Supplementary Online Material (SOM) for "Misinformation due to asymmetric information sharing"

by Berno Buechel, Stefan Kloessner, Fanyuan Meng and Anis Nassar

This supplementary online material (SOM) consists of the following sections:[1]

B Extension: Heterogeneous Relations

Denote by $\lambda_1(M^+)$ the largest eigenvalue of matrix M^+ and denote by c^+, d^+ the corresponding right and left eigenvector, normalized such that $\sum_{j=1}^n c_j^+ = 1 = \sum_{j=1}^n d_j^+$.[2] Likewise, let $\lambda_1(M^-)$ be the largest eigenvalue of matrix M^- and denote by c^-, d^- the corresponding normalized right and left eigenvector. Notice that these eigenvalues and eigenvectors now contain information not only about network asymmetry, but also about decay asymmetries, as the weights δ_{ij}^+ and δ_{ij}^- have already entered the matrices M^+ and M^-. When these matrices are considered as weighted networks, c^+ and c^- are called eigenvector centrality or right-hand eigenvector centrality of M^+ and M^- (Bonacich, 1987), while d^+ and d^- can be called left-hand eigenvector centrality (e.g. Golub and Sadler, 2016).

[1] We start with B since there is already an appendix A following the main text.

[2] The two eigenvectors c^+ and d^+ coincide in the special case that the matrix M^+ is symmetric.

B.1 Results of Extended Model

Consider first the special case of symmetry i.e. if $\delta^+_{ij} = \delta^-_{ij}$ and $a^+_{ij} = a^-_{ij}$ for all i, j. Then the results of Section 4 stay essentially unchanged. The only difference is that every appearance of (left and right) eigenvector centrality c in Proposition 1 has to be replaced by $d := d^+ = d^-$, the left eigenvector centrality. The important point of symmetry thus is symmetry with respect to positive and negative networks, but not symmetry of the matrices A or M. In other words, it is not essential that the network is undirected or that the discounting is symmetric in the sense that $\delta_{ij} = \delta_{ji}$; but that there is symmetry between positive and negative networks.

Extended Proposition 1 (Symmetry)**.** *Under symmetry, the long-run signal mix is a convex combination of the initial signals s_j with weights according to left-hand eigenvector centrality, i.e. for all i, $\lim_{t\to\infty} x_i(t) = \sum_{j=1}^n d_j s_j$. Therefore, the probability of 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 $\lim_{n\to\infty} \max_{j=1,\dots,n} d_{j,n} = 0$, then the probability of misinformation converges to zero, i.e. for all i, $\lim_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$.*

We now proceed with the most general case where signal processing is allowed to differ between the two networks.

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

1. *If $\lambda_1(M^+) < \lambda_1(M^-)$, then for all i and large t*

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

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

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

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

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

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

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

Extended Proposition 2 first of all shows that the key result obtained in our baseline model (Proposition 2) is robust to the broad generalization. Second, the crucial condition is now expressed in terms of the largest eigenvalues $\lambda_1(M^+)$ and $\lambda_1(M^-)$, which combine the network structures given in the matrices A^+ and A^- with the decay factors δ_{ij}^+ and δ_{ij}^-. For the analogon of Equation 4, we find:

$$x_i(t) \approx \underbrace{\frac{c_i^+}{c_i^-}}_{\text{centrality ratio}} \cdot \underbrace{\left(\frac{1+\lambda_1(M^+)}{1+\lambda_1(M^-)}\right)^t}_{\text{exponential decay}} \cdot \underbrace{\frac{\sum_{k=1}^n c_k^- d_k^-}{\sum_{k=1}^n c_k^+ d_k^+}}_{\text{“concentration ratio”}} \cdot \underbrace{\frac{\sum_{j=1}^n d_j^+ s_j}{1-\sum_{j=1}^n d_j^- s_j}}_{\text{signal averages}} \tag{B.1}$$

Thus, the centrality ratio is determined by the right eigenvectors c^+ and c^-, while the left eigenvectors d^+ and d^- capture the influence of initial signals. The concentration ratio, which considered the squared centralities in the baseline model, now uses the product of left and right eigenvector entries.

Extended Corollary 1 (Probability of Misinformation)**.** *1. If $\lambda_1(M^+) < \lambda_1(M^-)$, 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 $\lambda_1(M^+) > \lambda_1(M^-)$, 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 $\lambda_1(M^+) = \lambda_1(M^-)$, then an agent i's (long-run) probability of misinformation is bounded by

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

Thus, Corollary 1 also neatly generalizes, with $\lambda_1(M^+)$ and $\lambda_1(M^-)$ taking the roles of $\delta^+\lambda_1^+$ and $\delta^-\lambda_1^-$, respectively. The same holds true for the next extended proposition.

