Misinformation due to asymmetric information sharing[*]

Berno Buechel[♯†], Stefan Klößner[♭], Fanyuan Meng[‡], Anis Nassar[♯]

[♯]Department of Economics, University of Fribourg, 1700 Fribourg, Switzerland

[♭]Faculty of Educational and Social Sciences, University of Vechta, 49377 Vechta, Germany

[‡]Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland

March 10, 2023

Abstract

On social media platforms, true and false information compete. Importantly, some messages travel much further than others, even if they concern the same topic. This fact is not reflected in models of social learning (or opinion formation) in networks. Our model fills this gap by allowing different types of information to have different decay factors and to be shared with different networks of people, incorporating asymmetries in sharing behaviors. More "shareable" information then dominates in the long run. This yields a substantial probability of misinformation, in contrast to the special case of symmetry covered by the literature. Asymptotic learning requires a perfect balance between two types of asymmetry: the product of decay factor and largest eigenvalue in the respective signal sharing networks must coincide. Approaching this balance reduces the speed of convergence and enables social learning in the shorter term. Our analysis thus suggests that policy makers, who do not know the true state, aim to mitigate asymmetries in signal sharing, e.g. by weakening echo chambers or by fostering the shareability of cumbersome, boring messages.

Keywords: Misinformation, asymmetry, social networks, social learning, opinion dynamics, echo chambers

JEL Classification Codes: D83, D85

[*] We thank Andras Borsos, Herbert Dawid, Andreas Fahr, Ben Golub, Matthew Jackson, Giacomo Livan, Antoine Mandel, and the audiences of the Seventh Annual Conference on Network Science and Economics at the University of Chicago, the virtually-held Networks 2021 (joint Sunbelt and NetSci Conference), as well as the seminar participants at Bielefeld University, Central European University (CEU in Vienna), Corvinus University (Budapest), University of Fribourg, and at University Paris I – Pantheon-Sorbonne for helpful comments.

[†] Corresponding author. Email address: berno.buechel@unifr.ch, web: www.berno.info. Declarations of interest (all authors): none.

"If it doesn't spread, it's dead." (e.g. henryjenkins.org)

1 Introduction

Misinformation is thriving. This is problematic because it can lead people to make poor decisions and hence generate various socially harmful outcomes, such as lower vaccination coverage or dangerous political developments (Burki, 2019; Pennycook and Rand, 2021). While there is an ongoing public and scientific debate about how prevalent misinformation really is and to which extent it translates into decision making, there seems to be broad consensus that misinformation exists and should be curbed (Lazer et al., 2018; European Commission, 2018; Grinberg et al., 2019; Greene and Murphy, 2021). To evaluate adequate policies in the fight against misinformation, the way it spreads has to be properly understood. However, there are significant gaps between theoretical models on how people learn from each other and empirical studies on how information spreads.

Models of social learning, which analyze the formation of beliefs on a given issue in social networks, implicitly assume that all signals are shared in the same manner (see, Golub and Sadler, 2016; Grabisch and Rusinowska, 2020 for two surveys). In contrast, empirical studies of information diffusion on social media platforms provide ample evidence that some messages travel further and are shared in different parts of the network than other messages (e.g. Cheng et al., 2014; Goel et al., 2016; Del Vicario et al., 2016; Vosoughi et al., 2018; Johnson et al., 2020). The reason for such differences in "infectiousness" or *shareability* seems to be underlying characteristics concerning the form (text, image, video, ...) and content (emotionally loaded, surprising, disgusting, boring, ...) of a message, similar to the study of "newsworthiness" in communication studies. Importantly, differences in shareability also occur on the same issue, as it has been shown, e.g. regarding vaccination (Johnson et al., 2020). This is highly plausible if we consider that signals supporting and questioning a certain claim may differ in form and certainly differ in content, as e.g. messages supporting the claim that vaccines cause autism may create a different emotional response than messages questioning it.

The goal of this paper is to study how such asymmetries in signal sharing affect social learning and misinformation. We particularly consider two types of asymmetry when agents share *positive*, i.e. supporting, and *negative*, i.e. questioning, signals on a given issue: decay asymmetry and network asymmetry. First, positive and negative signals need not be shared to the same extent (*decay asymmetry*). This is particularly motivated by studies documenting that fact-checked false information travels significantly further than fact-checked true information (Vosoughi et al., 2018; Juul and Ugander, 2021). Second, positive and negative signals need not be shared with the same people (*network asymmetry*). This is particularly motivated by studies documenting that some type of information is shared more or less heavily in different parts of a given network, e.g. some people are sharing more rumors while others are sharing more scientific papers (Del Vicario et al., 2016; Zollo et al., 2017; Johnson et al., 2020). Decay asymmetry and network asymmetry embody differences in shareability.

We propose the first model of social learning that admits such differences. It builds on the social learning literature, where agents repeatedly share binary signals. Consider an original network that consists of nodes that are interpreted as agents and links that are interpreted as possible communication channels between these agents. The (sub)set of links which is used to share positive signals on a given issue is captured by a symmetric adjacency matrix and called the *positive signal sharing network*; likewise, we define the *negative signal sharing network* as the (sub)set of links which is used to share negative signals). Network asymmetry means technically that these two networks do not coincide, which occurs, e.g. if two agents share positive signals while they do not share negative signals on this issue. Moreover, we introduce decay asymmetry by two decay factors, one for positive signals and one for negative signals, which admits that the two types of signals may face different decay when shared. With this model, we investigate how network and decay asymmetry drive the probability of misinformation. A particular focus is how the results depend on network size, specifically, whether the probability of misinformation converges to zero when the network grows large (*asymptotic learning*).

We first address the benchmark scenario that both asymmetries are absent. In this scenario of symmetry, the probability of misinformation is bounded. Moreover, it converges to zero with increasing network size under a mild condition. The condition is well-known from the DeGroot model and means that there are no agents with excessive influence (Golub and Jackson, 2010). Our analysis of the general case then reveals that the symmetry benchmark is a special case. We establish a threshold condition that determines which mix of signals will be held in the long run. The condition compares the product of decay factor and largest eigenvalue between the positive signal sharing network and the negative signal sharing network. Asymptotic learning can only occur when the threshold condition is exactly met. Even slight asymmetry is sufficient to cause one signal type to fully dominate in the long run (as long as there is at least one of each signal initially), reminding of the saying “If it doesn’t spread, it’s dead.” As a consequence, the long-run probability of misinformation is substantial, even in large networks. The probability of misinformation essentially equals the ex ante probability that the state with the more shareable signals turns out to be true. We further show that agents can be ordered according to their ratio of eigenvector centralities in the two different networks. Those who lean more to the more shareable signal type, i.e. those who are relatively more central in the network that shares this signal type, are more prone to be misinformed. We then extend the model to allow agents to have pair-specific relationships, which accommodates heterogeneous subgroups, directed networks and idiosyncratic learning; and show that the results still hold.

While our results provide reasons for misinformation to thrive, they also indicate how misinformation could be mitigated. When asymmetries are mild, speed of convergence is low. As a consequence, information aggregation does take place and the probability of misinformation remains low in the short and medium term (before the more shareable signal starts to dominate). To mitigate network asymmetry, additional links in the network where the less shareable signal is shared help as they always increase its eigenvalue. However, for a given number of links, smaller denser groups have a (much) stronger effect on the largest eigenvalue than sparser larger groups: the echo chamber effect. Conversely, we show by numerical examples that the required decay asymmetry favoring one type of signal needed to compensate for the dominance of the other type of signal gets disproportionately large with the existence of such echo chambers. Our results provide a new perspective on measures against misinformation. They suggest as general policy implications that asymmetries in shareability should be identified and mitigated. Specifically, this might mean to hamper the spread of sensational, catchy messages, e.g. by breaking the corresponding echo chambers, and to foster the spread of cumbersome, boring messages, e.g. by making them more attractive to share.

The remainder of the paper is structured as follows. Section 2 relates it to the literature. We introduce the model in Section 3 and study the special case of symmetry as a benchmark in Section 4. Section 5 presents the main results. We illustrate our results in Section 6 and provide the extension in Section 7. We conclude by a discussion in Section 8.

2 Related Literature

Our paper belongs to the literature on social learning (or opinion dynamics), where agents repeatedly learn from their neighbors in a social network. If agents were fully Bayesian, then perfect information aggregation would occur in any connected network (DeMarzo et al., 2003; Mueller-Frank, 2013). However, Bayesian rationality is very demanding when it comes to learning in a network structure.[1] The classic model of repeated linear updating (DeGroot, 1974) has been studied intensively and it has been extended in many interesting directions (Friedkin and Johnsen, 1990; DeMarzo et al., 2003; Golub and Jackson, 2010, 2012; Buechel et al., 2015; Grabisch et al., 2019; Banerjee et al., 2019). A main focus is whether the long-run consensus opinion optimally aggregates the initially dispersed pieces of information. This is satisfied, at least asymptotically, if there are no individuals with excessive influence on the others (Golub and Jackson, 2010).[2] Our main contribution is to qualify this conclusion by additional conditions: there must be a balance between different asymmetries (embodied by the products of decay factor and largest eigenvalue which must coincide for both networks) and that an agent's ratio of centralities between the two networks must be moderate compared to the ratio of centrality concentration in the two networks. Both additional conditions happen to be satisfied under symmetry, an implicit assumption that is ubiquitous in the theoretical literature, but hard to justify empirically.

To our knowledge, none of the existing models accommodates asymmetric treatment of signals. Symmetric treatment of signals is related to so-called "label neutrality," which is an underlying assumption of the models in the literature and in fact a characterizing feature of the DeGroot model (Molavi et al., 2018). We contribute, to our best knowledge, the first model of social learning that relaxes label neutrality. We do so by addressing network or decay asymmetry in signal sharing. As in the DeGroot model positive and negative signals are mingled into continuous opinions, it is unclear to us how asymmetries could be meaningfully introduced. Instead, we base our model on the baseline model introduced in Sikder et al. (2020), which has the same properties as the DeGroot model in the special case of symmetry, but keeps track of all signals at all times.

As we show, the consequences are drastic as either decay or network asymmetry favoring one type of signal can be sufficient for misinformation. That is, not only is the society's long-run belief a suboptimal aggregation of the initial signals, but all members of the society are likely misinformed and hence would make the wrong decision with high probability. Similarly drastic conclusions are known in the literature only in the presence

[1] As experimental studies reveal, actual behaviors are less often consistent with Bayesian learning than they are with simpler updating rules such as repeated averaging (Corazzini et al., 2012; Friedkin and Bullo, 2017; Grimm and Mengel, 2020; Chandrasekhar et al., 2020).

[2] Other sufficient conditions are that there is at least one agent who is perfectly Bayesian (Mueller-Frank, 2014); and that agents additionally receive a private signal in *each* period and treat this signal in a Bayesian manner (Jadbabaie et al., 2012).

More generally, there are (relatively weak) conditions on non-Bayesian learning which are sufficient for learning the true underlying state in the long run (Molavi et al., 2018).

of forceful (or biased or stubborn) agents who, at some point, do not learn at all but still heavily influence the other agents (Acemoglu et al., 2010; Grabisch et al., 2018; Azzimonti and Fernandes, 2018; Rusinowska and Taalaibekova, 2019; Della Lena, 2019; Sikder et al., 2020). The policy implications that can be drawn from these models rather suggest acting on forceful agents and on fighting disinformation. Complementary to these insights, we argue that the way signals are shared is crucial for avoiding misinformation. Even in the absence of forceful agents, misinformation is likely and policy makers could act on signals' shareability to fight it.

Sharing behavior and signal accumulation in our model is closest to Sikder et al. (2020), in which agents repeatedly share binary signals in a rather naïve way but with Bayesian-style updating. These authors show that the introduction of confirmation bias leads to polarization. While Sikder et al. (2020) mostly focus on regular graphs, we first show for any connected network that misinformation in this baseline model is bounded and converges to zero under mild conditions. We then generalize their baseline model to any network structure and by introducing decay and network asymmetries depending on the type of information shared. Our model is also related to Fernandes (2019) and a model variation of Taalaibekova (2020, Chapter 3.3) who both consider agents who perceive or receive more positive or more negative signals without being aware of it. Despite biased reception or perception of signals, these agents still treat different signals in the same way, while our agents share positive and negative information asymmetrically.[3]

The justification to consider asymmetries stems from various strands of empirical literature. Most importantly, empirical studies of information diffusion on social media platforms including Facebook, LinkedIn, and Twitter document that the spread of information items depends on the type of information (Cheng et al., 2014; Goel et al., 2016; Vosoughi et al., 2018; Del Vicario et al., 2016; Johnson et al., 2020). The information type may not only concern its form (news, videos, pictures, ...) but also its content (e.g. emotional content, surprise, disgust, fear, ...), which both make some messages more "infectious" than others (Juul and Ugander, 2021). Moreover, there is a long history in communication studies investigating information value to determine which factors influence how "newsworthy" a certain news item is (e.g. Harcup and O'Neill, 2017). In particular, news stories are more attractive if they are surprising, or contain particularly good or bad news. Marketing studies have come to the same conclusions regarding the importance of emotional content, may it be positive or negative, for a post to "go viral" (e.g. Berger and Milkman, 2012). Finally, in the framework of the spread of competing

[3] Moreover, decay in a framework of naïve learning, has been considered recently by Grabisch et al. (2021) who consider decay of social influence, while we consider decay of information.

viruses, the rate at which people become infected and at which they recover determines which virus will dominate, a result that is also motivated by the spread of information (e.g. Prakash et al., 2012).

