

UNIVERSITY OF WARWICK

Economics

Opinion Dynamics with Differentiated Messaging

Andrew Harkins

October 2026
No: 1637

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Warwick Economics Research Papers

Opinion Dynamics with Differentiated Messaging*

Andrew Harkins†

October 30, 2026

Abstract

This paper develops a model of opinion dynamics in networks where message content can be differentiated by target. In contrast to previous approaches, this model studies the evolution of message content on the edges of the graph, rather than focusing on how beliefs evolve at nodes. This framework is then used to explore how *opinion conformity*, the act of tailoring one's stated opinion to the recipient, can impact convergence speed and polarization. Pairwise opinion conformity slows convergence to consensus, whereas anticonformity can lead to faster asymptotic convergence. It is also shown that group conformity can lead to persistent disagreement, even in a connected graph, providing a new perspective on the causes of opinion polarization within social networks.

JEL codes: D85; D83

Keywords: Opinion dynamics; social networks; conformity; polarization; social learning

1 Introduction

The question of whether opinions in a society naturally converge to a consensus or remain divided lies at the heart of the literature on social learning in networks. The canonical model of DeGroot (1974) has become a workhorse in this area, providing a tractable setting in which to study this question. Yet the DeGroot model abstracts away from an important feature of opinion formation, which is that individuals routinely express different opinions to different people. This paper relaxes that assumption by moving from a setting where individuals express a single public opinion to all neighbours, to one

*I thank participants at the CRETA Seminar at the University of Warwick, CTN 2026 Padua and the 12th Network Science & Economics conference for helpful comments and suggestions.

†Department of Economics, University of Warwick, CV4 7AL Coventry, United Kingdom. Email: andrew.harkins@warwick.ac.uk

where they tailor their message to each neighbour. Opinion sharing is modelled as a linear process on the edges of the graph, which admits a richer class of opinion dynamics whilst also preserving the structure that makes the DeGroot framework tractable.

This paper examines how differentiated messaging affects both the speed at which opinions converge to consensus and also the conditions under which persistent disagreement may arise. Since the number of edges in a graph is generally much larger than the number of nodes, edge-level opinion dynamics can operate in a higher-dimensional space than node-level dynamics. This allows for novel messaging rules such as the 'group conformity' rule, where agents send one message to their in-group neighbours and another message to their out-group neighbours.

Under the standard models of heuristic-based learning in networks, a group of agents who average the opinions of their neighbours will, under mild conditions, converge to the same consensus value. This result persists even in highly segregated networks, where the echo chamber effect of repeatedly hearing similar opinions only slows convergence to consensus, but does not prevent it. Models such as Golub and Jackson (2012) illustrate how network structure can increase the time it takes for a society to reach consensus, but persistent disagreement can only obtain under additional assumptions, such as stubbornness (Acemoglu et al., 2013), antagonism (Della Lena et al., 2026), or endogenous network structures (Polanski and Vega-Redondo, 2023). A key contribution of this paper is to show that disagreement can persist in a connected network where agents listen to and incorporate the opinions of their neighbours, but tailor their outgoing messages according to group identity.

Experimental work on learning in networks has drawn a distinction between Bayesian models of learning and heuristic-based approaches such as DeGroot learning. This experimental work has provided mixed support for the Bayesian model and has tended to favour heuristic-based models of learning. Results from a field experiment by Chandrasekhar et al. (2020) lend more support to the DeGroot learning model over Bayesian learning when viewed across their whole sample, with support for the Bayesian model only in a subset of the sample. Grimm and Mengel (2020) also find that the naive learning model tends to be a more accurate predictor of individual choices in their experiment. Interestingly, they suggest that the results of their experiment can be better explained by naive learning rules which account for some correlation of signals, but in a less sophisticated manner than a fully Bayesian learner. Further experimental evidence is provided by Choi et al. (2023) where subjects' behaviour aligned closely with the predictions of DeGroot learning. Choi et al. (2023) also show that conformity to the DeGroot prediction increased for later rounds of their experiment, suggesting that subjects may have been learning to behave in a manner more consistent with DeGroot.

Although the DeGroot model is tractable and has experimental support, the restric-

tion to linear updating rules can be criticised as being ad hoc. There have been attempts to address this by providing both axiomatic foundations (Molavi et al., 2018) and generalisations to non-linear aggregation rules (Cerreia-Vioglio et al., 2024). In addition to this concern, two other features of the model have attracted criticism. Firstly, there is a mechanical echo chamber effect which is introduced by the recursive nature of the naive learning rule. Secondly, the model has difficulty in accounting for long run disagreement between agents, despite widespread empirical evidence to the contrary (see Bowen et al. (2023) and the references therein).

The existence of an echo chamber effect is a well-known consequence of the DeGroot modelling approach and was described by DeMarzo et al. (2003) as a form of 'persuasion bias' whereby correlated messages are overweighted in the agent's belief update. Banerjee and Compte (2024) show how the impact of any transmission errors would be amplified by this effect in the DeGroot model and how agents may optimally choose to move away from the DeGroot learning heuristic towards alternative heuristics which anchor them more closely to their initial belief.

Relatively less attention has been paid to the assumption made by naive learning models that agents broadcast a single opinion to all neighbours simultaneously. Agents are assumed to communicate publicly to those who will listen, rather than privately along pairwise links. Furthermore, existing models often provide no distinction between the beliefs held by nodes and the messages that are sent. A notable exception is the paper of Buechel et al. (2015) which examines the consequences of opinion conformity for learning. They assume that nodes send identical messages to all neighbours, but that they are motivated to balance the authenticity of their message with conformity to their neighbours' positions. A similar model is that of Anufriev et al. (2023), which considers how incentives to tune message content to the beliefs of the audience can impact consensus and convergence speed in the DeGroot model. However, these papers do not allow for differentiation of messages by target.

In contrast to the above, both Galeotti et al. (2013) and Foerster (2019) consider the case where agents communicate privately in a network and may tailor their message based on the recipient. However, in those models the communication is strategic and agents are assumed to be Bayesian updaters.

The paper by Bowen et al. (2023) is closer in spirit to the present paper, but differs in some important respects. Agents in Bowen et al. (2023) can selectively share incoming messages based on the content and agent types, which is similar to the present model. The mechanism for persistent disagreement in their model relies on agents over-attributing silence to a lack of new information, combined with the presence of uneven friendship networks, e.g. a homophilous network where similar types are more likely to link. Neither feature is required in the present model, and therefore this model offers an alternative

perspective on how selective sharing of message content can lead to opinion polarization.

The structure and main contributions of the paper are as follows: Sections 2 and 3 introduce the model and explore cases where edge-level dynamics extend the results of existing node-level models. Section 4 motivates a particular choice of messaging rule driven by conformity incentives and then studies the impact of pairwise conformity on convergence speed. The main result in this section is that pairwise conformity preserves the consensus opinion but slows convergence, whereas anticonformity can increase the asymptotic convergence speed. Section 5 then analyses how group conformity can generate persistent disagreement in messages, even when the underlying graph is connected. The key mechanism is that group conformity can split the opinion dynamic into multiple independent 'conversations' by filtering messages according to type. Each agent is then exposed to different conversations based on their connections to different types.

2 Model

A finite set $N = \{1, 2, \dots, n\}$ of agents are embedded in a network where $n \geq 3$. The network is represented by a graph $\mathcal{G} = (N, E)$, where E is the set of directed edges and $m := |E|$. Assume that the underlying network is undirected, so $(i, j) \in E$ implies that $(j, i) \in E$. An undirected edge between i and j generates two directed edges which can be used to send messages. To lighten notation the directed edge (i, j) will be simply referred to as ij. Let $\mathcal{S}(ij) = i$ denote the source of the directed edge ij and let $\mathcal{T}(ij) = j$ denote the target of directed edge ij. For each agent $i \in N$, let $E_i^{\text{out}} = \{e \in E : \mathcal{S}(e) = i\}$ and $E_i^{\text{in}} = \{e \in E : \mathcal{T}(e) = i\}$ denote the set of edges out of i and into i, respectively.

Self links are permitted and therefore the neighbourhood of agent i, denoted by $N_i = \{j \in N : ij \in E\}$, may include agent i themselves. Let $d_i = |N_i|$ denote the degree of agent i in the network. We will assume throughout that there are no isolated agents in the network, that is $|N_i \setminus \{i\}| \geq 1$ for all $i \in N$.[1]

A vector $\mathbf{v}$ is taken to be a column vector by default, with $\mathbf{v}^T$ denoting the transpose of $\mathbf{v}$. The vector $\mathbf{1}_n$ denotes a vector of 1s of length n, where the subscript is dropped when it is clear from the context. Matrices are denoted by bold uppercase letters and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, with subscripts again omitted if clear from the context.

To model the process of social influence we assume that agents send messages to their neighbours each period, where time evolves in discrete steps and is indexed by $t \in \{0, 1, \dots\}$. A message which is sent from agent i to agent j at time t is denoted by $s_{ij}(t) \in \mathbb{R}$. Let $\mathbf{s}_i(t)$ denote the vector of messages sent by i and let $\mathbf{s}(t) = (\mathbf{s}_1(t), \dots, \mathbf{s}_n(t))$ denote the vector of all messages sent at time t. Let $\mathbf{r}_i(t)$ denote the vector of messages

[1] Isolated agents cannot contribute to message sending and therefore we restrict attention only to those agents who have links to other agents.

received by i at time t, where $r_{ij}(t) = s_{ji}(t)$. The vector $\mathbf{r}(t) = (\mathbf{r}_1(t), \ldots, \mathbf{r}_n(t))$ is therefore a permutation of $\mathbf{s}(t)$, where this permutation is encoded by a permutation matrix $\mathbf{P}$ such that $\mathbf{r}(t) = \mathbf{P}\mathbf{s}(t)$.

Each message s_{ij} can be interpreted as a statement of i's current opinion on some topic, for example, their level of support for a policy, their level of enthusiasm for a new product, or their level of trust in a political leader.

A key feature of this model which distinguishes it from past work is that s_{ij} and s_{ik} need not be identical for $j \neq k$. In the classical DeGroot framework, agents are assumed to broadcast identical opinions to all their neighbours and then to form a new opinion as a convex combination of the opinions they heard in the previous period (including possibly their own previous opinion). Agents in the present model are not restricted to sending identical opinions to each neighbour; rather, they can condition the message content based on the target node. This approach therefore shifts focus from the evolution of beliefs at nodes to the evolution of message content on edges.

The DeGroot model imposes a linear updating rule for node beliefs $x_i(t)$, such that a new opinion in period $t+1$ is given by $x_i(t+1) = \sum_{j \in N_i} T_{ij} x_j(t)$. Expressing this updating rule in vector form gives $\mathbf{x}(t+1) = \mathbf{T}\mathbf{x}(t)$ where the $n \times n$ matrix $\mathbf{T}$ is assumed to be nonnegative and row-stochastic. This approach maintains tractability of the model as standard results from the theory of nonnegative matrices and Markov chains can be applied. In the same spirit, we will assume that the outgoing messages from an agent i next period are a linear combination of incoming messages received this period. In particular, if we let $\mathbf{w}_{ij}$ denote a nonnegative vector of weights for a given neighbour $j \in N_i$, then we will assume that $s_{ij}(t+1) = \mathbf{w}_{ij}^T \mathbf{r}_i(t)$.

In line with the setting of the DeGroot model, we will assume that each vector $\mathbf{w}_{ij}$ sums to 1, so that outgoing messages are a convex combination of incoming messages. This can be understood as a requirement that agents can only share information that they heard from other agents (i.e. agents cannot fabricate information). This assumption allows us to better contrast the outcomes of this model with those of DeGroot and related models in terms of consensus and polarization.

If we take an agent i and fix an ordering of i's neighbours, say $N_i = \{j, k, \ldots\}$, then we can arrange the weight vectors $\mathbf{w}_{ij}^T, \mathbf{w}_{ik}^T, \ldots$ as the rows of a matrix $\mathbf{W}_i \in \mathbb{R}_+^{d_i \times d_i}$. This matrix $\mathbf{W}_i$ shall be referred to as i's *messaging rule*, where each $\mathbf{W}_i$ is square, nonnegative and row-stochastic. Ordering the set of outgoing edges as $E_i^{\text{out}} = \{ij, ik, \ldots\}$ and the set of incoming edges as $E_i^{\text{in}} = \{ji, ki, \ldots\}$, we then have a linear mapping from messages received at time t to messages sent at time $t+1$, given by

$$\mathbf{s}_i(t+1) = \mathbf{W}_i \mathbf{r}_i(t).$$

A simple example of a messaging rule for an agent with three neighbours is provided in Figure 1, where row labels denote the outgoing sent message and column labels denote the incoming received message. In this example, agent i uniformly averages the messages from all three neighbours and forwards this information on in the following period.

$$\mathbf{W}_i^{\mathrm{U}} = \begin{array}{c} \\ ij \\ ik \\ i\ell \end{array} \begin{array}{c} \begin{array}{ccc} ij & ik & i\ell \end{array} \\ \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array}$$

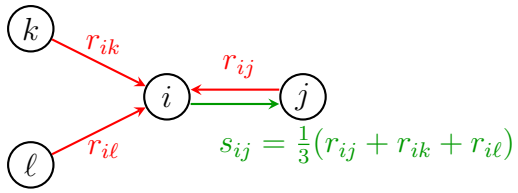


Figure 1: The uniform messaging rule

To formally define the uniform messaging rule, let $\mathbf{W}_i$ for an agent i be the $d_i \times d_i$ matrix given by

$$\mathbf{W}_i^{\mathrm{U}} = \frac{1}{d_i} \mathbf{1}_{d_i} \mathbf{1}_{d_i}^T. \tag{1}$$

This is consistent with the standard DeGroot model on an undirected and unweighted network, where each agent uniformly averages incoming messages to form their opinion, which is then shared identically with all neighbours.

A second example is provided in Figure 2 where agent i only relays information to agent j which does not contain the message content sent by agent j the previous period. This messaging rule is named the 'nonbacktracking' messaging rule, taking inspiration from models of random walks on networks where the walker is prevented from immediately backtracking along an edge that they have just traversed.[2]

$$\mathbf{W}_i^{\mathrm{NB}} = \begin{array}{c} \\ ij \\ ik \\ i\ell \end{array} \begin{array}{c} \begin{array}{ccc} ij & ik & i\ell \end{array} \\ \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array}$$

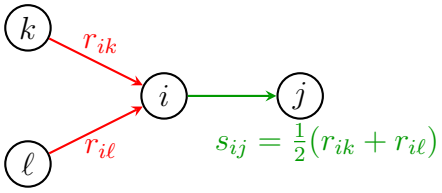


Figure 2: The nonbacktracking messaging rule

The nonbacktracking messaging rule can be formally defined in the case where $d_i \geq 2$ for all $i \in N$.[3] Letting e and f denote two edges in E_i^{out}, where e corresponds to sent message s_e and f corresponds to received message r_f, we can then define the nonbacktracking messaging rule as

[2]Nonbacktracking processes have been studied outside of the economics literature in the context of community detection by Krzakala et al. (2013), and network centrality by Grindrod et al. (2018) and Arrigo et al. (2019).

[3]If $d_i = 1$ then this agent only has a single source of messages and so any message out must necessarily be backtracking. Therefore, such an agent is left isolated when backtracking is forbidden.

$$\mathbf{W}_i^{\mathrm{NB}} = [W_{ef}] \quad \text{where} \quad W_{ef} = \begin{cases} \dfrac{1}{d_i - 1}, & \text{if } \mathcal{T}(e) \neq \mathcal{T}(f) \\ \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

This is an example of a *differentiated* messaging rule whereby different neighbours receive potentially different messages from node i at time t. Under this messaging rule, any backtracking content is filtered out by agents when sending new messages out the following period.

When considering alternative messaging rules, a natural question is whether the resulting opinion dynamics could also be represented using the existing frameworks. The following section answers this question and provides conditions under which the edge-level model will generate novel opinion dynamics compared to the existing node-level models.

3 Node-level vs. Edge-level Opinion Dynamics

To compare the two approaches to modelling opinion dynamics in networks, we will distinguish between a node-level opinion dynamic and an edge-level opinion dynamic. A node-level opinion dynamic is a sequence of opinion vectors $(\mathbf{x}(t))_{t\geq 0}$ with $\mathbf{x}(t) \in \mathbb{R}^n$ for all $t \geq 0$, where each $x_i(t)$ is the opinion of agent i at time t. Opinions evolve according to a linear process

$$\mathbf{x}(t+1) = \mathbf{T}\mathbf{x}(t), \tag{3}$$

where $\mathbf{T} \in \mathbb{R}_+^{n\times n}$ is row-stochastic with $T_{ji} = 0$ whenever $ij \notin E$. If $T_{ji} > 0$ then there is a directed edge from i to j (along which i sends a message to j) and T_{ji} is the weight which j places on i's opinion when forming their own opinion. The matrix $\mathbf{T}$ is often referred to as the 'listening matrix' in the literature, since $T_{ji} = 0$ indicates that node j does not listen to i's opinion. A key feature of node-level models is that each agent i expresses a *single* opinion $x_i(t)$ at time t and any opinion shared by i to their neighbours is the same value.

By contrast, an edge-level model focuses on the evolution of messages, where message content can differ along the outgoing edges from agent i. An edge-level opinion dynamic is therefore a sequence of message vectors $(\mathbf{s}(t))_{t\geq 0}$ with $\mathbf{s}(t) \in \mathbb{R}^m$ for all $t \geq 0$, where each $s_e(t)$ is interpreted as the message on directed edge e at time t.

To evolve messages on these directed edges through time, begin by stacking all outgoing and incoming messages into vectors $\mathbf{s}(t) \in \mathbb{R}^m$ and $\mathbf{r}(t) \in \mathbb{R}^m$ respectively. Then

collect the messaging rules $\mathbf{W}_i$ into a block diagonal matrix $\mathbf{W}$ which is defined as

$$\mathbf{W} := \operatorname{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_n).$$

We can then define the edge-level influence matrix as

$$\mathbf{M} := \mathbf{WP}.$$

Since $\mathbf{r}(t) = \mathbf{Ps}(t)$, it then follows that the edge-level opinion dynamic evolves via the linear process

$$\mathbf{s}(t+1) = \mathbf{M}\,\mathbf{s}(t). \tag{4}$$

In line with the assumptions of the DeGroot model, the matrix $\mathbf{M}$ is nonnegative and row-stochastic, since both $\mathbf{W}$ and $\mathbf{P}$ are nonnegative and row-stochastic.

