Ca' Foscari University of Venice

Department of Economics

Working Paper

Luca P. Merlino
Nicole Tabasso

Optimal Inspection of Rumors in Networks

ISSN: 1827-3580
No. 19/WP/2022

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 19/WP/2022
ISSN 1827-3580

Optimal Inspection of Rumors in Networks

Luca P. Merlino
University of Antwerp and ECARES, Université libre de Bruxelles

Nicole Tabasso
Ca' Foscari University of Venice ; University of Surrey, School of Economics

Abstract
We study the diffusion of a true and a false message when agents are (i) biased towards one of the messages and (ii) agents are able to inspect messages for veracity. Inspection of messages implies that a higher rumor prevalence may increase the prevalence of the truth. We employ this result to discuss how a planner may optimally choose information inspection rates of the population. We find that a planner who aims to maximize the prevalence of the truth may find it optimal to allow rumors to circulate.

Keywords
Social Networks, Rumors, Scrutiny

JEL Codes
D83, D85

Address for correspondence:
Nicole Tabasso
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: nicole.tabasso@unive.it

Optimal Inspection of Rumors in Networks

Luca P. Merlino and Nicole Tabasso*

July 6, 2022

Abstract

We study the diffusion of a true and a false message when agents are (i) biased towards one of the messages and (ii) agents are able to inspect messages for veracity. Inspection of messages implies that a higher rumor prevalence may increase the prevalence of the truth. We employ this result to discuss how a planner may optimally choose information inspection rates of the population. We find that a planner who aims to maximize the prevalence of the truth may find it optimal to allow rumors to circulate.

JEL classifications: D83, D85.
Keywords: Social Networks, Rumors, Scrutiny.

*Merlino: University of Antwerp and ECARES, Université libre de Bruxelles, Ave. F.D. Roosevelt 50, CP 114/04, Belgium; email: LucaPaolo.Merlino@uantwerpen.be. Tabasso: Ca'Foscari University of Venice, Venice, Italy and University of Surrey, School of Economics, Guildford, GU2 7XH, UK; email: nicole.tabasso@unive.it. Merlino gratefully acknowledges financial support from the Flemish Research Funds (FWO). Tabasso gratefully acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 793769.

1 Introduction

The diffusion of rumors, misinformation, or “fake news” has received considerable attention in recent years. Yet, such information generally diffuses simultaneously with correct information, and possible interactions are often overlooked in the quest to minimise rumor diffusion. In particular, the prevalence of the truth may be the socially more important variable. This is especially the case when believing the truth makes a person more likely to adopt a correct behavior, while believing the rumor implies the same action as an uniformed individual.[1]

The crucial question is whether there is any difference between fostering the truth and fighting the rumor. While intuitions suggests that these are different sides of the same coin, in this paper we show that this is not necessarily the case.

Indeed, the diffusion of information on social networks is a complex matter and various policies have been suggested to curb the spread of rumors. In the present paper we focus on one particular aspect, namely the rate at which individuals may inspect messages they receive for accuracy. This can be influenced by policy makers through various channels, such as guides on how to spot false message, or raising ”information literacy” of the population in general. Our main question of interest is the inspection rate that a benevolent planner would set, whose goal it is to maximise the proportion of correctly informed agents in society.

Specifically, we describe the diffusion of information using the SIS (*Susceptible-Infected-Susceptible*) framework, initially developed in epidemiology, where the network is modeled as the number of meetings each individual has per period. On this network, two messages pertaining to the true state of world diffuse via word of mouth. These messages are contradictory; one is correct about the true state, and the other not (the rumor). Individuals

[1]This situation occurs naturally whenever the truth requires a specific change in behavior, such as when a new disease is discovered.

are described by their type; two types exist in the population, each biased towards believing one of these messages, and individuals that do not inspect messages ignore discordant ones. Inspection is able to reveal instead the veracity of information. Consequently, irrespective of which message agents receive, if they inspect it, they become aware of the true state of the world. Finally, individuals only pass on information they believe to their neighbors.

This framework is highly tractable, which allows us to derive a clear relationship between information prevalence and inspection rates. Rumor prevalence, as intuition suggests, is strictly decreasing in inspection rates. In fact, high enough inspection rates are able to eradicate the rumor entirely. The prevalence of the truth is increasing in inspection rates whenever the rumor dies out, but whenever the rumor survives, truth prevalence is in fact increasing in rumor prevalence. Thus, an increase in inspection rates may in- or decrease the prevalence of the truth. For a planner whose objective it is to maximise this prevalence, the optimal policy depends on the budget they have available to induce inspection. We find that for either very low or very high budgets, it is always best for the planner to use all the budget, until the rumor is completely debunked, and beyond. However, for intermediate levels of the budget, it may be better to induce lower inspection rates, which allow the rumor to circulate.

We extend our results along various dimensions. First, we show that similar predictions obtain when the planner can target inspection rates to individuals whose type biases them against the truth. In particular, the planner may choose to diversify inspection rates across groups, even if a focus on the group biased towards the rumor would allow its complete eradication. Next, we show that similar results would hold if the planner is a networking platform whose objective is to maximise the overall volume of messages.

Our paper contributes to the economic literature that employs the SIS framework to model the diffusion of various states. Inspection in our paper acts very much like vaccination against a disease, as it inoculates individuals against believing a rumor. This relates us to papers that focus on strategic decisions to protect one against the diffusion of a disease (Chen and Toxvaerd, 2014; Goyal and Vigier, 2015; Toxvaerd, 2019; Talamàs and Vohra, 2020; Bizzarri, Panebianco and Pin, 2021). In particular, Galeotti and Rogers (2013) employ the SIS model to investigate how a planner would allocate vaccinations among two groups in the population. In contrast to these papers, our focus is not how protection affects the harmful state, but instead its impact on the prevalence of the truth, a positive state. Recent literature on opinion dynamics on random graphs (Akbarpour, Malladi and Saberi, 2018; Sadler, 2020) or the impact of word-of-mouth advertising on product adoption (Campbell, 2013; Carroni, Pin and Righi, 2020), captures adoption behavior in which the possibility of verification is absent.