All of these strands of empirical literature show that the extent to which some piece of information is shared depends on its properties that are summarized as "infectiousness", "newsworthiness," "virality", or simply *shareability.* Moreover, they suggest that shareability may well differ between messages that concern the same topic, which is the motivation of our model.

3 A Model of Asymmetric Signal Sharing

Ingredients. There are n agents $N = \{1, ..., n\}$ who talk about a binary issue. One classic example is whether vaccines cause autism; new examples are popping up every day. To model uncertainty, nature draws the true state with a commonly known prior probability and then each agent receives a signal. Specifically, the true state $\theta \in \{0, 1\}$ is drawn with prior probability $b^+ = P(\theta = 1) \in (0, 1)$. As a convention, we call state $\theta = 1$ the *positive* state and $\theta = 0$ the *negative* state. Each agent i independently receives signal $s_i \in \{0, 1\}$ which matches the true state with probability ρ, i.e. $P(s_i = 1|\theta = 1) = P(s_i = 0|\theta = 0) = \rho \in (\frac{1}{2}, 1)$. Conditional on the state, the signals are independent. Continuing with our convention, we call signal 1, i.e. a signal that indicates state $\theta = 1$, a *positive* signal and 0 a *negative* signal. These two types of signals may differ in relevant characteristics, e.g. information indicating that vaccine causes autism may be more or less emotionally loaded than information indicating the opposite state.

Time is discrete $t = 0, 1, 2, ...$ At time $t = 0$, the initial signals are received. At each time step $t > 0$, agents communicate with their neighbors in a social network, as specified below. We model this communication activity in a way that admits asymmetric treatment of positive and negative signals since sharing behavior between these two types of messages might differ.

When a signal is passed on from one agent to the next, it decays by $\delta^+ \in (0, 1]$ if it is positive and by $\delta^- \in (0, 1]$ if it is negative. Decay can be either (i) due to senders who only share a fraction δ^+ of their signals, (ii) due to the signal's ability to travel through the communication channel where a fraction $1 - \delta^+$ gets lost, or (iii) due to the recipients who discount the received signals by δ^+.[4] If, for any reason, $\delta^+ \neq \delta^-$, there is *decay*

[4] In Online Appendix C.2, we show that all three interpretations can be explicitly modeled and are captured in a reduced form in our model with one parameter δ^+. All Online Appendices are permanently available here: www.berno.info under the name MisInfo_SupplementaryOnlineMaterial.pdf. In an ex-

asymmetry.

Positive signals are shared in a network (N, A^+), negative signals in a network (N, A^-), where A^+ and A^- are both symmetric $n \times n$ matrices with entries 0 or 1, i.e. adjacency matrices representing each an undirected unweighted network. We assume that both these networks are connected, which implies that the matrices are irreducible. As an interpretation, consider one single original network consisting of all possible communication channels between the fixed set of agents. Then the (sub)set of links which are used to share positive signals are represented by A^+ and the (sub)set of links which are used to share negative signals are represented by A^-. If, for any reason, $A^+ \neq A^-$, there is *network asymmetry.*

Signal Accumulation. Given these ingredients, we can now formulate how signals are shared and accumulated. Let $N_i^+(t)$ denote number of positive signals of node i at time t and let $N^+(t)$ be the $(n \times 1)$-vector. The law of motion for positive signals is

$$N^+(t) = (I + \delta^+ A^+) N^+(t-1). \tag{1}$$

Hence, the number of agent i's positive signals in a given period is the number of positive signals i held in the previous period plus the number of positive signals that i's neighbors in network A^+ held in the previous period discounted by δ^+. For the negative signals the law of motion and the notation is fully analogous. $N^-(t)$ is the vector of negative signals and the law of motion is $N^-(t) = (I + \delta^- A^-) N^-(t-1)$. Let $N^+(0)$ and $N^-(0)$ be the vectors of initial signals, i.e. $N^+(0) = s$ and $N^-(0) = 1 - s$. Given the initial signals and the law of motion, we can compute the number of signals of any agent at any time. The signal accumulation for positive signals is

$$N^+(t) = (I + \delta^+ A^+)^t s. \tag{2}$$

Technically, the entries in the matrix $(I + \delta^+ A^+)^t$ can be considered as the number of walks of length t or smaller, when the network A^+ is augmented by self-loops (ones on the diagonal); thereby each walk is discounted by δ^+ at any step, except when using a self-loop. A less technical interpretation is that agents share all the positive signals that they have with their neighbors in the positive signal sharing network; and whenever signals are passed on only a fraction δ^+ fully arrives at the recipient. And analogously for negative signals.

tension of the model, we will study pair-specific decay factors, which admit heterogeneity in behavior of senders, communication channels, or recipients (Section 7).

Signal Mixes and Misinformation. An agent's *signal mix* is the fraction of positive signals that i holds, i.e.

$$x_i(t) := \frac{N_i^+(t)}{N_i^+(t) + N_i^-(t)}, \tag{3}$$

and the vector $x(t)$ contains all agents' signal mixes at time t.

To measure misinformation, we assess whether an agent has received more false signals than true signals, where a signal is called true if it indicates the true state and false if it indicates the other state. Hence, we consider an agent i to be *misinformed* if her signal mix satisfies $x_i(t) < 0.5$ when the true state is $\theta = 1$ or if $x_i(t) > 0.5$ when the true state is $\theta = 0$. Agents with $x_i(t) = 0.5$ are considered as misinformed with probability 0.5. While it is thinkable that an agent who is misinformed according to this definition still holds a belief that is accurate, this is implausible for prior belief $b^+ = 0.5$ and, for any prior belief, it becomes less and less plausible over time. When $x_i(t)$ converges to a value not equal to 0.5, then the difference between positive and negative signals grows arbitrarily large as more and more signals are acquired. Hence, an agent who is misinformed according to our definition, is prone to hold an inaccurate belief and to make a poor decision. We define $p_i^{\text{Mis}}(t)$ as agent $i's$ probability of misinformation at time t from an ex ante perspective, i.e. before the true state and the initial signals are realized. Initially, at $t = 0$, an agent is misinformed with probability $p_i^{\text{Mis}}(0) = 1 - \rho < 0.5$, as ρ is the probability that the received signal matches the true state. The long-run probability of misinformation is denoted by $p_i^{\text{Mis}}(\infty) := \lim_{t\to\infty} p_i^{\text{Mis}}(t)$ and will be characterized in the next sections.

Besides analyzing networks of given size n, we also study how the probability of misinformation depends on the network size. For this purpose, we consider sequences of growing networks ($n \to \infty$), where size-dependent quantities are denoted by an additional index n, e.g. adjacency matrices are denoted by A_n^+ and A_n^- and the long-run probabilities of misinformation are denoted by $p_{i,n}^{\text{Mis}}(\infty)$. At each element of such a sequence, the model as defined above applies, while the prior b^+ and the signal quality ρ stay unchanged and the realization of true state and signals is independent across networks.

4 Benchmark: Misinformation under Symmetry

To assess the effects of decay and network asymmetry, a natural benchmark is, of course, the special case of *symmetry*: $\delta^+ = \delta^- =: \delta$ and $A^+ = A^- =: A$. This is a generalization of the baseline model introduced in Sikder et al. (2020), which our model nests for $\delta = 1$. In the original article by Sikder et al. (2020), the focus is on regular networks (N, A), that are characterized by each node having the same degree.[5] Regular networks happen to minimize the probability of misinformation, as our Proposition 1 below implies. More importantly, this result shows that the probability of misinformation is always bounded under symmetry. Let λ_1 be the largest eigenvalue of A.[6] Let $c = (c_1, ..., c_n)^\top$ be its corresponding eigenvector, normalized such that its components sum to one, i.e. $Ac = \lambda_1 c$ and $\sum_i c_i = 1$.[7] Entries in the eigenvector c are a measure of *eigenvector centrality* (Bonacich, 1972; Friedkin, 1991).

Proposition 1 (Symmetry)**.** *Under symmetry, the long-run signal mix is a convex combination of the initial signals s_j with weights according to eigenvector centrality, i.e. for all i, $\lim_{t\to\infty} x_i(t) = \sum_{j=1}^{n} c_j s_j$. Therefore, the probability of long-run misinformation $p_i^{\text{Mis}}(\infty)$ is bounded from above by 0.5. Moreover, if for some sequence of growing networks, indexed by network size n, we have $\lim_{n\to\infty} \max_{j=1,\ldots,n} c_{j,n} = 0$, then the probability of long-run misinformation converges to zero, i.e. for all i, $\lim_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$.*

The proofs of this proposition and of all other results are collected in Appendix A. Proposition 1 means that misinformation under symmetry only occurs due to an unlucky distribution of signals. It depends on the relative influence of each agent on the long-run opinions, as measured by her entry in the largest eigenvector. In the best case, all agents, who are by assumption equally well-informed, are equally influential.[8] This is satisfied in regular graphs, in which by definition every agent has the same degree. Then the long-run signal mix of every agent exactly reflects the initial signal distribution, i.e. $\lim_{t\to\infty} x_i(t) = \frac{1}{n}\sum_{j=1}^{n} s_j$, which is just the mean of the initial signals. Hence, under symmetry and when the network is regular, naïve agents who accumulate the same signals over and over again can fare equally well as Bayesians would (as we discuss in Example C.1 in Online Appendix C.1). In the worst case, there is a group of agents who are overly influential (as discussed in Example C.2 in Online Appendix C.1).

The probability of misinformation in these two extreme cases is illustrated in Figure 1, with the lowest and highest line ("Best" and "Worst"). In addition, the figure reports

[5] Sikder et al. (2020) find polarization in an extension of their model, in which they introduce agents with confirmation bias.

[6] More precisely, λ_1 is the largest positive eigenvalue of A and other eigenvalues of A might exist which are as large in absolute value as λ_1. For ease of reading, we usually omit 'positive' when addressing largest positive eigenvalues.

[7] Recall that by assumption the adjacency matrix is symmetric. Hence, there is no need yet to distinguish between left-hand and right-hand eigenvectors.

[8] More generally, the requirement is a balance between idiosyncratic signal precision and social influence (see, e.g. Buechel et al., 2015).

the probability of misinformation from simulations with two classes of random networks: Erdös-Rényi (ER), in which the probability of every link is fixed (p), and Barabasi-Albert (BA) where in every step of constructing the network, m new links are created. The thickness of the two lines represents the variation of misinformation that covers 50% of all simulation runs (from 25th to 75th percentile). Selecting parameters (p and m) such that the two random networks have the same expected density, it is expected that the probability of misinformation tends to be higher in the BA random networks since their degree distribution is more skewed than that in ER random graphs. The simulations confirm this. Moreover, Figure 1 illustrates that misinformation for the two classes of random graphs is closer to the regular graph's. In particular, we observe that the probability of misinformation decreases with network size n. Relatedly, Proposition 1, provides a formal condition for asymptotic learning: In a sequence of growing networks, the probability of long-run misinformation vanishes if the largest eigenvector centrality converges to zero.

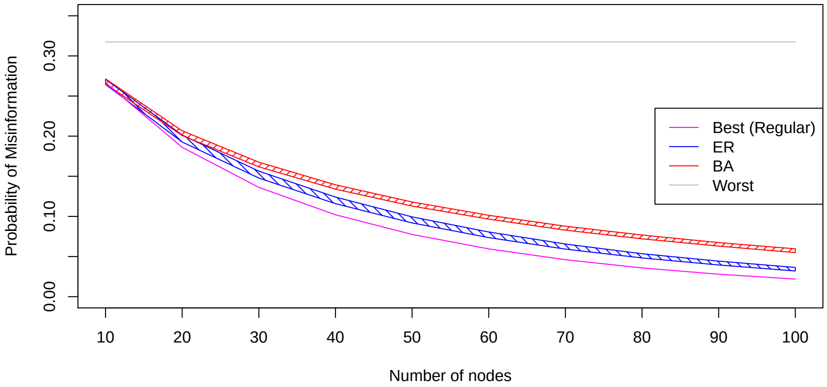

Figure 1: Misinformation under symmetry: comparing different network structures.

Notes: Signal precision $\rho = 0.6$. Number of nodes $n = 10, ..., 100$ on the x-axis. Random graph parameters are set such that asymptotic average degree is 6 in these simulation runs. 1,000 simulation runs per class of random network of a given size. Thickness of corresponding line represents variation covering 50% of all outcomes (from 25th to 75th percentile).

The results in this section closely resemble those of the common DeGroot model of naïve learning. In Online Appendix C.4, we describe the similarities and differences of these two models and their results in some detail. Importantly, in the DeGroot model positive and negative signals are mingled into continuous opinions, while our model of naïve learning keeps track of all signals. As a consequence, we can use our model to study asymmetries in signal sharing, as we do next.

