by Taylor's expansion

$$\ln(1-x) = \sum_{n=1}^{\infty} -\frac{x^n}{n}, \quad \text{for } |x| < 1.$$

We can the transform the equation (F1) by the following steps:

$$\begin{aligned}
(1-p)N &= R_0 + \sum_{n=0}^{\infty} \frac{1}{n+1}\Big(\frac{B_0}{U_0+B_0}\cdot\frac{R^*}{R_0}\Big)^n R^* - \sum_{n=0}^{\infty} \frac{1}{n+1}\Big(\frac{B_0}{U_0+B_0}\Big)^n R_0 \\
&= R_0 + R^*\frac{(U_0+B_0)R_0}{B_0R^*}\sum_{n=0}^{\infty} \frac{1}{n+1}\Big(\frac{B_0}{U_0+B_0}\cdot\frac{R^*}{R_0}\Big)^{n+1} - R_0\frac{U_0+B_0}{B_0}\sum_{n=0}^{\infty} \frac{1}{n+1}\Big(\frac{B_0}{U_0+B_0}\Big)^{n+1} \\
&= R_0 + R_0\frac{U_0+B_0}{B_0}\sum_{n=1}^{\infty} \frac{1}{n}\Big(\frac{B_0}{U_0+B_0}\cdot\frac{R^*}{R_0}\Big)^{n} - R_0\frac{U_0+B_0}{B_0}\sum_{n=1}^{\infty} \frac{1}{n}\Big(\frac{B_0}{U_0+B_0}\Big)^{n} \\
&= R_0 + R_0\frac{U_0+B_0}{B_0}\Big\{-\ln\Big(1-\frac{B_0}{U_0+B_0}\cdot\frac{R^*}{R_0}\Big)\Big\} - R_0\frac{U_0+B_0}{B_0}\Big\{-\ln\Big(1-\frac{B_0}{U_0+B_0}\Big)\Big\} \\
&= R_0 + R_0\frac{U_0+B_0}{B_0}\ln\frac{1-\frac{B_0}{U_0+B_0}}{1-\frac{B_0}{U_0+B_0}\cdot\frac{R^*}{R_0}} \\
&= R_0 + R_0\frac{U_0+B_0}{B_0}\ln\frac{U_0}{U_0+B_0-B_0R^*/R_0}.
\end{aligned}$$

One important thing is that the root of (4.4) satisfies (C2), which has

$$\frac{B_0}{U_0+B_0}\Big(\frac{R}{R_0}\Big)^\rho = \frac{\varphi}{1+\varphi} < 1,$$

where $\rho = \frac{\beta}{\sigma\kappa}$ and $\varphi = \frac{B}{U}$. Here, we let $\rho = 1$ for the prerequisite $\beta = \sigma\kappa_1$, $R = R^*$ for the final state, then we have

$$\frac{B_0}{U_0+B_0}\Big(\frac{R^*}{R_0}\Big) < 1,$$

$$R^* < R_0\frac{U_0+B_0}{B_0}.$$

This is an important property about the size of the root in equation (4.4).
Thus, we have

$$R_0 + R_0\frac{U_0+B_0}{B_0}\ln\frac{U_0}{U_0+B_0-B_0R_{21}^*/R_0} = (1-p)N$$

and

$$R_{21}^* < R_0\frac{U_0+B_0}{B_0}. \tag{F2}$$

Comparison between R_{21}^* and R_3^* For simplicity, let us denote

$$u = \frac{U_0}{B_0}; \quad v = \frac{(1-p)N}{R_0}; \quad x_{21}^* = \frac{R_{21}^*}{R_0}; \quad x_3^* = \frac{R_3^*}{R_0}.$$

It is clear that $u > 0$, $v > 1$. We can see that x_{21}^* satisfies:

$$1 + (1+u) \ln \frac{u}{1 + u - R^*/R_0} = v.$$

We may define an equation $g_3(x)$ by

$$g_3(x) = 1 + (1+u) \ln \frac{u}{1+u-x}.$$

Then, we have

$$g_3(x_{21}^*) = v.$$

We can easily find that $g_3'(x) > 0$. Therefore, if we can find that $g_3(x_3^*) > v = g_3(x_{21}^*)$, it will be equivalent to $R_3^* > R_{21}^*$.