Related to our work, Tabasso (2019) and Campbell, Leister and Zenou (2019) study the simultaneous diffusion of two types of information. However, in these papers the two information are non-contradictory and may be held by agents simultaneously.

We focus on the problem a planner faces when they are able to set inspection rates. This is a complementary problem to questions of strategic diffusion of messages (Bloch, Demange and Kranton, 2018; Kranton and McAdams, 2020), or individuals' incentives to exert effort in verification (Merlino, Pin and Tabasso, 2022).

The paper proceeds as follows. Section 2 introduces the model. Section 3 presents the results and Section 5.1 the planner's problem. Section 6 concludes. All proofs are in the appendix.

2 The Model

We consider an infinite population of mass 1, whose members are indexed by i. Time, indexed by t, is continuous. There exist two messages $m \in \{0, 1\}$ that diffuse simultaneously on the network.

These messages convey information about the state of the world, $\Phi \in \{0,1\}$. Without loss of generality, we assume that the true state of the world, unknown to the agents, is $\Phi = 0$. Hence, we refer to $m = 0$ as the "truth", and $m = 1$ the "rumor".

A link between two agents i and j signifies a meeting between them. The set of meetings can be represented by a communication network. This network is realized independently in every period. Formally, we model the *mean-field approximation* of the system.

Each agent i has k meetings at t, also denoted the degree of the agent, which is constant over time. Agents are classified as being either in state S (*Susceptible*) on in state I (*Infected*). Specifically, agents are in state S if they are unaware of both messages, or if they ignore a message they have received. They transition into state I when they receive a message which they do not ignore. We discuss the concept of ignoring messages in more detail below. Agents in state I die at rate δ and are replaced by identical agents in state S.[2]

We denote by ν the per contact transmission rate of m. This is affected, for example, by communication technology, but also by the number of topics presently being discussed among individuals.[3]

Neither ν nor δ are affected by the type of an agent or the message received.

Agents in state I transmit either the message they themselves received (if they believe it to be correct), or—in case the agent has inspected the message they received—the truth, independently of the message they received. An interpretation of this is that agents cannot transmit their opinion but only some evidence supporting it (possibly wrong). Hence, they cannot forge such

[2]In many scenarios, δ will conceivably be very small, given the speed at which information diffuses, relative to lifetimes. Our model can accommodate arbitrarily small values of $\delta > 0$.

[3]It has been documented in, e.g., Leskovec, Backstrom and Kleinberg (2009) and Weng et al. (2012) that the total number of tweets on Twitter appears roughly constant over time, although the number of topics discussed varies, i.e., memes crowd each other out.

evidence if they did not receive it from one of their social contacts, or they did not acquire it via verification.[4]

We assume that mass $x \in [0,1]$ of the population are of type $b = 0$ and mass $1-x$ are of type $b = 1$. Agents' types matter for whether or not they ignore a message they receive.[5] After receiving message m, an agent of type b believes it in the case that either, *(i)* the message is in line with their type, $m = b$, or *(ii)* the agent has inspected it and therefore understands that the message is correct. In effect, while agents' types restricts the message to which they are susceptible, this restriction is overcome by inspection.[6] We assume that a fraction $\alpha \in (0,1)$ of the population inspects messages. We treat this variable as the main policy instrument through which a planner may affect the diffusion of both the rumor and the truth. It can be influenced through, e.g., affording information literacy space on the educational curriculum, or (digital) campaigns and guides on how to spot misinformation.

We define $\rho^{\alpha}_{b,m,t}$ ($\rho^{1-\alpha}_{b,m,t}$) as the proportion of type b agents who believe message m after (not) having inspected it, for $b \in \{0,1\}$.[7] Note that due to susceptibility to messages, it is the case that $\rho^{\alpha}_{0,1,t} = \rho^{1-\alpha}_{0,1,t} = \rho^{1-\alpha}_{1,0,t} = \rho^{\alpha}_{1,1,t} = 0$. A randomly chosen contact of an agent believes message $m \in \{0,1\}$, at time t with probability $\theta_{m,t}$ given by

$$\theta_{0,t}(\alpha) = x[\alpha\rho^{\alpha}_{0,0,t} + (1-\alpha)\rho^{1-\alpha}_{0,0,t}] + (1-x)\alpha\rho^{\alpha}_{1,0,t}, \quad (1)$$

$$\theta_{1,t}(\alpha) = (1-x)(1-\alpha)\rho^{1-\alpha}_{1,1,t}. \quad (2)$$

$\theta_{0,t}$ and $\theta_{1,t}$ are also the overall truth and rumor prevalence in the population at time t.

[4]See Merlino, Pin and Tabasso (2022) for a model where agents transmit their opinions, i.e., messages that they might have not themselves acquired.

[5]Such groupings according to believes have been shown to exist for a variety of topics online (e.g., Bessi et al. 2016; Samantray and Pin 2019).

[6]Our assumption on beliefs implies that no agent holds conflicting beliefs after being exposed to multiple, conflicting, messages over time.

[7]With a slight abuse of notation we suppress the dependence of the various ρ's on the number of meetings, k, which is the same for all agents.

We assume that the per contact transmission rate, ν, is sufficiently small that an agent in state S becomes aware of message m at rate $k\nu\theta_{m,t}$ through meeting k neighbors, for $m \in \{0,1\}$. This framework allows us to model information diffusion as a set of differential equations:

$$\begin{aligned}
\frac{\partial \rho^{\alpha}_{0,0,t}}{\partial t} &= x\alpha(1-\rho^{\alpha}_{0,0,t})k\nu[\theta_{0,t}+\theta_{1,t}] - x\alpha\rho^{\alpha}_{0,0,t}\delta, && (3)\\
\frac{\partial \rho^{1-\alpha}_{0,0,t}}{\partial t} &= x(1-\alpha)(1-\rho^{1-\alpha}_{0,0,t})k\nu\theta_{0,t} - x(1-\alpha)\rho^{1-\alpha}_{0,0,t}\delta, && (4)\\
\frac{\partial \rho^{\alpha}_{1,0,t}}{\partial t} &= (1-x)\alpha(1-\rho^{\alpha}_{1,0,t})k\nu[\theta_{0,t}+\theta_{1,t}] - (1-x)\alpha\rho^{\alpha}_{1,0,t}\delta, && (5)\\
\frac{\partial \rho^{1-\alpha}_{1,1,t}}{\partial t} &= (1-x)(1-\alpha)(1-\rho^{1-\alpha}_{1,1,t})k\nu\theta_{1,t} - (1-x)(1-\alpha)\rho^{1-\alpha}_{1,1,t}\delta. && (6)
\end{aligned}$$