5 Results

5.1 Key Result

We now study how misinformation in the long run is affected by asymmetric treatment of signals. Decay asymmetry is captured simply by its two parameters δ^+ and δ^-. For network asymmetry, the adjacency spectrum of the two matrices A^+ and A^- matters. Analogously to before, let λ_1^+ be the largest eigenvalue of A^+. Let $c^+ = (c_1^+, ..., c_n^+)^\top$ be its corresponding eigenvector, normalized such that its components sum to one, i.e. $A^+c^+ = \lambda_1^+ c^+$ and $\sum_i c_i^+ = 1$. And likewise λ_1^- and c^- are the largest eigenvalue and its normalized eigenvector of A^-, i.e. $A^-c^- = \lambda_1^- c^-$ and $\sum_i c_i^- = 1$. Entries in the eigenvectors c^+, c^- are the respective eigenvector centrality.

Proposition 2 (Key Result)**.** *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal.*

1. *If $\delta^+\lambda_1^+ < \delta^-\lambda_1^-$, then for all i and large t:*

$$x_i(t) \approx \frac{c_i^+}{c_i^-}\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t \frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2} \frac{\sum_{j=1}^n c_j^+ s_j}{1-\sum_{j=1}^n c_j^- s_j}$$

such that $\lim_{t\to\infty} x_i(t) = 0$.

2. *If $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$, then for all i and large t:*

$$x_i(t) \approx 1 - \frac{c_i^-}{c_i^+}\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^t \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}$$

such that $\lim_{t\to\infty} x_i(t) = 1$.

3. *If $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$, then for all i:*

$$\lim_{t\to\infty} x_i(t) = \frac{1}{1 + \frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}} \in (0,1).$$

The proposition states that the combination of decay factor and largest eigenvalue, $\delta^+\lambda_1^+ \lesseqgtr \delta^-\lambda_1^-$, determine which signals dominate in the long run, given that there is at least one of each signals. Intuitively, the condition for Case 1, requires that the decay factor for positive signals δ^+ is low compared to the decay factor with negative signals δ^-, which means that positive signals are shared to a lower extent; and that the eigenvalue λ_1^+ of the positive signal sharing network A^+ is low compared to λ_1^-, which has the interpretation that the agents are generally better connected in the negative signal sharing network A^-. In the knife-edge case, Case 3, decay and network asymmetry compensate each other: $\frac{\delta^+}{\delta^-} = \frac{\lambda_1^-}{\lambda_1^+}$; and the distribution of initial signals matters for the long-run signal mixes. This case also nests the symmetry benchmark. While in the symmetric benchmark the common decay factor δ did not affect the long-run opinions, decay factors δ^+ and δ^- are crucial for the case distinction under asymmetry.

In Case 1, there are four factors that determine asymptotic behavior:

$$x_i(t) \approx \underbrace{\frac{c_i^+}{c_i^-}}_{\text{centrality ratio}} \cdot \underbrace{\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t}_{\text{exponential decay}} \cdot \underbrace{\frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2}}_{\text{concentration ratio}} \cdot \underbrace{\frac{\sum_{j=1}^n c_j^+ s_j}{1-\sum_{j=1}^n c_j^- s_j}}_{\text{signal averages}} \qquad (4)$$

The first is agent-specific, the three latter factors are common for all agents in a given society. The first factor shows that agent i's characteristics enter her asymptotic signal mix $x_i(t)$ via her ratio of eigenvector centrality in the positive signal sharing network over her centrality in the negative signal sharing network: $\frac{c_i^+}{c_i^-}$. We will refer to this as i's *centrality ratio.* The second factor shows the exponential decay process, which depends on the (information) decay factors δ^+ and δ^-, as well as on the largest eigenvalues of the two networks. The next factor is the inverse ratio of the sums of squared centralities. This can be considered as the relative concentration (or centralization) of the networks. Indeed, the larger the sum of squared centralities, the more centrality is concentrated on a few agents, while this sum is minimal in regular networks (where every agent's centrality is equal to $\frac{1}{n}$). The last factor is determined by weighted averages of the initial signals, whereas the weights are the eigenvector centralities. The centrality ratio will determine opinion diversity, the exponential decay will determine speed of convergence, the weighted signal averages will determine the levels of the signal mixes. For Case 2, the factor decomposition is analogous. In Case 3, opinion diversity is also determined by the centrality ratio, whereas misinformation depends on the concentration ratio, as further discussed below.

5.2 Implications

Through the key result, we can derive implications regarding the probability of misinformation, the diversity of opinions, and the speed of convergence.

Probability of Misinformation. Trivially, when all signals happen to be correct, which occurs with probability ρ^n, there is never misinformation, while when all signals happen to be false, which occurs with probability $(1-\rho)^n$, there is always misinformation. All remaining cases are covered by Proposition 2. As a consequence, we receive the following main result.

Corollary 1 (Probability of Misinformation)**.** *1. If $\delta^+\lambda_1^+ < \delta^-\lambda_1^-$, then each agent i's probability of long-run misinformation is*

$$p_i^{\text{Mis}}(\infty) = (1-b^+)(1-\rho)^n + b^+(1-\rho^n),$$

which is essentially b^+ for large networks.

2. If $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$, then each agent i's probability of long-run misinformation is

$$p_i^{\text{Mis}}(\infty) = (1-b^+)(1-\rho^n) + b^+(1-\rho)^n,$$

which is essentially $1-b^+$ for large networks.

3. If $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$,

then an agent i's probability of long-run misinformation is bounded by

$$p_i^{\text{Mis}}(\infty) \leq \max\{(1-b^+)(1-\rho)^n+b^+(1-\rho^n), (1-b^+)(1-\rho^n)+b^+(1-\rho)^n\} < \max\{b^+, 1-b^+\}.$$

The intuition behind the first two parts of Corollary 1 is simple. Consider Case 1. According to Proposition 2, negative signals dominate in the long run given that there is at least one negative signal initially. Hence, the probability of long-run misinformation equals the the probability that the positive state is realized while there is at least one negative signal, $b^+(1-\rho^n)$, plus the probability that the negative state is realized while there are only positive signals, $(1-b^+)(1-\rho)^n$. For growing network size n, this probability converges (quickly) to b^+. That is, when negative signals are more shareable (Case 1), then the long-run probability of misinformation is essentially the probability that the true state is the positive state.

For equal prior $b^+ = 0.5$, a setting that many models in the literature focus on, the probability of misinformation is essentially 0.5 in both Case 1 and Case 2. Hence, there is substantial misinformation independent of whether positive or negative signals are more shareable. The reason is that, despite the initially informative distribution of signals, one of the two signals will fully dominate in the long run (“if it doesn’t spread, it’s dead”), while both states are equally likely. For $b^+ > 0.5$ ($b^+ < 0.5$) misinformation is even more likely in Case 1 (in Case 2). The interpretation is that the positive state has a higher ex ante probability of being true ($b^+ > 0.5$), while negative signals have a higher shareability (Case 1). Such a scenario can naturally arise, e.g. because signals indicating the ex ante less likely state are more *sensational* and hence tend to be shared more and decay less. The opposite scenario is also well thinkable: It could be that the ex ante more likely state is indicated by signals which are more *plausible* or for some other reason are shared more and decay less. However, even then the probability of misinformation is converging to $b^+ > 0$ or $1 - b^+ > 0$ and hence does not converge to zero for large networks.

Concerning Case 3, Corollary 1 provides an upper bound for the probability of misinformation, showing that it is always weakly lower than in the worse case of Case 1 or Case 2.[9]

We now study the conditions under which the probability of misinformation converges to zero for large networks. Consider a sequence of growing networks, indexed by network size n, as defined in the last paragraph of Section 3. Proposition 3 provides a necessary condition and a set of sufficient conditions for asymptotic learning.

Proposition 3 (Asymptotic Learning)**.** *Consider a sequence of growing networks, indexed by network size n.*

1. *If some agent i's probability of long-run misinformation converges to zero ($\lim\limits_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$), then the products of decay factor and largest eigenvalue of the two networks coincide from a certain point in time on, i.e. there is a natural number n^* such that $\delta_n^+\lambda_1(A_n^+) = \delta_n^-\lambda_1(A_n^-)$ for all $n > n^*$.*

2. *For a fixed agent i the probability of long-run misinformation converges to zero ($\lim\limits_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$) if the following three conditions are all satisfied:*

   (i) The products of decay factor and largest eigenvalue of the two networks coincide from a certain point in time on, i.e. there is a natural number n^* such that $\delta_n^+\lambda_1(A_n^+) = \delta_n^-\lambda_1(A_n^-)$ for all $n > n^*$.

[9] Indeed, there are networks where agents in Case 3 are equally prone to misinformation as in Case 1 (Case 2) because they always converge to a signal mix below (above) 0.5 given that there is at least one signal of each type signal initially.

(ii) Maximal eigenvector centrality in both networks converges to zero, i.e. $\lim_{n\to\infty} \max_{j=1,\ldots,n} c^+_{j,n} = 0$ *and* $\lim_{n\to\infty} \max_{j=1,\ldots,n} c^-_{j,n} = 0$.

(iii) The agent's centrality ratio over the two networks' concentration ratio has all accumulation points in the open interval $\left(\frac{1-\rho}{\rho}, \frac{\rho}{1-\rho}\right)$*, i.e. there is a positive real number* $\varepsilon > 0$ *and an integer* n^{**} *such that for all* $n \geq n^{**}$*,*

$$\gamma_{i,n} := \frac{\frac{c^+_{i,n}}{c^-_{i,n}}}{\frac{\sum_{k=1}^n (c^+_{k,n})^2}{\sum_{k=1}^n (c^-_{k,n})^2}} \in \left[\frac{1-\rho}{\rho} + \varepsilon, \frac{\rho}{1-\rho} - \varepsilon\right]. \tag{5}$$

In its first part Proposition 3 shows that asymptotic learning necessarily requires that the condition of the knife-edge case, Case 3, is always satisfied after some point. Otherwise, we would have infinitely many occurrences of Case 1 or Case 2, which would imply that the sequence $p^{\text{Mis}}_{i,n}(\infty)$ would not converge to zero. In other words: If the two types of asymmetry are not in perfect balance, then the probability of misinformation does not vanish for any agent. The second part of Proposition 3 provides sufficient conditions on top of this first requirement: (ii) all agents' influence is vanishing; and (iii) the agent's centrality ratio, in relation to the networks' concentration ratio, is in a moderate range. To understand the last condition, consider an agent i with low centrality ratio (relative to the concentration ratio), $\gamma_{i,n} \leq 1$, which means that this agent leans more towards negative signals than others. Now, if $\gamma_{i,n} < \frac{1-\rho}{\rho} < 1$, her tendency to negative signals is not moderate and she will behave like all agents in Case 1, letting negative signals dominate, and be misinformed with probability b^+ in large networks. If, in contrast, her tendency is moderate, $\frac{1-\rho}{\rho} < \gamma_{i,n} < 1$, and the two other conditions are satisfied, she will incorporate signals of both kinds and eventually learn the true state with high probability. Analogously for $\gamma_{i,n} > 1$. The condition on moderate centrality ratios is less demanding for large ρ and it is always satisfied in the special case of symmetry, which implies $\gamma_{i,n} = 1$; and the condition (i), that we are in Case 3, in the first place.

While long-run misinformation is bounded in the symmetric benchmark and converges to zero for most networks when they become large (Section 4), introducing even slight asymmetry is sufficient to induce substantial levels of long-run misinformation even for large networks, as Corollary 1 and Proposition 3 in this subsection show.

Opinion Diversity. Proposition 2 already indicates that individual differences between agent's beliefs are driven by their centrality ratio $\frac{c^+_i}{c^-_i}$. The following corollary of it establishes this relation, considering two agents' ratios of positive over negative signals,

$\frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)}$.[10] Clearly, if an agent's ratio of positive over negative signals is above another agent's ratio, then her signal mix is higher, i.e. closer to the positive state.

Corollary 2 (Centrality Ratios and Opinion Diversity)**.** *Suppose that the initial distribution of signals contains at least one positive and at least one negative signal. Then the ratio of two agents' ratios of positive over negative signals converges to these agents' ratio of centrality ratios, i.e.*

$$\lim_{t\to\infty} \frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)} = \frac{c_i^+}{c_i^-} \Big/ \frac{c_j^+}{c_j^-}. \tag{6}$$

Hence, an agent i with higher centrality ratio than another agent j has a higher asymptotic signal mix, i.e. if $\frac{c_i^+}{c_i^-} > \frac{c_j^+}{c_j^-}$ (or equivalently, $\gamma_i > \gamma_j$ when using the definition from Equation 5), then for large t, $x_i(t) > x_j(t)$.

Corollary 2 applies to all three cases of Proposition 2. It means that, even if signal mixes converge to consensus in Cases 1 and 2, they are ordered by their centrality ratios. In Case 3, this order also holds and opinion diversity even persists in the limit, as Example 1 below will show.