$$x_3^* = \frac{R_3^*}{R_0} = 1 + u \ln \frac{v+u}{1+u}.$$

Substitute this expression of x_3^* in $g_3(x)$, we have

$$g_3(x_3^*) = 1 + (1+u) \ln \frac{u}{1 + u \ln \frac{v+u}{1+u}} = 1 + (1+u) \ln \frac{1}{1 - \ln \frac{v+u}{1+u}} = 1 - (1+u) \ln \ln \frac{\mathrm{e}(1+u)}{v+u}.$$

Notice that in order to make the term $\ln \ln \frac{\mathrm{e}(1+u)}{v+u}$ meaningful, it requires

$$\begin{aligned} \ln \frac{\mathrm{e}(1+u)}{v+u} &> 0 \\ \frac{\mathrm{e}(1+u)}{v+u} &> 1 \\ \mathrm{e}(1+u) &> v+u \\ v &< \mathrm{e} + (\mathrm{e}-1)u. \end{aligned}$$

This means that if $v \geq \mathrm{e} + (\mathrm{e}-1)u$, then the expression of $g_3(x_3^*)$ will be meaningless. In fact, if we let $v \geq \mathrm{e} + (\mathrm{e}-1)u$, then we have

$$x_3^* = 1 - u \ln \frac{1+u}{v+u} \geq 1 + u.$$

The inequality $x_3^* \geq 1 + u$ means that R_3^* satisfies

$$R_3^* \geq R_0 \Big(1 + \frac{U_0}{B_0}\Big) = R_0 \frac{U_0 + B_0}{B_0}.$$

From (F2), we have

$$R_{21}^* < R_0 \frac{U_0 + B_0}{B_0} \leq R_3^*, \quad \text{when} \quad v \geq \mathrm{e} + (\mathrm{e}-1)u.$$

As for the case with $1 < v < \mathrm{e} + (\mathrm{e}-1)u$, we may define a function

$$g_4(u,v) = 1 - (1+u) \ln \ln \frac{\mathrm{e}(1+u)}{v+u} - v,$$

with $u > 0$, $1 < v < \mathrm{e} + (\mathrm{e}-1)u$. It comes from $g_4(u,v) = g_3(x_3^*) - v$. If $g_4(u,v) > 0$, it means that $g_3(x_3^*) > v$ and $g_3(x_3^*) > g_3(x_{21}^*)$. The value of $g_4(u,v)$ at the boundary $v = 1$ is

$$g_4(u,1) = 1 - 1 = 0.$$

We can also find that

$$\begin{aligned}\frac{\partial g_4(u,v)}{\partial v} &= -(1+u)\frac{1}{\ln\frac{\mathrm{e}(1+u)}{v+u}}\frac{v+u}{\mathrm{e}(1+u)}\frac{-\mathrm{e}(1+u)}{(v+u)^2}-1\\ &= \frac{1}{1+\ln\frac{1+u}{v+u}}\frac{1+u}{v+u}-1\\ &> \frac{1}{\frac{1+u}{v+u}}\frac{1+u}{v+u}-1\\ &= 0.\end{aligned}$$

This implies that for any $1 < v < \mathrm{e} + (\mathrm{e} - 1)u$, we have $g_4(u, v) > 0$. Thus, $g_3(x_3^*) > v$. Since we have $g_3(x_{21}^*) - v$ and $g_3'(x) > 0$, we can conclude that $x_3^* > x_{21}^*$ is true for the case $1 < v < \mathrm{e} + (\mathrm{e} - 1)u$. Therefore, we have proved that

$$R_{21}^* < R_3^*$$

is always satisfied for any cases. Then, the social damage X_{21}^* from equation (4.4) and X_3^* form (4.5) have:

$$X_{21}^* > X_3^*.$$

Appendix G Proof of Theorem 5.1

Since X monotonically increasing and bounded above ($X \le (1-p)N$),

$$X \to X^* \quad \text{as} \quad t \to \infty. \tag{G1}$$

As for the final state of U and B, it depends on the value of κ_3. Firstly, we use proof by contradiction to show that $B \to 0$.