These expressions keep track of how many agents enter and leave each group at each point in time. Take for example expression (3) for truth prevalence among information literate agents of type 0. The first term describes the mass of these agents that newly believe message $m=0$: they are the proportion of inspecting type $b=0$ agents ($x\alpha$) that did not yet believe message 0 before time t ($1-\rho^{\alpha}_{0,0,t}$); in each period they meet k others, of whom $\theta_{0,t}+\theta_{1,t}$ are in state I, and communicate with them with probability ν. The second (negative) term, indicates that a proportion δ of the agents of this group die. The interpretation of the other expressions is similar.

3 Diffusion of Truth and Rumor

We are interested in the steady state of the system, where the above differential equations are equal to zero. We remove the time subscript t to indicate the steady state value of variables.

Defining the diffusion rate λ as $\lambda = \nu k/\delta$, the conditions for a steady state are

$$\rho_{0,0}^{\alpha} = \rho_{1,0}^{\alpha} = \frac{\lambda[\theta_0 + \theta_1]}{1 + \lambda[\theta_0 + \theta_1]}, \tag{7}$$

$$\rho_{0,0}^{1-\alpha} = \frac{\lambda\theta_0}{1 + \lambda\theta_0}, \tag{8}$$

$$\rho_{1,1}^{1-\alpha} = \frac{\lambda\theta_1}{1 + \lambda\theta_1}. \tag{9}$$

From equations (9) and (2), a positive steady state for θ_1 is given by

$$\theta_1 = (1-\alpha)(1-x) - \frac{1}{\lambda}, \tag{10}$$

which is larger than zero if and only if $\alpha < 1 - 1/[\lambda(1-x)]$. For any $\lambda \geq 0$ there exists a steady state in which $\theta_0 = \theta_1 = 0$. However, by standard arguments (see, e.g., Jackson (2008)), the value of θ_1 in (10) is the unique positive steady state for any value of $\lambda > 0$, and it is globally stable.

Lemma 1 *If $\alpha < 1 - 1/[\lambda(1-x)]$, the rumor exhibits a positive steady state that is decreasing in α and x, and increasing in λ. The prevalence of the rumor is unaffected by the prevalence of the truth.*

Indeed, by (10), the rumor exhibits a positive prevalence only if either a sufficient share of the population is of the type that is biased in its favor, or if a sufficient share of the population fails to inspect messages.

The steady state value of the rumor is not affected by the prevalence of the truth. The steady state of the truth itself, instead, does depend on the diffusion of the rumor. Substituting equations (7) and (8) into the expression for θ_0, we find that a steady state for θ_0 is a fixed point $\theta_0 = H(\theta_0)$ of the following expression:

$$H(\theta_0) = \alpha\frac{\lambda(\theta_0 + \theta_1)}{1 + \lambda(\theta_0 + \theta_1)} + x(1-\alpha)\frac{\lambda\theta_0}{1 + \lambda\theta_0}. \tag{11}$$

The first part of this expression represents the influence of inspecting agents—

for them, receiving either message results in believing that the true state of the world is 0. The second is the additional impact on truth prevalence of those individuals of type 0 who receive $m = 0$. In fact, whenever the rumor does not exhibit a positive prevalence, i.e., whenever $\alpha \geq 1 - 1/[\lambda(1-x)]$, equation (11) simplifies further and the prevalence of the truth is

$$\theta_0 = \alpha(1-x) + x - \frac{1}{\lambda}, \tag{12}$$

which by the same arguments as before is the unique positive steady state of θ_0 (positive whenever $\alpha > 1/(1-x)[1/\lambda - x]$) and globally stable. It is immediate from equation (12) that, once a rumor has been eradicated, further increases in the inspection rate are beneficial as they increase the prevalence of the truthful message. The steady state of θ_0 when both truth and rumor exhibit a positive prevalence is more intricate. The following Proposition highlights its main characteristics.

Proposition 1 *The truth exhibits a unique globally stable steady state θ_0, which is positive, for any x if either of the following conditions is satisfied:*

1. $\alpha > 1/(1-x)[1/\lambda - x]$.
2. $\alpha \in (0, 1 - 1/[\lambda(1-x)])$.

Whenever the conditions imply that the rumor is also endemic, this steady state is strictly increasing in θ_1.

Proposition 1 highlights that rumors may indeed have a positive impact for truth prevalence through either one of two channels. First, whenever $\lambda < 1/(0.5 - x)$, there exists a range of α where the truth *only* survives because some agents heard the rumor and inspected it, thus discovering the truth. Second, the prevalence of the truth is, *ceteris paribus*, increasing in the prevalence of the rumor. Indeed, equation (11) shows that $H(\theta_0)$ is increasing in θ_1 everywhere. Hence, far from hurting the diffusion of the truth, the rumor in fact may benefit its diffusion.

4 Optimal Inspection

The individual and social costs that arise from misinformation is being repeatedly highlighted in public discussions. However, a planner's budget to induce inspection is likely limited. The preceding analysis allows us to address the question of how a policy maker may optimally use a finite budget A to set an optimal inspection rate α. For simplicity, we assume that the unit cost of inducing inspection is one.

In general, a benevolent planner may have different objectives in setting inspection rates. An obvious one is to minimise the diffusion of the rumor. In this case, the optimal policy is straightforward, as shown in the following corollary. Define α' as the optimal inspection rate when the planner wishes to eradicate the rumor. Then,

Corollary 1 *Let a policy maker's aim be to minimise the diffusion of the rumor. They are able to fully eradicate the rumor if and only if* $A \geq 1 - 1/(\lambda(1-x))$*. Whenever* A *is lower than this value, it is optimal for the policy maker to set* $A = \alpha'$*.*

Notably, as communication increases (either through easier technology, or more meetings), a higher rate of message inspection will be necessary to eliminate the rumor.