The influence matrix $\mathbf{M}$ is the counterpart of the listening matrix $\mathbf{T}$ in a node-level model, since it encodes the influence that messages can have on other messages. Specifically, the element M_{fe} represents the direct influence that the message sent along edge e has on the message sent along edge f next period. For node-level models, opinions spread across the network by updating the content of the opinions held at each node. In contrast, for edge-level models, opinions spread across the network by updating the content of the messages being sent along each edge of the graph.

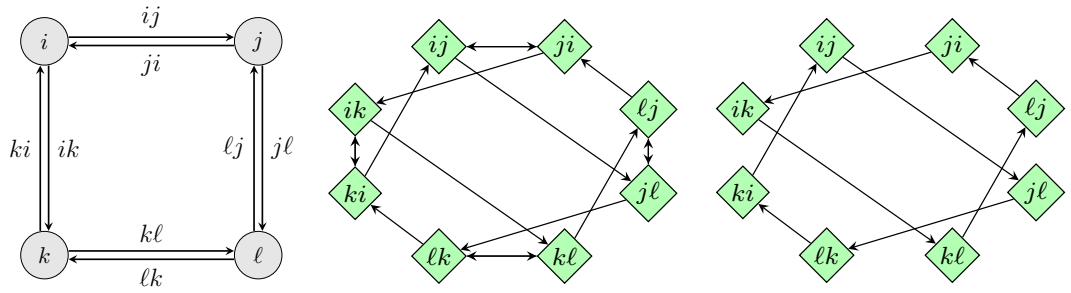
Whether a message on a given edge can influence the content of another message depends on both the structure of $\mathcal{G}$ and the particular messaging rules used by the agents. This motivates the following graph representation of message influence:

Definition 1. *Let $\mathcal{G} = (N, E)$ be the agent graph and let $\mathbf{M}$ be the edge-level influence matrix induced by the messaging rules $\mathbf{W}_1, \ldots, \mathbf{W}_n$. The* message graph *is the directed graph $\mathcal{G}_M = (E, \mathcal{E})$ whose nodes are the directed edges of $\mathcal{G}$, where $(e, f) \in \mathcal{E}$ if and only if $M_{fe} > 0$.*

In the original agent graph $\mathcal{G}$, node i is adjacent to node j whenever it is possible to move from i to j in a single step. In the message graph $\mathcal{G}_M$, a message on $ij \in E$ is adjacent to a message on $k\ell \in E$ whenever it is possible to move from ij to $k\ell$ in a single step. Since a messaging rule only permits influence from message e to f when $\mathcal{T}(e) = \mathcal{S}(f)$, the message graph is always (weakly) a subgraph of the directed line graph of $\mathcal{G}$, which is denoted by $\mathcal{L}(\mathcal{G})$.[4]

For any node-level opinion dynamic, each agent i broadcasts a single opinion $x_i(t)$ to all neighbours, so the influence of i's opinion is mediated by paths in the agent graph $\mathcal{G}$.

[4]The directed line graph $\mathcal{L}(\mathcal{G})$ of graph $\mathcal{G}$ is a graph where the nodes are the directed edges of the original graph $\mathcal{G}$. The nodes of $\mathcal{L}(\mathcal{G})$ are connected by arcs from $e \in E$ to $f \in E$ if and only if $\mathcal{T}(e) = \mathcal{S}(f)$.



(a) Agent graph

(b) Message graph for the uniform messaging rule

(c) Message graph for the nonbacktracking messaging rule

Figure 3: Underlying agent graph and two corresponding message graphs.

For any edge-level opinion dynamic, each directed edge e carries its own message $s_e(t)$, so influence is instead mediated by paths in the message graph $\mathcal{G}_M$.

A feature of the message graph is that it is induced by the particular messaging rules chosen, and therefore different messaging rules can remove some arcs relative to $\mathcal{L}(\mathcal{G})$. Figure 3 displays two message graphs for different messaging rules on the same agent graph $\mathcal{G}$. Panel (b) demonstrates that under the uniform messaging rule the message on ki can influence the messages on ij and ik in one step, and messages such as $j\ell$ or $k\ell$ in two steps via intermediate messages in the message graph. Panel (c) shows that if all agents use the nonbacktracking messaging rule then some arcs between edges are removed relative to the uniform messaging rule. For example, the message on ki can no longer directly influence the content of the message on ik, since it involves backtracking. Moreover, the message graph is now comprised of two disconnected components which cannot influence one another.

A directed graph is *strongly connected* if for every ordered pair of nodes (i, j) there exists a directed path from i to j. When the graph of positive weights in $\mathbf{T}$ is strongly connected then the opinion of any agent can (directly or indirectly) influence the opinion of any other agent in a node-level model. Likewise, when the message graph $\mathcal{G}_M$ is strongly connected then the content of any message can (directly or indirectly) influence the content of any other message.

To prove results on the convergence of edge-based opinion dynamics, let $\mathcal{G} = (N, E)$ and $\mathcal{G}_M = (E, \mathcal{E})$ have associated influence matrices $\mathbf{T}$ and $\mathbf{M}$. A nonnegative matrix is *irreducible* if the graph induced by its positive weights is strongly connected; thus $\mathbf{M}$ is irreducible if $\mathcal{G}_M$ is strongly connected. As is well-known in the literature, if a nonnegative and row-stochastic matrix $\mathbf{Z}$ is irreducible, then there exists a unique and strictly positive left eigenvector $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}^T\mathbf{Z} = \boldsymbol{\pi}^T$ and $\boldsymbol{\pi}^T\mathbf{1} = 1$.

A directed graph is *aperiodic* if the greatest common divisor of the lengths of its simple directed cycles is 1. A nonnegative matrix is said to be aperiodic if the graph

induced by its positive weights is aperiodic. Thus, $\mathbf{M}$ is aperiodic if $\mathcal{G}_M$ is aperiodic. As is also well-known, if a nonnegative and row-stochastic matrix $\mathbf{Z}$ is both irreducible and aperiodic then

$$\lim_{t\to\infty} \mathbf{Z}^t \;=\; \mathbf{1}\boldsymbol{\pi}^T, \tag{5}$$

and so the powers of $\mathbf{Z}$ converge to a rank-one matrix. Applying this fact to the edge to edge influence matrix $\mathbf{M}$ gives the following result:

Lemma 3.1. *Let $\mathbf{M} \in \mathbb{R}^{m\times m}$ be a nonnegative row-stochastic influence matrix. If $\mathbf{M}$ is irreducible and aperiodic, then for any initial message vector $\mathbf{s}(0) \in \mathbb{R}^m$*

$$\lim_{t\to\infty} \mathbf{s}(t) = \mathbf{1}\,\boldsymbol{\pi}^T\mathbf{s}(0),$$

where $\boldsymbol{\pi}^T\mathbf{M} = \boldsymbol{\pi}^T$, $\boldsymbol{\pi} \in \mathbb{R}^m_{++}$ and $\boldsymbol{\pi}^T\mathbf{1} = 1$.

Proof. See Appendix A. □

Under the assumptions of Lemma 3.1, the content of each message converges to a constant value $c = \boldsymbol{\pi}^T\mathbf{s}(0)$ where $\boldsymbol{\pi}$ is the unique left eigenvector of $\mathbf{M}$ associated with the eigenvalue 1. The value of this consensus message therefore depends on the initial messages at $t = 0$ and the weights of each edge according to $\boldsymbol{\pi}$.

Assumption 1. *The graph $\mathcal{G}$ is strongly connected and aperiodic.*

To guarantee that $\mathcal{G}_M$ is strongly connected and aperiodic, it is sufficient to assume that Assumption 1 holds and that each messaging rule $\mathbf{W}_i$ allows all incoming messages to influence all outgoing messages. This condition is formalised using the definition and result below:

Definition 2. *A collection $\mathcal{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)$ of messaging rules is* strictly positive *if every entry of each $\mathbf{W}_i$ is strictly positive.*

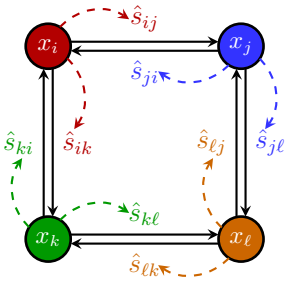
Proposition 3.2. *Suppose that $\mathcal{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)$ is strictly positive. If $\mathcal{G}$ satisfies Assumption 1, then $\mathbf{M}$ is irreducible and aperiodic.*

Proof. See Appendix A. □

Restricting attention to collections of messaging rules which are strictly positive will ensure that opinions converge to a consensus message whenever a node-based DeGroot dynamic would converge to a consensus. The following subsection now compares the edge-based and node-based models in more detail.

$$\mathbf{L} = \begin{array}{c} \\ ij \\ ik \\ ji \\ j\ell \\ ki \\ k\ell \\ \ell j \\ \ell k \end{array} \begin{array}{c} \begin{array}{cccc} i & j & k & \ell \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array}$$

(a)



(b)

Figure 4: (a) The lifting matrix $\mathbf{L}$ with labelled rows and columns; (b) lifting node-level opinions to edges

3.1 Lifted Dynamics

Under a node-level model, agent i's opinion at time t will influence agent j's opinion at time $t+1$ if $T_{ji} > 0$. To translate this dynamic to an edge-based model, one can interpret agent i as sending a message along every outgoing edge at time t that is equal to their current opinion $x_i(t)$.

The opinion of each node is therefore 'lifted' onto their outgoing edges and then transmitted to neighbours. Let us then define the *lifting matrix* $\mathbf{L} \in \mathbb{R}^{m \times n}$ as

$$L_{e,i} = \begin{cases} 1 & \text{if } \mathcal{S}(e) = i, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Since each edge can only have a single source node, each row of $\mathbf{L}$ contains a single nonzero entry. For the node-level dynamic $(\mathbf{x}(t))_{t\geq 0}$ with influence matrix $\mathbf{T}$, its corresponding edge-level dynamic $(\hat{\mathbf{s}}(t))_{t\geq 0}$ is found by lifting each node's opinion onto their outgoing edges at each t. Figure 4 illustrates the lifting process for a simple graph where these lifted messages are given by

$$\hat{\mathbf{s}}(t) = \mathbf{L}\mathbf{x}(t).$$

By lifting these node-level opinions onto edges it is possible to compare the trajectories of node-level and edge-level opinion dynamics directly. The question is then whether a given edge-level dynamic is simply a different representation of some node-level dynamic or genuinely novel. More specifically, if we are given an edge-level dynamic $(\mathbf{s}(t))_{t\geq 0}$, when does there exist a lifted node-level dynamic $(\hat{\mathbf{s}}(t))_{t\geq 0}$ which represents the same sequence of messages, given the initial conditions $\mathbf{s}(0) = \hat{\mathbf{s}}(0)$? The following result gives necessary and sufficient conditions for this representation.

Lemma 3.3. *Let $(\mathbf{x}(t))_{t\geq 0}$ be a node-level dynamic with initial opinion $\mathbf{x}(0) \in \mathbb{R}^n$ and let $(\mathbf{s}(t))_{t\geq 0}$ be an edge-level dynamic with initial opinion $\mathbf{s}(0) = \mathbf{L}\mathbf{x}(0)$. Then*

$$\mathbf{s}(t) = \hat{\mathbf{s}}(t) = \mathbf{L}\mathbf{x}(t)$$

for all $t > 0$ and every $\mathbf{x}(0)$ if and only if each $\mathbf{W}_i$ has identical rows with $W_{ik,ij} = T_{ij}$ for all $j, k \in N_i$.

Proof. See Appendix A. □

Since agents in a node-level model broadcast the same opinion to all neighbours, each $\mathbf{W}_i$ must have identical rows to represent the same dynamic. The key insight of Lemma 3.3 is that once agents begin to tailor their messages by recipient, then the opinion dynamics can no longer be represented by a node-level model of the form $\mathbf{x}(t+1) = \mathbf{T}\mathbf{x}(t)$.

Lemma 3.3 provides a way to obtain an equivalent edge-level dynamic from a node-level dynamic. To do this we construct an influence matrix $\mathbf{M}$ such that

$$M_{ef} = \begin{cases} T_{\mathcal{S}(e)\mathcal{S}(f)} & \text{if } \mathcal{S}(e) = \mathcal{T}(f) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Under this construction we ensure that each outgoing message contains an identical mixture of incoming messages. In particular, for an outgoing message on edge ik, the weight placed on r_{ij} is $W_{ik,ij} = M_{ik,ji} = T_{ij}$. Therefore, each outgoing message on edge ik places the same weight on r_{ij} as agent i places on j's opinion in a node-based model.

For example, the DeGroot model on an undirected and unweighted graph has influence matrix $\mathbf{T}$ given by

$$\mathbf{T}^{\mathrm{DG}} = [T_{ij}^{\mathrm{DG}}] \quad \text{where} \quad T_{ij}^{\mathrm{DG}} = \begin{cases} \dfrac{1}{d_i}, & \text{if } ij \in E \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, by setting each $W_{ik,ij} = \frac{1}{d_i}$ it is possible to replicate the opinion dynamics of the baseline DeGroot model. The sequence of messages generated by the equation $\hat{\mathbf{s}}^{\mathrm{DG}}(t) = \mathbf{L}(\mathbf{T}^{\mathrm{DG}})^t\mathbf{x}(0)$ would be identical to that generated by messaging rules $\mathcal{W}^{\mathrm{U}} = (\mathbf{W}_1^{\mathrm{U}}, \ldots, \mathbf{W}_n^{\mathrm{U}})$ corresponding to the uniform messaging rule shown in Figure 1. The uniform messaging rule is in that sense equivalent to the DeGroot dynamic on an undirected and unweighted network, since it replicates the same sequence of messages.

3.2 Doubly Stochastic Messaging Rules

When messages are not differentiated by target then any resulting opinion dynamic can be analysed using node-based models. To explore opinion dynamics which are not representable within the existing frameworks, agents must differentiate the content of the message conditional on the target. This subsection introduces a class of differentiated messaging rules which proves to be a natural extension of the node-level DeGroot process on an undirected graph.

The assumption that the influence matrix $\mathbf{M}$ is row-stochastic is standard in models of non-Bayesian learning and opinion dynamics in networks, but now consider the special case where $\mathbf{1}^T\mathbf{W}_i = \mathbf{1}^T$ for each $\mathbf{W}_i$, so that the columns of $\mathbf{W}_i$ also sum to 1. This can be interpreted as a requirement that no source of opinion is over- or under-represented in aggregate when considering all messages that are sent by i. If each $\mathbf{W}_i$ is column-stochastic then the average message received by i at time t, given by $\frac{1}{d_i}\mathbf{1}^T\mathbf{r}_i(t)$, is equal to the average message sent by i at time $t+1$, given by $\frac{1}{d_i}\mathbf{1}^T\mathbf{s}_i(t+1) = \frac{1}{d_i}\mathbf{1}^T\mathbf{W}_i\mathbf{r}_i(t)$. Therefore, such *doubly stochastic* messaging rules will only aggregate messages without biasing the average message content.

The nonbacktracking messaging rule defined in (2) is one example of a differentiated messaging rule which is doubly stochastic. This messaging rule addresses a criticism of the standard DeGroot model, by avoiding the direct echo chamber effect where opinions from j would be immediately relayed back by i the following period.[5]

Any doubly stochastic messaging rule can also be extended to include private opinions, represented by a message contained on a self-loop. Appendix B.1 illustrates such a messaging rule with private opinions, where a parameter controls the persistence of the agent's private opinion relative to the opinions of other agents.

When comparing alternative messaging rules, it is important to distinguish between differences in transient dynamics and differences in the long run consensus opinion. For a fixed initial condition $\mathbf{s}(0)$, the opinions generated by any two doubly stochastic rules will generally differ in the short run, for example when comparing the uniform rule with the nonbacktracking rule. However, as the result in Proposition 3.4 shows, if opinions converge to a consensus then they will ultimately converge to the same consensus value in the long run.

Proposition 3.4. *Suppose that* $\mathbf{M}$ *is primitive. The consensus message is* $c = \frac{1}{m}\sum_{e\in E} s_e(0)$ *for every* $\mathbf{s}(0) \in \mathbb{R}^m$ *if and only if* $\mathbf{W}_i$ *is doubly stochastic for all* $i \in N$.

Proof. See Appendix A. □

[5] An indirect echo chamber effect may still arise in this case if there is clustering in the network. If i and j have a neighbour in common then the opinions of i and j may already be correlated before any messages between the pair. The nonbacktracking messaging rule therefore only addresses the direct echo chamber effect.

Proposition 3.4 demonstrates that the lifted DeGroot dynamic discussed in Section 3.1 is a special case of a more general family of doubly stochastic messaging rules where agents aggregate messages in an unbiased manner. Doubly stochastic rules preserve the mean message at each iteration, so it follows that the stationary distribution should put equal weight on all initial messages. The above result does not restrict each agent to use the same doubly stochastic messaging rule, and moreover any convex combination of doubly stochastic messaging rules is also doubly stochastic.

Lifting each node's initial opinion to their outgoing edges means that this opinion is replicated d_i times in the message vector $\mathbf{s}(0)$. The influence of each node on the consensus message is therefore proportional to their degree whenever they are using doubly stochastic messaging rules. This agrees exactly with the result of Golub and Jackson (2010), where each agent has influence in proportion to their degree. Golub and Jackson (2010) then show how a large society of agents who aggregate information by repeatedly averaging the opinions of their neighbours can end up converging on the true state, provided that their influence shrinks to zero as $n \to \infty$.

Given that each node has the same influence on the consensus message as in Golub and Jackson (2010), a straightforward extension of their wisdom of crowds result is provided in Appendix B.2, under the assumption of doubly stochastic messaging rules.

Extending their result to this broader class of heuristics highlights a connection between the initial seeding of messages and wisdom. Since doubly stochastic rules will uniformly aggregate all initial messages on edges, the influence of high degree agents on the consensus message only comes in via their influence on the initial message vector $\mathbf{s}(0)$. Although not explored further in this paper, this suggests that allowing agents to send different initial messages to different neighbours (e.g. as in Bowen et al. (2023)) may weaken the restrictions on network structure which are required for wisdom.