Case of $\kappa_3 = 0$

When $\kappa_3 = 0$, we have

$$\frac{dU}{dt} + \frac{dB}{dt} = -\gamma\kappa_2 AB. \tag{G2}$$

If $B \to B^* > 0$, from

$$\frac{dA}{dt} = \alpha BS - \delta A \quad \text{and} \quad S + A = pN,$$

we have

$$\begin{aligned}\alpha B^* S^* - \delta A^* &= 0\\ \alpha B^*(pN - A^*) - \delta A^* &= 0\\ \alpha B^* pN &= \delta A^* + \alpha B^* A^*\\ A^* &= \frac{\alpha B^*}{\delta + \alpha B^*}pN.\end{aligned}$$

We have shown that if $B \to B^* > 0$, the, $A \to A^* > 0$. From (G2), we have

$$\lim_{t\to\infty}\frac{dU}{dt} + \frac{dB}{dt} = -\gamma\kappa_2 A^* B^* < 0.$$

This violates the final state. Hence, it is impossible that B converges to a positive value B^*,

$$B \to 0 \quad \text{as} \quad t \to \infty. \tag{G3}$$

Since $B^* = 0$, we have

$$A^* = \frac{\alpha B^*}{\delta + \alpha B^*}pN = \frac{\alpha \cdot 0}{\delta + \alpha \cdot 0}pN = 0,$$

which means that

$$A \to 0, \quad S \to pN \quad \text{as} \quad t \to \infty. \tag{G4}$$

As for U, we use the proof by contradiction again. From the system (5.1) and the condition $\kappa_3 = 0$, we have

$$\frac{dB}{dt} = \{\beta U - \gamma\kappa_2 A\}B.$$

We may assume that $U \to U^* > 0$. As we already know that $A \to 0$, there exists a finite time T such that

$$\beta U - \gamma\kappa_2 A > 0, \quad \text{for any} \quad t > T.$$

Note that T is finite, this means that the value of B at time T is positive. From the argument above, we have

$$B > B(T), \quad \text{for any} \quad t > T.$$

However, we already proved that $B \to 0$ in (G3). This contradiction implies that $U \to U^* > 0$ is false, U can not converge to a positive value U^*. Therefore,

$$U \to 0 \quad \text{as} \quad t \to \infty. \tag{G5}$$

The argument above have proved the final state of case $\kappa_3 = 0$.

Case of $\kappa_3 > 0$

In the case that $\kappa_3 > 0$, the method is a little bit different. From the system (5.1), we have

$$\lim_{t\to\infty} \frac{dX}{dt} = \lim_{t\to\infty} \gamma(\kappa_2 A + \kappa_3 X)B \geq \lim_{t\to\infty} \gamma\kappa_3 XB \text{ and } \lim_{t\to\infty} \frac{dX}{dt} = 0.$$

Since $X \to X^* > 0$, B must go to 0, which means

$$B \to 0 \quad \text{as} \quad t \to \infty. \tag{G6}$$

Using similar argument as (G4), we can also obtain the result that $A \to 0$, $S \to pN$.