Another plausible objective for a planner is to maximise the prevalence of the truth. Indeed, in many scenarios that individuals face, such as how to act to minimise the chance of being infected with a disease, it is important to spread the correct guidelines on how to behave optimally. Being aware of the truth might entail taking an action that has positive externalities. For example, being aware that AIDS is a sexually transmitted disease makes it more likely to have protected sexual contacts rather than unprotected ones. Thus, in these contexts, a benevolent planner will have the objective to maximise the diffusion of the truth.

As Proposition 1 suggest, the diffusion of the rumor plays a non-trivial role in the diffusion of the truth. The following Proposition establishes general results about the optimal use of a planner's budget A when setting the optimal inspection rate α^* in order to maximize θ_0.

Proposition 2 *Let the planner have budget A available to set information inspection rates and let their objective be to maximise the prevalence of the truth in the population. Then, there exist a value $\bar{\lambda}$ such that*

i) For all values of λ and x there exists a value $\underline{A}$ such that for all $A < \underline{A}$ it is optimal for the planner to set $\alpha^* = A$.

ii) For all values of λ and x there exists a value $\bar{A}$ such that for all $A > \bar{A}$ it is optimal for the planner to set $\alpha^* = A$.

iii) For $\lambda < \bar{\lambda}$, there exists a range of $\underline{A} \leq A \leq \bar{A}$ such that it is optimal for the planner to set $\alpha^* < A$. In particular, whenever A is just sufficient to fully eradicate the rumor, it may instead be optimal for the planner to allow the rumor to survive.

Proposition 2 establishes that it may indeed be optimal policy for a planner to *not* fully eradicate the rumor, even if that was possible. At the same time, whenever it is optimal to eradicate the rumor, it is also optimal to spend the entire budget on inspecting messages. The case where a planner may purposefully allow the rumor to propagate occurs for certain, relatively low, diffusion rates. For optimal policy, this implies that there may be scenarios where a planner might wish to allow a rumor to propagate at low diffusion rates, but the optimal decision changes to full eradication when the diffusion rates increases.[8]

Figure 1 depicts the diffusion of the truth as a function of the inspection rate in an example when $\lambda < \bar{\lambda}$. In particular, due to the ambiguous role

[8]Note that, as the budget necessary to fully eradicate the rumor depends on the diffusion rate, it may no longer be possible to do so with a higher diffusion rate, even if that were now optimal.

of the rumor in the diffusion of the truth, the diffusion of the truth does not have a monotonic relationship with inspection rates. This has counter-intuitive consequences on the optimal inspection rate. When the planner's budget is very low, all of it should be spent. Similarly, when the budget is sufficiently high, the rumor should be eradicated. However, for intermediate values of the budget, not all of it should be spent. While in general, the optimal inspection rate depends on the parameters of the model, in this example, the inspection rate α should be set at $\underline{A}$ for any budget A such that $\underline{A} < A < \bar{A}$.

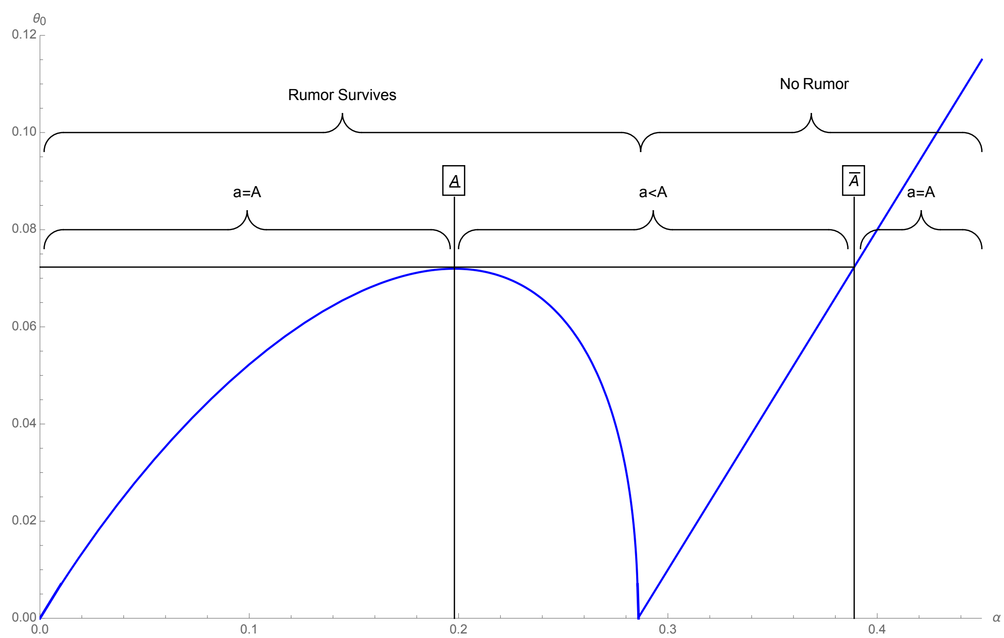

Figure 1: Steady state prevalence of the truth, θ_0, as a function of α, for $\lambda = 2$ and $x = 0.3$.

Overall, our results paint an intricate picture regarding how to optimally set information inspection rates. In particular, the planner will need to be informed of the exact diffusion rate. Note that, as the diffusion rate increases—as might have been the case with the rise of online social networks—there are multiple effects on truth and rumor prevalence. First of all, the prevalence of both messages will, *ceteris paribus*, increase. Second, this raises the budget necessary to eradicate the rumor. Finally, it may lead to a situation where, instead of optimally tolerating the rumor, it now is optimal to eradicate it.

Note that the rise of online social networks may not only have increased communication, but potentially also the number of topics that are discussed. As a result, the overall effect on the diffusion rate of the message the policy-maker is interested in, and that we have described in this model, might well be negative.