Minimal opinion diversity is given in all networks where $\frac{c_i^+}{c_i^-}$ is constant across agents, which holds in particular, if $A^+ = A^-$. Corollary 2 then implies that all agents approach the same signal mix and opinions converge to consensus. We have seen this in the symmetry benchmark (Section 4), where the positive and negative signal sharing network coincide by assumption. In contrast, strong opinion diversity emerges in networks where the centrality ratios differ strongly across agents. For example, consider two star networks (N, A^+) and (N, A^-) with different agents at the center. The ratio of one center is $\frac{c_i^+}{c_i^-} = \frac{\sqrt{n-1}}{1}$, the ratio of the other center $\frac{c_j^+}{c_j^-} = \frac{1}{\sqrt{n-1}}$. Hence, by Corollary 2, $\frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)}$ converges to $n-1$. That is, agent i has asymptotically $n-1$ times more positive signals over negative signals than agent j. Again, this holds in all of the three cases, even if the differences in signal mix vanish for large t.

With respect to misinformation, Corollary 2 first implies that if, for instance, the true state is positive ($\theta = 1$) and some agent i is misinformed in the long run $x_i(\infty) < 0.5$, then all agents with lower centrality ratios must be as well. Moreover, agents with relatively low centrality in the positive signals network, i.e. $\gamma_i < 1$, are more prone to be misinformed in the long run if $b^+ >> 0.5$. Comparative-statics further show that i's probability of misinformation is then weakly increasing in b^+: as she leans towards negative signals,

[10] Notice that agent i's ratio of positive over negative signals, $\frac{N_i^+(t)}{N_i^-(t)}$, can equivalently be written as $\frac{x_i(t)}{1-x_i(t)}$.

she is more often misinformed when the positive state is ex ante more likely.[11] The implications are analogous for agents with a relatively high centrality in the positive signals network. Finally, the order of the signal mixes will have consequences for the probability of misinformation in the shorter term as well.

Speed of Convergence. The motivation for studying misinformation is that agents make decisions, which may be based on inaccurate information. If the point of decision making is not far in the future, then the short or medium term opinion dynamics matter. Speed of convergence can also be measured with the help of Proposition 2. For instance, in Case 1, the speed of convergence is governed by the speed that the exponential decay factor $\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t$ converges to 0 (see also Eq. (4)). Looking for the half-life, as it is standard for exponential decay processes, we define $t_{1/2}$ as the number of periods it takes for this quantity to fall to one half of its initial value. Thus, we obtain half-life in Cases 1 and 2 as another Corollary of Proposition 2.

Corollary 3 (Speed of Convergence)**.** *Suppose that $\delta^+\lambda_1^+ \neq \delta^-\lambda_1^-$. Then half-life is*

$$t_{1/2} = \frac{\log(0.5)}{\log(\tau)}, \qquad \textit{with } \tau := \frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}. \tag{7}$$

Half-life will be large when $\delta^+\lambda_1^+$ and $\delta^-\lambda_1^-$ are close to each other, i.e. when we are close to Case 3. In some sense, Case 3 can be considered as unlikely because it is a special case of the parameter space. Still, it is important to study this case for at least two reasons. First, in the absence of asymmetry – the special case that has been studied in the literature – we are in Case 3, as it was discussed in Section 4. Second, under certain conditions, opinions in the short and medium term are well approximated by the analysis of Case 3, even if this does not hold in the long run.[12]

Concerning misinformation, this might mean that the low levels of misinformation, which are typical for Case 3, can be held for a while before the unfavorable long-run properties of Cases 1 and 2 unfold.[13] An illustration helps.

[11] It is possible that the probability of misinformation is increasing (decreasing) in b^+ for all i. Typically for Case 3, however, it will be increasing in b^+ for some agents and decreasing in b^+ for others, as some agents' centrality ratio will be above the concentration ratio and the other agents' will be below.

[12] Conditions are discussed in Online Appendix C.3.

[13] Moreover, it seems to be a rule of thumb that this is particularly true for those with γ_i tending to the signals which will not dominate in the long run: agents with moderate or moderately small γ_i in Case 2 and agents with moderate or moderately large γ_i in Case 1. The reasons is that, by Corollary 2, these agents' signal mixes tend to be less extreme than those of the agents who already lean toward the state that has the more shareable signals.

6 Illustration

6.1 Illustration of Key Result and Implications

Example 1 illustrates Proposition 2 and Cororallies 1-3.

Example 1. *Consider five agents $N = \{1, 2, 3, 4, 5\}$. The positive signal sharing network (N, A^+) is the complete network, i.e. $a^+_{ij} = 1$ for all $i \neq j$. The negative signal sharing network (N, A^-) is a star network with Agent 1 at the center, i.e. $a^-_{1j} = a^-_{j1} = 1$ for all $j = 2, 3, 4, 5$ and $a^-_{ij} = 0$ else.*[14] *We fix the decay factor of negative signals to be $\delta^- = 0.8$, whereas we vary the decay factor of the positive signals.*

Proposition 2 distinguishes three cases based on the condition $\delta^+\lambda^+_1 \lesseqgtr \delta^-\lambda^-_1$. In the complete network, we have $\lambda^+_1 = n-1 = 4$. In the star network, we have $\lambda^-_1 = \sqrt{n-1} = 2$. Hence, we are in Case 3 of Proposition 2 if $\delta^+ = 0.4$, in Case 1 if $\delta^+ < 0.4$, and in Case 2 if $\delta^+ > 0.4$ (indeed, $\delta^+\lambda^+_1 = 0.4 \cdot 4 = 0.8 \cdot 2 = \delta^-\lambda^-_1$ yields Case 3). The corresponding dynamics of signal mixes is illustrated in Figure 2 for the signal distribution $s = (0, 0, 1, 1, 1)$ and for values of parameter δ^+ close to the equality threshold of 0.4. The first four panels belong to Case 1 of Proposition 2, where signal mixes converge to 0. The last four panels belong to Case 2, where signal mixes converge to 1. The middle panel induces Case 3 by setting $\delta^+ = 0.4$. Interestingly, in that case ($\delta^+ = 0.4$) Agent 1, the center of the star network, receives more negative signals than positive signals as $x_1(\infty) < 0.5$, which does not hold for the other agents as $x_i(\infty) > 0.5$ for $i = 2, 3, 4, 5$.

To illustrate opinion diversity, established in Corollary 2, we compute eigenvector centralities: $c^+_{1,2,3,4,5} = \frac{1}{5}$, $c^-_1 = \frac{1}{3}$ and $c^-_{2,3,4,5} = \frac{1}{6}$. The centrality ratios $\frac{c^+_i}{c^-_i}$ for the four agents are hence $(\frac{3}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5})$. In words, the first agent has a smaller relative importance in the positive network relative to the negative network than all other agents. The reason is that this agent, as the center of the star network, is better connected in the negative signal sharing network than the others, while all agents are equally well-connected in the positive signal sharing network, the complete network. By Corollary 2, this implies that this agent's signal mix will asymptotically always be closer to 0 than those of the other agents: $x_1(t) < x_i(t)$, for $i = 2, 3, 4, 5$. In all panels of Figure 2 this can be observed as Agent 1's signal mix is below the others' signal mix, even if they converge to 1 or 0. Finally, observe that, in line with Corollary 3, convergence is slower the closer the parameter δ^+ to the critical value that induces Case 3.

[14] For illustration purpose, one can think of five colleagues who discuss the rumour that their company has been sold to a competitor. All of them share information that speak against this theory while they exchange information supporting this claim with only one person, Agent 1.

Let us now consider misinformation, as established by Corollary 1. For this purpose we relax the assumption of a specific signal realization and return to the ex ante perspective. Figure 3 shows the probability of misinformation over time for the setting $b^+ = 0.6$ (i.e. the positive state is slightly more likely than the negative state) and information quality $\rho = 0.8$. In Case 3 the probability of misinformation decreases quickly and remains on the low level. This is due to learning of each other's signals with an aggregation that is balanced. In contrast, in Case 1 the probability of misinformation will settle on $(1 - b^+)(1 - \rho)^n + b^+(1 - \rho^n)$, which is generally close to b^+ and only reduced by the possibility that all initial signals are correct (which happens here relatively frequently as $n = 5$ and $\rho = 0.8$). Likewise in Case 2, the probability of misinformation will settle below $1 - b^+$, which is better than in Case 1 because the positive state, which is reached in Case 2, has a higher ex ante probability: $b^+ = 0.6$. In both Case 1 and Case 2, the long-run probability of misinformation is substantial. Importantly, even in Case 1 and 2, there is a low level of misinformation that is quickly reached. How long this can be sustained, then depends on the extent of asymmetry: the more moderate the asymmetry, i.e. the closer we are to the threshold (which is here obtained at $\delta^+ = 0.4$), the longer the low level of misinformation can be sustained. Moreover, we can observe that Agent 1, who tends to negative signals, tends to more often misinformed than the other agents in Case 1, where negative signals dominate eventually, but not in Case 2, where positive signals dominate eventually.

Since Example 1 is restricted to $n = 5$, we finally look at some larger networks.

6.2 Decay Asymmetry vs. Network Asymmetry

To illustrate how additional links in the positive signal sharing network A^+ can compensate for decay favoring negative signals (or inversely), we use the following construction. We generate a graph with 250 random links on $n = 100$ agents.[15] We keep it and call it A if it is connected. The network of negative signals is then fixed to $A^- = A$. The network of positive signals A^+ also consists of all links in A, but has α^+ additional links. Because, by construction, A^- is a subgraph of A^+ (i.e. for each entry, $a^+_{ij} \geq a^-_{ij}$, and for some entry $a^+_{ij} > a^-_{ij}$) and both are connected, we have $\lambda^+_1 > \lambda^-_1$. There is hence network asymmetry favoring positive signals. We also introduce decay asymmetry, this time favoring negative signals, by fixing $\delta^- = 1$ and setting $\delta^+ \in [0, 1)$.

While decay asymmetry is immediate, the intensity of network asymmetry depends

[15] These networks are like Erdős–Rényi random graphs with average degree of five.

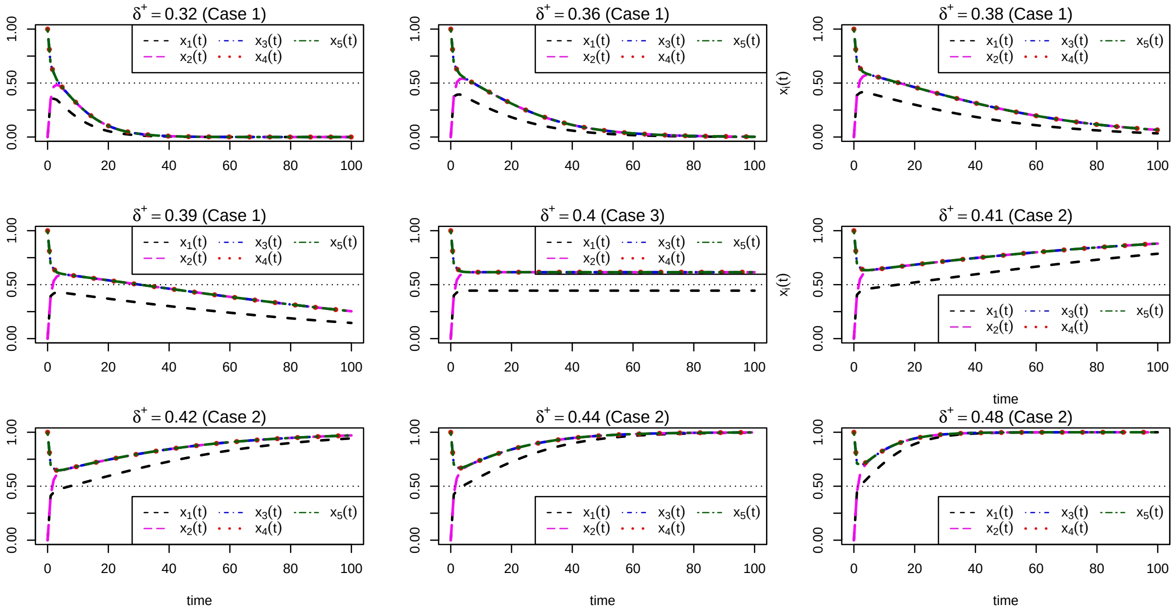

Figure 2: Illustration of results.

Notes: Dynamics of signal mixes in Example 1 over time. Agent 1, illustrated by the black dashed line, is the center in the negative signal sharing network. The first four panels with $\delta^+ < 0.4$ belong to Case 1 of Proposition 2. The middle panel with $\delta^+ = 0.4$ yields Case 3. The last four panels with $\delta^+ > 0.4$ belong to Case 2.

not only on how many positive links are added (α^+), but also on how they are added. We compare two link generation processes: random and echo chamber. In the random link generation process, we add the α^+ additional links at random. In the echo chamber process, each additional link is created in order to generate an increasingly large echo chamber. The first additional link is generated between two random nodes, the second between these two nodes and a third. The third link will either increase the connectedness of the first three nodes, or, if all links are already present, create a link between a fourth node and the first three, and so on until α^+ links are added to the positive signal sharing network.