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

1. If some agent i's long-run probability of misinformation shrinks to zero, i.e. if $\lim_{n\to\infty} p_{i,n}^{\text{Mis}}(\infty) = 0$, then the largest eigenvalues of the two networks coincide from a certain point in time on, i.e. there is a natural number n^* such that $\lambda_1(M_n^+) = \lambda_1(M_n^-)$ for all $n > n^*$.

2. *For a fixed agent i, the probability of long-run misinformation converges to zero, i.e. $\lim_{n\to\infty} p_{i,n}^{\mathrm{Mis}}(\infty) = 0$, if the following three conditions are satisfied:*
   - (i) *The largest eigenvalues of the two networks coincide from a certain point in time on, i.e. there is a natural number n^* such that $\lambda_1(M_n^+) = \lambda_1(M_n^-)$ for all $n > n^*$.*
   - (ii) *Maximal left-hand eigenvector centrality converges to zero, i.e. we have both $\lim_{n\to\infty} \max_{j=1,\ldots,n} d_{j,n}^+ = 0$ and $\lim_{n\to\infty} \max_{j=1,\ldots,n} d_{j,n}^- = 0$.*
   - (iii) *The agent's centrality ratio over the two networks' concentration ratio has all accumulation points within 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 $n \geq n^{**}$,*

$$\gamma_{i,n} := \frac{\frac{c_{i,n}^+}{c_{i,n}^-}}{\frac{\sum_{j=1}^n c_{j,n}^+ d_{j,n}^+}{\sum_{j=1}^n c_{j,n}^- d_{j,n}^-}} \in \left[\frac{1-\rho}{\rho} + \varepsilon, \frac{\rho}{1-\rho} - \varepsilon\right]. \tag{B.2}$$

For completeness' sake, we also present the generalizations of Corollaries 2 and 3, which hold almost unchanged.

Extended 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.*

$$\lim_{t\to\infty} \frac{N_i^+(t)}{N_i^-(t)} \Big/ \frac{N_j^+(t)}{N_j^-(t)} = \frac{c_i^+}{c_i^-} \Big/ \frac{c_j^+}{c_j^-}. \tag{B.3}$$

Hence, an agent i with higher centrality ratio than another agent j has a higher asymptotic signal mix, i.e. if $\frac{c_i^+}{c_i^-} > \frac{c_j^+}{c_j^-}$, then for large t, $x_i(t) > x_j(t)$.

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

$$t_{1/2} = \frac{\log(0.5)}{\log(\tau)}, \quad \text{with } \tau := \frac{1 + \min\{\lambda_1(M^+), \lambda_1(M^-)\}}{1 + \max\{\lambda_1(M^+), \lambda_1(M^-)\}} \tag{B.4}$$

B.2 Proofs for Results of Extended Model

B.2.1 Proof of Extended Proposition 1

According to Case 3 of Extended Proposition 2, which will be proven below, and due to symmetry, we have:

$$\lim_{t\to\infty} x_i(t) = \frac{1}{1 + \frac{1-\sum_{j=1}^n d_j s_j}{\sum_{j=1}^n d_j s_j}} = \sum_{j=1}^n d_j s_j.$$

If the true state equals 1, $\theta = 1$, then, depending on ρ, the probability of long-run misinformation is given by

$$p_1(\rho) := P\left(\sum_{j=1}^{n} d_j s_j < \frac{1}{2}\right) + \frac{1}{2} P\left(\sum_{j=1}^{n} d_j s_j = \frac{1}{2}\right),$$

and all s_j will be iid $B(1,\rho)$-distributed. In case of $\theta = 0$, however, the probability of long-run misinformation is given by

$$p_0(1-\rho) := P\left(\sum_{j=1}^{n} d_j s_j > \frac{1}{2}\right) + \frac{1}{2} P\left(\sum_{j=1}^{n} d_j s_j = \frac{1}{2}\right),$$

with all s_j being iid $B(1, 1-\rho)$-distributed.