5 Extensions

5.1 Targeted Inspection

In our model, we assume that message susceptibility of agents is restricted by their type, and that this restriction is overcome through inspection of messages. From equations (10) and (12), we can see that if no agent was biased towards the rumor, it would die out and the truth would achieve its maximum prevalence. This raises the question whether truth prevalence could be increased by instigating inspection of messages particularly among those agents whose type biases them towards believing the rumor. Online guides on how to spot misinformation, or information literacy campaigns, may be tailored and placed accordingly.

Given a budget A and assuming for simplicity that to induce inspection in both groups has the same unitary cost, the problem is the following:

$$\begin{aligned}
\max \quad & \theta_0 && (13)\\
\text{s.t.} \quad & H(\theta_0) = \theta_0 && (14)\\
& x\alpha_0 + (1-x)\alpha_1 \leq A && (15)\\
& \alpha_0, \alpha_1 \in (0,1). && (16)
\end{aligned}$$

In the following, we constrain ourselves to scenarios where $\lambda > 1/(1-x)$, as otherwise the rumor always dies out, independently of inspection rates. For technical reasons, we also focus on budgets $A < x$; this implies that for positive rumor prevalence it is always optimal for the planner to use all their budget.[9] Under these conditions, we are able to prove existence and uniqueness of a solution to the above problem, as shown in Lemma 2.

Lemma 2 *There always exists a unique solution to problem* (13)*.*

Again, how to optimally allocate the budget across the types of agents depends crucially on the size of budget the planner has at their disposal, as well as on the diffusion rate. We first address the question of whether it is necessarily optimal to eradicate the rumor entirely. Note that a prerequisite is a sufficient budget, in particular, $A \geq 1 - x - 1/\lambda$.

Proposition 3 *Let the planner's budget be sufficient to fully eradicate the rumor if $\alpha_1 = A$. Then, $\alpha_1^* = A$ only when it is optimal to eradicate the rumor. Otherwise, $\alpha_0^* > 0$.*

Proposition 3 reiterates that a planner may, purposefully, allow a rumor to propagate. In particular, it is possible that it is not optimal to eradicate the rumor, but to instead divest resources to the group biased towards the truth. In the proof of Proposition 3 we show that, as in the baseline case of a unique α, this is the case if the diffusion rate is relatively low.

We can further detail how the planner's budget influences the optimal allocation of resources to induce inspection.

Proposition 4 *The optimal inspection rates across groups depend on the available budget as follows:*

i) For all λ, there exists a value $\underline{A}$ such that for all $A < \underline{A}$, it is optimal to set $\alpha_0 = 0$ and induce inspection only in group 1*.*

[9] As this condition implies a budget sufficient to allow all agents of type 0 to inspect messages, we do not perceive this as particularly stringent.

ii) *For all λ, there exists a value $\bar{A}$ such that for all $A > \bar{A}$, it is optimal to set $\alpha_0 = 0$ and induce inspection only in group* 1*.*

iii) *There exist $\tilde{\lambda}$ such that for $\lambda < \tilde{\lambda}$, there exists a range of A with $\underline{A} < A < \bar{A}$ where it is optimal to set $\alpha_0 > 0$.*

This Proposition delivers a similar message to our results when inspection cannot be targeted. In particular, while it is always optimal to focus all available resources on targeting agents biased towards the rumor if the budget is either very low or very large, this strategy may no longer be optimal when the budget takes on some intermediate values.

5.2 Optimal Inspection by Platforms

Instead of being set by a social planner, inspection rates may be influenced by policies that online messaging platforms control. There are, for example, various guides on how to spot mis- and disinformation that are published by online providers. Platforms, however, tend to benefit from the total volume of messages that are exchanged on them, irrespective of their veracity. We therefore question what inspection rates a platform would set whose objective it is to maximise the total volume of both truth- and untruthful messages. In the context of our model, this is equal to total information prevalence in the economy, $\theta = \theta_0 + \theta_1$.

Total prevalence in our model is given by the fixed point that solves

$$H(\theta) = \alpha \frac{\lambda\theta}{1+\lambda\theta} + x(1-\alpha)\frac{\lambda\theta_0}{1+\lambda\theta_0} + \theta_1 \tag{17}$$

Our next Proposition establishes how the optimal inspection rate $\tilde{\alpha}^*$ of the platform relates to the optimal rate the planner would choose, α^*. To allow this comparison, we assume that both planner and platform would have the same budget A available and that they both face a cost of unity of inducing inspection.

Proposition 5 *Let the platform have budget A available to set information inspection rates and let their objective be to maximise overall information prevalence. Then, there exists a value $\tilde{A} > \bar{A}$ such that for all $A \geq \tilde{A}$, $\tilde{\alpha}^* = \alpha^* = A$. For all $A < \tilde{A}$, the optimal inspection rate of the platform is weakly lower than the one of the planner.*

The fact that a player who aims to maximize the prevalence of information overall will in general choose lower inspection rates than one whose objective is only to maximize the prevalence of the truth is unsurprising. It is a direct consequence of the fact that rumor prevalence is linearly decreasing in inspection, an effect which is added to the effect of inspection on truth prevalence. More interesting is the fact that, for high enough budgets, both objectives are met at the same inspection rate. This is due to the fact that, as inspection rates increase, fewer messages are ignored. In fact, in the proof of Proposition 5 we show that overall information prevalence is maximized when all agents inspect messages.

6 Conclusions

In this paper we model how a true and a false message spread in a population of biased agents who become aware of the veracity of messages they receive if they inspect them.

In this framework, we find that the presence of a false message creates truth, in the sense that a larger prevalence of the rumor leads to a larger prevalence of the truth. As a result, it is possible that increased inspection rates, particularly among agents biased towards the rumor, lead to a lower prevalence of the truth. We employ this result to show that a central planner may optimally choose to allow a rumor to perpetuate in the network, even if they have sufficient resources to eradicate it. In fact, we show that the question of how to optimally allocate resources to induce inspection across the two groups is non-trivial, and depends on the exact value of the budget the planner has at their disposal as well as the diffusion rate of messages.