In Figure 4, we show the ratio of decay factors needed to exactly compensate network asymmetry. Network asymmetry is determined by a given ratio of links between A^+ and A^-, using the two different link generation processes. In the lower left (upper right) of the parameter space we are in Case 1 (Case 2) where negative (positive) signals eventually dominate. The purple and black line show the points in the parameter space where the threshold condition is met with equality. Near these lines misinformation can be

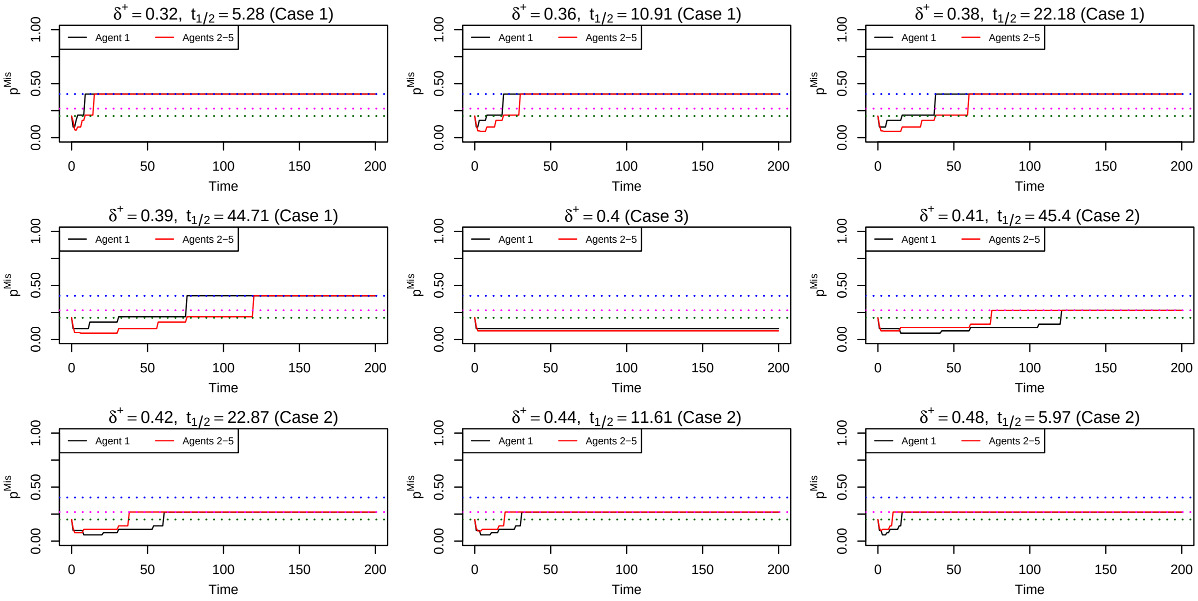

Figure 3: Probability of Misinformation in Example 1.

Notes: Probability misinformation in Example 1 over time. Agent 1 is the center in the negative signal sharing network. Agent 2-5 are in structurally equivalent positions. In each panel, the dotted lines indicate the levels of the probability of misinformation initially and in the long run in Cases 1 and 2.

small in the short and medium term, while far away from this lines the probability of misinformation is substantial, apart from the very first periods.

For example, if there are 20% more links in the positive network than in the negative network and all links are created at random, then positive signals must face a decay of roughly 0.8 to compensate it, as $\frac{\delta+}{\delta-} \approx 0.8$ at the position where the bright line is at 1.2. If, however, the 20% additional positive links are arranged in an echo chamber, then it takes twice the decay for negative signals compared with positive signals to achieve a balanced sharing, as $\frac{\delta+}{\delta-} \approx 0.5$ at the position where the black line is at 1.2. Hence, network asymmetry produced by echo chambers causes misinformation and is hard to compensate.

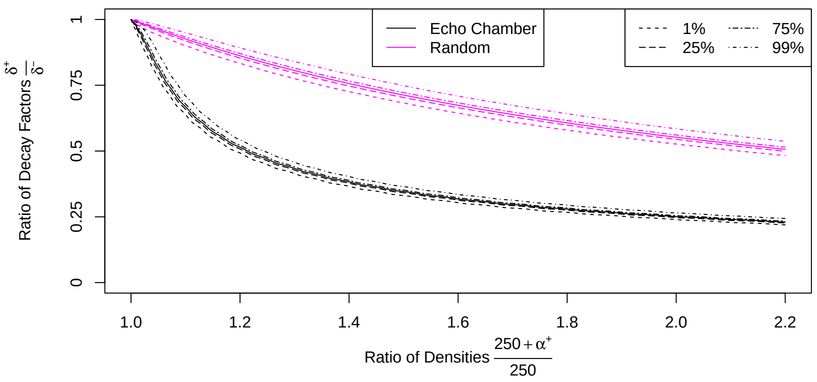

Figure 4: Compensating decay asymmetry with network asymmetry.

Notes: Results based on 10'000 replications per setting. Each line represents a different additional link generation process. The lines draw the median ratios of decay asymmetry that exactly compensate the ratio of links between the A^+ and A^- networks, generating Case 3 of Proposition 2. For parameters at the lower left (upper right) of a given line negative (positive) signals dominate. Parameters close to a given line foster information aggregation in the short and medium term.

7 Extension: Heterogeneous Relations

In this section, we extend the model by defining decay factors that are pair-specific. Formally, we consider two weighted graphs (N, M^+) and (N, M^-), where M^+ and M^- are $n \times n$ matrices with entries $m_{ij}^+ = \delta_{ij}^+ \in (0,1]$ or $m_{ij}^+ = 0$ and $m_{ij}^- = \delta_{ij}^- \in (0,1]$ or $m_{ij}^- = 0$. Like in the baseline model, a positive entry $\delta_{ij}^+ > 0$ is the decay of information for the pair ij, i.e. the fraction of positive signals that agent i receives when communicating with agent j. Signal accumulation becomes $N^+(t) = (I+M^+)N^+(t-1)$ for positive signals and likewise $N^-(t) = (I+M^-)N^-(t-1)$ for negative signals. We extend the assumption of connectedness by assuming here that both networks are strongly connected. In all other aspects, the extended model is defined as the baseline model, introduced in Section 3.

This extension makes the model more flexible since it can accommodate many forms of heterogeneity. First, it allows for every pair i, j to have a different quality of communication and hence a pair-specific decay factor. Second, networks can be directed such that some agent i receives signals from another agent j, while j does not receive signals of i, i.e. $m_{ij}^+ = \delta_{ij} > 0$ and $m_{ji}^+ = 0$. Third, it allows an agent i to face a stronger decay as a receiver of information, i.e. $\delta_{ij} < \delta_{kj}$ for all linked j, k. This can incorporate in particular differences in discounting of received signals. Fourth, it allows for some agent i to face a stronger decay than others as a sender of information, i.e. $\delta_{ji} < \delta_{jk}$ for and all linked j, k. This could incorporate in particular differences in which a fraction of signals is shared. These variations can be applied at the individual level to single agents, or more broadly to groups of agents.

Even though this extension carries a rich array of new applications, our results neatly generalize, as we show in Online Appendix B. A noteworthy difference is that left and right eigenvectors do not coincide. Hence, we define right-hand eigenvector centrality c^+, c^-, that is originally defined (Bonacich, 1987) and left-hand eigenvector centrality d^+, d^- that determines social influence in the DeGroot model (Golub and Sadler, 2016). Interestingly, we have $\lim_{t\to\infty} \frac{N_i^+(t)}{\sum_j N_j^+(t)} \to c_i^+$ and $\lim_{t\to\infty} \frac{N_i^-(t)}{\sum_j N_j^-(t)} \to c_i^-$, showing that the right-hand eigenvectors capture which fraction of all positive, respectively negative, signals in the society is asymptotically held by agent i. In contrast, the left-hand eigenvector d_i^+ expresses which fraction of on average asymptotically held signals originated from agent i's initial signal. In other words: the right eigenvectors c^+, c^- determine the relations between the asymptotic signal mixes of agents, while the left eigenvectors d^+, d^- determine the effect of agents' initial signals on the asymptotic signal mix, as Extended Proposition 2 and Extended Corollary 2 in the Online Appendix B demonstrate.

The generalization shows that it is not essential that the network is undirected or that discounting is similar for different agents, while symmetry or asymmetry between positive and negative networks is still crucial.

8 Discussion

8.1 Policy Implications

Our main result imply that the fight against misinformation is challenging. Slight asymmetry is sufficient for substantial misinformation in the long run.[16] Countervailing measures that do not reach the threshold condition, $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$, eventually lead to the same result. Countervailing measures that go beyond the threshold condition lead to substantial misinformation as well. Achieving a perfect balance between sharing of negative and

[16] "Substantial" means that it does not converge to zero for large networks. The best case with substantial misinformation is the scenario where the state which has higher ex ante probability, say the positive state ($b^+ > 0.5$), also has more shareable signals ($\delta^+\lambda_1^+ > \delta^-\lambda_1^-$). Then the probability of long-run misinformation is still $1 - b^+$ in large networks.

positive signals (i.e. hitting the threshold $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$) seems unrealistically hard.[17] Our analysis, however, also shows that it is still worthwhile to strive for such a balance. When asymmetries are reduced, $\delta^+\lambda_1^+ \approx \delta^-\lambda_1^-$, speed of convergence to one dominant signal is slow, which enables information aggregation and low levels of misinformation in the short and medium term. In the fight against misinformation hence the first step is to identify asymmetry of the positive and negative messages regarding the issue. Consider an issue for which positive signals are less shareable than negative signals, i.e. $\delta^+\lambda_1^+ << \delta^-\lambda_1^-$, because, say, they have a higher emotional content, are more sensational, are less boring, are more plausible, come in a simpler form, or for some other reason. Then we can tackle misinformation by moderating the asymmetry. Potential counter-measures address the form of the message, the platforms' feed algorithms, and the literacy of users.

Concerning the form, an example of particular importance are scientific publications. Most of them still exist only in article form and their complexity regularly constitutes a high barrier for the general public. In the meantime, concurring information is created in ways that is simple and easily shareable online via social media or messaging apps, such as images, videos and news-like posts. Providers of scientific information are thus not always fighting with the appropriate tools and should consider how to tailor their message in a way that allows them to keep its informative integrity, while being easily shareable. This could improve sharing of positive signals in both dimensions, reducing decay ("$\delta^+ \uparrow$") and increasing the number of people it is shared with ("$\lambda^+ \uparrow$"), in order to moderate the stipulated asymmetry above (that positive signals were less shareable than negative signals).

Through feed algorithms, social media platforms strongly affect which information we are confronted with. Such feed algorithms could reduce the push of messages that are sensational, simple etc. and already shared heavily, but instead, more often select messages that are rather boring, complicated, etc., and hence have not yet been shared heavily. Users would then be confronted more often with opposing information and might later than earlier commit to a one-sided view of the issue. However, such policies have to be implemented and monitored with care as they can backfire.[18] More easily, feed algorithms that support echo chambers where only one type of information shared can be altered to reduce network asymmetry.

[17] And even if this balance was achieved, there are additional conditions for vanishing misinformation in the limit: moderate centrality ratios and vanishing influence, see Proposition 3.

[18] Users who have been exposed to debunking information regarding conspiracy theories have been observed to consequently increase the intensity of their engagement with conspiracy posts (Zollo et al., 2017).

Finally, to fight asymmetric sharing of signals, people's sharing behavior can be affected by giving nudges and by education. Experiments find that simply asking people to pause and reflect before sharing may decrease their intention to share false information (Fazio, 2020). Stressing the importance of accuracy can have a similar effect (Pennycook et al., 2021). Educating people to higher information literacy, which emphasizes the ability to navigate and locate information, to recognize opinion pieces or to search databases (Livingstone et al., 2008), can make them less prone to misinformation (Jones-Jang et al., 2021). Interestingly, a recent study finds that increasing acceptance of information from credible news sources can have a much higher impact than reducing acceptance of information from dubious news sources (Acerbi et al., 2022).

These suggestions for how to counter-act misinformation, are all from the perspective of a policy maker that does not know the true state. If it would, the same measures could be used to promote the correct signals and to hamper the spread of the false signals. Of course, another approach is to increase the accuracy of information and to fact-check messages, given there is such a technology. In our model this would result in a higher signal quality ρ, which is slightly reducing misinformation, but qualitatively as problematic as long as there are asymmetries.

8.2 Limitations

Separating decay and signal type. Our model assumes decay and networks to be fully dependent on signal type. Agents share all positive, and respectively negative, signals to the same extent and with the same people. More realistically, decay and signal sharing networks would be idiosyncratic to the signal and only partly correlated to signal type. For instance, negative signals on some issue might generally be more shareable (and hence decay less and be shared with more people), but some positive signals might also be highly shareable.