4 Conformity and the Speed of Convergence

The previous section demonstrated that unbiased messaging rules lead to a consensus message which places equal weight on all initial messages. Although these messaging rules lead to the same consensus opinion, different doubly stochastic rules will typically have different transient dynamics and different rates of convergence.

Golub and Jackson (2012) show how bottlenecks in the network due to homophily can limit the convergence speed of the consensus building process. Differentiated messaging introduces another mechanism which may influence the convergence speed, by affecting the mixing of opinions in the message graph. Holding network structure fixed, how do differentiated messaging rules compare to the uniform rule in terms of amplifying or attenuating the effects of any structural bottlenecks in the network?

When agents select different messages for different neighbours it is natural to ask why they might wish to differentiate their messages at all. If agents are simply motivated to accurately represent what they just heard, then the uniform messaging rule should be used. One reason to differentiate messages by target is provided by the literature on *audience tuning* in social psychology.

Audience tuning is the act of modifying a message to match the attitude of a listener and is well-studied following the work of Higgins and Rholes (1978).[6] Economists have documented similar effects experimentally, where individuals have been shown to tune their messages in order to ingratiate themselves with the recipient (Robin et al., 2014), or to avoid punishment (Dvorak et al., 2025). This idea has also been explored within the context of naive learning models by Buechel et al. (2015) and Anufriev et al. (2023), however these existing models restrict agents to broadcasting an identical message to all neighbours. The appropriate audience in their setting is taken to be the whole neighbourhood of each agent, but when agents can differentiate their message by target, then the appropriate audience is the individual recipient.

When messages are differentiated more coarsely, for example by an agent's type or group rather than on an individual basis, then the group's average opinion becomes the relevant reference point. Conformity of this type is also well-studied, both in psychology and economics, and may be driven by factors such as social pressure (Asch, 1955), reputational concerns (Akerlof, 1980; Bernheim, 1994) or a direct preference to match the norm of their reference group (Zafar, 2011). Group identity may also influence which incoming messages the sender views as credible and therefore which are deemed to be most informative (Zou and Xu, 2023).

For the remaining sections of the paper we therefore model agents as being driven by these social and psychological motives to send messages which conform to the opinions of a reference group. Let $\bar{r}_i(t-1) := \frac{1}{d_i}\sum_{j' \in N_i} r_{ij'}(t-1)$ denote the average message received at $t-1$ and let $\mathcal{R}_{ij}(t-1)$ denote a reference opinion which i is motivated to conform to when communicating with j. This reference opinion is defined to be a convex combination of incoming messages, so that $\mathcal{R}_{ij}(t-1) = {\mathbf{w}'_{ij}}^T \mathbf{r}_i(t-1)$ where $\mathbf{w}'_{ij}$ is a vector of convex weights.

Each agent $i \in N$ then has a utility function given by

$$u_i(\mathbf{s}_i(t)) = \sum_{j \in N_i} \Big[B_{ij} - \gamma_{ij}\big(s_{ij}(t) - \mathcal{R}_{ij}(t-1)\big)^2 - (1-\gamma_{ij})\big(s_{ij}(t) - \bar{r}_i(t-1)\big)^2 \Big], \quad (8)$$

[6] Participants in Higgins and Rholes (1978) were presented with information regarding a fictional target individual and asked to describe this individual to a recipient. Participants were observed to tune their messages to different audiences, e.g. by describing the individual as 'confident' rather than 'conceited' when the receiver was known to already view the target individual favourably. These results were replicated and extended in several further works; see Higgins (1999) for an overview.

where B_{ij} is an exogenous benefit to interacting with agent j and the parameter $\gamma_{ij} \in [0,1]$ measures the importance of conformity along edge ij. Maximising (8) with respect to each $s_{ij}(t)$ then results in

$$s_{ij}(t) = \gamma_{ij}\,\mathcal{R}_{ij}(t-1) \;+\; (1-\gamma_{ij})\bar{r}_i(t-1).$$

In the baseline case where $\gamma_{ij} = 0$ for all $ij \in E$, agents send messages which replicate the standard DeGroot opinion dynamic on an undirected network. An increase in γ_{ij} means that the agent increasingly emphasises the tailored mix of opinions $\mathcal{R}_{ij}$ over the mean incoming message $\bar{r}_i(t-1)$. There are two complementary interpretations of what γ_{ij} measures; firstly, it measures the extent to which agents are deviating from the benchmark case of uniform messaging towards differentiated messaging; secondly, since $\bar{r}_i(t-1)$ is the agent's belief in the standard DeGroot framework, γ_{ij} measures the degree of performative messaging which agent i undertakes and therefore their preference for conformity over an accurate representation of the information that they hold.

4.1 Pairwise Opinion Conformity

To study the implications of opinion conformity on convergence speed, suppose that the reference opinion is $\mathcal{R}_{ij}(t-1) = r_{ij}(t-1)$ and that all agents have the same conformity parameter $\gamma_{ij} = \gamma$ for all $ij \in E$. When $\gamma > 0$ each recipient j receives a message that is disproportionately tailored toward j's own previous statement rather than a neutral aggregation of what i has heard. As $\gamma \to 1$ the opinions heard from each neighbour are reflected back (i.e. $s_{ij}(t+1) \approx r_{ij}(t)$) and the mixing of opinion across neighbouring edges is prevented in the limit. For $\gamma \in (0,1)$, let the *pairwise conformity* messaging rule be defined as

$$\mathbf{W}_i^{\mathrm{PC}}(\gamma) = \gamma \mathbf{I}_{d_i} + (1-\gamma)\frac{1}{d_i}\mathbf{1}_{d_i}\mathbf{1}_{d_i}^T.$$

Here we restrict attention to the case where $\gamma \in (0,1)$, even though a negative γ may still yield a nonnegative $\mathbf{W}_i^{\mathrm{PC}}(\gamma)$.[7]

It is well-known (Golub and Jackson, 2012) that the convergence speed of consensus dynamics on networks is related to the size of the second largest eigenvalue modulus (SLEM) of the influence matrix. Existing results, which are closely related to results in the theory of Markov chains, provide tight bounds on convergence speed in terms of the SLEM. A key difference between the present edge-based model and the node-based models of Golub and Jackson (2012) and Cheng et al. (2026) is that now the eigenvalues of $\mathbf{M}$ are determined by both the structure of the underlying network *and* the structure of the messaging rules used.

[7] Discussion of this issue is deferred until Section 4.2.

Fixing a particular network structure, we may then ask how opinion conformity affects convergence speed and how this interacts with network structure. Proposition 4.1 provides an explicit solution for the SLEM of the influence matrix in terms of the conformity parameter γ and the SLEM of the matrix $\mathbf{T} = \mathbf{T}^{\mathrm{DG}}$. A similar convergence speed result to Golub and Jackson (2012) and Cheng et al. (2026) is then provided in Proposition 4.3, which directly relates the consensus time to both the strength of conformity and underlying network structure.

To present these results, let $\lambda(\mathbf{Z})$ denote an eigenvalue of an $n \times n$ matrix $\mathbf{Z}$ and when these eigenvalues are real order them such that $\lambda_1(\mathbf{Z}) \geq \lambda_2(\mathbf{Z}) \geq \cdots \geq \lambda_n(\mathbf{Z})$. Let μ be reserved specifically for the eigenvalues of $\mathbf{M}$ with $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$ when real. For the matrices considered in this section we can then define the SLEM of $\mathbf{Z}$ as $\lambda_*(\mathbf{Z}) := \max\{|\lambda(\mathbf{Z})| : \lambda \neq 1\}$ and the SLEM of $\mathbf{M}$ as $\mu_* := \max\{|\mu(\mathbf{M})| : \mu \neq 1\}$.

To determine the eigenvalues governing our pairwise conformity dynamic, define the matrix

$$\mathbf{M}_\gamma := \left(\gamma \mathbf{I} + (1-\gamma)\mathbf{W}^{\mathrm{U}}\right)\mathbf{P},$$

where $\mathbf{W}^{\mathrm{U}} = \mathrm{diag}(\mathbf{W}_1^{\mathrm{U}}, \ldots, \mathbf{W}_n^{\mathrm{U}})$ and $\gamma \in (-1, 1)$. Restricting attention to $\gamma \in (0, 1)$ gives us the influence matrix under pairwise conformity.

Proposition 4.1. *Suppose that Assumption 1 holds and let $\lambda_* = \lambda_*(\mathbf{T}^{\mathrm{DG}})$. The second largest eigenvalue modulus (SLEM) of $\mathbf{M}_\gamma$ is given by*

$$\mu_*(\gamma) = \begin{cases} \dfrac{(1-\gamma)\lambda_* + \sqrt{(1-\gamma)^2\lambda_*^2 + 4\gamma}}{2}, & \gamma \geq \bar{\gamma}, \\ \sqrt{|\gamma|}, & \gamma < \bar{\gamma}, \end{cases}$$

where

$$\bar{\gamma} = -\frac{1 - \sqrt{1-\lambda_*^2}}{1 + \sqrt{1-\lambda_*^2}} \leq 0.$$

Proof. See Appendix A. □

Corollary 4.2. *μ_* is increasing in γ for $\gamma \in (\bar{\gamma}, 1)$.*

The above result provides the SLEM under the more general case of $\gamma \in (-1, 1)$, but if we restrict attention to $\gamma \in [0, 1)$ then the relevant solution is

$$\mu_*(\gamma) = \frac{(1-\gamma)\lambda_* + \sqrt{(1-\gamma)^2\lambda_*^2 + 4\gamma}}{2}. \tag{9}$$

When there is no differentiated messaging (i.e. when $\gamma = 0$) then $\mu_* = \lambda_*$ and so the speed of convergence is only determined by the network structure itself.

To establish a direct link between the SLEM provided in (9) and the convergence speed of the pairwise conformity opinion dynamic, we can follow Golub and Jackson (2012) in defining a measure of *consensus distance*. This allows us to quantify how far the agents are from reaching consensus at any time period t, for some fixed influence matrix $\mathbf{M}$ and initial condition $\mathbf{s}(0)$. Let the consensus distance be defined as

$$\mathrm{CD}(t; \mathbf{M}, \mathbf{s}(0)) := \|\mathbf{M}^t \mathbf{s}(0) - c\mathbf{1}\|_2,$$

where, under doubly stochastic messaging rules, the value c is given by $c = \frac{1}{m}\mathbf{1}^T\mathbf{s}(0)$. The measure of consensus distance defined in Golub and Jackson (2012) uses a weighted norm that weights each agent's distance from consensus by their degree, which is the number of messages they send per period. The present paper simply uses the ℓ_2 norm, but defined over the vector of all messages, which is therefore equivalent.

When $\mathbf{M}$ is doubly stochastic then it is a mean preserving transformation at each t and $\mathrm{CD}(t; \mathbf{M}, \mathbf{s}(0))$ is exactly the standard deviation of messages at time t, scaled by $\sqrt{m}$. The worst-case consensus distance at time t is then given by

$$\mathrm{CD}^*(t; \mathbf{M}) := \max_{\|\mathbf{s}(0)\|_2 = 1,\, \mathbf{s}(0) \perp \mathbf{1}} \mathrm{CD}(t; \mathbf{M}, \mathbf{s}(0)),$$

where the maximum is taken over all initial conditions which are orthogonal to the consensus vector.[8] Then following Golub and Jackson (2012), define the *consensus time* as

$$\mathrm{CT}(\varepsilon; \mathbf{M}) := \min\{t : \mathrm{CD}^*(t; \mathbf{M}) \leq \varepsilon\}.$$

The measure of consensus time is closely related to the notion of mixing time in Markov chains, where the most useful bounds typically arise in the case of reversible Markov chains.[9] The nature of the message sending process presents some challenges in determining more general bounds in this directed messaging model, because the associated Markov chain is generally non-reversible.[10]

However, by restricting attention to the pairwise conformity case, we can obtain similar consensus time results to Golub and Jackson (2012) and Cheng et al. (2026) for values

[8] Setting the initial opinion vector to sum to zero ensures that the consensus message is $c = 0$, under doubly stochastic messaging rules. This restriction is without loss of generality and simplifies the analysis since the initial disagreement at $t = 0$ is now $\|\mathbf{s}(0) - c\mathbf{1}\|_2 = \|\mathbf{s}(0)\|_2 = 1$.

[9] A Markov chain with a finite state space and transition matrix $\mathbf{M}$ is reversible with respect to a stationary distribution $\boldsymbol{\pi}$ if and only if it satisfies the detailed balance condition, that is, if for all pairs of states (edges) ij, jk we have

$$\pi_{ij} M_{ij,jk} = \pi_{jk} M_{jk,ij}.$$

[10] A message on edge jk can be influenced by ij when $M_{jk,ij} > 0$, but the message on ij is typically *not* influenced by jk. Unless $i = k$ the entry $M_{ij,jk}$ will be zero and the detailed balance condition cannot be satisfied.

of γ between 0 and 1:

Proposition 4.3. *Suppose that each agent uses the pairwise conformity messaging rule with $\gamma \in (0,1)$. If Assumption 1 holds then $\lim_{t\to\infty} \mathbf{s}(t) = c\mathbf{1}$ and the consensus time $\mathrm{CT}(\varepsilon;\mathbf{M})$ satisfies*

$$\left\lceil \frac{\log(1/\varepsilon)}{\log(1/\mu_*)} \right\rceil \leq \mathrm{CT}(\varepsilon;\mathbf{M}) \leq 1 + \left\lceil \frac{\log(1/\varepsilon)}{\log(1/\mu_*)} \right\rceil,$$

for $0 < \varepsilon < 1$.

Proof. See Appendix A. □

The consensus time of opinion dynamics under pairwise conformity is therefore approximately given by

$$\mathrm{CT}(\varepsilon;\mathbf{M}) \approx \frac{\log(1/\varepsilon)}{\log(1/\mu_*)}, \tag{10}$$

where this approximation differs from the true consensus time by less than two periods. This approximation to consensus time is increasing in μ_*, where from Proposition 4.1 we recall that μ_* encodes information about both the network structure and the messaging rule. Here μ_*, and therefore the right hand side of (10), is increasing in the strength of the conformity incentive γ due to Corollary 4.2.

A final observation to note is that when $\lambda_* > 0$ and we evaluate the first derivative of (9) at $\gamma = 0$, we see that it is positive and equal to $(1-\lambda_*^2)/\lambda_*$. This shows that some level of anticonformity, where agents *underweight* the previous opinion of the recipient and amplify the opinions of other neighbours, will lower the SLEM. Whether this lower SLEM in the $\gamma < 0$ case can translate to faster convergence to consensus in some broader sense will be discussed next.

4.2 Asymptotic Convergence Speed under Anticonformity

In this subsection we extend the model to allow for negative γ in the case of pairwise conformity. This analysis retains the assumption that row and column sums are equal to 1, but relaxes the earlier nonnegativity assumption on $\mathbf{W}_i$. If agents place a negative weight on the previous opinions of their neighbour then outgoing messages need not be convex combinations of previously heard messages. In fact, agents are in a sense able to manipulate opinions in a simple manner under this extension to the main model.[11]

By extending the pairwise conformity rule to the case of anticonformity and allowing for γ to be negative, we lose the connection between μ_* and consensus time which is

[11] A similar approach is taken by Della Lena et al. (2026) in their model of group based antagonism, where group identity determines whether opinions are conformed to rather than actively opposed.

provided by Proposition 4.3. Although these finite time bounds on convergence behaviour will no longer apply, the value of μ_* can still characterise the asymptotic convergence speed of the opinion dynamic as $t \to \infty$, even when $\gamma < 0$.

An alternative measure of the speed of convergence which is used in the literature on social learning is the asymptotic rate at which the error of the learning process approaches zero. For example, Molavi et al. (2018) measure how quickly the log error decreases as t grows large, where the error is taken to be the belief which is placed on the incorrect state of the world. Here we consider the same approach, but instead taking the error to be $\mathrm{CD}^*(t; \mathbf{M})$ which is the worst-case distance from consensus.

Let the *asymptotic convergence speed* be defined as

$$\rho_{\text{asym}} := -\lim_{t\to\infty} \frac{1}{t} \log\left(\mathrm{CD}^*(t; \mathbf{M})\right), \tag{11}$$

where taking negative logs ensures that a faster decay of the worst-case consensus distance leads to an *increase* in the asymptotic convergence speed. Therefore, the measure ρ_{asym} captures the average exponential rate of decay in worst-case consensus distance over the long run.

To apply this measure in the case of anticonformity ($\gamma < 0$), we still need to guarantee convergence to a unique consensus message. The matrix $\mathbf{M}$ could have mixed signs and therefore the results of Perron-Frobenius theory cannot be applied. The explicit formula for the SLEM of $\mathbf{M}$ provided by Proposition 4.1 is therefore instrumental in deriving the following result.

Theorem 4.4. *Suppose that all agents use the pairwise conformity rule but assume that* $\gamma \in (-1, 0]$. *If Assumption 1 holds then*

- *The limit* $\lim_{t\to\infty} \mathbf{M}^t$ *exists and is given by* $\lim_{t\to\infty} \mathbf{M}^t = \frac{1}{m}\mathbf{1}\mathbf{1}^T$.
- *The asymptotic convergence speed is* $\rho_{\text{asym}}(\mathbf{M}) = -\log(\mu_*)$.
- *The minimiser of* $\mu_*(\gamma)$ *is* $\gamma^* = \bar{\gamma} \leq 0$. *Moreover,* $\bar{\gamma} = 0$ *if and only if* $\lambda_*(\mathbf{T}) = 0$.

Proof. See Appendix A. □

Theorem 4.4 guarantees convergence to a single consensus opinion under anticonformity, provided that $|\gamma| < 1$. The level of γ which maximises convergence speed is negative unless the underlying agent graph already has $\lambda_*(\mathbf{T}) = 0$, in which case the uniform messaging rule is optimal. A key implication of this is that some level of anticonformity is always optimal unless the network is a complete graph with self-loops present for every agent.