As for U, we use a different method by considering the following relation:

$$\frac{dU}{dX} = \frac{-\beta BU}{\gamma(\kappa_2 A + \kappa_3 X)B} = -\frac{\beta U}{\gamma(\kappa_2 A + \kappa_3 X)} \geq -\frac{\beta U}{\gamma\kappa_3 X}.$$

$$\frac{dU}{U} \geq -\frac{\beta}{\gamma\kappa_3}\frac{dX}{X}.$$

Then, we have this relationship:

$$d\ln U \geq -\frac{\beta}{\gamma\kappa_3} d\ln X. \tag{G7}$$

We may consider a time $t' > 0$. At that time $t = t'$, the state of system is

$$(U, B, X, S, A) = (U', B', X', S', A'),$$

where the value $X' > 0$. Then consider the time $t > t'$, from (G7), we have

$$\begin{aligned}
\ln U - \ln U' &\geq -\frac{\beta}{\gamma\kappa_3}(\ln X - \ln X') \\
\frac{U}{U'} &\geq \left(\frac{X}{X'}\right)^{-\frac{\beta}{\gamma\kappa_3}} \\
U &\geq U'\left(\frac{X}{X'}\right)^{-\frac{\beta}{\gamma\kappa_3}}
\end{aligned}$$

This equation holds for the final state ($t \to \infty$). Hence, we have the following relationship:

$$U^* \geq U' \Big(\frac{X^*}{X'}\Big)^{-\frac{\beta}{\gamma\kappa_3}}.$$

This value of lower bound depends on the selection of time t'. Any way, it shows that there exists a lower bound of U^*, which is positive. Since U in monotonically decreasing and have a positive lower bound, we can say that

$$U \to U^* \quad \text{as} \quad t \to \infty. \tag{G8}$$

From (G1), (G3), (G4) and (G5), (G6) and (G8), we have shown the final state for the case $\kappa_3 > 0$:

$$(U, B, X, S, A) \to (U^*, 0, X^*, pN, 0), \quad \text{as } t \to \infty,$$

where the value of U^* follows

$$\begin{cases} U^* = 0, & \text{if } \kappa_3 = 0; \\ U^* > 0, & \text{if } \kappa_3 > 0. \end{cases}$$

Appendix H Proof of Theorem 6.1

In the system (3.1), we can see that X is monotonically increasing and bounded above, which means

$$X \to X^* \quad \text{as} \quad t \to \infty. \tag{H1}$$

As for the final state of other variables, it also depends on the value of κ_3.

Case of $\kappa_3 > 0$

It is easier to consider the case $\kappa_3 > 0$ in the beginning. From the equation

$$\frac{dU}{dt} = -\beta BU - \sigma(\kappa_1 R + \kappa_2 A + \kappa_3 X)U \leq -\sigma\kappa_3 XU,$$

we can see that $U \to 0$. Otherwise, dU/dt will be negative in the final state. Since

$$\frac{dB}{dt} = \{\beta U - \gamma(\kappa_1 R + \kappa_2 A + \kappa_3 X)\}B \leq (\beta U - \gamma\kappa_3 X)B,$$

if $B \to B^*$, it causes

$$\lim_{t\to\infty} \frac{dB}{dt} \leq \lim_{t\to\infty} (\beta U - \gamma\kappa_3 X)B = -\gamma\kappa_3 X^* B^* < 0.$$

It violates the final state, which means $B \to B^*$ is impossible. Hence, we have $B \to 0$.

Then, we consider the sum of R and Y:

$$\frac{dR}{dt} + \frac{dY}{dt} = \sigma(\kappa_1 R + \kappa_2 A + \kappa_3 X)U.$$

It is positive in finite time, since all the variables are positive in finite time. Therefore, the sum $R+Y$ is monotonically increasing and bounded above, it will converge to a positive value. From the system,

$$\frac{dY}{dt} = \delta R - \alpha BY.$$

Since $\lim_{t\to\infty} \frac{dY}{dt} = 0$,

$$\begin{aligned} \delta R^* - \alpha B^* Y^* &= 0 \\ R^* &= \frac{\alpha}{\delta} B^* Y^* \\ R^* &= 0, \end{aligned}$$

since we know that $B^* = 0$. This show that $R \to 0$. We already have the sum of $R + Y$ will converge to a positive value, it implies $Y \to Y^* > 0$.

We already proved the convergence of U, B, R, X and Y. As for S and A, since $B \to 0$, we can get $A \to 0$ and $S \to 0$ by using the similar arguments as Appendix (G).