Extending our model to allow for idiosyncratic decay factors and signal sharing networks would generalize our main condition $\delta^+\lambda_1^+ \lesseqgtr \delta^-\lambda_1^-$ to considering not the common decay factors and networks, but the maximally realized product of decay factor and eigenvalue in this equation, i.e. the highest product among all positive signals and the highest product among all negative signals. While gaining realism, this extension would not fundamentally change our analysis and conclusions.

More Sophisticated Sharing Behavior. In our model, agents are assumed to be naïve and the way they share signals is mechanic. Recently, alternative approaches to model the sequential sharing of messages by more sophisticated agents have been introduced (e.g. Papanastasiou, 2020; Acemoglu et al., 2022). Incorporating asymmetry into such models is left for future research.

Creation of Information. We study how given information on an issue is used without studying the creation of information. First, knowledge is the product of scientific efforts of many institutions, which will add new information into the system. This is not captured in our model, but it is a reason why we emphasize knowledge usage in the short and medium term, as in the long term, new information arrives. Second, messages are produced by various institutions (companies, media outlets, governments, private persons,...) with various goals. Both aspects, arrival of new information and existence of forceful agents, have already been incorporated in the previous literature on social learning.

9 Conclusion

We have investigated asymmetric sharing of positive (supporting) and negative (questioning) information as a source of misinformation. Our results established that differences in the decay factor and in the connectedness, as measured by the largest eigenvalues, determine the long-run state a society reaches. Even slight asymmetries lead to substantial levels of misinformation in the long run. Though it is clear from our discussion that there is no silver bullet in the fight against misinformation, we can conclude that countermeasures should consider signal-dependent sharing behaviors and moderate asymmetries in signal sharing. This aspect has not yet been studied in the (theoretical) literature on social learning, but may deserve more attention in the future.

A Appendix: Proofs

This appendix contains the proofs of all results stated in the main text.[19]

A.1 Proof of Proposition 1

To prove Proposition 1, we apply Proposition 2, which is proven below. Using the result of Case 3 of Proposition 2 and $c^+ = c^- = c$, we find that

$$\begin{aligned}\lim_{t\to\infty} x_i(t) &= \frac{1}{1 + \dfrac{c_i^-}{c_i^+}\dfrac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\dfrac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}} = \frac{1}{1 + \dfrac{1-\sum_{j=1}^n c_j s_j}{\sum_{j=1}^n c_j s_j}} \\ &= \frac{1}{\dfrac{\sum_{j=1}^n c_j s_j + 1 - \sum_{j=1}^n c_j s_j}{\sum_{j=1}^n c_j s_j}} = \sum_{j=1}^n c_j s_j.\end{aligned}$$

We now show that the probability of misinformation is bounded by 0.5, i.e. for $\theta = 1$, we have to show that

$$p_i^{\text{Mis}}(\infty) = P(x_i(\infty) < 0.5) + 0.5P(x_i(\infty) = 0.5) \leq 0.5 \text{ for all } i,$$

while for $\theta = 0$, we must prove

$$p_i^{\text{Mis}}(\infty) = P(x_i(\infty) > 0.5) + 0.5P(x_i(\infty) = 0.5) \leq 0.5 \text{ for all } i.$$

Defining $T_j := s_j$ when $\theta = 1$ and $T_j := 1 - s_j$ when $\theta = 0$, this amounts to showing that

$$P\left(\sum_{j=1}^n c_j T_j < \frac{1}{2}\right) + \frac{1}{2}P\left(\sum_{j=1}^n c_j T_j = \frac{1}{2}\right) \leq \frac{1}{2}, \tag{A.1}$$

as, for all i, $x_i(\infty) = \sum_{j=1}^n c_j s_j$ equals $\sum_{j=1}^n c_j T_j$ when $\theta = 1$ and $1 - \sum_{j=1}^n c_j T_j$ when $\theta = 0$. Due to their construction, the T_j are iid $B(0, \rho)$-distributed, independent of the true state θ. In order to prove Equation (A.1), we also define the following quantities: $p_l(\rho) := P\left(\sum_{j=1}^n c_j T_j < \frac{1}{2}\right)$, $p_m(\rho) := P\left(\sum_{j=1}^n c_j T_j = \frac{1}{2}\right)$, and $p_u(\rho) := P\left(\sum_{j=1}^n c_j T_j > \frac{1}{2}\right)$, which obviously sum to unity: $p_l(\rho) + p_m(\rho) + p_u(\rho) = 1$ for all ρ. Using these, Equation (A.1) can be restated as $p_l(\rho) + \frac{1}{2}p_m(\rho) \leq \frac{1}{2}$, which is equivalent to $2p_l(\rho) + p_m(\rho) \leq 1 = p_l(\rho) + p_m(\rho) + p_u(\rho)$, which is in turn equivalent to $p_l(\rho) \leq p_u(\rho)$. This part of the proof can therefore be completed by showing that indeed $p_l(\rho) \leq p_u(\rho)$ for all $\rho > \frac{1}{2}$. Due to

[19] Further appendices are provided as Supplementary Online Material, which is permanently available here: `www.berno.info` under the name MisInfo_SupplementaryOnlineMaterial.pdf.

$\sum_{j=1}^n c_j(1-T_j) = \sum_{j=1}^n c_j - \sum_{j=1}^n c_j T_j = 1 - \sum_{j=1}^n c_j T_j$, the condition $\sum_{j=1}^n c_j T_j < \frac{1}{2}$ is equivalent to $\sum_{j=1}^n c_j(1-T_j) > \frac{1}{2}$. As $1-T_j$ takes the values 0 and 1 with probabilities ρ and $1-\rho$, respectively, we find that

$$p_l(\rho) = P\left(\sum_{j=1}^n c_j T_j < \frac{1}{2}\right) = P\left(\sum_{j=1}^n c_j(1-T_j) > \frac{1}{2}\right) = p_u(1-\rho) \leq p_u(\rho),$$

where the inequality at the end holds true because $1-\rho < \rho$, due to $\rho > \frac{1}{2}$. We thus have shown that $p_l(\rho) \leq p_u(\rho)$, concluding the major part of the proof.

Finally, concerning the result for network size growing to infinity, the proof proceeds along the lines of the proof of Lemma 1 of Golub and Jackson (2010), a well-established result on the wisdom of crowds. In particular, for the term $\sum_{j=1}^n c_{j,n} T_{j,n}$, we first find that its expected value is ρ, as all $T_{j,n}$ have expectation ρ and $\sum_{j=1}^n c_{j,n} = 1$. Similarly, the term's variance equals $\rho(1-\rho)\sum_{j=1}^n c_{j,n}^2$, as the $T_{j,n}$ are all independent with variance $\rho(1-\rho)$. Due to $\sum_{j=1}^n c_{j,n}^2 \leq \max_{j=1}^n c_{j,n} \sum_{j=1}^n c_{j,n} = \max_{j=1}^n c_{j,n}$ together with assumption $\max_{j=1}^n c_{j,n} \overset{n\to\infty}{\longrightarrow} 0$, this variance converges to 0. With the help of Chebyshev's inequality, this ensures that $\sum_{j=1}^n c_{j,n} T_{j,n}$ converges in probability to $\rho > 0.5$. Thus, the probability of misinformation vanishes asymptotically.

A.2 Proof of Proposition 2

To begin the proof, we use the eigendecompositions of the real symmetric matrices A^+ and A^-, writing them as $A^+ = Q^+\Lambda^+\left(Q^+\right)^\top$ and $A^- = Q^-\Lambda^-\left(Q^-\right)^\top$, respectively, with Q^+ and Q^- being orthogonal matrices whose columns are eigenvectors of A^+ and A^-, and Λ^+ and Λ^- being diagonal matrices whose entries are the eigenvalues of A^+ and A^-. From this, we have that $I + \delta^+ A^+ = Q^+\left(I + \delta^+\Lambda^+\right)\left(Q^+\right)^\top$ and $I + \delta^- A^- = Q^-\left(I + \delta^-\Lambda^-\right)\left(Q^-\right)^\top$ as well as $\left(I + \delta^+ A^+\right)^t = Q^+\left(I + \delta^+\Lambda^+\right)^t\left(Q^+\right)^\top$ and $\left(I + \delta^- A^-\right)^t = Q^-\left(I + \delta^-\Lambda^-\right)^t\left(Q^-\right)^\top$ for all t. Overall, this delivers

$$\left(I + \delta^+ A^+\right)^t = \sum_{i=1}^n \left(1 + \delta^+\lambda_i^+\right)^t q_i^+\left(q_i^+\right)^\top, \left(I + \delta^- A^-\right)^t = \sum_{i=1}^n \left(1 + \delta^-\lambda_i^-\right)^t q_i^-\left(q_i^-\right)^\top,$$

with q_i^+ and q_i^- $(i = 1, \ldots, n)$ denoting the eigenvectors of A^+ and A^-, respectively. Denoting the vector of initial signals by s, we thus get:

$$N^+(t) = \left(I + \delta^+ A^+\right)^t N^+(0) = \left(I + \delta^+ A^+\right)^t s = \sum_{j=1}^n \left(1 + \delta^+\lambda_j^+\right)^t q_j^+\left(q_j^+\right)^\top s,$$

$$N^-(t) = \left(I + \delta^- A^-\right)^t N^-(0) = \left(I + \delta^+ A^+\right)^t (\mathbb{1} - s) = \sum_{j=1}^n \left(1 + \delta^-\lambda_j^-\right)^t q_j^-\left(q_j^-\right)^\top (\mathbb{1} - s).$$

From this, we get for the numbers of positive signals at time t,

$$N^+(t) = \left(1+\delta^+\lambda_1^+\right)^t \left(q_1^+ \left(q_1^+\right)^\top s + \sum_{j=2}^{n} \left(\frac{1+\delta^+\lambda_j^+}{1+\delta^+\lambda_1^+}\right)^t q_j^+ \left(q_j^+\right)^\top s\right), \qquad \text{(A.2)}$$

and for the numbers of negative signals at time t,

$$N^-(t) = \left(1+\delta^-\lambda_1^-\right)^t \left(q_1^- \left(q_1^-\right)^\top (\mathbb{1}-s) + \sum_{j=2}^{n} \left(\frac{1+\delta^-\lambda_j^-}{1+\delta^-\lambda_1^-}\right)^t q_j^- \left(q_j^-\right)^\top (\mathbb{1}-s)\right). \qquad \text{(A.3)}$$

Due to Perron-Frobenius theory, it is clear that $\dfrac{1+\delta^+\lambda_j^+}{1+\delta^+\lambda_1^+}$ and $\dfrac{1+\delta^-\lambda_j^-}{1+\delta^-\lambda_1^-}$ are both smaller than 1 for $j = 2, \ldots, n$, implying

$$\left(1+\delta^+\lambda_1^+\right)^{-t} N^+(t) \overset{t\to\infty}{\longrightarrow} q_1^+ \left(q_1^+\right)^\top s,\ \left(1+\delta^+\lambda_1^-\right)^{-t} N^-(t) \overset{t\to\infty}{\longrightarrow} q_1^- \left(q_1^-\right)^\top (\mathbb{1}-s).$$

Now, using that $q_1^+ = \dfrac{c^+}{||c^+||} = \dfrac{c^+}{\sqrt{\sum\limits_{k=1}^{n}\left(c_k^+\right)^2}}$ and $q_1^- = \dfrac{c^-}{||c^-||} = \dfrac{c^-}{\sqrt{\sum\limits_{k=1}^{n}\left(c_k^-\right)^2}}$, we finally get:

$$\left(1+\delta^+\lambda_1^+\right)^{-t} N^+(t) \overset{t\to\infty}{\longrightarrow} c^+ \frac{\left(c^+\right)^\top s}{\sum\limits_{k=1}^{n}\left(c_k^+\right)^2} = c^+ \frac{\sum\limits_{j=1}^{n} c_j^+ s_j}{\sum\limits_{k=1}^{n}\left(c_k^+\right)^2}, \qquad \text{(A.4)}$$

$$\left(1+\delta^-\lambda_1^-\right)^{-t} N^-(t) \overset{t\to\infty}{\longrightarrow} c^- \frac{\left(c^-\right)^\top (\mathbb{1}-s)}{\sum\limits_{k=1}^{n}\left(c_k^-\right)^2} = c^- \frac{\sum\limits_{j=1}^{n} c_j^- (1-s_j)}{\sum\limits_{k=1}^{n}\left(c_k^-\right)^2} = c^- \frac{1-\sum\limits_{j=1}^{n} c_j^- s_j}{\sum\limits_{k=1}^{n}\left(c_k^-\right)^2}. \qquad \text{(A.5)}$$

With these general results at hand, we can address the three cases considered in Proposition 2.