To gain some further intuition for why anticonformity can speed up asymptotic convergence, suppose that we have a d-regular graph where $d \geq 3$. Then each message $s_{ij}(t+1)$ is given by

$$s_{ij}(t+1) = \gamma r_{ij}(t) + \frac{1-\gamma}{d} \sum_{k \in N_i} r_{ik}(t) = \frac{1+(d-1)\gamma}{d} r_{ij}(t) + \frac{1-\gamma}{d} \sum_{k \in N_i \setminus \{j\}} r_{ik}(t).$$

By setting

$$\gamma = \gamma^{\text{NB}} := -\frac{1}{d-1}$$

we can recover the nonbacktracking rule defined in (2). Decreasing γ further means the weight on $r_{ij}(t)$ becomes negative, which can be interpreted figuratively as agents playing "devil's advocate" by taking the opposite stance to their neighbour whilst also amplifying the opinions of all other neighbours.

The special case of a d-regular network allows us to express the optimal γ in a simple way. The optimal γ^* depends on the SLEM of $\mathbf{T}$ in the following manner:

$$\gamma^* \begin{cases} > \gamma^{\text{NB}} & \text{if } \lambda_*(\mathbf{T}) < \frac{2\sqrt{d-1}}{d}, \\ = \gamma^{\text{NB}} & \text{if } \lambda_*(\mathbf{T}) = \frac{2\sqrt{d-1}}{d}, \\ < \gamma^{\text{NB}} & \text{if } \lambda_*(\mathbf{T}) > \frac{2\sqrt{d-1}}{d}. \end{cases}$$

Graphs which naturally have faster mixing of opinions will have lower $\lambda_*(\mathbf{T})$ and therefore γ^* is closer to 0, which corresponds to the benchmark case of uniform weights on each incoming message. On the other hand, graphs which have bottlenecks will have a high $\lambda_*(\mathbf{T})$, which means that the optimal γ^* is more negative and potentially involves negative weights on $r_{ij}(t)$.

The intuition for why negative messaging weights can be optimal becomes apparent in the d-regular case. First suppose that $\lambda_*(\mathbf{T}) = \lambda_2(\mathbf{T})$ and let $\mathbf{A}$ denote the unweighted adjacency matrix of $\mathcal{G}$. Then since $\mathbf{T} = \frac{1}{d}\mathbf{A}$ and consequently $d\lambda_2(\mathbf{T}) = \lambda_2(\mathbf{A})$, a high value of $\lambda_2(\mathbf{T})$ indicates that the network can be partitioned into two groups which have many links within group and few links across groups.[12] Convergence to a consensus is therefore slowed if the network structure permits the opinions of isolated groups to persist

[12]This follows from the Courant-Fischer theorem which characterises $\lambda_2(\mathbf{A})$ as

$$\lambda_2(\mathbf{A}) = \max_{\mathbf{v} \perp \mathbf{1}, \mathbf{v} \neq 0} \frac{\mathbf{v}^T \mathbf{A} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \max_{\mathbf{v} \perp \mathbf{1}, \mathbf{v} \neq 0} \frac{\sum_{(i,j) \in E} v_i v_j}{\sum_i v_i^2}.$$

This means there exists a mean-zero vector which partitions the agents into two groups using the positive and negative entries of that vector. A high value of $\lambda_2(\mathbf{A})$ therefore implies that it is easy to partition the network into two loosely connected groups with many links between agents of the same sign under $\mathbf{v}$.

for long periods of time.

When the message $s_{ij}(t+1)$ uses negative messaging weights on $r_{ij}(t)$ this can lead to better diffusion of opinions across bottlenecks in the network via two mechanisms:

- Firstly, it reverses the stance of agent j and incorporates this into the message which j receives and thus forwards on to others. If j is mostly surrounded by neighbours with similar opinions, then this message from i disrupts the local consensus within that isolated group.

- Secondly, it amplifies the opinions of $k \in N_i \setminus \{j\}$. When there is a small number of shared neighbours between i and j, for example if they are in different groups, the amplification of these opinions permits faster diffusion.

A similar intuition can be provided for the case where $\lambda_*(\mathbf{T}) = |\lambda_n(\mathbf{T})|$. As discussed in Bramoullé et al. (2014), when $|\lambda_n(\mathbf{A})|$ is large then the network is close to bipartite and can be split into two groups with links predominantly *across* groups but not within. Opinion dynamics in these graphs can lead to oscillations, where group opinions are reflected back and forth with slow convergence. Negative message weights can play a similar disruptive role in this case, dampening the oscillatory behaviour and ensuring a faster convergence to consensus.

5 Group Conformity and Polarization

This section focuses on an intermediate case between the pairwise conformity rules studied in the previous section and the neighbourhood-wide conformity which is studied in the existing literature. Rather than conforming to an individual's opinion, conformity may be driven by group identity, in line with the experimental literature cited in Section 4. Suppose now that each node $i \in N$ has a publicly observable type $\theta_i \in \{A, B\}$, for example, gender, age group, political ideology or religious affiliation. When agent i sends a message to j, the reference opinion $\mathcal{R}_{ij}$ in (8) can now be conditioned on either i or j's type. For example, when i and j are both A types then i's message may skew more towards the average opinion of A type neighbours.

5.1 The Group Conformity Messaging Rule

Partition the set of agents N into $N^A := \{i \in N : \theta_i = A\}$ and $N^B := \{i \in N : \theta_i = B\}$. If $\theta_i = \theta$ then N^θ is agent i's in-group and $N \setminus N^\theta$ is agent i's out-group. Let the number of A and B neighbours for a node i be denoted by $d_i^A = |N_i \cap N^A|$ and $d_i^B = |N_i \cap N^B|$ respectively. Whenever $d_i^\theta > 0$ let $\bar{r}_i^\theta(t)$ denote the average message received by i from

neighbours of type $\theta \in \{A, B\}$:

$$\bar{r}_i^{\theta}(t) = \frac{1}{d_i^{\theta}} \sum_{k \in N_i \cap N^{\theta}} r_{ik}(t).$$

Suppose that the reference opinion under the utility function defined in (8) is now given by $\mathcal{R}_{ij} = \bar{r}_i^{\theta_i}$, so that for a positive γ_{ij} the message content is skewed towards the in-group's most recent opinion.[13] Letting $p_i = d_i^{\theta_i}/d_i$ denote the proportion of in-group neighbours, the resulting message is therefore:

$$s_{ij}(t) = \begin{cases} [\gamma_{ij} + (1-\gamma_{ij})p_i]\,\bar{r}_i^A(t-1) + [(1-\gamma_{ij})(1-p_i)]\,\bar{r}_i^B(t-1), & \text{if } \theta_i = A, \\ [\gamma_{ij} + (1-\gamma_{ij})p_i]\,\bar{r}_i^B(t-1) + [(1-\gamma_{ij})(1-p_i)]\,\bar{r}_i^A(t-1), & \text{if } \theta_i = B. \end{cases}$$

The message sent by i to j is a convex combination of the in-group and out-group average messages, with weights given by $\gamma_{ij} + (1-\gamma_{ij})p_i$ and $(1-\gamma_{ij})(1-p_i)$ respectively. A natural assumption to make is that the conformity incentive γ_{ij} should be allowed to differ by recipient type, for example due to peer pressure or status concerns. Therefore, define the in-group conformity parameter as γ_i^{in}, so that the weight put on $\bar{r}_i^{\theta_i}$ when i sends a message to any $j \in N^{\theta_i}$ is then given by $\alpha_i := \gamma_i^{\text{in}} + (1-\gamma_i^{\text{in}})p_i$. Similarly, when i sends a message to any $j \notin N^{\theta_i}$ the weight placed on $\bar{r}_i^{\theta_i}$ can be denoted by $\beta_i := \gamma_i^{\text{out}} + (1-\gamma_i^{\text{out}})p_i$.[14]

For an agent $i \in N^A$, the message sent along edge ij under the *group conformity* messaging rule can then be written as:

$$s_{ij}(t) = \begin{cases} \alpha_i \bar{r}_i^A(t-1) + (1-\alpha_i)\bar{r}_i^B(t-1), & \theta_j = A, \\ \beta_i \bar{r}_i^A(t-1) + (1-\beta_i)\bar{r}_i^B(t-1), & \theta_j = B. \end{cases} \tag{12}$$

In the case where $i \in N^B$, we can define the rule analogously by interchanging A and B. Figure 5 below displays an example of this messaging rule for an agent $i \in N^A$ with $d_i^A = 3$ and $d_i^B = 2$.

Message content is therefore conditioned on both the sender's type and the receiver's type. Since the same message is sent to all out-group members, this can be interpreted as a single 'external' message which is expressed to the out-group, whereas in-group members receive a single 'internal' message.

[13] We assume throughout this section that each node has at least one same-type neighbour. When there are no out-group neighbours then the out-group term is omitted and the messaging rule is uniform.

[14] An alternative specification would be to consider the reference opinion $\mathcal{R}_{ij}$ as the *out-group's* opinion, under a similar 'audience tuning' motivation to that presented in Section 4. The results presented in this section and in Proposition B.2 can be extended to that case.

$$\mathbf{W}_i^{\mathrm{GC}}(\alpha_i, \beta_i) = \begin{array}{c|ccccc} & A & A & A & B & B \\ \hline A & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{2}(1-\alpha_i) & \frac{1}{2}(1-\alpha_i) \\ A & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{2}(1-\alpha_i) & \frac{1}{2}(1-\alpha_i) \\ A & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{3}\alpha_i & \frac{1}{2}(1-\alpha_i) & \frac{1}{2}(1-\alpha_i) \\ B & \frac{1}{3}\beta_i & \frac{1}{3}\beta_i & \frac{1}{3}\beta_i & \frac{1}{2}(1-\beta_i) & \frac{1}{2}(1-\beta_i) \\ B & \frac{1}{3}\beta_i & \frac{1}{3}\beta_i & \frac{1}{3}\beta_i & \frac{1}{2}(1-\beta_i) & \frac{1}{2}(1-\beta_i) \end{array}$$

Figure 5: Group Conformity Messaging Rule

If $\gamma_i^{\text{in}}, \gamma_i^{\text{out}} \in (0,1)$ for all $i \in N$ then $\mathcal{W}^{\mathrm{GC}} = (\mathbf{W}_1^{\mathrm{GC}}, \ldots, \mathbf{W}_n^{\mathrm{GC}})$ is strictly positive and Proposition 3.2 implies that messages converge to a single consensus value. However, the matrix $\mathbf{W}$ is no longer guaranteed to be doubly stochastic as in the previous sections, so the influence of each initial opinion on the consensus message (encoded in the stationary vector $\boldsymbol{\pi}$) will no longer be uniform across edges. These consensus weights will generally depend on the details of the network structure, however Appendix B.4 explores a simplified case where $d_i = d$ and $p_i = p$ for all $i \in N$ and all agents of the same type have identical conformity parameters.

The main implications for this special case can be summarised as follows: firstly, the consensus message c is a weighted average of the mean messages sent within and between groups at $t = 0$; secondly, the influence of group θ's mean internal message at $t = 0$, denoted by $\bar{s}_{\theta\theta}(0)$, is increasing in both $\gamma_\theta^{\text{in}}$ and $\gamma_\theta^{\text{out}}$; and lastly, the value of c approaches $\bar{s}_{\theta\theta}(0)$ as $\gamma_\theta^{\text{in}} \to 1$. This implies that group θ can steer the final consensus towards $\bar{s}_{\theta\theta}(0)$ by creating strong incentives for in-group opinion conformity.

5.2 Persistent Disagreement

As discussed in Section 4, opinion conformity may reflect an exogenous preference to conform or it may arise due to social pressure and reputational concerns. In this subsection we consider the case where the incentives to conform to the in-group's opinion are strong when sending messages within group, but more moderate when sending messages across groups. Here we focus on the limiting case where $\gamma_i^{\text{in}} = 1$ for all agents $i \in N$, while maintaining the previous assumption that $\gamma_i^{\text{out}} \in (0,1)$. We shall see in this subsection that disagreement can persist in the long run, even when there is a moderate exchange of opinions across groups.

For $\theta, \theta' \in \{A, B\}$, let $E_{\theta\theta'} := \{ij \in E : i \in N^\theta,\ j \in N^{\theta'}\}$ denote the set of edges between θ and θ' types. The vector $\mathbf{s}_{\theta\theta'}(t)$ denotes the corresponding messages on these edges at time t. If $\gamma_i^{\text{in}} = 1$ for all agents $i \in N$ then no message content on non-AA edges can influence AA edges, and analogously for BB edges. We can then consider these

within-group messages as a closed *conversation*, where no influence on the discussion is felt from outside that group.

Lemma 5.1. *Suppose that* $\gamma_i^{\text{in}} = 1$ *and* $\gamma_i^{\text{out}} \in (0,1)$ *for all* $i \in N$*. If both* E_{AA} *and* E_{BB} *are nonempty, then the influence matrix* $\mathbf{M}$ *is reducible under the group conformity messaging rule.*

Proof. See Appendix A. □

To analyse the resulting opinion dynamics, maintain the assumption that the underlying agent graph is strongly connected so that any polarization in message content is driven by the messaging rules rather than by disconnectedness of the overall graph. In addition, we make the following assumption:

Assumption 2. *The subgraphs* $\mathcal{G}_A = (N^A, E_{AA})$ *and* $\mathcal{G}_B = (N^B, E_{BB})$ *are strongly connected and aperiodic.*

This assumption ensures that absent any influence from the out-group, each in-group is sufficiently connected that they can arrive at an internal consensus opinion. We can therefore view this as an assumption on the cohesiveness of each group. Under this assumption, there can be no group which acts as a bridge between different factions of the other group.

Proposition 5.2. *Suppose that* $\gamma_i^{\text{in}} = 1$ *and* $\gamma_i^{\text{out}} \in (0,1)$ *for all* $i \in N$*, and that Assumption 1 and Assumption 2 hold. If each agent uses the group conformity messaging rule then the long run messages on within-type edges are*

$$\mathbf{s}_{AA}(\infty) = c_{AA}\mathbf{1}, \qquad \mathbf{s}_{BB}(\infty) = c_{BB}\mathbf{1},$$

where

$$c_{AA} = \frac{1}{|E_{AA}|}\sum_{e \in E_{AA}} s_e(0), \qquad c_{BB} = \frac{1}{|E_{BB}|}\sum_{e \in E_{BB}} s_e(0).$$

Furthermore, the long run messages on cross-type edges $\mathbf{s}_{AB}(\infty)$ *and* $\mathbf{s}_{BA}(\infty)$ *are convex combinations of* c_{AA} *and* c_{BB}*.*

Proof. See Appendix A. □

Under the conditions of Proposition 5.2, the process reaches consensus *within* each conversation (AA and BB), but these consensus values generically differ (i.e. $c_{AA} \neq c_{BB}$ for generic initial conditions). Therefore, persistent disagreement in message content arises across types even if the agents are still communicating with one another.

The internal consensus values c_{AA} and c_{BB} are a uniform average of the group's initial messages sent along internal edges. If individuals send the same opinion to all

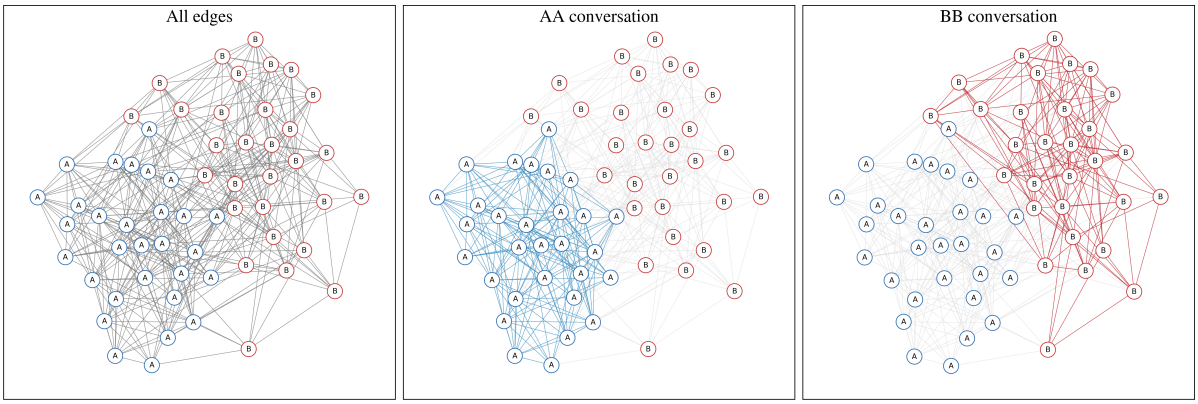


Figure 6: Closed Conversations - Stochastic Block Model ($n = 60$, $p^{in} = 0.35$, $p^{out} = 0.1$).

their neighbours at $t = 0$, then their initial opinion carries more weight when they have a higher degree within their in-group. Proposition 5.2 also demonstrates that the initial messages sent along cross-group links at $t = 0$ have no influence on any of the messages sent as $t \to \infty$. Therefore, individuals with a higher value of p_i will have a larger impact on their group's internal consensus message, holding all else equal.

Agents who use the group conformity rule are still influenced by neighbours from the out-group, but only when communicating with out-group members. Since the fraction of an agent's same-type neighbours will vary by position in the network, they are exposed to potentially different incoming messages even as t grows large. Therefore the cross-group messages which are sent in the limit will generally differ both across and within groups.

This variation is driven by differences in the direct and indirect exposure of each agent to the other group. Agents who have a larger proportion of in-group neighbours (larger p_i) will receive an average message which is close to their reference opinion, and therefore their message to the out-group is close to their in-group message. Conversely, having a large proportion of out-group friends means that the agent sends a more moderate message back along these cross-group edges. There are also indirect effects of having out-group neighbours who *themselves* have many cross-group neighbours, which further moderate the messages sent back and forth along the AB and BA edges.

Figure 7 provides an example of these dynamics, where messages are initialised by drawing node-level opinions from $\mathrm{U}[-1, 0]$ for A types and $\mathrm{U}[0, 1]$ for B types and then lifting each node's initial opinion to their outgoing edges. Within-type messages converge to separate consensus messages, whereas cross-type messages are non-uniform and determined by both exposure to the out-group and γ_i^{out} which here is set to $1/2$ for all $i \in N$. The second panel displays the average incoming message received by each node, which in the context of the DeGroot model is the node's belief. Although beliefs are not explicitly modelled here, the second panel shows how the average message content to which each node is exposed can vary, even after many rounds of opinion sharing.