Case of $\kappa_3 = 0$

In the case that $\kappa_3 = 0$, we will use proof by contradiction. Firstly, let us assume that $U \to U^* > 0$. From

$$\frac{dU}{dt} = \{-\beta B - \sigma(\kappa_1 R + \kappa_2 A)\}U,$$

we have $-\beta B - \sigma(\kappa_1 R + \kappa_2 A) \to 0$. Otherwise, it will be impossible for $U \to U^*$. This implies

$$B \to 0 \quad \text{and} \quad \kappa_1 R + \kappa_2 A \to 0.$$

However, as shown in

$$\frac{dB}{dt} = \{\beta U - \gamma(\kappa_1 R + \kappa_2 A)\}B, \tag{H2}$$

it is impossible to make $B \to 0$. This is because we already have $U \to U^* > 0$ and $kappa_1 R + \kappa_2 A \to 0$. Hence, there exists a time T such that

$$\beta U - \gamma(\kappa_1 R + \kappa_2 A) > 0, \quad \text{for any} t > T.$$

The value $B > 0$ in finite time, therefore, for any $t > T$, $dB/dt > 0$. It is impossible to make $B \to 0$. This contradiction implies our assumption $U \to U^* > 0$ is false,

$$U \to 0 \quad \text{as} \quad t \to \infty.$$

Then, we can consider the B. If $B \to B^* > 0$, it brings out $R \to R^* > 0$, $A \to A^* > 0$ (using the similar arguments as case $\kappa_3 > 0$). However, according to equation (H2),

$$\lim_{t\to\infty} \frac{dB}{dt} = \lim_{t\to\infty} \{\beta U - \gamma(\kappa_1 R + \kappa_2 A)\}B = -\gamma(\kappa_1 R^* + \kappa_2 A^*)B^* < 0,$$

since we already know $U \to 0$. This violates the definition of final state. Therefore, the assumption $B \to B^* > 0$ is false,

$$B \to 0 \quad \text{as} \quad t \to \infty.$$

We already show that $X \to X^*$, $U \to 0$ and $B \to 0$. As for the convergence of other states, it can be proven in the same argument as the case $\kappa_3 > 0$.

Appendix I Application of fast process assumption

When the proportion of sophisticated members p is sufficiently large that makes $1/(1-p) \gg 1$, the risk of unsophisticated members seems converge to a certain value. We will apply fast process assumption to verify this property for the case that reformed members releasing no counter information ($\kappa_3 = 0$). In this case, the rejoinders R and A are responsible for the spreading of counter information.

The fast process assumption is based on the assumption that state transition between some states are much faster than in other states. For example, we may assume that the state transition of sophisticated members are much faster, since their state transition is not caused by "change of belief". They change from "unconcerned" state to rejoinder is like an activation of voluntary behaviour, it can be much easier than the unsophisticated members who need to accept new information. Also, we may assume the expected duration of sophisticated rejoinder state is also short. For these sophisticated members, the counter information is only an alarm that do not bring new knowledge or new ideas. It may not occupy their brain for long time and easily be forgotten. Hence, it is reasonable that the states transition of sophisticated members are easier and faster. With this assumption, we regard the state "unconcerned"(S) and "sophisticated rejoinder" (A) are fast process, and apply the quasi-stationary state approximation (QSSA) method.

From the model (3.1), we assume $dS/dt \approx 0$ and $dA/dt \approx 0$:

$$\alpha BS - \delta A \approx 0.$$

Then we can get the expression of A:

$$A \approx \frac{\alpha B}{\alpha B + \delta} pN. \tag{I1}$$

Under the fast process assumption, the sophisticated rejoinder population A can be expressed by the density of believers B and other parameters. As B changes, the value of A changes much quickly to keep the value of A close to the equilibrium $\frac{\alpha B}{\alpha B+\delta} pN$. We can use it to reduce the dimension of the original model, and conduct analysis on a simplified system. Note that the fast process assumption requires α and δ are much larger than other parameters.