1. For $\left(\dfrac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^{-t} x_i(t)$, we get when $\delta^+\lambda_1^+ < \delta^-\lambda_1^-$ (Case 1):

$$\begin{aligned}
\left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^{-t} x_i(t) &= \left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^{-t} \frac{N_i^+(t)}{N_i^+(t)+N_i^-(t)} \\
&= \frac{\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t)}{\left(1+\delta^-\lambda_1^-\right)^{-t}\left(N_i^+(t)+N_i^-(t)\right)} \\
&= \frac{\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t)}{\left(\dfrac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^{t}\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t) + \left(1+\delta^-\lambda_1^-\right)^{-t} N_i^-(t)}
\end{aligned}$$

$$\xrightarrow{t\to\infty} \frac{c_i^+ \dfrac{\sum\limits_{j=1}^n c_j^+ s_j}{\sum\limits_{k=1}^n \left(c_k^+\right)^2}}{c_i^- \dfrac{1-\sum\limits_{j=1}^n c_j^- s_j}{\sum\limits_{k=1}^n \left(c_k^-\right)^2}} = \frac{c_i^+}{c_i^-}\frac{\sum\limits_{k=1}^n (c_k^-)^2}{\sum\limits_{k=1}^n (c_k^+)^2}\frac{\sum\limits_{j=1}^n c_j^+ s_j}{1-\sum\limits_{j=1}^n c_j^- s_j}.$$

2. For $\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}(1-x_i(t))$, we get when $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$ (Case 2):

$$\begin{aligned}
\left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}(1-x_i(t)) &= \left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{-t}\frac{N_i^-(t)}{N_i^+(t)+N_i^-(t)} \\
&= \frac{\left(1+\delta^-\lambda_1^-\right)^{-t} N_i^-(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t}\left(N_i^+(t)+N_i^-(t)\right)} \\
&= \frac{\left(1+\delta^-\lambda_1^-\right)^{-t} N_i^-(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t) + \left(\frac{1+\delta^-\lambda_1^-}{1+\delta^+\lambda_1^+}\right)^{t}\left(1+\delta^-\lambda_1^-\right)^{-t} N_i^-(t)} \\
&\xrightarrow{t\to\infty} \frac{c_i^- \dfrac{1-\sum\limits_{j=1}^n c_j^- s_j}{\sum\limits_{k=1}^n \left(c_k^-\right)^2}}{c_i^+ \dfrac{\sum\limits_{j=1}^n c_j^+ s_j}{\sum\limits_{k=1}^n \left(c_k^+\right)^2}} = \frac{c_i^-}{c_i^+}\frac{\sum\limits_{k=1}^n (c_k^+)^2}{\sum\limits_{k=1}^n (c_k^-)^2}\frac{1-\sum\limits_{j=1}^n c_j^- s_j}{\sum\limits_{j=1}^n c_j^+ s_j}.
\end{aligned}$$

3. Finally, when $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$ (Case 1), we get:

$$\begin{aligned}
x_i(t) &= \frac{\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t)}{\left(1+\delta^+\lambda_1^+\right)^{-t} N_i^+(t) + \left(1+\delta^-\lambda_1^-\right)^{-t} N_i^-(t)} \\
&\xrightarrow{t\to\infty} \frac{c_i^+ \dfrac{\sum\limits_{j=1}^n c_j^+ s_j}{\sum\limits_{k=1}^n \left(c_k^+\right)^2}}{c_i^+ \dfrac{\sum\limits_{j=1}^n c_j^+ s_j}{\sum\limits_{k=1}^n \left(c_k^+\right)^2} + c_i^- \dfrac{1-\sum\limits_{j=1}^n c_j^- s_j}{\sum\limits_{k=1}^n \left(c_k^-\right)^2}} = \frac{1}{1+\dfrac{c_i^-}{c_i^+}\dfrac{\sum\limits_{k=1}^n (c_k^+)^2}{\sum\limits_{k=1}^n (c_k^-)^2}\dfrac{1-\sum\limits_{j=1}^n c_j^- s_j}{\sum\limits_{j=1}^n c_j^+ s_j}}.
\end{aligned}$$

A.3 Proof of Corollary 1

1. Due to the first part of Proposition 2, there are only two important cases that have to be looked at: either all initial signals are equal to 1, entailing $\lim_{t\to\infty} x_i(t) = 1$, or at least one initial signal is 0, entailing $\lim_{t\to\infty} x_i(t) = 0$. If the true state is 0, which happens with probability $1 - b^+$, then long-run misinformation occurs if all initial signals are equal to 1, and if the true state is 1, which happens with probability b^+, then long-run misinformation occurs if at least one initial signal is 0. Thus, the probability of misinformation is given by the sum of $(1-b^+)(1-\rho)^n$, the probability of state 0 occurring and leading to misinformation, and $b^+(1-\rho^n)$, the probability of state 1 occurring and leading to misinformation.

2. Due to the second part of Proposition 2, there are only two important cases that have to be looked at: either all initial signals are equal to 0, entailing $\lim_{t\to\infty} x_i(t) = 0$, or at least one initial signal is 1, entailing $\lim_{t\to\infty} x_i(t) = 1$. If the true state is 0, which happens with probability $1 - b^+$, then long-run misinformation occurs if at least one initial signal is equal to 1, and if the true state is 1, which happens with probability b^+, then long-run misinformation occurs if all initial signal are 0. Thus, the probability of misinformation is given by the sum of $(1-b^+)(1-\rho^n)$, the probability of state 0 occurring and leading to misinformation, and $b^+(1-\rho)^n$, the probability of state 1 occurring and leading to misinformation.

3. In order to prove

$$p_i^{\text{Mis}}(\infty) \leq \max\{(1-b^+)(1-\rho)^n + b^+(1-\rho^n), (1-b^+)(1-\rho^n) + b^+(1-\rho)^n\},$$

we proceed as follows: we first rearrange the right-hand side as

$$(1-\rho)^n + \Big(1-\rho^n-(1-\rho)^n\Big)\max\{b^+, 1-b^+\}$$

and observe that this term can be interpreted as the sum of the probability of all agents' initial signals being incorrect, $(1-\rho)^n$, and the product of the probability of the initial signal distribution being non-trivial, $1-\rho^n-(1-\rho)^n$, and the maximum of the probabilities of the state being equal to 1 and 0, respectively. Our task thus reduces to proving that the conditional probability of long-run misinformation given non-trivially distributed initial signals is bounded by $\max\{b^+, 1-b^+\}$. Conditional on non-trivial signals and the state $\theta = 1$, the probability of misinformation is given by

$$P\left(\frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j} > 1\right) + \frac{1}{2}P\left(\frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j} = 1\right),$$

while, conditional on $\theta = 0$, it is given by:

$$P\left(\frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j} < 1\right) + \frac{1}{2}P\left(\frac{c_i^-}{c_i^+}\frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2}\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j} = 1\right).$$

Defining $T_j := 1$ if $s_j = \theta$ and $T_j := 0$ if $s_j \neq \theta$, the two conditional probabilities above can be rearranged as

$$\begin{aligned} p_1(\rho) := P&\left(\sum_{j=1}^n \left(\frac{c_i^- c_j^-}{\sum_{k=1}^n \left(c_k^-\right)^2} + \frac{c_i^+ c_j^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) T_j < \frac{c_i^-}{\sum_{k=1}^n \left(c_k^-\right)^2}\right) \\ &+ \frac{1}{2}P\left(\sum_{j=1}^n \left(\frac{c_i^- c_j^-}{\sum_{k=1}^n \left(c_k^-\right)^2} + \frac{c_i^+ c_j^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) T_j = \frac{c_i^-}{\sum_{k=1}^n \left(c_k^-\right)^2}\right) \end{aligned}$$

for $\theta = 1$, while $\theta = 0$ leads to

$$\begin{aligned} p_0(\rho) := P&\left(\sum_{j=1}^n \left(\frac{c_i^- c_j^-}{\sum_{k=1}^n \left(c_k^-\right)^2} + \frac{c_i^+ c_j^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) T_j < \frac{c_i^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) \\ &+ \frac{1}{2}P\left(\sum_{j=1}^n \left(\frac{c_i^- c_j^-}{\sum_{k=1}^n \left(c_k^-\right)^2} + \frac{c_i^+ c_j^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) T_j = \frac{c_i^+}{\sum_{k=1}^n \left(c_k^+\right)^2}\right) \end{aligned}$$

and the probability of misinformation conditional on non-trivially distributed initial signals is given by $b^+ p_1(\rho) + (1-b^+)p_0(\rho)$. For $\rho = 0.5$, the T_j have the same distribution as $1 - T_j$, and replacing T_j by $1 - T_j$ in one of the equations above shows that $p_1(0.5) + p_0(0.5) = 1$. For $\rho = 0.5$, we therefore have

$$\begin{aligned} b^+ p_1(\rho) + (1-b^+)p_0(\rho) &= b^+ p_1(0.5) + (1-b^+)p_0(0.5) \\ &\leq \max\{b^+, 1-b^+\}p_1(0.5) + \max\{b^+, 1-b^+\}p_0(0.5) \\ &= \max\{b^+, 1-b^+\}\left(p_1(0.5) + p_0(0.5)\right) = \max\{b^+, 1-b^+\}. \end{aligned}$$

This concludes the proof for $\rho = 0.5$. For $\rho > 0.5$, the assertion follows from the fact that both $p_1(\rho)$ and $p_0(\rho)$ are decreasing in ρ, as the distribution of the T_j is shifted to the right when ρ increases.

The last inequality, finally, follows easily from b^+ and $1 - b^+$ both being bounded by $\max\{b^+, 1-b^+\}$ and from $1 - \rho^n + (1-\rho)^n < 1$, with the last inequality holding due to $\rho > \frac{1}{2}$.

A.4 Proof of Proposition 3

1. In order to prove the assertion, it is sufficient to show the following: If some agent i's probability of long-run misinformation converges to zero ($\lim_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$), then the sets $N_1 := \{n : \delta^+\lambda_1(A_n^+) < \delta^-\lambda_1(A_n^-)\}$ and $N_2 := \{n : \delta^+\lambda_1(A_n^+) > \delta^-\lambda_1(A_n^-)\}$ are both finite. Finiteness of the two sets will imply $\delta^+\lambda_1(A_n^+) = \delta^-\lambda_1(A_n^-)$ for all $n \geq n^* := \max(N_1 \cup N_2) + 1$.

   First, let us assume that N_1 was not finite, i.e., Case 1 would occur infinitely often. Then N_1 would contain an infinite sub-sequence $n_1 < n_2 < n_3 < ...$ such that $\delta^+\lambda_1(A_{n_m}^+) < \delta^-\lambda_1(A_{n_m}^-)$ for all $m = 1, \ldots, \infty$. For this sub-sequence, part i of Corollary 1 would imply that i's long-run probability of misinformation for network size n_m is $p_{i,n_m}^{\text{Mis}}(\infty) = (1-b^+)(1-\rho)^{n_m} + b^+(1-\rho^{n_m})$. This quantity would converge to $b^+ > 0$ for $m \to \infty$. The existence of a sub-sequence that converges to a value different from 0 contradicts the assumption that the original sequence converges to 0. Thus, N_1 must be finite.

   In a completely analogous way, one can see that N_2 not being finite would imply the existence of a sub-sequence $n_1 < n_2 < n_3 < ...$ such that $\delta^+\lambda_1(A_{n_m}^+) > \delta^-\lambda_1(A_{n_m}^-)$ for all $m = 1, \ldots, \infty$. In this case, we would have $p_{i,n_m}^{\text{Mis}}(\infty) = (1-b^+)(1-\rho^{n_m}) + b^+(1-\rho)^{n_m}$ for i's long-run probability of misinformation for network size n_m, and this quantity would thus converge to $1 - b^+ > 0$ for $m \to \infty$, again contradicting the assumption that the original sequence converges to 0.

2. Condition (i) and Proposition 2 imply

   $$x_i(\infty) = \frac{1}{1 + \frac{c_{i,n}^- \sum_{k=1}^n (c_{k,n}^+)^2}{c_{i,n}^+ \sum_{k=1}^n (c_{k,n}^-)^2} \frac{1 - \sum_{j=1}^n c_{j,n}^- s_j}{\sum_{j=1}^n c_{j,n}^+ s_j}}$$

   As the s_j are iid $B(1,\rho)$-distributed when $\theta = 1$ and iid $B(1, 1-\rho)$-distributed when $\theta = 0$, the variances of $\sum_{j=1}^n c_{j,n}^- s_j$ and $\sum_{j=1}^n c_{j,n}^+ s_j$ are in any case equal to $\rho(1-\rho)\sum_{j=1}^n \left(c_{j,n}^-\right)^2$ and $\rho(1-\rho)\sum_{j=1}^n \left(c_{j,n}^+\right)^2$, respectively. Due to $\sum_{j=1}^n \left(c_{j,n}^-\right)^2 \leq \max_{j=1,\ldots,n} c_{j,n}^- \sum_{j=1}^n c_{j,n}^- = \max_{j=1,\ldots,n} c_{j,n}^-$ and $\sum_{j=1}^n \left(c_{j,n}^+\right)^2 \leq \max_{j=1,\ldots,n} c_{j,n}^+ \sum_{j=1}^n c_{j,n}^+ = \max_{j=1,\ldots,n} c_{j,n}^+$, condition (ii) implies that these variances shrink to 0. As $\sum_{j=1}^n c_{j,n}^- = 1$ and $\sum_{j=1}^n c_{j,n}^+ = 1$ imply that the expected values of $\sum_{j=1}^n c_{j,n}^- s_j$ and $\sum_{j=1}^n c_{j,n}^+ s_j$ equal ρ when $\theta = 1$ and $1-\rho$ when $\theta = 0$, we find that $\sum_{j=1}^n c_{j,n}^- s_j$ and $\sum_{j=1}^n c_{j,n}^+ s_j$ converge in probability to ρ when $\theta = 1$ and to $1-\rho$ when $\theta = 0$.