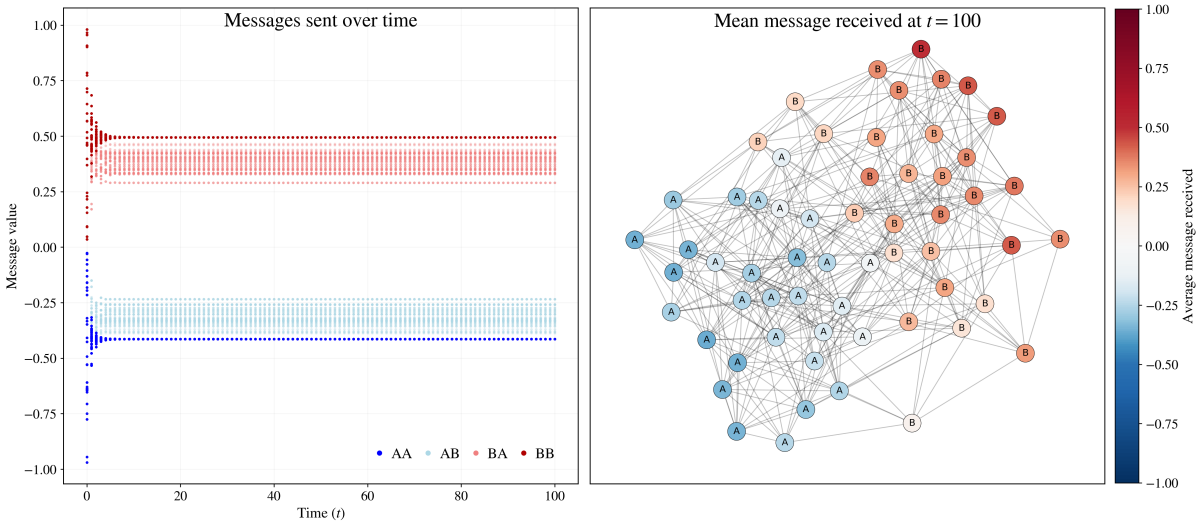


Figure 7: Message dynamics ($\gamma_i^{\text{in}} = 1; \gamma_i^{\text{out}} = 1/2$) - Stochastic Block Model ($n = 60$, $p^{in} = 0.35$, $p^{out} = 0.1$).

5.3 Comparison to Other Models of Polarization

The persistent disagreement established by Proposition 5.2 relates to disagreement in messages, since node-level beliefs are not explicitly studied in the model. Existing models typically study disagreement or polarization in beliefs, and so to compare the results of this section to those of the prior literature we must focus on the differences in long run message content. This subsection now discusses the relationship between this model and previous approaches, in terms of both the observable pattern of disagreement and the mechanism which drives it.

To identify the sources of disagreement in message content within previous literature, we can focus on one or several of the following mechanisms:

(i) Disconnectedness of the network.

(ii) Antagonism.

(iii) Stubbornness.

(iv) Homophily.

Although not their main focus, DeMarzo et al. (2003) show how isolated pockets of 'extremists' who are only influenced by each other's opinion can prevent convergence to consensus. The results in the present model are distinguished from theirs in two ways: firstly, polarization exists here even though the underlying graph is strongly connected; and secondly, agents are still influenced by neighbours from the out-group when sending their 'external' message to the out-group. Therefore communication and internalisation of opinions across groups still occur.

The bounded-confidence model of Hegselmann and Krause (2002) provides a separate explanation of polarization via disconnectedness of the network. There are some similarities to the present results such as the emergence of long run disagreement, however the model of Hegselmann and Krause (2002) assumes global knowledge of the opinions of all agents. The present model is entirely local, as nodes only need to observe each neighbour's type and the content of their messages.

A related recent contribution is Della Lena et al. (2026). In their model agents have access to exogenous sources of opinion and use negative weights to actively oppose the views of their out-group. With no external information sources, the agents in their model arrive at a within-group consensus which is diametrically opposed to the view of their out-group. Although their mechanism for polarization differs, the prediction of stable long run disagreement without anchoring to an exogenous source of information is shared with the present model. A key distinction is that the present model displays diversity in message content both across and within groups, precisely because agents vary their messages by target.[15]

Stubbornness is another source of long run disagreement within models of opinion dynamics. Acemoglu et al. (2013) consider subsets of stubborn agents who never change their opinion (e.g. media nodes) and who act as sources of opinion flow into the network, in a similar manner to the closed groups of extremists in DeMarzo et al. (2003). In-group conversations play a similar role to this in the present model, but some important differences between the dynamics remain. Firstly, the group conformity rule under $\gamma_i^{\text{in}} = 1$ leads the in-group to converge on an endogenous internal consensus, rather than remaining at their exogenously fixed view. Secondly, all agents in the present model are repeatedly influenced by the opinions of their neighbours, but only when sending a subset of messages.

Other models which generate polarization under naive learning in networks rely on homophily to drive some of the results, such as Melguizo (2019). The result of Bowen et al. (2023) relies on the combined assumptions that agents systematically misinterpret messages and have unbalanced neighbourhood compositions which favour certain types. Although homophily can play a role in increasing the level of message polarization within the current model, it is not required for Proposition 5.2 to hold.

[15] Opinion diversity both across and within groups is known as 'strong diversity' (Grabisch et al., 2023), in contrast to opinion clustering at the group level. For example, Cookson and Niessner (2020) study disagreement between different types of financial investors (e.g. fundamental vs. technical) in an online social network and document disagreement in opinions regarding particular investments both across and within groups, with across group disagreement larger in magnitude.

6 Conclusion

This paper has developed a tractable framework for studying opinion dynamics on networks in which message content can be differentiated by target. By shifting the focus from beliefs held on nodes to messages contained on edges, these opinion dynamics take place in a higher-dimensional state space than existing node-level approaches. This permits us to study a richer class of messaging rules, such as the pairwise and group conformity messaging rules. Doubly stochastic messaging rules were taken as a natural class to study, since these induce the same limiting consensus message as the standard DeGroot model, extending known wisdom of crowds results.

Pairwise opinion conformity results in the same consensus message as the DeGroot model, but convergence to this consensus is slower. Further analysis demonstrates that anticonformity, defined as the propensity to underweight or even negatively weight a neighbour's previous opinion in your response, can increase the asymptotic convergence speed. In cases where the network displays slow mixing of opinion, e.g. due to homophily, the optimal level of anticonformity has agents negatively weighting each neighbour's previous opinion when sending a message back the following period. This suggests that antagonism in opinion sharing can directly combat the effects of homophily if a society is slow to reach a consensus.

Lastly, the group conformity rule was shown to lead to persistent disagreement in message content, under the assumption that agents only relay in-group messages to other in-group members. There is diversity of message content both across and within groups, and polarization persists despite the fact that agents continue to exchange and incorporate messages across groups.

Whether observed polarization, e.g. in the media or in online social networks, reflects a genuine difference in beliefs or is merely performative in nature is a topic of debate (Bullock et al., 2015). The second aspect of this debate is less studied in the literature on networks and therefore this model makes a contribution by providing a formal account of how conformity to in-group norms can prevent a convergence to consensus, within a simple model of opinion dynamics.

The notion of an echo chamber is often taken within the literature to be a structural feature of the network. Levy and Razin (2019) suggest that echo chambers can be thought of as one consequence of endogenous segregation. This model shows that some features of echo chambers, such as opinion polarization, can be present without the network itself being segregated. Even agents in a complete network can experience persistent disagreement via the selective sharing of messages. These results therefore question the effectiveness of policy remedies aimed at reducing polarization by increasing integration within the network.

A Proofs

Proof of Lemma 3.1. As is well-known (e.g. Seneta (2006), Thm. 4.2), if a nonnegative and row-stochastic matrix $\mathbf{M}$ is irreducible and aperiodic then it is primitive and the following holds:

$$\lim_{t\to\infty} \mathbf{M}^t = \mathbf{1}\boldsymbol{\pi}^T, \tag{13}$$

where $\boldsymbol{\pi}^T\mathbf{M} = \boldsymbol{\pi}^T$, $\boldsymbol{\pi} \in \mathbb{R}^m_{++}$ and $\boldsymbol{\pi}^T\mathbf{1} = 1$. Therefore, by (4) we have that $\lim_{t\to\infty} \mathbf{s}(t) = \lim_{t\to\infty} \mathbf{M}^t\mathbf{s}(0) = \mathbf{1}\boldsymbol{\pi}^T\mathbf{s}(0)$ and the result that $\lim_{t\to\infty} s_e(t) = c$ for all $e \in E$ follows, where $c = \boldsymbol{\pi}^T\mathbf{s}(0)$. □

Proof of Proposition 3.2. Take any pair of edges $e, f \in E$. A message on e can only influence a message on f if $\mathcal{T}(e) = \mathcal{S}(f)$. Therefore $M_{fe} > 0$ implies $\mathcal{T}(e) = \mathcal{S}(f)$. Under strict positivity, if $\mathcal{T}(e) = \mathcal{S}(f)$ then $M_{fe} > 0$. Hence $e \to f$ in $\mathcal{G}_M$ if and only if $\mathcal{T}(e) = \mathcal{S}(f)$ and $\mathcal{G}_M$ coincides with $\mathcal{L}(\mathcal{G})$ the directed line graph of $\mathcal{G}$.

We can apply the following known result to establish that $\mathbf{M}$ is irreducible:

Theorem (Aigner (1967), Thm. 4)**.** *If $\mathcal{G}$ is a directed graph on $n \geq 3$ nodes, then $\mathcal{G}$ is strongly connected if and only if $\mathcal{L}(\mathcal{G})$ is strongly connected.*

Finally, to show that $\mathbf{M}$ is aperiodic under Assumption 1, note that every simple cycle $(i_0, i_1, i_2, \ldots, i_{k-1}, i_0)$ in $\mathcal{G}$ corresponds to a simple cycle of the same length in $\mathcal{G}_M$, given by $(i_0i_1, i_1i_2, \ldots, i_{k-1}i_0)$. As there are simple cycles of length 2 in $\mathcal{G}$ (e.g. $i \to j \to i$) and simple cycles of odd length (including self-loops of length 1) in $\mathcal{G}$ due to its aperiodicity, then the same must be true of $\mathcal{G}_M$ and hence $\mathbf{M}$ is aperiodic. □

Proof of Lemma 3.3. Suppose that each $\mathbf{W}_i$ has identical rows with $W_{ik,ij} = T_{ij}$ for all $j, k \in N_i$. At $t = 0$, $\mathbf{s}(0) = \mathbf{L}\mathbf{x}(0)$ implies that $s_{ji}(0) = r_{ij}(0) = x_j(0)$. Therefore

$$s_{ik}(1) = \sum_{j\in N_i} W_{ik,ij} r_{ij}(0) = \sum_{j\in N_i} T_{ij} x_j(0) = x_i(1).$$

Now suppose that $\mathbf{s}(t) = \mathbf{L}\mathbf{x}(t)$ for $t \geq 1$. Then $r_{ij}(t) = s_{ji}(t) = x_j(t)$ and therefore

$$s_{ik}(t+1) = \sum_{j\in N_i} W_{ik,ij} r_{ij}(t) = \sum_{j\in N_i} T_{ij} x_j(t) = x_i(t+1).$$

For the only if direction, suppose that at $t = 1$ we have $\mathbf{MLx}(0) = \mathbf{LTx}(0)$. Since $(\mathbf{ML} - \mathbf{LT})\mathbf{x}(0) = 0$ must hold for every $\mathbf{x}(0) \in \mathbb{R}^n$ this implies that $\mathbf{ML} = \mathbf{LT}$, and therefore element by element $(ML)_{e,j} = (LT)_{e,j}$.

For edge $e = ik$

$$(ML)_{ik,j} = \sum_{f \in E} M_{ik,f} L_{f,j} = \sum_{f:\mathcal{S}(f)=j} M_{ik,f} = M_{ik,ji},$$

where the final equality holds because the only edge with source equal to j that can influence the message on edge ik is ji.

The eth row of $\mathbf{L}$ has a 1 in the ith column if $\mathcal{S}(e) = i$ and a zero otherwise. Therefore for edge $e = ik$ the entry $(LT)_{ik,j}$ is given by

$$(LT)_{ik,j} = \sum_{\ell \in N} L_{ik,\ell} T_{\ell j} = T_{ij}.$$

Consequently, $\mathbf{MLx}(0) = \mathbf{LTx}(0)$ implies that $M_{ik,ji} = W_{ik,ij} = T_{ij}$, which holds for every edge $ik \in E$ and every $j \in N_i$. □

Proof of Proposition 3.4. This result is equivalent to showing that the stationary distribution $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{M}$ is uniform if and only if $\mathbf{W}_i$ is doubly stochastic for all $i \in N$.

Recall that $\mathbf{M} = \mathbf{WP}$ where $\mathbf{W}$ is block diagonal with blocks $\mathbf{W}_i$. Therefore $\mathbf{W}$ is doubly stochastic if and only if each block $\mathbf{W}_i$ is doubly stochastic. $\mathbf{P}$ is a permutation matrix which permutes the columns of $\mathbf{W}$ to form $\mathbf{M}$. Consequently, $\mathbf{M}$ is column-stochastic if and only if $\mathbf{W}$ is column-stochastic, since $\mathbf{M}$ is just a column permutation of $\mathbf{W}$. Moreover, by permuting the columns of $\mathbf{W}$ the row sums remain unchanged, therefore $\mathbf{M}$ is row-stochastic if and only if $\mathbf{W}$ is row-stochastic.

For the next step suppose that $\mathbf{M}$ is doubly stochastic. Then it is immediate that $\boldsymbol{\pi} = \frac{1}{m}\mathbf{1}$ solves $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{M}$ by the definition of column-stochasticity. Conversely, suppose that $\boldsymbol{\pi} = \frac{1}{m}\mathbf{1}$, then by stationarity $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{M}$. This implies that $\mathbf{1}^T = \mathbf{1}^T \mathbf{M}$ and hence taking the fth component yields $1 = \sum_e M_{e,f}$ for every column f. So $\mathbf{M}$ is column-stochastic and therefore doubly stochastic. □

Proof of Proposition 4.1. The proof of this proposition proceeds in several steps. We first establish some convenient facts regarding the lifting matrix $\mathbf{L}$. Then we use this result to solve the characteristic equation for a subset of the eigenvalues of $\mathbf{M}_\gamma$. Finally, we show that the second largest of these eigenvalues (in modulus) is μ_*.

Step 1: Let $\mathbf{A}$ denote the unweighted adjacency matrix of the symmetric graph $\mathcal{G}$ and let $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}$ denote the row-normalised version of this adjacency matrix, where $\mathbf{D} := \mathrm{diag}(d_1, \ldots, d_n)$ denotes the degree matrix.

Lemma A.1. *For the lifting matrix* $\mathbf{L}$ *defined as in* (6) *let* $\mathbf{R} := \mathbf{PL}$ *and let* $\bar{\mathbf{J}} := \mathrm{diag}\big(d_1^{-1}\mathbf{1}_{d_1}\mathbf{1}_{d_1}^T, \ldots, d_n^{-1}\mathbf{1}_{d_n}\mathbf{1}_{d_n}^T\big)$. *Then:*

(i) $\mathbf{L}^T\mathbf{L} = \mathbf{R}^T\mathbf{R} = \mathbf{D}$,

(ii) $\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T = \bar{\mathbf{J}}$,

(iii) $(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{R} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{R}^T\mathbf{L} = \mathbf{T}$.

Proof. Parts (i) and (iii) are standard results in the literature (Arrigo et al., 2019), but are proven here for completeness. By the definition of $\mathbf{R}$ it follows that $\mathbf{R}^T\mathbf{R} = \mathbf{L}^T\mathbf{P}^T\mathbf{P}\mathbf{L} = \mathbf{L}^T\mathbf{L}$, where the ijth entry of $\mathbf{L}^T\mathbf{L}$ is

$$(\mathbf{L}^T\mathbf{L})_{i,j} = \sum_{e\in E} L_{e,i}L_{e,j} = \begin{cases} d_i & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

since $L_{e,k} = 1$ if $\mathcal{S}(e) = k$, otherwise it is 0. Therefore $\mathbf{L}^T\mathbf{L} = \mathbf{R}^T\mathbf{R} = \mathbf{D}$, proving part (i).

For any $e, f \in E$,

$$\left(\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\right)_{e,f} = \sum_{i\in N} \frac{1}{d_i} L_{e,i}L_{f,i} = \begin{cases} \dfrac{1}{d_i} & \text{if } \mathcal{S}(e) = \mathcal{S}(f) = i, \\ 0 & \text{otherwise.} \end{cases}$$

When the directed edges of the network are ordered by their source node then this matrix is block diagonal, with ith block given by $d_i^{-1}\mathbf{1}_{d_i}\mathbf{1}_{d_i}^T$. Therefore,

$$\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T = \bar{\mathbf{J}},$$

which proves part (ii).

Lastly for any $i, j \in N$,

$$(\mathbf{L}^T\mathbf{R})_{i,j} = \sum_{e\in E} L_{e,i}R_{e,j} = \begin{cases} 1 & \text{if } ij \in E, \\ 0 & \text{otherwise,} \end{cases}$$

since $R_{e,j} = 1$ if $\mathcal{T}(e) = j$, otherwise it is 0.

This implies that $\mathbf{L}^T\mathbf{R} = \mathbf{A}$ and since the graph is undirected $\mathbf{R}^T\mathbf{L} = \mathbf{A}^T = \mathbf{A}$. Part (iii) then follows by the definition of $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}$.