For $\rho = 0.5$, s_j and $1 - s_j$ have the same distribution, which together with $\sum_{j=1}^{n} d_j = 1$ and $p_1(0.5) + p_0(0.5) = 1$ implies $p_1(0.5) = p_0(0.5) = 0.5$. As p_1 is decreasing in its argument, while p_0 is increasing in its argument, and $\rho > 0.5$, the probability of misinformation is always bounded by 0.5 from above.

Finally, notice that $E(x_i(\infty)) = E\left(\sum_{j=1}^{n} d_{j,n} s_j\right) = \rho$ when $\theta = 1$ and $E(x_i(\infty)) = E\left(\sum_{j=1}^{n} d_{j,n} s_j\right) = 1 - \rho$ when $\theta = 0$. For both states, the variance of $x_i(\infty) = \sum_{j=1}^{n} d_{j,n} s_j$ is given by $\rho(1-\rho)\sum_{j=1}^{n} d_j^2$. For this variance, we find

$$\rho(1-\rho)\sum_{j=1}^{n} d_j^2 \le \rho(1-\rho) \max_{j=1,\dots,n} d_{j,n} \sum_{j=1}^{n} d_{j,n} = \rho(1-\rho) \max_{j=1,\dots,n} d_{j,n} \overset{n\to\infty}{\longrightarrow} 0.$$

Overall, thus, $x_i(\infty)$ converges in probability to $\rho > 0.5$ when $\theta = 1$ and to $1 - \rho < 0.5$ when $\theta = 0$, entailing that the probability of misinformation shrinks to 0 in any case.

B.2.2 Proof of Extended Proposition 2

In order to prove the assertions, we will show that

$$\left(1 + \lambda_1(M^+)\right)^{-t} N^+(t) \overset{t\to\infty}{\longrightarrow} c^+ \frac{\left(d^+\right)^\top s}{\sum\limits_{k=1}^{n} c_k^+ d_k^+} = c^+ \frac{\sum\limits_{j=1}^{n} d_j^+ s_j}{\sum\limits_{k=1}^{n} c_k^+ d_k^+}, \tag{B.5}$$

$$\left(1 + \lambda_1(M^-)\right)^{-t} N^-(t) \overset{t\to\infty}{\longrightarrow} c^- \frac{\left(d^-\right)^\top (\mathbb{1} - s)}{\sum\limits_{k=1}^{n} c_k^- d_k^-} = c^- \frac{1 - \sum\limits_{j=1}^{n} d_j^- s_j}{\sum\limits_{k=1}^{n} c_k^- d_k^-}. \tag{B.6}$$

With Equations (B.5) and (B.6) at hand, the assertions of Proposition 2 then follow from exactly the same arguments as those given in the proof of Proposition 2 after Equations (A.4) and (A.5).

As the essential parts of Equations (B.5) and (B.6) coincide, determining the limits of $(1+\lambda_1(M^+))^{-t} N^+(t)$ is completely analogous to determining the limit of $(1+\lambda_1(M^-))^{-t} N^-(t)$. Thus, we will do this in one sweep by looking at the limit of $(1+\lambda_1(M))^{-t}(I+M)^t$, where M and stands for M^+ and M^-, respectively. The proof will thus be complete when showing that

$$(1+\lambda_1(M))^{-t}(I+M)^t \overset{t\to\infty}{\longrightarrow} \frac{c\,d^\top}{c^\top d}, \tag{B.7}$$

where c (d) stands for c^+ and c^- (d^+ and d^-), respectively.[3] In order to prove Equation (B.7), we first rewrite M using its Jordan normal form: $M = SJS^{-1}$, where J is a block diagonal matrix

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix}$$

formed of Jordan blocks J_i ($i = 1, \ldots, p$), which are either scalars consisting of eigenvalues λ_i of M or have the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}.$$

Due to the network being strongly connected and M containing only non-negative entries, M is irreducible and Perron-Frobenius theory allows to infer that the spectral radius of M is a simple eigenvalue of M. We will assume without loss of generality that this value, $\lambda_1(M)$, corresponds to the matrix J_1. As $\lambda_1(M)$ is the spectral radius of M, we also know that $|\lambda_i| \leq \lambda_1(M)$ for all $i > 1$. From all this, by setting $\widetilde{\lambda}_i := \frac{1+\lambda_i}{1+\lambda_1(M)}$ we find that $\widetilde{M} := \frac{1}{1+\lambda_1(M)}(I+M) = S\widetilde{J}S^{-1}$, with

$$\widetilde{J} = \begin{pmatrix} 1 & & & \\ & \widetilde{J}_2 & & \\ & & \ddots & \\ & & & \widetilde{J}_p \end{pmatrix}, \quad \widetilde{J}_i = \begin{pmatrix} \widetilde{\lambda}_i & 1 & & \\ & \widetilde{\lambda}_i & \ddots & \\ & & \ddots & 1 \\ & & & \widetilde{\lambda}_i \end{pmatrix} (i = 2, \ldots, p).$$

Additionally, we know that $\widetilde{\lambda}_i < 1$ due to $|\lambda_i| \leq \lambda_1(M)$ and λ_i being different from $\lambda_1(M)$.