Our results challenge the intuition that making it easier to assess the veracity of information must necessarily be beneficial to society. We purposefully focus on a planner's problem. The introduction of equilibrium considerations and agents' incentives in this framework (Kremer, 1996; Talamàs and Vohra, 2020; Merlino, Pin and Tabasso, 2022) seems a promising avenue of future research.

References

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi. 2018. "Just a few seeds more: value of network information for diffusion." Mimeo.

Bessi, Alessandro, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. "Homophily and polarization in the age of misinformation." *The European Physical Journal Special Topics*, 225(10): 2047–2059.

Bizzarri, Matteo, Fabrizio Panebianco, and Paolo Pin. 2021. "Epidemic dynamics with homophily, vaccination choices, and pseudoscience attitudes." *arXiv preprint arXiv:2007.08523*.

Bloch, Francis, Gabrielle Demange, and Rachel Kranton. 2018. "Rumors and social networks." *International Economic Review*, 59(2): 421–48.

Campbell, Arthur. 2013. "Word-of-mouth communication and percolation in social networks." *The American Economic Review*, 103(6): 2466–2498.

Campbell, Arthur, C. Matthew Leister, and Yves Zenou. 2019. "Social Media and Polarization." Mimeo.

Carroni, Elias, Paolo Pin, and Simone Righi. 2020. "Bring a friend! Privately or Publicly?" *Management Science*, 66(5): 2269–2290.

Chen, Frederick, and Flavio Toxvaerd. 2014. "The economics of vaccination." *Journal of Theoretical Biology*, 363: 105–117.

Galeotti, Andrea, and Brian W Rogers. 2013. "Strategic immunization and group structure." *American Economic Journal: Microeconomics*, 5(2): 1–32.

Goyal, Sanjeev, and Adrien Vigier. 2015. "Interaction, Protection and Epidemics." *Journal of Public Economics*, 125: 64–69.

Jackson, Matthew O. 2008. *Social and Economic Networks.* Princeton University Press.

Kranton, Rachel, and Davis McAdams. 2020. "Social Networks and the Market for News." Mimeo.

Kremer, Michael. 1996. "Integrating behavioral choice into epidemiological models of AIDS." *The Quarterly Journal of Economics*, 111(2): 549–573.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. "Meme-tracking and the dynamics of the news cycle." 497–506, ACM.

Merlino, Luca P., Paolo Pin, and Nicole Tabasso. 2022. "Debunking Rumors in Networks." *American Economic Journal: Microeconomics*, forthcoming.

Sadler, Evan. 2020. "Diffusion games." *American Economic Review*, 110(1): 225–70.

Samantray, Abhishek, and Paolo Pin. 2019. "Credibility of climate change denial in social media." *Palgrave Communications*, 5(1): 1–8.

Tabasso, Nicole. 2019. "Diffusion of Multiple Information: On Information Resilience and the Power of Segregation." *Games and Economic Behavior*, 118: 219–40.

Talamàs, Eduard, and Rakesh Vohra. 2020. "Free and perfectly safe but only partially effective vaccines can harm everyone." *Games and Economic Behavior*, 122: 277 – 289.

Toxvaerd, Flavio. 2019. "Rational disinhibition and externalities in prevention." *International Economic Review*, 60(4): 1737–1755.

Weng, Lilian, Alessandro Flammini, Alessandro Vespignani, and Filippo Menczer. 2012. "Competition among memes in a world with limited attention." *Scientific Reports*, 2.

A Proofs

Proof of Proposition 1. Note first that $H(\theta_0)$ is strictly concave in θ_0, as it is a combination of two strictly concave functions in θ_0. Furthermore, $H(1) < 1$ and, for $\theta_1 > 0$, $H(0) > 0$. This implies that, for $\alpha > 0$, whenever $\theta_1 > 0$, $H(\theta_0)$ crosses from above θ_0 exactly once and there is a unique solution of $H(\theta_0) = \theta_0 > 0$. Alternatively, whenever $\theta_1 = 0$, truth prevalence is given by equation (12). This concludes the proof of Proposition 1. ■

Proof of Proposition 2. First, note that by equation (12) the prevalence of the truth is equal to $\theta_0 = 1 - 1/\lambda$ whenever $\alpha = 1$. This is the steady state prevalence of information in the standard SIS model and the maximum attainable value. This implies, by continuity, that there always exists a value $\bar{A}$, such that it is optimal to set $\alpha = A$ whenever $A > \bar{A}$.

Next, assume that the planner's budget is not sufficiently large to fully eradicate the rumor. Hence, the steady state truth prevalence is given by equation

(11). Defining $G = \theta_0 - H(\theta_0)$, the effect of α on θ_0 is given by

$$\frac{d\theta_0}{d\alpha} = -\frac{-\frac{\partial H}{\partial \alpha}}{1 - \frac{\partial H}{\partial \theta_0}}.$$

As $H(\theta_0)$ is strictly concave in θ_0, we know that at the steady state, $\partial H(\theta_0)/\partial\theta_0 < 1$. Hence, $d\theta_0/d\alpha > 0$ if and only if $\partial H(\theta_0)/\partial\alpha > 0$, where

$$\frac{\partial H(\theta_0)}{\partial \alpha} = \frac{\lambda(\theta_0+\theta_1)}{1+\lambda(\theta_0+\theta_1)} - x\frac{\lambda\theta_0}{1+\lambda\theta_0} - \alpha(1-x)\frac{\lambda}{[1+\lambda(\theta_0+\theta_1)]^2},$$

which is positive if and only if

$$[1+\lambda(\theta_0+\theta_1)]\,[\theta_0(1-x)\,[1+\lambda(\theta_0+\theta_1)]+\theta_1] > \alpha(1-x)(1+\lambda\theta_0). \quad \text{(A-1)}$$

Independent of the value of α, the left-hand side of (A-1) is always positive whenever some information is endemic. However, the right-hand side is zero whenever $\alpha = 0$. Again by continuity, this implies that setting $\alpha = A$ is optimal for low values of A, i.e., for all $A < \underline{A}$.