□

Step 2: From Lemma A.1 the influence matrix $\mathbf{M}$ is given by

$$\mathbf{M}_\gamma = (\gamma\mathbf{I}_m + (1-\gamma)\bar{\mathbf{J}})\mathbf{P} = \gamma\mathbf{P} + (1-\gamma)\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{R}^T.$$

This step now follows Kempton (2016) in solving the characteristic equation via the generalised matrix determinant lemma (see e.g. Meyer (2000)). The characteristic equation is given by

$$\det(\mu\mathbf{I}_m - \gamma\mathbf{P} - (1-\gamma)\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{R}^T) = 0. \tag{14}$$

Take the case where $\mu^2 \neq \gamma^2$. Then $\mu\mathbf{I}_m - \gamma\mathbf{P}$ is invertible and via the generalised matrix determinant lemma the left hand side of (14) can be written as

$$\det(\mu\mathbf{I}_m - \gamma\mathbf{P})\det(\mathbf{I}_n - (1-\gamma)(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{R}^T(\mu\mathbf{I}_m - \gamma\mathbf{P})^{-1}\mathbf{L}).$$

It can be easily verified that the inverse matrix in the second determinant is given by

$$(\mu\mathbf{I}_m - \gamma\mathbf{P})^{-1} = \frac{1}{\mu^2-\gamma^2}(\mu\mathbf{I}_m + \gamma\mathbf{P}).$$

After applying this inverse and using Lemma A.1, the second determinant becomes

$$\det(\mathbf{I}_n - \frac{1-\gamma}{\mu^2-\gamma^2}((\mathbf{L}^T\mathbf{L})^{-1}\mathbf{R}^T(\mu\mathbf{I}_m + \gamma\mathbf{P})\mathbf{L})) = \det(\mathbf{I}_n - \frac{1-\gamma}{\mu^2-\gamma^2}(\mu\mathbf{T} + \gamma\mathbf{I}_n)).$$

Taking the factor $(\mu^2-\gamma^2)^{-1}$ outside the determinant and collecting terms inside gives

$$(\mu^2-\gamma^2)^{-n}\det((\mu^2-\gamma)\mathbf{I}_n - (1-\gamma)\mu\mathbf{T}).$$

It is well-known that the matrix $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}$ is similar to the symmetric matrix $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ via the similarity transformation $\mathbf{D}^{1/2}\mathbf{T}\mathbf{D}^{-1/2}$. Thus $\mathbf{T}$ has real eigenvalues and is diagonalisable, therefore $\mathbf{T} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$ with $\mathbf{U}\mathbf{V}^T = \mathbf{I}_n$. This leads to

$$(\mu^2-\gamma^2)^{-n}\det(\mathbf{U}((\mu^2-\gamma)\mathbf{I}_n-(1-\gamma)\mu\mathbf{\Lambda})\mathbf{V}^T) = (\mu^2-\gamma^2)^{-n}\prod_{i=1}^{n}(\mu^2-\gamma-(1-\gamma)\mu\lambda_i). \tag{15}$$

Therefore, every eigenvalue μ of $\mathbf{M}_\gamma$ satisfying $\mu^2 \neq \gamma^2$ is of the form

$$\mu = \frac{(1-\gamma)\lambda_i \pm \sqrt{(1-\gamma)^2\lambda_i^2 + 4\gamma}}{2}, \tag{16}$$

for some λ_i.

Any remaining eigenvalues of $\mathbf{M}_\gamma$ must therefore be of the form $\mu = \pm\gamma$. For any pair of solutions (μ_i^+,μ_i^-) to (16), possibly complex valued, it holds that

$$\begin{aligned} \mu_i^+\mu_i^- &= -\gamma \\ |\mu_i^+\mu_i^-| &= |\mu_i^+||\mu_i^-| = |\gamma| \end{aligned} \tag{17}$$

and therefore either $|\mu_i^+| \geq \sqrt{|\gamma|} > |\gamma|$ or $|\mu_i^-| \geq \sqrt{|\gamma|} > |\gamma|$, for $\gamma \in (-1, 0)$ and $\gamma \in (0, 1)$. If $\gamma = 0$ then this is at least weakly dominated by a solution to (16). As a result

$$\mu_* = \max_{i \geq 2} \left| \frac{(1-\gamma)\lambda_i \pm \sqrt{(1-\gamma)^2\lambda_i^2 + 4\gamma}}{2} \right|.$$

Step 3: To find μ_* consider two separate cases with $i \geq 2$ assumed throughout. First take the case where the discriminant $(1-\gamma)^2\lambda_i^2 + 4\gamma$ is negative and therefore μ_i is a complex eigenvalue. Any complex eigenvalues of a real matrix come in conjugate pairs, $\mu_i = a + b\,\mathrm{i}$ and $\overline{\mu_i} = a - b\,\mathrm{i}$. Using the fact that $|\mu_i| = |a + b\,\mathrm{i}| = \sqrt{a^2 + b^2}$ is equal to $|\overline{\mu_i}| = |a - b\,\mathrm{i}| = \sqrt{a^2 + (-b)^2}$, combined with (17), we see that $|\mu_i| = |\overline{\mu_i}| = \sqrt{|\gamma|}$ for any non-real eigenvalues of $\mathbf{M}_\gamma$.

When $(1-\gamma)^2\lambda_i^2 + 4\gamma \geq 0$ the eigenvalues μ_i are real. Using (17), either $|\mu_i^+| > \sqrt{|\gamma|}$, $|\mu_i^-| > \sqrt{|\gamma|}$, or $|\mu_i^+| = |\mu_i^-| = \sqrt{|\gamma|}$. Therefore, whenever the discriminant is nonnegative for some λ_i, there always exists a real eigenvalue which weakly dominates any non-real eigenvalue in modulus.

Therefore, if there exists a real μ_i then

$$\mu_* = \max\left\{ \left| \frac{(1-\gamma)\lambda_2 + \sqrt{(1-\gamma)^2\lambda_2^2 + 4\gamma}}{2} \right|, \left| \frac{(1-\gamma)\lambda_n - \sqrt{(1-\gamma)^2\lambda_n^2 + 4\gamma}}{2} \right| \right\}.$$

Recalling that $\lambda_* = \max\{|\lambda_2|, |\lambda_n|\}$, we conclude that

$$\mu_* = \begin{cases} \dfrac{(1-\gamma)\lambda_* + \sqrt{(1-\gamma)^2\lambda_*^2 + 4\gamma}}{2}, & \gamma \geq \bar{\gamma}, \\ \sqrt{|\gamma|}, & \gamma < \bar{\gamma}. \end{cases}$$

□

Proof of Proposition 4.3. The approach to the first part of the proof is similar to Golub and Jackson (2012) but under a different weighted norm. This first part proves a more general result that if $\mathbf{W}$ is positive definite then $\mathbf{M}$ is similar to a symmetric matrix and therefore $\mathbf{M}$ has all real eigenvalues.

Note that the block diagonal matrix $\mathbf{W}$ is positive definite if and only if each block is positive definite. If $\mathbf{W}_i = \mathbf{W}_i^{\mathrm{PC}}$ for each i, then for any nonzero $\mathbf{z} \in \mathbb{R}^{d_i}$ we have

$$\mathbf{z}^T\mathbf{W}_i^{\mathrm{PC}}\mathbf{z} = \gamma\mathbf{z}^T\mathbf{z} + (1-\gamma)\frac{1}{d_i}(\mathbf{z}^T\mathbf{1})^2 > 0,$$

and so $\mathbf{W}$ is positive definite for $\gamma \in (0, 1)$. If $\mathbf{W}$ is positive definite then a symmetric

$\mathbf{W}^{1/2}$ exists and is given by

$$\mathbf{W}^{1/2} = \left(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\right)^{1/2} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{V}^T.$$

Since $\mathbf{v}_1(\mathbf{W}) = \mathbf{1}$ and $\lambda_1(\mathbf{W}) = 1$, it follows from symmetry that $\mathbf{W}^{1/2}$ has both row sums and column sums equal to 1. Now for $\mathbf{M} := \mathbf{WP}$ we note that

$$\mathbf{W}^{-1/2}\mathbf{M}\mathbf{W}^{1/2} = \mathbf{W}^{1/2}\mathbf{P}\mathbf{W}^{1/2},$$

which means that $\mathbf{M}$ is similar to a symmetric matrix, since both $\mathbf{W}^{1/2}$ and $\mathbf{P}$ are symmetric. Hence $\mathbf{M}$ has real eigenvalues $(\mu_1, \ldots, \mu_m)$ and under a primitive $\mathbf{M}$ we have $\mu_1 = 1 > \mu_2 \geq \cdots \geq \mu_m > -1$.

If we then define the $\mathbf{W}^{-1}$ weighted norm as $\|\mathbf{v}\|_{\mathbf{W}^{-1}} := \sqrt{\mathbf{v}^T\mathbf{W}^{-1}\mathbf{v}}$, it is possible to prove the standard bound

$$\|\mathbf{M}^t\mathbf{s}(0) - c\mathbf{1}\|_{\mathbf{W}^{-1}} \leq \mu_*^t\|\mathbf{s}(0) - c\mathbf{1}\|_{\mathbf{W}^{-1}},$$

where $\mathbf{s}(0) - c\mathbf{1}$ is the error vector at $t = 0$. Iterating forward one period, $\mathbf{s}(1) - c\mathbf{1} = \mathbf{WP}(\mathbf{s}(0) - c\mathbf{1})$ since

$$\mathbf{WP}(\mathbf{s}(0) - c\mathbf{1}) = \mathbf{WPs}(0) - c\mathbf{1} = \mathbf{s}(1) - c\mathbf{1}.$$

Then letting $\mathbf{z}(0) := \mathbf{W}^{-1/2}(\mathbf{s}(0) - c\mathbf{1})$, we can establish that

$$\begin{aligned}
\|\mathbf{s}(1) - c\mathbf{1}\|_{\mathbf{W}^{-1}}^2 &= (\mathbf{s}(0) - c\mathbf{1})^T\mathbf{PWW}^{-1}\mathbf{WP}(\mathbf{s}(0) - c\mathbf{1}) \\
&= (\mathbf{s}(0) - c\mathbf{1})^T\mathbf{PWP}(\mathbf{s}(0) - c\mathbf{1}) \\
&= (\mathbf{z}(0)^T\mathbf{W}^{1/2})\mathbf{PWP}(\mathbf{W}^{1/2}\mathbf{z}(0)) \\
&= \mathbf{z}(0)^T(\mathbf{W}^{1/2}\mathbf{P}\mathbf{W}^{1/2})(\mathbf{W}^{1/2}\mathbf{P}\mathbf{W}^{1/2})\mathbf{z}(0).
\end{aligned}$$

Since $\mathbf{1}^T\mathbf{z}(0) = \mathbf{1}^T(\mathbf{s}(0) - c\mathbf{1}) = 0$, by standard results (Golub and Jackson, 2012) we have

$$\|(\mathbf{s}(1) - c\mathbf{1})\|_{\mathbf{W}^{-1}} = \|\mathbf{z}(1)\|_2 \leq \mu_*\|\mathbf{z}(0)\|_2 = \mu_*\|(\mathbf{s}(0) - c\mathbf{1})\|_{\mathbf{W}^{-1}},$$

where it is noted that $\mathbf{W}^{1/2}\mathbf{P}\mathbf{W}^{1/2}$ and $\mathbf{M}$ have identical eigenvalues. If we iterate on powers of $\mathbf{W}^{1/2}\mathbf{P}\mathbf{W}^{1/2}$ then for any $t \geq 0$ we have

$$\|(\mathbf{s}(t) - c\mathbf{1})\|_{\mathbf{W}^{-1}} = \|\mathbf{z}(t)\|_2 \leq \mu_*^t\|\mathbf{z}(0)\|_2 = \mu_*^t\|(\mathbf{s}(0) - c\mathbf{1})\|_{\mathbf{W}^{-1}}. \tag{18}$$

Then to translate this bound on the error into one in the ℓ_2 norm, note that $\mathbf{W}^{-1}$ has eigenvalues weakly greater than 1, and therefore $\|\mathbf{s}(t) - c\mathbf{1}\|_2 \leq \|\mathbf{s}(t) - c\mathbf{1}\|_{\mathbf{W}^{-1}}$. Moreover,

since $\mathbf{s}(t) - c\mathbf{1} = \mathbf{W}(\mathbf{r}(t-1) - c\mathbf{1})$ this implies that $\|\mathbf{s}(t) - c\mathbf{1}\|_{\mathbf{W}^{-1}} = \|\mathbf{r}(t-1) - c\mathbf{1}\|_{\mathbf{W}}$ for $t \geq 1$. Since $\mathbf{W}$ has largest eigenvalue equal to 1 then

$$\|\mathbf{r}(t-1) - c\mathbf{1}\|_{\mathbf{W}} \leq \|\mathbf{r}(t-1) - c\mathbf{1}\|_2 = \|\mathbf{s}(t-1) - c\mathbf{1}\|_2$$

where the final equality holds because $\mathbf{1}^T\mathbf{r}(t-1) = \mathbf{1}^T\mathbf{s}(t-1)$ when $\mathbf{W}$ is column-stochastic and $\|\mathbf{r}(t-1)\|_2 = \|\mathbf{P}\mathbf{s}(t-1)\|_2 = \|\mathbf{s}(t-1)\|_2$ since $\mathbf{P}$ is a permutation matrix. Applying this to (18) implies that for $t \geq 1$

$$\begin{aligned}
\|\mathbf{s}(t) - c\mathbf{1}\|_2 &\leq \|\mathbf{s}(t) - c\mathbf{1}\|_{\mathbf{W}^{-1}} \\
&\leq \mu_*^{t-1}\|\mathbf{s}(1) - c\mathbf{1}\|_{\mathbf{W}^{-1}} \\
&= \mu_*^{t-1}\|\mathbf{r}(0) - c\mathbf{1}\|_{\mathbf{W}} \\
&\leq \mu_*^{t-1}\|\mathbf{r}(0) - c\mathbf{1}\|_2
\end{aligned}$$

and therefore the bound

$$\|\mathbf{s}(t) - c\mathbf{1}\|_2 \leq \mu_*^{t-1}\|\mathbf{s}(0) - c\mathbf{1}\|_2. \tag{19}$$

Finally, to provide the consensus time result we require that

$$\|\mathbf{s}(t) - c\mathbf{1}\|_2 \leq \varepsilon$$

which, by applying (19), we can guarantee to hold when

$$\mu_*^{t-1}\|\mathbf{s}(0) - c\mathbf{1}\|_2 \leq \varepsilon.$$

Now rearranging and taking logs of both sides gives

$$(t-1)\log(\mu_*) \leq \log\left(\frac{\varepsilon}{\|\mathbf{s}(0) - c\mathbf{1}\|_2}\right).$$

Since $c = 0$ for an admissible initial condition, $\|\mathbf{s}(0) - c\mathbf{1}\|_2 = 1$. Then noting that $\log(\mu_*) < 0$ we have $t \geq 1 + \frac{\log(\varepsilon)}{\log(\mu_*)} = 1 + \frac{\log(1/\varepsilon)}{\log(1/\mu_*)}$ and we can write the consensus time as

$$\mathrm{CT}(\varepsilon; \mathbf{M}) \leq 1 + \left\lceil \frac{\log(1/\varepsilon)}{\log(1/\mu_*)} \right\rceil.$$

For the lower bound, let $\mathbf{v}_*$ be a unit right eigenvector with an eigenvalue $\widehat{\mu}$, where $|\widehat{\mu}| = \mu_*$. Then $\mathbf{1}^T\mathbf{v}_* = \mathbf{1}^T\mathbf{M}\mathbf{v}_* = \widehat{\mu}\mathbf{1}^T\mathbf{v}_*$ implies that $\mathbf{1}^T\mathbf{v}_* = 0$ because $\widehat{\mu} \neq 1$. By

setting $\mathbf{s}(0) = \mathbf{v}_*$ and noting that $c = 0$ in this case, we get

$$\mathrm{CD}(t; \mathbf{M}, \mathbf{v}_*) = \|\mathbf{M}^t \mathbf{v}_* - c\mathbf{1}\|_2 = \|\mathbf{M}^t \mathbf{v}_*\|_2 = \mu_*^t.$$

Then observe that $\mu_*^t \leq \varepsilon$ if and only if

$$t \geq \frac{\log(1/\varepsilon)}{\log(1/\mu_*)},$$

and since $\mathrm{CT}(\varepsilon; \mathbf{M})$ is the maximum over all admissible initial conditions, it follows that

$$\mathrm{CT}(\varepsilon; \mathbf{M}) \geq \left\lceil \frac{\log(1/\varepsilon)}{\log(1/\mu_*)} \right\rceil.$$

□

Proof of Corollary 4.2. From (9), μ_* satisfies

$$\mu_*^2 - \gamma - (1-\gamma)\mu_*\lambda_* = 0.$$

By implicit differentiation we have

$$2\mu_* \frac{d\mu_*}{d\gamma} + \lambda_*\mu_* - (1-\gamma)\lambda_* \frac{d\mu_*}{d\gamma} - 1 = 0.$$

Then by rearranging and again using (9) to substitute in for $2\mu_*$ we observe that

$$\frac{d\mu_*}{d\gamma} = \frac{1 - \lambda_*\mu_*}{\sqrt{(1-\gamma)^2\lambda_*^2 + 4\gamma}},$$

which is strictly positive for $\gamma \in (\bar{\gamma}, 1)$, since $1 > \lambda_*\mu_*$ and the discriminant is positive in that case. □

Proof of Theorem 4.4. If Assumption 1 holds then we know that $\lambda_1(\mathbf{T}) = 1$ and $\lambda_*(\mathbf{T}) < 1$. Moreover, irreducibility implies (via the Perron-Frobenius theorem) that $\lambda_1(\mathbf{T}) = 1$ is a simple eigenvalue, that is, it appears only once as a root of the characteristic polynomial. Furthermore, by setting $\mu = 1$ in (15) and then setting this equal to zero we observe that $\mu = 1$ can only be an eigenvalue of $\mathbf{M}$ when $\lambda_i = 1$. Since the other root of (16) is $-\gamma$ when $\lambda_i = 1$, the eigenvalue $\mu = 1$ is also a simple eigenvalue of $\mathbf{M}$ if $\gamma \in (-1, 1)$.

By Proposition 4.1, if $\gamma < \bar{\gamma}$ then $\mu_*(\gamma) = \sqrt{|\gamma|} < 1$, whereas if $\gamma \geq \bar{\gamma}$ then

$$\mu_*(\gamma) = \frac{(1-\gamma)\lambda_* + \sqrt{(1-\gamma)^2\lambda_*^2 + 4\gamma}}{2}. \tag{20}$$

Note that μ_* in (20) is strictly increasing in λ_* and that $\lambda_* = 1$ would imply that $\mu_* = 1$.

Therefore $\mu_* < 1$ whenever $\gamma \in (-1, 1)$ because $\lambda_* < 1$ if Assumption 1 is satisfied.