[3] Additionally, Equation (B.7) implies that for Cases 1 and 2, the formulas given in the main text for speed of convergence remain valid in the generalized setting of Proposition 2.

Taking all this together, we find that

$$(1+\lambda_1(M))^{-t}\,(I+M)^t = \widetilde{M}^t = S\widetilde{J}^t S^{-1} = S\begin{pmatrix} 1 & & & \\ & \widetilde{J}_2^t & & \\ & & \ddots & \\ & & & \widetilde{J}_p^t \end{pmatrix} S^{-1}.$$

With respect to the terms $\widetilde{J}_i^t$, it is well known (and easy to prove) that for large t

$$\widetilde{J}_i^t = \begin{pmatrix} \widetilde{\lambda}_i^t & \binom{t}{1}\widetilde{\lambda}_i^{t-1} & \binom{t}{2}\widetilde{\lambda}_i^{t-2} & \cdots \\ & \widetilde{\lambda}_i^t & \ddots & \vdots \\ & & \ddots & \binom{t}{1}\widetilde{\lambda}_i^{t-1} \\ & & & \widetilde{\lambda}_i^t \end{pmatrix},$$

implying that all $\widetilde{J}_i^t$ shrink to 0 due to $|\widetilde{\lambda}_i| < 1$, entailing that the rate of convergence of $\widetilde{M}^t$ is essentially being determined by $\max\{\widetilde{\lambda}_i : i = 2,\ldots,p\}$.[4] For the limit of $\widetilde{M}^t$, we thus have: $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} Se_1e_1^\top S^{-1}$. Setting $u := Se_1$ and $v := S^{-\top}e_1$, we can rewrite this as $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} u\,v^\top$. The following derivations show that u is a right eigenvector of M for $\lambda_1(M)$, while v is a corresponding left eigenvector:

$$Mu = MSe_1 = SJS^{-1}Se_1 = SJe_1 = S\lambda_1(M)e_1 = \lambda_1(M)u,$$

$$\begin{aligned} v^\top M &= \left(S^{-\top}e_1\right)^\top M = e_1^\top S^{-1}M = e_1^\top S^{-1}SJS^{-1} = e_1^\top JS^{-1} = \lambda_1(M)e_1^\top S^{-1} \\ &= \lambda_1(M)\left(S^{-\top}e_1\right)^\top = \lambda_1(M)v^\top. \end{aligned}$$

As the left and right eigenvectors of M for $\lambda_1(M)$ are unique up to multiplying by a constant, $uv^\top$ and $cd^\top$ differ only by a constant: $uv^\top = \alpha cd^\top$ for some constant α. Now, from $uv^\top uv^\top = Se_1e_1^\top S^{-1}Se_1e_1^\top S^{-1} = Se_1e_1^\top e_1e_1^\top S^{-1} = Se_1e_1^\top S^{-1} = uv^\top$, we find that $uv^\top uv^\top = \alpha cd^\top \alpha cd^\top$ must equal $uv^\top = \alpha cd^\top$, thus we have $\alpha^2 d^\top c = \alpha$, implying $\alpha = \frac{1}{d^\top c}$ and $\widetilde{M}^t \overset{t\to\infty}{\longrightarrow} \frac{cd^\top}{c^\top d}$, proving Equation (B.7) and concluding the proof.

B.2.3 Proof of Extended Corollary 1

The proofs are analogous to the ones of Corollary 1, building on Extended Proposition 2 instead of Proposition 2.

B.2.4 Proof of Extended Proposition 3

The proof is completely analogous to that of Proposition 3, building on Extended Proposition 2 instead of Proposition 2 and substituting centralities $c_{j,n}^+$ and $c_{j,n}^-$ by $d_{j,n}^+$ and $d_{j,n}^-$ where appropriate.

[4] This implies that the formulas for Case 3 discussed in C.3 still apply to the generalized setting.