Finally, assume that the planner's budget was sufficient to just eradicate the rumor, i.e., $A = 1 - 1/\lambda(1-x)$. If the planner set $\alpha = A$, the rumor would be eradicated and the prevalence of the truth would be $\theta_0 = 1 - 2/\lambda$. At $\alpha = A$ and zero rumor prevalence, equation (A-1) simplifies to

$$\theta_0(1+\lambda\theta_0) > A.$$

Given the exact values of A and therefore θ_0, this condition is violated—meaning that θ_0 is decreasing in α—whenever

$$\lambda < 2 + \left[2 - \frac{1}{1-x}\right]^{\frac{1}{2}}.$$

Thus, the truth prevalence would increase if instead of eradicating the rumor, the planner was to set $\alpha < A$ and allow the rumor to survive. This completes the proof. $\blacksquare$

Proof of Lemma 2 Given that it's optimal for the planner to use all its budget whenever $A \leq x$, the planner's problem (13) can be rewritten as

$$\max_{\alpha_0 \in [0,1]} \quad \theta_0 \tag{A-2}$$

$$\text{s.t.} \quad \theta_0^3 \lambda^2 + \theta_0^2 B + \theta_0 C + D = 0. \tag{A-3}$$

where $B = \lambda(1+\lambda-2A\lambda-2\lambda x+2\alpha_0\lambda x)$, $C = \lambda(1-A-x(1-\alpha_0))(1-A\lambda-\lambda x+\alpha_0\lambda x)$ and $D = A(1-\lambda+\lambda A+\lambda x-\alpha_0\lambda x)$. Dividing by λ^2 and after some algebra, the constraint can be rewritten as $f(\alpha_0, \theta_0) = \theta_0^3\lambda^2+\theta_0^2 b+\theta_0 c+d = 0$ with $b = -(2\alpha_1(1-x)+2x-1-1/\lambda)/\lambda^2$, $c = -(1-\alpha_1)(1-x)(\alpha_1(1-x)+x-1/\lambda)/\lambda^2$ and $d = -\theta_1/\lambda$. Note that $f(\alpha_0, \theta_0)$ is continuous in α_0. Furthermore, given that $1 > 0$, b can be either positive or negative and $c, d < 0$, by Descartes' rule of sign, $f(\theta_0) = 0$ admits at most one positive real solution. If the solution is negative, $\theta_0^* = 1$. If it is bigger than one, $\theta_0^* = 1$. This proves existence and uniqueness. $\blacksquare$

Proof of Proposition 3. By the implicit function theorem, it is optimal to increase verification rates of group 1 whenever

$$\frac{\partial H(\theta_0)}{\partial \alpha_1} = (1-x)\left[\frac{\lambda\theta_0}{1+\lambda\theta_0} - A\frac{\lambda k}{[1+\lambda(\theta_0+\theta_1)]^2}\right] > 0, \tag{A-4}$$

i.e., whenever

$$\theta_0\left[1+\lambda(\theta_0+\theta_1)\right]^2 > A(1+\lambda\theta_0). \tag{A-5}$$

To show that it is not necessarily optimal to always fully eradicate the rumor, we consider the limit of equation (A-5) as $A \to 1-x-1/\lambda$, in which case the rumor may be eradicated by setting $\alpha_1 = 1$. In this case, $\theta_1 \to 0$ and

$\theta_0 \to 1 - 2/\lambda$, which implies that (A-5) becomes

$$\lambda + \frac{3}{\lambda} > 4 - x \tag{A-6}$$

and it is optimal to not eradicate the rumor and set $\alpha_0 > 0$ instead whenever this is not satisfied. This happens whenever

$$\lambda \in \left(\frac{4 - x - \left[(4-x)^2 - 12\right]^{1/2}}{2}, \frac{4 - x + \left[(4-x)^2 - 12\right]^{1/2}}{2} \right). \tag{A-7}$$

This proves that it is not always necessarily optimal to eradicate the rumor, even if possible.

To prove the optimality of setting $\alpha_0 = 0$ whenever it is optimal to eradicate the rumor, note that under $\theta_1 = 0$, the prevalence of the truth becomes

$$\theta_0 = x + (1-x)\alpha_1 - \frac{1}{\lambda}.$$

Given that α_0 is irrelevant for θ_0 once the rumor is eradicated, it is obvious that—if eradicating the rumor is the optimal strategy for the planner—it is optimal to set $\alpha_0 = 0$. ■

Proof of Proposition 4. To prove the first point of Proposition 4, we refer to equation (A-5). Whenever the truth exhibits a positive prevalence (which means, whenever either $x > 0$ and/or $\alpha_1 > 0$), the left hand side of this equation is positive. This implies that, for any $x > 0$, we can always find a value A that is small enough for the condition to be satisfied for all values of α_1, and hence setting $\alpha_0 = 0$ will be optimal. Conversely, whenever $x = 0$, the entire population is biased towards the rumor and hence all resources have to be spent on α_1. This completes the proof of the first point.

For the second, note again that, if $A > 1 - x - 1/\lambda$, the rumor can be completely eradicated and α_0 has no effect on the prevalence of the truth anymore. In particular, if $A \geq 1 - x$, all agents of type 1 may verify and the prevalence of the truth becomes $\theta_0 = 1 - 1/\lambda$, the highest possible value it can take. Thus, there always exists a value of A that is large enough to make it optimal for the planner to eradicate the rumor, which automatically implies the optimality of setting $\alpha_0 = 0$.

For the final point, we study how θ_0 changes with α_0 more generally. We need to consider two cases. First, we show that investing in α_0 can lead to a positive prevalence of the truth if it did not survive at $\alpha_0 = 0$. Second, we show that it can also be optimal when the truth would have positive prevalence also with $\alpha_0 = 0$.

To show that α_0 can foster a positive prevalence of the truth, set $A = 1-x-1/\lambda$. By setting $\alpha_0 = 0$, the rumor is eradicated and $\theta_0^* = 1-2/\lambda$. Hence, the truth exhibits positive prevalence if and only if $\lambda \geq 2$. Let instead $\alpha_0 > 0$. Then $\theta_1 = (1-x)(1-\alpha_1) - 1/\lambda > 0$. Now, $\theta_0^* > 0$ whenever $\lambda > 1/(1-x)$. Hence, if $\lambda \in (1/(1-x), 2]$ and $x < .5$, some rumor prevalence benefits the diffusion of the truth.