Now to show that $\lim_{t\to\infty} \mathbf{M}^t = \frac{1}{m}\mathbf{1}\mathbf{1}^T$ note that $\mathbf{M}\frac{1}{m}\mathbf{1}\mathbf{1}^T = \frac{1}{m}\mathbf{1}\mathbf{1}^T\mathbf{M} = \frac{1}{m}\mathbf{1}\mathbf{1}^T$ and so

$$\mathbf{M}^t = \left((\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T) + \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)^t = \left(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)^t + \frac{1}{m}\mathbf{1}\mathbf{1}^T. \tag{21}$$

The limit $\lim_{t\to\infty} \mathbf{M}^t$ is therefore equal to $\frac{1}{m}\mathbf{1}\mathbf{1}^T$ if $\rho(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T) < 1$, where $\rho(\cdot)$ is the spectral radius.

If $\mathbf{v} \neq 0$ denotes an eigenvector of $\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ with eigenvalue ν, then the fact that $\mathbf{1}^T(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T) = 0$ implies that

$$\mathbf{1}^T(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T)\mathbf{v} = \nu\mathbf{1}^T\mathbf{v} = 0.$$

Hence either $\nu = 0$, or $\mathbf{1}^T\mathbf{v} = 0$, and thus for any eigenvector $\mathbf{v} \perp \mathbf{1}$ we see that

$$(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T)\mathbf{v} = \mathbf{M}\mathbf{v} = \nu\mathbf{v}.$$

From this we conclude that $\mathbf{M}$ and $\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ share the same eigenvalues for any $\mathbf{v} \perp \mathbf{1}$. As the consensus eigenvalue 1 is simple, subtracting $\frac{1}{m}\mathbf{1}\mathbf{1}^T$ from $\mathbf{M}$ therefore changes the consensus eigenvalue from 1 to 0 and leaves all other eigenvalues unchanged. This implies that

$$\rho(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T) = \mu_* < 1$$

and

$$\lim_{t\to\infty} \mathbf{M}^t = \frac{1}{m}\mathbf{1}\mathbf{1}^T.$$

To prove the second bullet point, note that (21) implies that

$$\left\|\left(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)^t \mathbf{s}(0)\right\|_2 = \left\|\left(\mathbf{M}^t - \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)\mathbf{s}(0)\right\|_2 = \mathrm{CD}(t; \mathbf{M}, \mathbf{s}(0)).$$

Therefore,

$$\mathrm{CD}^*(t; \mathbf{M}) = \left\|\left(\mathbf{M} - \frac{1}{m}\mathbf{1}\mathbf{1}^T\right)^t\right\|_2.$$

Applying Gelfand's formula to $\mathrm{CD}^*(t; \mathbf{M})^{1/t}$ and taking logs yields the result.

To prove the third bullet point, if $\gamma < \bar{\gamma} \leq 0$ then $\mu_*(\gamma) = \sqrt{|\gamma|} = \sqrt{-\gamma}$, which is strictly decreasing in γ. If $\gamma \geq \bar{\gamma}$ then we invoke Corollary 4.2.

Therefore, $\mu_*(\gamma)$ is strictly decreasing for $\gamma \in (-1, \bar{\gamma})$ and strictly increasing for $\gamma \in (\bar{\gamma}, 1)$. When $\gamma = \bar{\gamma}$ then

$$(1 - \gamma)^2\lambda_*^2 + 4\gamma = 0$$

and therefore

$$\frac{(1-\bar{\gamma})\lambda_*}{2} = \sqrt{-\bar{\gamma}}.$$

Thus the value of $\mu_*(\gamma)$ in the two cases coincides when $\gamma = \bar{\gamma}$ and $\bar{\gamma}$ is the unique minimiser of $\mu_*(\gamma)$ for $\gamma \in (-1, 1)$.

It remains to note that

$$\bar{\gamma} = -\frac{1-\sqrt{1-\lambda_*^2}}{1+\sqrt{1-\lambda_*^2}}$$

is equal to 0 only if $\lambda_* = 0$.

□

Proof of Lemma 5.1. From (12), outgoing messages on AA edges depend only on messages received from other A types. Similarly, messages on BB edges only depend on messages received from other B types and no influence can enter E_{AA} and E_{BB} from outside those sets. If both sets are nonempty then the message graph $\mathcal{G}_M$ cannot be strongly connected and therefore $\mathbf{M}$ is reducible. □

Proof of Proposition 5.2. To apply insights from the theory of absorbing Markov chains first permute the rows and columns of $\mathbf{M}$ to obtain

$$\left(\begin{array}{cc|cc} 0 & \mathbf{M}_{AB,BA} & \mathbf{M}_{AB,AA} & 0 \\ \mathbf{M}_{BA,AB} & 0 & 0 & \mathbf{M}_{BA,BB} \\ \hline 0 & 0 & \mathbf{M}_{AA,AA} & 0 \\ 0 & 0 & 0 & \mathbf{M}_{BB,BB} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{\Psi} \\ \hline 0 & \mathbf{\Sigma} \end{array}\right).$$

Each nonzero element of $\mathbf{Q}$ is the weight placed on an incoming cross-type message when sending messages to agents of the other type. For an A type, this weight is given via (12) as $(1/d_i^B)(1-\beta_i) = (1/d_i^B)(1-p_i)(1-\gamma_i^{\text{out}})$.

The maximum row sum of $\mathbf{Q}$ is therefore $\|\mathbf{Q}\|_\infty = \max_{i\in N}(1-p_i)(1-\gamma_i^{\text{out}}) < 1$ and the spectral radius of $\mathbf{Q}$, denoted by $\rho(\mathbf{Q})$, must satisfy $\rho(\mathbf{Q}) \leq \|\mathbf{Q}\|_\infty < 1$. As a consequence we have that $\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots = (\mathbf{I}-\mathbf{Q})^{-1}$. We then have the following limit:

$$\lim_{t\to\infty} \begin{pmatrix} \mathbf{Q} & \mathbf{\Psi} \\ 0 & \mathbf{\Sigma} \end{pmatrix}^t = \begin{pmatrix} 0 & (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi}\mathbf{\Sigma}^\infty \\ 0 & \mathbf{\Sigma}^\infty \end{pmatrix}$$

where, under Assumption 2, the powers of the matrices $\mathbf{M}_{AA,AA}$ and $\mathbf{M}_{BB,BB}$ converge to rank-one matrices and so

$$\lim_{t\to\infty} \mathbf{\Sigma}^t := \mathbf{\Sigma}^\infty = \begin{pmatrix} \frac{1}{|E_{AA}|}\mathbf{1}\mathbf{1}^T & 0 \\ 0 & \frac{1}{|E_{BB}|}\mathbf{1}\mathbf{1}^T \end{pmatrix}.$$

The weights on within-type initial messages are uniform since $\mathbf{M}_{AA,AA}$ and $\mathbf{M}_{BB,BB}$ are doubly stochastic matrices.

The long run within-type message vectors are then given by

$$\begin{pmatrix} \mathbf{s}_{AA}(\infty) \\ \mathbf{s}_{BB}(\infty) \end{pmatrix} = \begin{pmatrix} \frac{1}{|E_{AA}|}\mathbf{1}\mathbf{1}^T & 0 \\ 0 & \frac{1}{|E_{BB}|}\mathbf{1}\mathbf{1}^T \end{pmatrix} \begin{pmatrix} \mathbf{s}_{AA}(0) \\ \mathbf{s}_{BB}(0) \end{pmatrix} = \begin{pmatrix} c_{AA}\mathbf{1} \\ c_{BB}\mathbf{1} \end{pmatrix}.$$

For cross-type messages we have

$$\begin{pmatrix} \mathbf{s}_{AB}(\infty) \\ \mathbf{s}_{BA}(\infty) \end{pmatrix} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi}\mathbf{\Sigma}^{\infty} \begin{pmatrix} \mathbf{s}_{AA}(0) \\ \mathbf{s}_{BB}(0) \end{pmatrix} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi} \begin{pmatrix} c_{AA}\mathbf{1} \\ c_{BB}\mathbf{1} \end{pmatrix}.$$

It remains to show that cross-type messages are convex combinations of within-type messages, or equivalently that $(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi}$ is nonnegative and row-stochastic. Each nonzero element of $\mathbf{\Psi}$ is the weight placed on incoming within-type messages when sending messages to agents of the other type. For an agent of type θ, this weight is given by β_i/d_i^θ for all agents with $d_i^\theta > 0$. Therefore, $(\mathbf{I}-\mathbf{Q})^{-1}$ and $\mathbf{\Psi}$ are both nonnegative matrices since $(\mathbf{I}-\mathbf{Q})^{-1}$ is the convergent sum of nonnegative matrices.

To show that $(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi}$ must be row-stochastic, note that

$$\mathbf{Q}\mathbf{1} + \mathbf{\Psi}\mathbf{1} = \mathbf{1}$$

since the matrix $\mathbf{M}$ is row-stochastic and these matrices correspond to the first $|E_{AB}| + |E_{BA}|$ rows of $\mathbf{M}$. Therefore,

$$\mathbf{\Psi}\mathbf{1} = (\mathbf{I}-\mathbf{Q})\mathbf{1},$$

and so

$$(\mathbf{I}-\mathbf{Q})^{-1}\mathbf{\Psi}\mathbf{1} = \mathbf{1}$$

as required. □

B Additional Results

B.1 A Doubly Stochastic Rule with Private Opinions

Self-loops allow agents to retain private opinions at each t which influence (and are influenced by) external messages. An example of a doubly stochastic messaging rule which includes private opinions is illustrated in Figure 8 below.

$$\mathbf{W}_i' = \left(\begin{array}{c|ccc} \delta & \frac{1-\delta}{d_i-1} & \cdots & \frac{1-\delta}{d_i-1} \\ \hline \frac{1-\delta}{d_i-1} & & & \\ \frac{1-\delta}{d_i-1} & & \frac{d_i-2+\delta}{d_i-1}\,\mathbf{W}_i & \\ \vdots & & & \\ \frac{1-\delta}{d_i-1} & & & \end{array}\right)$$

Figure 8: Augmented block messaging rule with a private opinion.

Here it is assumed that $\mathbf{W}_i$ is a doubly stochastic matrix corresponding to the external messages sent and received by agent i. Creating $\mathbf{W}_i'$ then involves adding a self-loop from i to i in the agent graph $\mathcal{G}$, which augments $\mathbf{W}_i$ with an additional row and column. The addition of a row and column allows for agent i to send a message to themselves at each period and have the previous period's self-message influence outgoing messages. The parameter $\delta \in (0,1)$ then controls the persistence of agent i's private opinion relative to the influence of other agents.

Since we assume that $\mathbf{W}_i$ is a doubly stochastic matrix of dimension $(d_i - 1) \times (d_i - 1)$, it is then easily verified that $\mathbf{W}_i'$ is also doubly stochastic.

B.2 DeGroot Equivalence and Wisdom

This appendix subsection considers the implications of Proposition 3.4 for so-called *wisdom of crowds* results, following the framework and insights of Golub and Jackson (2010). To extend their results from a node-based setting to the richer class of messaging rules studied in the present framework we briefly revisit the main features of their model.

In Golub and Jackson (2010) each node starts with the initial opinion $x_i(0) = \omega + \varepsilon_i$, where $x_i(0)$ is assumed to be in the interval $[0,1]$. The objective true state of the world is ω, which is unknown to the agents, and ε_i is an independently drawn noise term with expected value $\mathbb{E}[\varepsilon_i] = 0$ and variance of at least $\sigma^2 > 0$.

The authors consider how beliefs evolve on sequences of networks $(\mathcal{G}_n)_{n=1}^{\infty}$ with row-stochastic listening matrices $\mathbf{T}_n$ and whether beliefs converge to the true state as $n \to \infty$.

For each $\mathbf{T}_n$ in the sequence $(\mathbf{T}_n)_{n=1}^{\infty}$, the vector of consensus beliefs is given by

$$\bar{\mathbf{x}}_n = (\bar{\boldsymbol{\pi}}_n^T \mathbf{x}(0))\mathbf{1} = c\mathbf{1},$$

where the vector $\mathbf{x}(0)$ is the initial random belief determined by the true state ω and the idiosyncratic errors ε_i. The final consensus belief is then a weighted average of the initial beliefs $\mathbf{x}(0)$ where the weights are given by $\bar{\boldsymbol{\pi}}_n$, which is the stationary distribution of their node to node listening matrix $\mathbf{T}_n$. A key result of their paper is that the society is *wise* in the sense that each consensus belief $\bar{x}_i$ converges in probability to the true state ω if and only if each agent's influence on the stationary distribution becomes vanishingly small as $n \to \infty$. In undirected networks this is equivalent to the condition that

$$\max_{1 \leq i \leq n} \bar{\pi}_{i,n} = \max_{1 \leq i \leq n} \frac{d_i(\mathcal{G}_n)}{\sum_j d_j(\mathcal{G}_n)} \overset{n \to \infty}{\longrightarrow} 0.$$

To connect the notion of wisdom in Golub and Jackson (2010) to the setting of differentiated messaging on edges, we will assume that each agent i in graph $\mathcal{G}_n$ has a belief map $b_{i,n} : \mathbb{R}^{d_i(\mathcal{G}_n)} \to \mathbb{R}$ such that $x_{i,n}(t) = b_{i,n}(\mathbf{r}_i(t))$ for $t \geq 1$.

Assumption 3. *The belief map $b_{i,n}$ is a continuous function such that $b_{i,n}(c\mathbf{1}) = c$ for any $c \in \mathbb{R}$.*

Therefore, if i receives the same message from all of their neighbours (including any message on a self-loop) then i's belief must be equal to this message.[16] We can now define a notion of wisdom which is similar to Golub and Jackson (2010):

Definition 3. *Let $(\mathcal{G}_n)_{n=1}^{\infty}$ be a sequence of undirected graphs and let $(\mathbf{M}_n)_{n=1}^{\infty}$ be a corresponding sequence of edge to edge transition matrices, where each $\mathbf{M}_n$ is assumed to be primitive. Let $\bar{x}_{i,n} = \lim_{t \to \infty} b_{i,n}(\mathbf{r}_i(t))$ denote the limiting belief of node i under $\mathbf{M}_n$. The sequence $(\mathbf{M}_n)_{n=1}^{\infty}$ is* wise *if*

$$\operatorname*{plim}_{n \to \infty} \max_{i \leq n} |\bar{x}_{i,n} - \omega| = 0.$$

Now consider a sequence of undirected graphs $(\mathcal{G}_n)_{n=1}^{\infty}$ where $\mathcal{G}_n$ is a graph on n nodes with m_n directed edges. For each n assume that $\mathbf{M}_n$ is primitive and doubly stochastic.[17] Initialise the opinion dynamics on edges by lifting $x_{i,n}(0) = \omega + \varepsilon_i$ from nodes to outgoing edges, i.e. by setting

$$\mathbf{s}_n(0) = \mathbf{L}_n \mathbf{x}_n(0). \tag{22}$$

[16] This is referred to as *unanimity* by Molavi et al. (2018).

[17] From Proposition 3.4, each $\mathbf{M}_n$ is doubly stochastic if and only if each $\mathbf{W}_{i,n}$ is doubly stochastic.

Since $\mathbf{M}_n$ is primitive and doubly stochastic, its unique stationary distribution is uniform on edges. Hence this opinion dynamic converges to the constant message

$$c_n = \frac{1}{m_n}\mathbf{1}_{m_n}^T \mathbf{L}_n \, \mathbf{x}_n(0) = \sum_{i=1}^{n} \frac{d_i(\mathcal{G}_n)}{\sum_{j=1}^{n} d_j(\mathcal{G}_n)} \, x_{i,n}(0),$$

where the final equality follows from the fact that lifting a node i's initial opinion to their incident edges replicates their opinion d_i times. Therefore, by continuity of $b_{k,n}$ for each $k \in N$,

$$\bar{x}_{k,n} = \lim_{t\to\infty} b_{k,n}(\mathbf{r}_k(t)) = b_{k,n}\big(c_n \mathbf{1}\big) = c_n.$$

Thus the induced limiting belief is the same weighted average of initial signals as in the DeGroot model on an undirected and unweighted graph $\mathcal{G}_n$. A corollary of Proposition 3.4 then follows immediately from Golub and Jackson (2010):

Corollary B.1. *Let $(\mathcal{G}_n)_{n=1}^{\infty}$ be a sequence of undirected graphs and let $(\mathbf{M}_n)_{n=1}^{\infty}$ be a corresponding sequence of edge to edge influence matrices. For each n, suppose that $\mathbf{M}_n$ is primitive and doubly stochastic. Initialise messages as in (22) and let beliefs be given by maps $b_{i,n}$ satisfying Assumption 3. Then the sequence $(\mathbf{M}_n)_{n=1}^{\infty}$ is* wise *if and only if*

$$\max_{1\le i\le n} \frac{d_i(\mathcal{G}_n)}{\sum_{j=1}^{n} d_j(\mathcal{G}_n)} \longrightarrow 0.$$

B.3 Positive Definiteness and Bounds on Consensus Time

As is evident from the proof of Proposition 4.3, the finite time bounds on the consensus time hold in the more general case where $\mathbf{W}$ is a positive definite matrix.[18] Since $\mathbf{W}$ is block diagonal, $\mathbf{W}$ is positive definite if and only if each block $\mathbf{W}_i$ is positive definite. This appendix subsection now discusses the interpretation of positive definiteness of $\mathbf{W}$ within the context of the model.

To understand what positive definiteness of $\mathbf{W}$ equates to intuitively, it is helpful to consider the covariance between messages received by i at period t and messages sent by i at period $t+1$. Letting $\tilde{\mathbf{v}}(t) := \mathbf{v}(t) - \bar{v}(t)\mathbf{1}$ denote the demeaned form of any vector $\mathbf{v}(t)$, the covariance is given by

$$\begin{aligned} \mathrm{Cov}(\mathbf{r}_i(t), \mathbf{s}_i(t+1)) &:= \frac{1}{d_i}\sum_{j\in N_i} \left(r_{ij}(t) - \bar{r}_i(t)\right)\left(s_{ij}(t+1) - \bar{s}_i(t+1)\right) \\ &= \frac{1}{d_i}\tilde{\mathbf{r}}_i(t)^T \mathbf{W}_i \tilde{\mathbf{r}}_i(t). \end{aligned} \tag{23}$$

The final equality in (23) comes from the fact that $\bar{s}_i(t+1) = \bar{r}_i(t)$ when the messaging rule

[18] Here we define a positive definite matrix as a symmetric matrix $\mathbf{Z}$ satisfying $\mathbf{v}^T\mathbf{Z}\mathbf{v} > 0$ for all $\mathbf{v} \neq \mathbf{0}$.

is doubly stochastic as unbiased messaging preserves the mean message at each iteration. If $\mathbf{W}$ is positive definite (and thus symmetric) then it must be doubly stochastic under the assumption that each row sums to 1.