B.2.5 Proof of Extended Corollary 2

The proof of this corollary is perfectly analogous to the one of Corollary 2, building on Extended Proposition 2 instead of Proposition 2.

B.2.6 Proof of Extended Corollary 3

The proof is completely analogous to the one of Corollary 3, building on Extended Proposition 2 instead of Proposition 2, see also footnote 3.

C Further Appendices

C.1 Additional Examples for the Benchmark Case of Symmetry

To illustrate occurrence of misinformation in the setting of symmetry, as discussed in Section 4, we study two extreme examples: A regular graph in Example C.1 and a network with a clique of five in Example C.2. oth examples are also used in Figure 1 in the main text as "best" and "worst'" networks.

Example C.1 (Regular network)**.** *Consider network* (N, A) *that is connected and regular of degree* k*, i.e. every agent has exactly* k *links.*

Regularity of degree k *implies that the largest eigenvalue is* $\lambda_1 = k$ *and eigenvector centrality is* $c = (\frac{1}{n}, ..., \frac{1}{n})$*. From Proposition 1,* $\lim_{t \to \infty} x_i(t) = \frac{1}{n} \sum_{j=1}^{n} s_j$*, which is just the mean of the initial signals. This is a remarkable observation: Under symmetry and when the network is regular, the long-run signal mix of every agent exactly reflects the initial signal distribution.*

The only source of misinformation is hence that the initial draw of signals is "unlucky" (i.e. it happens to consist of many signals that do not reflect the true state). For instance, let n *be odd. Then the probability of misinformation is* $p_i^{\text{Mis}}(\infty) \sum_{r=0}^{\frac{n-1}{2}} \binom{n}{r} \rho^r (1-\rho)^{n-r}$*, which equals the probability that the minority of* n *independent signals is correct. To have concrete numerical examples, let the quality of each initial signal be* $\rho = 0.6$*. Then* $p_i^{\text{Mis}}(\infty) = 0.267$ *for* $n = 9$ *agents and* $p_i^{\text{Mis}}(\infty) = 0.022$ *for* $n = 99$ *agents. Observe that the probability of misinformation in regular graphs goes to zero for growing* n*.*

Observe finally the comparison to Bayesian learners. Suppose for a moment that all agents are proper Bayesian learners in the following sense: they account for the repetition of signals and form their beliefs according to Bayes' rule using each independent signal only once. In a connected network, these Bayesian learners will update until they have received each initial signal and then form their belief based on exactly the same signal mix as our much more naïve agents form in the long run when the network is regular (see, e.g. DeMarzo et al. (2003), Theorem 3).

Example C.2 (Network with clique of five)**.** *Consider the network* (N, A) *depicted in Figure C.1. This network consists of* $n = 10$ *agents. Five of them,* $1, ..., 5$*, form a clique, i.e. the network restricted to these agents is complete; the others are arranged in a line.*

The normalized eigenvector corresponding to the largest eigenvalue is

$$c = (19.42\%, 18.41\%, 18.41\%, 18.41\%, 18.41\%, 5.12\%, 1.35\%, 0.36\%, 0.09\%, 0.02\%)$$

The nodes are labeled according to their entry in this eigenvector with 1 *having the largest entry and* 10 *the lowest. Observe that the five members of the clique obtain the highest*

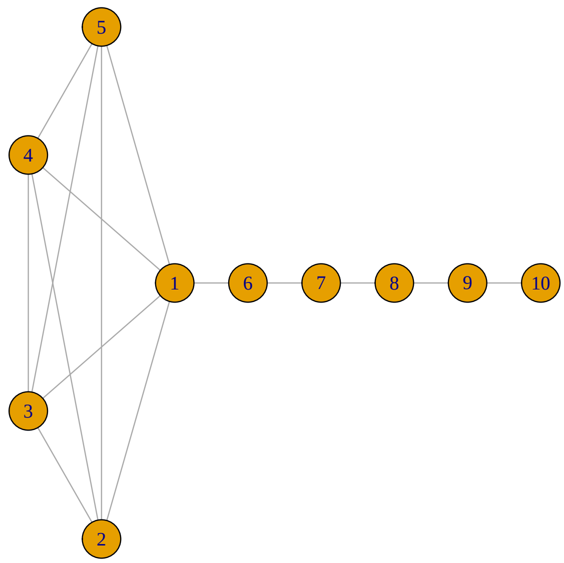

Figure C.1: A network with a clique of five and all other agents arranged in a line. Eigenvector centrality is:

$$c = (19.42\%, 18.41\%, 18.41\%, 18.41\%, 18.41\%, 5.12\%, 1.35\%, 0.36\%, 0.09\%, 0.02\%)$$

eigenvector entries. In fact, any three of their entries sum up to more than half of all entries. Hence, if it happens that at least three out of the five agents $1, ..., 5$ *receive the wrong signal, say* 0*, we have* $\sum_{j=1}^{n} c_j s_j < 0.5$ *for* $\theta = 1$ *and hence misinformation prevails (by Proposition 1).*[5] *The probability to have such a draw of signals and in fact the probability of misinformation is* $(1-\rho)^5 + 5\rho(1-\rho)^4 + 10\rho^2(1-\rho)^3$*, e.g. for* $\rho = 0.6$*, it is* $p^{Mis}(x_i(\infty)) = 0.31744$*. There are many more such networks (with the same expected level of misinformation) for* $n = 10$*, but there is no network with higher probability of misinformation.*

More importantly, we can construct networks with a clique of five and all others arranged in a line for all $n > 7$*. The probability of misinformation is unchanged, as we checked for* n *up to 1'000 by using programming language R. The eigenvector centralities converge to* $c_1 = 19.41919\%$ *and* $c_{2,\ldots,5} = 18.40593\%$ *for the members of the clique. Hence, misinformation still happens when at least three out of these five receive the wrong signal. Thus, we observe that as the number of nodes grow, the probability of misinformation need not go to zero, as there are networks with a substantial probability of misinformation.*

The example shows that a small group of people who are well-connected among themselves may have a disproportional large influence on the long-run signal mixes and hence

[5] And likewise for $\theta = 0$, as under symmetry, the realized state does affect the probability of misinformation.

can be a cause for misinformation under symmetry. However, under symmetry the probability of misinformation is always bounded and it converges to zero for large networks under a standard condition, as shown by Proposition 1 in the main text.

C.2 Different Interpretations for the Decay Factor

With respect to the decay factors, there are in fact three different interpretations, as we will explain in more detail below. In a nutshell, these interpretations are (i) the sender only shares part of her signals, (ii) the communication channel does not transmit 100% of the signals, and (iii) the recipient discounts part of the received signals. In order to improve readability, we will in this subsecion omit all '+' and '-' superscripts, thus A may stand for A^+ and A^-, respectively, $N(t)$ may denote either the numbers of positive signals $N^+(t)$ or that of negative signals $N^-(t)$, δ will be either δ^+ or δ^- and so on.

In order to showcase all the explanations given above for the existence of decay factors, we might consider the following very general model: by $N_{(s)}(t)$, we denote the numbers of signals that agents send out to their neighbours, and we write $N_{(s)}(t) = \delta_{(s)}N(t)$ to model that agents do not communicate all their signals to their neighbors, with $\delta_{(s)} \in (0, 1]$ capturing the share of signals that agents are willing to transmit. We then denote by $N_{(t)}(t)$ the numbers of signals that are transmitted between the agents, and by modeling $N_{(t)}(t) = \delta_{(t)}AN_{(s)}(t)$, with $\delta_{(t)} \in (0, 1]$ describing the share of signals that are successfully transmitted by the communication channel. Finally, we use $N_{(p)}(t)$ to denote the numbers of signals that agents are actually processing when updating their signals from time t to $t + 1$. Here, by setting $N_{(p)}(t) = \delta_{(p)}N_{(t)}(t)$, the discounting of received signals by agents would be described by $\delta_{(p)} \in (0, 1]$. Taken together and defining $\delta := \delta_{(p)}\delta_{(t)}\delta_{(s)}$, agents process

$$N_{(p)}(t) = \delta_{(p)}N_{(t)}(t) = \delta_{(p)}\delta_{(t)}AN_{(s)}(t) = \delta_{(p)}\delta_{(t)}A\delta_{(s)}N(t) = \delta_{(p)}\delta_{(t)}\delta_{(s)}AN(t) = \delta AN(t), \tag{C.1}$$

which is exactly the formula that we use in our main model. By doing so, we are able to model any of the three interpretations, by setting two of the three factors to 1 and allowing only one to be smaller than 1: e.g. setting $\delta_{(p)} = 1$, $\delta_{(t)} = 1$, and $\delta_{(s)} < 1$ leads to a model where the decay factor δ captures that agents share only some part of the signals they receive. Furthermore, our model also allows variations where two or even all three effects are at play.