We also assume the unsophisticated rejoinder R is a fast process, and have $dR/dt \approx 0$:

$$\frac{dR}{dt} = \sigma(\kappa_1 R + \kappa_2 A)U - \delta R + \alpha BY \approx 0.$$

Since $U+B+R+X+Y = (1-p)N$ always holds, $Y = (1-p)N - (U+B+R+X)$, we can get

$$R \approx \frac{\sigma AU + \alpha B\{(1-p)N - U - B - X\}}{\alpha B + \delta - \sigma U}.$$

As for the concentration of counter information $\kappa_1 R + \kappa_2 A$, we have

$$\kappa_1 R + \kappa_2 A \approx \kappa_1 \alpha B \frac{\left(1 + \frac{\kappa_2 - \kappa_1}{\kappa_1} p\right)N - (U+B+X)}{\alpha B + \delta - \sigma \kappa_1 U}.$$

The expression includes three variables U, B, X, we may denote it as a function $g(U, B, X)$. Thus, we obtain a model based on the QSSA with only three variables U, B, X:

$$\begin{aligned}
\frac{dU}{dt} &= -\beta BU - \sigma g(U,B,X)U; \\
\frac{dB}{dt} &= \beta BU - \gamma g(U,B,X)B; \\
\frac{dX}{dt} &= \gamma g(U,B,X)B.
\end{aligned}$$

Let us consider the limit of $g(U, B, X)$ as $p \to 1$ which makes $1/(1-p) \gg 1$.

$$\lim_{p\to 1} g(U,B,X) = \frac{\kappa_2 \alpha N}{\delta} B = \omega B,$$

where $\omega = \kappa_2 \alpha N/\delta$ is a constant. This implies that as $p \to 1$, the concentration of counter information converges to ωB, which is proportional to the density of believers.

Then we can construct a system with this limiting condition. Here we use the term $(\hat{U}, \hat{B}, \hat{X})$ to represent the limiting system of $p \to 1$.

$$\begin{aligned}
\frac{d\hat{U}}{dt} &= -\beta \hat{B}\hat{U} - \sigma\omega \hat{B}\hat{U}; \\
\frac{d\hat{B}}{dt} &= \beta \hat{B}\hat{U} - \gamma\omega \hat{B}^2; \\
\frac{d\hat{X}}{dt} &= \gamma\omega \hat{B}^2.
\end{aligned}$$

It can be found that the equation

$$\frac{d\hat{U}}{d(\hat{B}+\hat{X})} = -\frac{\beta+\sigma\omega}{\beta}$$

holds for any $t \geq 0$. Since we know that $\hat{U} \to 0$, $\hat{B} \to 0$, we could derive the final size

$$\hat{X}^* = \frac{\beta}{\beta+\sigma\omega}\hat{U}_0 + \hat{B}_0$$

which is the final size of the limit system. If we let $\hat{B}_0/\hat{U}_0 \to 0$, the result becomes

$$\hat{X}^* \approx \frac{\beta}{\beta+\sigma\omega}(1-p)N = \frac{(1-p)N}{1+\kappa_2\alpha\sigma N/\beta\delta}.$$

As $p \to 1$, we have

$$X^* \to \hat{X}^* \approx \frac{(1-p)N}{1+\kappa_2\alpha\sigma N/\beta\delta}$$

and the risk of unsophisticated members

$$\frac{X^*}{(1-p)N} \to \frac{\hat{X}^*}{(1-p)N} \approx \frac{1}{1+\kappa_2\alpha\sigma N/\beta\delta}.$$

In this part, we use the limiting system which is the limit for sufficiently large p that $1/(1-p) \gg 1$ based on the fast process assumption. The risk in the limiting system converge to a value that can be expressed by the parameters κ_2, α, β, σ, δ and N. This implies that as $p \to 1$, the risk of the QSSA system converges to this value. This is a positive value, and hence the risk cannot be suppress to sufficiently small. Interestingly, this value includes almost all the parameters except γ. It implies that the limit of risk is independent with the value of γ.