   When $\theta = 1$, long-run misinformation is avoided if $x_i(\infty) > \frac{1}{2}$, which is equivalent to

$$\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j} < \frac{c_{i,n}^{+}}{c_{i,n}^{-}} \frac{\sum_{k=1}^{n} (c_{k,n}^{-})^2}{\sum_{k=1}^{n} (c_{k,n}^{+})^2} = \gamma_{i,n}. \tag{A.6}$$

Condition (iii) implies that for n large enough, there exists $\varepsilon > 0$ such that $\gamma_{i,n} \geq \frac{1-\rho}{\rho} + \varepsilon$. As $\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j}$ converges in probability to $\frac{1-\rho}{\rho}$ when $\theta = 1$, the probability of $\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j}$ being smaller than $\frac{1-\rho}{\rho} + \varepsilon$ converges to 1, and thus the long-run probability of misinformation shrinks to 0 when $\theta = 1$.

When $\theta = 0$, long-run misinformation is avoided if $x_i(\infty) < \frac{1}{2}$, which is equivalent to

$$\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j} > \frac{c_{i,n}^{+}}{c_{i,n}^{-}} \frac{\sum_{k=1}^{n} (c_{k,n}^{-})^2}{\sum_{k=1}^{n} (c_{k,n}^{+})^2} = \gamma_{i,n}. \tag{A.7}$$

Condition (iii) implies that for n large enough, there exists $\varepsilon > 0$ such that $\gamma_{i,n} \leq \frac{\rho}{1-\rho} - \varepsilon$. As $\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j}$ converges in probability to $\frac{\rho}{1-\rho}$ when $\theta = 0$, the probability of $\frac{1-\sum_{j=1}^{n} c_{j,n}^{-} s_j}{\sum_{j=1}^{n} c_{j,n}^{+} s_j}$ being larger than $\frac{\rho}{1-\rho} - \varepsilon$ converges to 1, and thus the long-run probability of misinformation shrinks to 0 when $\theta = 0$, which concludes the proof.

A.5 Proof of Corollary 2

First of all, notice that $\frac{N_i^{+}(t)}{N_i^{-}(t)} \Big/ \frac{N_j^{+}(t)}{N_j^{-}(t)}$ can be written as $\frac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))}$. We thus have to prove that $\lim_{t\to\infty} \frac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))} = \frac{c_i^{+} c_j^{-}}{c_i^{-} c_j^{+}}$, which we will do by successively tackling the three cases given in Proposition 2. For Case 1, we know that $\delta^{+}\lambda_1^{+} < \delta^{-}\lambda_1^{-}$ and

$$x_i(t) \left(\frac{1+\delta^{-}\lambda_1^{-}}{1+\delta^{+}\lambda_1^{+}} \right)^t \xrightarrow{t\to\infty} \frac{c_i^{+}}{c_i^{-}} \frac{\sum_{k=1}^{n} (c_k^{-})^2}{\sum_{k=1}^{n} (c_k^{+})^2} \frac{\sum_{l=1}^{n} c_l^{+} s_l}{1-\sum_{l=1}^{n} c_l^{-} s_l},$$

as well as $x_i(t) \xrightarrow{t\to\infty} 0$ for all i. Trivially, thus, $1 - x_i(t)$ and $1 - x_j(t)$ each converge to 1. For

$$\frac{x_i(t)}{x_j(t)} = \frac{x_i(t) \left(\frac{1+\delta^{-}\lambda_1^{-}}{1+\delta^{+}\lambda_1^{+}} \right)^t}{x_j(t) \left(\frac{1+\delta^{-}\lambda_1^{-}}{1+\delta^{+}\lambda_1^{+}} \right)^t},$$

we then find that it converges to

$$\frac{\frac{c_i^{+}}{c_i^{-}} \frac{\sum_{k=1}^{n} (c_k^{-})^2}{\sum_{k=1}^{n} (c_k^{+})^2} \frac{\sum_{l=1}^{n} c_l^{+} s_l}{1-\sum_{l=1}^{n} c_l^{-} s_l}}{\frac{c_j^{+}}{c_j^{-}} \frac{\sum_{k=1}^{n} (c_k^{-})^2}{\sum_{k=1}^{n} (c_k^{+})^2} \frac{\sum_{l=1}^{n} c_l^{+} s_l}{1-\sum_{l=1}^{n} c_l^{-} s_l}} = \frac{\frac{c_i^{+}}{c_i^{-}}}{\frac{c_j^{+}}{c_j^{-}}},$$

which proves the assertion for Case 1.

For Case 2, we know that $\delta^+\lambda_1^+ > \delta^-\lambda_1^-$ and

$$(1 - x_i(t)) \left(\frac{1 + \delta^+\lambda_1^+}{1 + \delta^-\lambda_1^-}\right)^t \overset{t\to\infty}{\longrightarrow} \frac{c_i^-}{c_i^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{\sum_{l=1}^n c_l^- s_l}{1 - \sum_{l=1}^n c_l^+ s_l}.$$

as well as $x_i(t) \overset{t\to\infty}{\longrightarrow} 1$ for all i. Trivially, thus, $x_i(t)$ and $x_j(t)$ each converge to 1. For

$$\frac{1 - x_j(t)}{1 - x_i(t)} = \frac{(1 - x_j(t)) \left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t}{(1 - x_i(t)) \left(\frac{1+\delta^+\lambda_1^+}{1+\delta^-\lambda_1^-}\right)^t},$$

we then find that it converges to

$$\frac{\frac{c_j^-}{c_j^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{\sum_{l=1}^n c_l^- s_l}{1 - \sum_{l=1}^n c_l^+ s_l}}{\frac{c_i^-}{c_i^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{\sum_{l=1}^n c_l^- s_l}{1 - \sum_{l=1}^n c_l^+ s_l}} = \frac{\frac{c_i^+}{c_i^-}}{\frac{c_j^+}{c_j^-}},$$

which proves the assertion for Case 2.

Finally, for the Case 3, we know that for all i

$$\lim_{t\to\infty} x_i(t) = \frac{1}{1 + \frac{c_i^-}{c_i^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{1 - \sum_{l=1}^n c_l^- s_l}{\sum_{l=1}^n c_l^+ s_l}} \in (0, 1).$$

This immediately implies that

$$\lim_{t\to\infty} 1 - x_i(t) = \frac{\frac{c_i^-}{c_i^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{l=1}^n c_l^+ s_l}}{1 + \frac{c_i^-}{c_i^+} \frac{\sum_{k=1}^n (c_k^+)^2}{\sum_{k=1}^n (c_k^-)^2} \frac{1 - \sum_{j=1}^n c_j^- s_j}{\sum_{l=1}^n c_l^+ s_l}} \text{ as well as}$$

$$\lim_{t\to\infty} \frac{x_i(t)}{1 - x_i(t)} = \frac{c_i^+}{c_i^-} \frac{\sum_{k=1}^n (c_k^-)^2}{\sum_{k=1}^n (c_k^+)^2} \frac{\sum_{l=1}^n c_l^+ s_l}{1 - \sum_{l=1}^n c_l^- s_l}$$

for all i, from which it easily follows that $\lim\limits_{t\to\infty} \frac{x_i(t)(1-x_j(t))}{x_j(t)(1-x_i(t))} = \frac{c_i^+ c_j^-}{c_i^- c_j^+}$.

A.6 Proof of Corollary 3

We have already established that there is exponential decay according to $\left(\frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}\right)^t$. Therefore, half-life $t_{1/2}$ is determined by $\left(\frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}\right)^{t_{1/2}} = 0.5$, which leads to $t_{1/2} = \frac{\log(0.5)}{\log(\tau)}$, with $\tau = \frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}$.

References

Acemoglu, D., Ozdaglar, A., and ParandehGheibi, A. (2010). Spread of (mis)information in social networks. *Games and Economic Behavior*, 70(2):194–227.

Acemoglu, D., Ozdaglar, A., and Siderius, J. (2022). A model of online misinformation. *mimeo*.

Acerbi, A., Altay, S., and Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *The Harvard Kennedy School Misinformation Review*, 3.

Azzimonti, M. and Fernandes, M. (2018). Social Media Networks, Fake News, and Polarization. NBER Working Papers 24462, National Bureau of Economic Research, Inc.

Banerjee, A., Breza, E., Chandrasekhar, A. G., and Mobius, M. (2019). Naive Learning with Uninformed Agents. NBER Working Papers 25497, National Bureau of Economic Research, Inc.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.

Buechel, B., Hellmann, T., and Klößner, S. (2015). Opinion dynamics and wisdom under conformity. *Journal of Economic Dynamics and Control*, 52:240 – 257.

Burki, T. (2019). Vaccine misinformation and social media. *The Lancet Digital Health*, 1(6):e258–e259.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32.

Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. (2014). Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936.

Corazzini, L., Pavesi, F., Petrovich, B., and Stanca, L. (2012). Influential listeners: An experiment on persuasion bias in social networks. *European Economic Review*, 56(6):1276–1288.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

Della Lena, S. (2019). Non-Bayesian Social Learning and the Spread of Misinformation in Networks. Working Papers 2019:09, Department of Economics, University of Venice "Ca' Foscari".

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968.

European Commission (2018). Tackling online disinformation: a European approach. Technical Report 236.

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, 1(2).

Fernandes, M. (2019). Confirmation bias in social networks. *Available at SSRN 3504342*.

Friedkin, N. E. (1991). Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6):1478–1504.

Friedkin, N. E. and Bullo, F. (2017). How truth wins in opinion dynamics along issue sequences. *Proceedings of the National Academy of Sciences*, 114(43):11380–11385.

Friedkin, N. E. and Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206.

Goel, S., Anderson, A., Hofman, J., and Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1):180–196.

Golub, B. and Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.

Golub, B. and Sadler, E. D. (2016). Learning in social networks. *Available at SSRN 2919146*.

Grabisch, M., Mandel, A., and Rusinowska, A. (2021). On the design public debate in social networks. *Paris School of Economics*, mimeo.

Grabisch, M., Mandel, A., Rusinowska, A., and Tanimura, E. (2018). Strategic influence in social networks. *Mathematics of Operations Research*, 43(1):29–50.

Grabisch, M., Poindron, A., and Rusinowska, A. (2019). A model of anonymous influence with anti-conformist agents. *Journal of Economic Dynamics and Control*, 109(C).

Grabisch, M. and Rusinowska, A. (2020). A survey on nonstrategic models of opinion dynamics. *Games*, 11(4).

Greene, C. M. and Murphy, G. (2021). Quantifying the effects of fake news on behavior: Evidence from a study of covid-19 misinformation. *Journal of Experimental Psychology: Applied*, 27(4):773.

Grimm, V. and Mengel, F. (2020). Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

Harcup, T. and O'Neill, D. (2017). What is news? News values revisited (again). *Journalism studies*, 18(12):1470–1488.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-Bayesian social learning. *Games and Economic Behavior*, 76(1):210–225.

Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S., and Lupu, Y. (2020). The online competition between pro-and anti-vaccination views. *Nature*, pages 1–4.

Jones-Jang, S. M., Mortensen, T., and Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, 65(2):371–388.

Juul, J. L. and Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46).

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Livingstone, S., Van Couvering, E., Thumin, N., Coiro, J., Knobel, M., Lankshear, C., and Leu, D. (2008). Converging traditions of research on media and information literacies. *Handbook of Research on New Literacies*, pages 103–132.

Molavi, P., Tahbaz-Salehi, A., and Jadbabaie, A. (2018). A theory of non-Bayesian social learning. *Econometrica*, 86(2):445–490.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. *Theoretical Economics*, 8(1):1–40.

Mueller-Frank, M. (2014). Does one Bayesian make a difference? *Journal of Economic Theory*, 154(C):423–452.

Papanastasiou, Y. (2020). Fake news propagation and detection: A sequential model. *Management Science*, 66(5):1826–1846.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855):590–595.

Pennycook, G. and Rand, D. (2021). Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *The Harvard Kennedy School Misinformation Review*, 2.

Prakash, B. A., Beutel, A., Rosenfeld, R., and Faloutsos, C. (2012). Winner takes all: Competing viruses or ideas on fair-play networks. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 1037–1046, New York, NY, USA. Association for Computing Machinery.

Rusinowska, A. and Taalaibekova, A. (2019). Opinion formation and targeting when persuaders have extreme and centrist opinions. *Journal of Mathematical Economics*, 84(C):9–27.

Sikder, O., Smith, R. E., Vivo, P., and Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, 10(1):1–11.

Taalaibekova, A. (2020). *Diffusion of opinions and innovations among limitedly forward-looking individuals*. PhD thesis, UCL-Université Catholique de Louvain.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., and Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLoS ONE*, 12(7):e0181821.