If some of the above phenomena are no longer homogeneous across agents, but agent-specific, we might preserve the general structure, but replace the scalar quantities $\delta_{(p)}$, $\delta_{(t)}$, and $\delta_{(s)}$ by matrices $\Delta_{(p)}$, $\Delta_{(t)}$, and $\Delta_{(s)}$. In this case, the equation describing sharing only parts of available signals becomes $N_{(s)}(t) := \Delta_{(s)}N(t)$, and $\Delta_{(s)}$ will be a diagonal matrix which captures the agent-specific factors describing which share of their signals agents do actually share. Similarly, the generalized equation for the discounting of received signals becomes $N_{(p)}(t) = \Delta_{(p)} N_{(t)}(t)$, with the diagonal matrix $\Delta_{(p)}$ capturing the agent-specific factors used for ignoring some part of the signals transmitted to the agents. Finally, with the non-diagonal matrix $\Delta_{(t)}$ whose entries ${\Delta_{(t)}}_{ij}$ determine the share of signals that are successfully transmitted from agent j to agent i, the equation for the transmitted signals becomes $N_{(t)}(t) := (\Delta_{(t)} \circ A) N_{(s)}(t)$, with $\Delta_{(t)} \circ A$ denoting the Hadamard product of $\Delta_{(t)}$ and A. Equation (C.1) then generalizes to

$$N_{(p)}(t) = \Delta_{(p)} N_{(t)}(t) = \Delta_{(p)} \left(\Delta_{(t)} \circ A\right) N_{(s)}(t) = \Delta_{(p)} \left(\Delta_{(t)} \circ A\right) \Delta_{(s)} N(t) \tag{C.2}$$

Due to properties of diagonal matrices and the Hadamard product, $\Delta_{(p)} \left(\Delta_{(t)} \circ A\right) \Delta_{(s)}$ turns out to be identical to $\left(\Delta_{(p)} \Delta_{(t)} \Delta_{(s)}\right) \circ A$, implying that we finally have $N_{(p)}(t) = (\Delta \circ A)\, N(t)$, with $\Delta := \Delta_{(p)} \Delta_{(t)} \Delta_{(s)}$, which amounts to the formula we use in our extended model. The components Δ_{ij} thus simultaneously capture agent i's possible discounting (${\Delta_{(p)}}_i$), the communication between i and j not working properly (${\Delta_{(t)}}_{ij}$), and agent j not sharing all signals (${\Delta_{(s)}}_j$). Similar to above, our generalized model therefore also allows for one, two, or even all three of these effects being at play.

C.3 Conditions for Case 3 to Approximate Case 1 and 2

Proposition 2 provides a case distinction with three cases. In the main text, we argue that Case 3 is sometimes a good approximation for the short-run and medium-run dynamics of Cases 1 or 2. Here we elaborate on the conditions for this approximation to work.

The first condition is that parameters are reasonably close to Case 3, i.e. $\delta^+\lambda_1^+$ being close to $\delta^-\lambda_1^-$. We can observe this in Example 1 in Figure 2, where values of δ^+ which are close to the critical value of 0.4 induce similar dynamics. There is however a second condition, which happens to be satisfied in Example 1. For explaining that condition, let us reconsider equations (A.2) and (A.3) and define $\tau^{\text{Case 3}} := \max\{\frac{|1+\delta^+\lambda_i^+|}{1+\delta^+\lambda_1^+}, \frac{|1+\delta^-\lambda_i^-|}{1+\delta^-\lambda_1^-}, i = 2, \ldots, n\}$, where λ_i^+ and λ_i^- denote all but the largest eigenvalues of A^+ and A^-, respectively. The second condition is that $\tau^{\text{Case 3}}$ is substantially smaller than $\frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}$. Intuitively, the first condition assures that the importance of the largest eigenvalues persists sufficiently long (they always matter in Case 3, but vanish in the long run of the two others) and the second condition assures that the importance of all other eigenvalues vanishes sufficiently fast. As a rule of thumb, when slow convergence occurs in Cases 1 or 2, resulting in large values for actual half-life $t_{1/2} = \log(0.5) \Big/ \log\left(\frac{1+\min\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}{1+\max\{\delta^+\lambda_1^+, \delta^-\lambda_1^-\}}\right)$, and at the same time the half-life formula for Case 3, $\frac{\log(0.5)}{\log(\tau^{\text{Case 3}})}$ produces a much lower "pseudo half-life," then the formula given for Case 3 of Proposition 2 may provide a better approximation for the relevant misinformation in the short or medium term than the actual long-term limit of either 0 or 1.