To show that α_0 can foster the prevalence of the truth also when $\theta_0 > 0$ even when $\alpha_0 = 0$, let us compare the prevalence of the truth and the rumor when $\alpha_1 > 0$ and $\alpha_1 = 0$, i.e., when the planner invests the budget in both groups versus in only the group biases in favor of the rumor.

Consider now the following function:

$$
\begin{aligned}
G(\theta_0) &= H(\theta_0, \alpha_0) - H(\theta_0, \alpha_0 = 0) = \frac{\alpha_1 \lambda \theta_0 (1-x)}{1+\lambda\theta_0} + \frac{A}{1+\lambda\theta_0} - A \\
&+ A\left[-\frac{1}{\lambda(1+\theta_0 - \alpha_1(1-x) - x)} + \frac{1}{\lambda(1-A+\theta_0-x)}\right] \\
&= -\frac{\lambda\theta_0(A - \alpha_1(1-x))}{1+\lambda\theta_0} + \frac{A}{\lambda}\left[\frac{1}{1-A+\theta_0-x} - \frac{1}{1+\theta_0-\alpha_1(1-x)-x}\right]
\end{aligned}
$$

It is easy to see that $G(0) > 0$, $G(1) < 0$ and $dG(\theta_0)/d\theta_0 < 0$ since both the first term and the term in the squared parenthesis are decreasing in θ_0. Notice that the highest is the value θ_0' such that $G(\theta_0') = 0$, the more it is likely that $\alpha_0 > 0$ is optimal, because when $H(\theta_0, \alpha_0)$ for $\alpha_0 > 0$ and $H(\theta_0, \alpha_0 = 0)$

cross at a high value, then it is likely that $H(\theta_0, \alpha_0)$ crosses the 45-degrees line at a higher value of θ_0, implying a higher equilibrium prevalence of the truth when $\alpha_0 > 0$. So, the comparative statics with respect to θ_0' tells us when $\alpha_0 > 0$ is more likely to be optimal.

Since $dG(\theta_0)/d\theta_0 < 0$ and, by the implicit function theorem,

$$\frac{\partial \theta_0'}{\partial \lambda} = -\frac{dG(\theta_0)/d\lambda}{dG(\theta_0)/d\theta_0},$$

we just need to study $dG(\theta_0)/d\lambda$. Then:

$$\frac{dG(\theta_0)}{d\lambda} = -\frac{\theta_0(A - \alpha_1(1-x))}{(1+\lambda\theta_0)^2} - \frac{A}{\lambda^2} < 0$$

Hence, $\alpha_0 > 0$ is more likely to be optimal the lower λ. Regarding the effect of A, we find that:

$$\frac{dG(\theta_0)}{dA} = \frac{G(\theta_0)}{A} - \frac{\alpha_1\lambda\theta_0(1-x)}{A(1+\lambda\theta_0)} - \frac{1}{(1-A+\theta_0-x)^2}$$

Since at θ_0', $G(\theta_0') = 0$, at this point $dG(\theta_0')/dA < 0$. To further study the effect of α_0 on θ_0, note that by the implicit function theorem,

$$\frac{\partial \theta_0'}{\partial \alpha_0} = -\frac{dF(\theta_0)/d\alpha_0}{dF(\theta_0)/d\theta_0},$$

where $F(\theta_0) = \theta_0^3\lambda^2 + \theta_0^2 B + \theta_0 C + D$ as defined in (A-2). Since $dF(\theta_0)/d\theta_0 > 0$ at θ_0^* as F crosses the x-axis from below, α_0 increases θ_0 if $F(\theta_0)/d\alpha_0 < 0$.

Then,

$$\begin{aligned}
\frac{dF(\theta_0)}{d\alpha_0} &= \frac{1}{\lambda^2}\left[2x\theta_0^2 + \theta_0\frac{x}{\lambda}(1+\lambda-2A\lambda-2(1-\alpha_0)\lambda x) - \frac{Ax}{\lambda}\right] \\
&= \lambda^2 x\left[2\theta_0^2 - \theta_0\left(x+(2x-1)(1-\alpha_1)-\frac{1}{\lambda}\right) - \frac{A\alpha_0}{\lambda}\right] \\
&= -\theta_0\left[x(\alpha_1(1-x)+x-\frac{1}{\lambda}) - (1-\alpha_1)(1-x)x\right] \\
&= -\theta_0\left[(2\alpha_1-1)(1-x)x + x - \frac{1}{\lambda}\right].
\end{aligned}$$

Hence, when $dF(\theta_0)/d\alpha_0$ is decreasing in x, $dF(\theta_0)/d\alpha_0 < 0$ is more likely the higher x.

This concludes the proof of Proposition 4. ■

Proof of Proposition 5. From the facts that $\theta = \theta_0 + \theta_1$ and that θ_1 is decreasing in α linearly, it is immediate that maximization of θ requires a lower α than maximization of θ_0 whenever $\theta_1 > 0$. At the same time, note that the fixed point of equation (17) can be written as

$$\theta = \alpha\frac{\lambda\theta}{1+\lambda\theta} + (1-\alpha)\left[x\frac{\lambda\theta_0}{1+\lambda\theta_0} + (1-x)\frac{\lambda\theta_1}{1+\lambda\theta_1}\right], \tag{A-8}$$

and also note that, if $\alpha = 1$, total information prevalence would be identical to the one in the standard SIS model, i.e., $\theta = 1 - 1/\lambda$. We can see from equation (A-8) that, for all other values of α, θ will be lower than this. By continuity of θ, there must exist a budget $\tilde{A}$ such that above it, it is optimal for the platform to set inspection rates equal to the budget. By the fact that $\theta = \theta_0$ when the rumor dies out, this choice is identical for the platform and the planner. Finally, as $\theta > \theta_0$ whenever $\theta_1 > 0$, it must be the case that $\tilde{A} > \bar{A}$. This concludes the proof of Proposition 5. ■