C.4 Comparison with DeGroot Model

In the following, we will discuss the similarities and differences of our model as compared to the classical DeGroot model, with respect to model set-up, conditions for convergence, reaching a consensus, and eigenvector centrality.

With respect to the set-up of our model, the most striking difference to the DeGroot model is that we separate the evolution of positive and of negative signals, while in the DeGroot model this distinction is not possible. However, the evolution of each type of signals (positive or negative) resembles the DeGroot model, in the sense that values at time $t+1$ are linear functions of values at time t. While the weights of these regression-type recursions are restricted to be non-negative in our model as well in the DeGroot model, our model for the signals' evolution does not require convex combinations, i.e. the weights do not have to sum up to unity, in contrast to the DeGroot case. For the special case of symmetry in the sense that $A^+ = A^- =: A$ as well as $\delta^+ = \delta^- =: \delta$ and denoting $I + \delta A$ by W, it is easy to see that $N^+(t) = W^t s$, $N^-(t) = W^t(\mathbb{1} - s)$, and $N(t) = W^t \mathbb{1}$, implying

$$x_i(t) = \frac{e_i^\top W^t s}{e_i^\top W^t \mathbb{1}}, \tag{C.3}$$

with e_i denoting the i-th unit vector. Furthermore, Equation (C.3) implies that the updating of the signal mixes $x(t)$ in our model may be written as

$$x(t) = \widetilde{W}(t) x(t-1) \tag{C.4}$$

with the row-stochastic matrices $\widetilde{W}(t)$ having entries $\widetilde{w}(t)_{ij} = w_{ij} \frac{\kappa_j(t-1)}{\kappa_i(t)}$, with w_{ij} denoting the entries of $W = I + \delta A$ and $\kappa(t) := W^t \mathbb{1}$.[6] Equation (C.4) thus is a representation of the updating process in our model as a generalized DeGroot model, where in general the updating matrices $\widetilde{W}(t)$ are not constant, but change over time. Furthermore, if the networks described by A are regular, then these updating matrices will in fact not depend on time, and the updating formula (C.4) will actually become a DeGroot model.[7]

With regard to conditions ensuring convergence, it is well-known that values converge in the DeGroot model if and only if every set of nodes is strongly connected and closure is also aperiodic. This is in fact very similar to our model, where we receive convergence by assuming that the whole society is strongly connected and by observing that the matrices W^+ and W^- are aperiodic. Aperiodicity of the matrices $W^+ = I + \delta^+ A^+$ and $W^- = I + \delta^- A^-$ is guaranteed, as they both involve the identity matrix I.

For the DeGroot model, a consensus emerges whenever the model converges, where the consensus is a convex combination of the initial values. This is different for our model,

[6] A similar representation of $x(t)$ also holds for the extended model, it can be developed by simply replacing all appearances of δA by M.

[7] This observation has been analogously made in Sikder et al. (2020).

which in contrast to the DeGroot model, allows for the signal mixes to converge to the truth or to complete misinformation, depending on the networks' eigenvalues and the decay parameters. In addition, in our model, signal mixes will converge to consensus if $\delta^+\lambda^+ = \delta^-\lambda^-$ and $\frac{c_i^-}{c_i^+}$ does not depend on i: the latter condition is equivalent to $c^- = c^+$, i.e. in our main as well as in the extended model (Section 7), the networks must have identical centralities. This actually can happen even though A^+ and A^- are different, with an example being the case of both networks being regular, but of different degrees.

With respect to eigenvector centrality in DeGroot models, the left-hand eigenvector determines the weights for the asymptotically emerging consensus (Jackson, 2010; Golub and Sadler, 2016). In our model, however, centralities only play this role when the relation $\delta^+\lambda_1^+ = \delta^-\lambda_1^-$ holds (i.e. in case 3). In this case, eigenvector centralities play two roles (in our main model): they influence the long-run signal mix of all agents commonly through $\frac{1-\sum_{j=1}^n c_j^- s_j}{\sum_{j=1}^n c_j^+ s_j}$, reflecting the initial signals' impact, and they influence the agents' individual long-run signal mixes through the centrality ratios $\frac{c_i^-}{c_i^+}$. In our extended model, the former role (the part common to all agents) is played by the left-hand eigenvectors d^+ and d^-, while the latter role (the individual part) is still played by the (right-hand) eigenvector centralities c^+ and c^-.