This within agent covariance measure is equal to zero if agents use the uniform messaging rule, since $s_{ij}(t+1) = \bar{s}_i(t+1) = \bar{r}_i(t)$ for all $j \in N_i$.[19] The expression in (23) also reveals that the covariance between received and sent messages is positive whenever $\mathbf{v}^T \mathbf{W}_i \mathbf{v} > 0$ for any $\mathbf{v} \neq 0$. If $\mathbf{W}_i$ is a symmetric matrix, then $\text{Cov}(\mathbf{r}_i(t), \mathbf{s}_i(t+1)) > 0$ for any non-constant $\mathbf{r}_i(t)$ if and only if $\mathbf{W}_i$ is positive definite. Therefore, there is a direct link between covariance of messages sent and received by each agent and the positive definiteness of each $\mathbf{W}_i$.

Furthermore, the pairwise conformity rule $\mathbf{W}_i^{\text{PC}}(\gamma)$ is a symmetric positive definite matrix whenever $\gamma \in (0, 1]$. Indeed we observe for any $\tilde{\mathbf{r}}_i(t) \neq \mathbf{0}$ that

$$\begin{aligned}\text{Cov}(\mathbf{r}_i(t), \mathbf{s}_i(t+1)) &= \frac{1}{d_i}\tilde{\mathbf{r}}_i(t)^T \left(\gamma \mathbf{I}_{d_i} + (1-\gamma)\frac{1}{d_i}\mathbf{1}_{d_i}\mathbf{1}_{d_i}^T\right)\tilde{\mathbf{r}}_i(t) \\ &= \gamma \text{Var}(\mathbf{r}_i(t)) > 0.\end{aligned}$$

Therefore the conformity parameter γ directly controls the covariance between received and sent messages, scaling up the covariance in proportion to the variance of messages received.

B.4 Group Conformity under Identical Neighbourhoods

To solve for each group's influence on the final consensus in closed form, assume a simple case where the network is d-regular with an equal number of A and B types, with $p_i = p$ for each $i \in N$ where $p \in (0, 1)$. To simplify further, assume that agents of each type $\theta \in \{A, B\}$ have identical group conformity preferences, so that $\alpha_i = \alpha_\theta$ and $\beta_i = \beta_\theta$ for each $i \in N$.

Under these assumptions the influence matrix $\mathbf{M}$ can then be represented in block form as

$$\mathbf{M} = \begin{pmatrix} \alpha_A \mathbf{M}_{AA,AA} & 0 & (1-\alpha_A)\mathbf{M}_{AA,BA} & 0 \\ \beta_A \mathbf{M}_{AB,AA} & 0 & (1-\beta_A)\mathbf{M}_{AB,BA} & 0 \\ 0 & (1-\beta_B)\mathbf{M}_{BA,AB} & 0 & \beta_B \mathbf{M}_{BA,BB} \\ 0 & (1-\alpha_B)\mathbf{M}_{BB,AB} & 0 & \alpha_B \mathbf{M}_{BB,BB} \end{pmatrix}. \tag{24}$$

Each subblock $\mathbf{M}_{\cdot,\cdot}$ in (24) now denotes a nonnegative row-stochastic submatrix induced by the underlying network structure and the messaging rule defined in (12).

[19] The equality $\text{Cov}(\mathbf{r}_i(t), \mathbf{s}_i(t+1)) = 0$ holds for any non-differentiated messaging rule, since all neighbours are sent identical messages and $s_{ij}(t+1) = \bar{s}_i(t+1)$ for all $j \in N_i$.

Proposition B.2. *Suppose that Assumption 1 holds. Suppose further that $d_i = d$ and $p_i = p$ for all $i \in N$, where pd and $(1-p)d$ are both integers. If each agent i of type θ_i uses the group conformity messaging rule $\mathbf{W}_i^{\mathrm{GC}}(\alpha_i, \beta_i)$ with $\alpha_i = \alpha_{\theta_i} \in (0,1)$ and $\beta_i = \beta_{\theta_i} \in (0,1)$, then*

(i) The influence matrix $\mathbf{M}$ admits a unique stationary distribution, given by

$$\boldsymbol{\pi}^T = (\boldsymbol{\pi}_{AA}^T, \boldsymbol{\pi}_{AB}^T, \boldsymbol{\pi}_{BA}^T, \boldsymbol{\pi}_{BB}^T),$$

where stationary weights are constant within each block.

(ii) The limiting consensus message c is given by

$$c = \frac{1}{2 + \frac{\beta_A}{1-\alpha_A} + \frac{\beta_B}{1-\alpha_B}} \left(\bar{s}_{AA}(0) \frac{\beta_A}{1-\alpha_A} + \bar{s}_{AB}(0) + \bar{s}_{BA}(0) + \bar{s}_{BB}(0) \frac{\beta_B}{1-\alpha_B} \right),$$

where $\bar{s}_{\theta\theta'}(0)$ is the mean message from type θ to θ' at $t = 0$.

Proof of Proposition B.2. Proposition 3.2 guarantees a unique $\boldsymbol{\pi} \in \mathbb{R}^m_{++}$ such that $\boldsymbol{\pi}^T \mathbf{M} = \boldsymbol{\pi}^T$. Using (24), the stationarity condition $\boldsymbol{\pi}^T \mathbf{M} = \boldsymbol{\pi}^T$ is equivalent to the following system:

$$\boldsymbol{\pi}_{AA}^T = \alpha_A \, \boldsymbol{\pi}_{AA}^T \mathbf{M}_{AA,AA} + \beta_A \, \boldsymbol{\pi}_{AB}^T \mathbf{M}_{AB,AA}, \tag{25}$$
$$\boldsymbol{\pi}_{AB}^T = (1-\beta_B) \, \boldsymbol{\pi}_{BA}^T \mathbf{M}_{BA,AB} + (1-\alpha_B) \, \boldsymbol{\pi}_{BB}^T \mathbf{M}_{BB,AB}, \tag{26}$$
$$\boldsymbol{\pi}_{BA}^T = (1-\alpha_A) \, \boldsymbol{\pi}_{AA}^T \mathbf{M}_{AA,BA} + (1-\beta_A) \, \boldsymbol{\pi}_{AB}^T \mathbf{M}_{AB,BA}, \tag{27}$$
$$\boldsymbol{\pi}_{BB}^T = \beta_B \, \boldsymbol{\pi}_{BA}^T \mathbf{M}_{BA,BB} + \alpha_B \, \boldsymbol{\pi}_{BB}^T \mathbf{M}_{BB,BB}. \tag{28}$$

Isolating $\boldsymbol{\pi}_{AA}^T$ and $\boldsymbol{\pi}_{BB}^T$ in (25) and (28) yields

$$\boldsymbol{\pi}_{AA}^T = \beta_A \boldsymbol{\pi}_{AB}^T \mathbf{M}_{AB,AA} \left(\mathbf{I} - \alpha_A \mathbf{M}_{AA,AA} \right)^{-1},$$
$$\boldsymbol{\pi}_{BB}^T = \beta_B \boldsymbol{\pi}_{BA}^T \mathbf{M}_{BA,BB} \left(\mathbf{I} - \alpha_B \mathbf{M}_{BB,BB} \right)^{-1}.$$

Substituting into (26) and (27) then gives

$$\boldsymbol{\pi}_{AB}^T = \boldsymbol{\pi}_{BA}^T \left[(1-\beta_B) \mathbf{M}_{BA,AB} + \beta_B \mathbf{M}_{BA,BB} (1-\alpha_B) \left(\mathbf{I} - \alpha_B \mathbf{M}_{BB,BB} \right)^{-1} \mathbf{M}_{BB,AB} \right],$$
$$\boldsymbol{\pi}_{BA}^T = \boldsymbol{\pi}_{AB}^T \left[\beta_A \mathbf{M}_{AB,AA} (1-\alpha_A) \left(\mathbf{I} - \alpha_A \mathbf{M}_{AA,AA} \right)^{-1} \mathbf{M}_{AA,BA} + (1-\beta_A) \mathbf{M}_{AB,BA} \right].$$

Then from (26) we have

$$\boldsymbol{\pi}_{AB}^T = \boldsymbol{\pi}_{AB}^T \left[\beta_A \mathbf{M}_{AB,AA} (1-\alpha_A) \left(\mathbf{I} - \alpha_A \mathbf{M}_{AA,AA} \right)^{-1} \mathbf{M}_{AA,BA} + (1-\beta_A) \mathbf{M}_{AB,BA} \right] \left[(1-\beta_B) \mathbf{M}_{BA,AB} + \beta_B \mathbf{M}_{BA,BB} (1-\alpha_B) \left(\mathbf{I} - \alpha_B \mathbf{M}_{BB,BB} \right)^{-1} \mathbf{M}_{BB,AB} \right]. \tag{29}$$

The matrices $\mathbf{M}_{AA,AA}$, $\mathbf{M}_{BB,BB}$, $\mathbf{M}_{AB,BA}$ and $\mathbf{M}_{BA,AB}$ are doubly stochastic, since they provide an unbiased aggregation of messages from one group back along the same undirected edges. The assumption that $d_i = d$ and $p_i = p$ for all $i \in N$ means $\mathbf{M}_{AB,AA}$ and $\mathbf{M}_{BA,BB}$ contain nonzero entries equal to $\frac{1}{dp}$ and these matrices therefore have column sums $\frac{d(1-p)}{dp} = \frac{1-p}{p}$. Moreover, the matrices $\mathbf{M}_{AA,BA}$ and $\mathbf{M}_{BB,AB}$ contain nonzero entries $\frac{1}{d(1-p)}$ which leads to the column sums $\frac{dp}{d(1-p)} = \frac{p}{1-p}$.

It can then be observed that the matrix

$$\frac{p}{1-p}\mathbf{M}_{AB,AA}(1-\alpha_A)\left(\mathbf{I}-\alpha_A\mathbf{M}_{AA,AA}\right)^{-1}\frac{1-p}{p}\mathbf{M}_{AA,BA}$$

is doubly stochastic, since for $\alpha_A \in (0,1)$ we have

$$\left(\mathbf{I}-\alpha_A\mathbf{M}_{AA,AA}\right)^{-1} = \sum_{k=0}^{\infty} {\alpha_A}^k \mathbf{M}_{AA,AA}^k,$$

and therefore both the rows and columns of $\left(\mathbf{I}-\alpha_A\mathbf{M}_{AA,AA}\right)^{-1}$ sum to $\frac{1}{1-\alpha_A}$.

The matrix in the first square bracket of (29) is therefore a convex combination of doubly stochastic matrices. An identical argument applies to the matrix in the second square bracket, which is also doubly stochastic. Therefore a constant $\boldsymbol{\pi}_{AB}^T$ is a left eigenvector of the matrix in (29) with eigenvalue 1.

Since $\mathbf{M}$ is irreducible, a constant eigenvector is the only positive solution to (29) and consequently,

$$\boldsymbol{\pi}_{AB}^T = \kappa\mathbf{1}^T$$

for some $\kappa \neq 0$. If $\boldsymbol{\pi}_{AB}^T$ is constant, it then follows that

$$\begin{aligned}\boldsymbol{\pi}_{BA}^T &= \kappa\mathbf{1}^T\left[\beta_A\mathbf{M}_{AB,AA}(1-\alpha_A)\left(\mathbf{I}-\alpha_A\mathbf{M}_{AA,AA}\right)^{-1}\mathbf{M}_{AA,BA} + (1-\beta_A)\mathbf{M}_{AB,BA}\right]\\ &= \kappa\mathbf{1}^T.\end{aligned}$$

Substituting into the equations for $\boldsymbol{\pi}_{AA}^T$ and $\boldsymbol{\pi}_{BB}^T$ gives

$$\begin{aligned}\boldsymbol{\pi}_{AA}^T &= \frac{\beta_A}{1-\alpha_A}\frac{1-p}{p}\kappa\mathbf{1}^T,\\ \boldsymbol{\pi}_{BB}^T &= \frac{\beta_B}{1-\alpha_B}\frac{1-p}{p}\kappa\mathbf{1}^T.\end{aligned}$$

Therefore the stationary vector $\boldsymbol{\pi}$ is constant within block. To pin down the constant κ, assume the normalisation $\boldsymbol{\pi}^T\mathbf{1} = 1$. The group sizes must be equal since $|N^A|d(1-p) =$

$|E_{AB}| = |E_{BA}| = |N^B|d(1-p)$, and so

$$\begin{aligned}\boldsymbol{\pi}^T\mathbf{1} &= \kappa\left(\frac{\beta_A}{1-\alpha_A}\frac{1-p}{p}\frac{n}{2}dp + \frac{n}{2}d(1-p) + \frac{n}{2}d(1-p) + \frac{\beta_B}{1-\alpha_B}\frac{1-p}{p}\frac{n}{2}dp\right)\\ &= \kappa\frac{m}{2}(1-p)\left(2 + \frac{\beta_A}{1-\alpha_A} + \frac{\beta_B}{1-\alpha_B}\right)\end{aligned}$$

which gives

$$\kappa = \left(\frac{m(1-p)}{2}\left(2 + \frac{\beta_A}{1-\alpha_A} + \frac{\beta_B}{1-\alpha_B}\right)\right)^{-1}.$$

To solve for the consensus message c we then find

$$\begin{aligned}c = \boldsymbol{\pi}^T\mathbf{s}(0) = \left(2 + \frac{\beta_A}{1-\alpha_A} + \frac{\beta_B}{1-\alpha_B}\right)^{-1}\Bigg(&\frac{\beta_A}{1-\alpha_A}\frac{\mathbf{1}^T\mathbf{s}_{AA}(0)}{(m/2)p}\\ &+ \frac{\mathbf{1}^T\mathbf{s}_{AB}(0)}{(m/2)(1-p)} + \frac{\mathbf{1}^T\mathbf{s}_{BA}(0)}{(m/2)(1-p)} + \frac{\beta_B}{1-\alpha_B}\frac{\mathbf{1}^T\mathbf{s}_{BB}(0)}{(m/2)p}\Bigg).\end{aligned}$$

This can be rewritten as

$$c = \left(2 + \frac{\beta_A}{1-\alpha_A} + \frac{\beta_B}{1-\alpha_B}\right)^{-1}\left(\bar{s}_{AA}(0)\frac{\beta_A}{1-\alpha_A} + \bar{s}_{AB}(0) + \bar{s}_{BA}(0) + \bar{s}_{BB}(0)\frac{\beta_B}{1-\alpha_B}\right).$$

□

The influence of group θ's initial messages depends on the parameters α_θ and β_θ in an intuitive way. A stronger focus on conforming to the in-group norm leads to a higher influence of $\bar{s}_{\theta\theta}$ on the final consensus. The mean internal message $\bar{s}_{\theta\theta}(0)$ at $t = 0$ has a larger influence than the mean cross-group message $\bar{s}_{\theta\theta'}(0)$ at $t = 0$ whenever $\beta_\theta > 1-\alpha_\theta$. Equivalently, by writing the weight on $\bar{s}_{\theta\theta}(0)$ relative to $\bar{s}_{\theta\theta'}(0)$ as

$$\frac{\beta_\theta}{1-\alpha_\theta} = \frac{p + (1-p)\gamma_\theta^{\text{out}}}{1-p-(1-p)\gamma_\theta^{\text{in}}} = \frac{\frac{p}{1-p} + \gamma_\theta^{\text{out}}}{1-\gamma_\theta^{\text{in}}},$$

the within-group average has more influence on the final consensus value when

$$\frac{p}{1-p} > 1 - \gamma_\theta^{\text{in}} - \gamma_\theta^{\text{out}}.$$

This condition holds whenever the network displays homophily ($p > 1/2$), with the relative influence of $\bar{s}_{\theta\theta}(0)$ increasing in both in-group conformity ($\gamma_\theta^{\text{in}}$) and out-group conformity ($\gamma_\theta^{\text{out}}$). Furthermore, holding all other parameters fixed, the consensus message c approaches the mean internal message $\bar{s}_{\theta\theta}(0)$ as $\gamma_\theta^{\text{in}} \to 1$.

References

Acemoglu, Daron, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar (2013). "Opinion Fluctuations and Disagreement in Social Networks". *Mathematics of Operations Research* 38.1, pp. 1–27 (pp. 2, 28).

Aigner, Martin (1967). "On the Linegraph of a Directed Graph". *Mathematische Zeitschrift* 102, pp. 56–61 (p. 30).

Akerlof, George A. (1980). "A Theory of Social Custom, of which Unemployment may be One Consequence". *Quarterly Journal of Economics* 94.4, pp. 749–775 (p. 15).

Anufriev, Mikhail, Kirill Borissov, and Mikhail Pakhnin (2023). "Dissonance minimization and conversation in social networks". *Journal of Economic Behavior and Organization* 215, pp. 167–191 (pp. 3, 15).

Arrigo, Francesca, Desmond J. Higham, and Vanni Noferini (2019). "Non-backtracking PageRank". *Journal of Scientific Computing* 80.3, pp. 1419–1437 (pp. 6, 32).

Asch, Solomon E. (1955). "Opinions and Social Pressure". *Scientific American* 193.5, pp. 31–35 (p. 15).

Banerjee, Abhijit and Olivier Compte (2024). "Consensus and Disagreement: Information Aggregation under (Not So) Naive Learning". *Journal of Political Economy* 132.8, pp. 2790–2829 (p. 3).

Bernheim, B. Douglas (1994). "A Theory of Conformity". *Journal of Political Economy* 102.5, pp. 841–877 (p. 15).

Bowen, T. Renee, Danil Dmitriev, and Simone Galperti (2023). "Learning from Shared News: When Abundant Information Leads to Belief Polarization". *Quarterly Journal of Economics* 138.2, pp. 955–1000 (pp. 3, 14, 28).

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours (2014). "Strategic Interaction and Networks". *American Economic Review* 104.3, pp. 898–930 (p. 22).

Buechel, Berno, Tim Hellmann, and Stefan Klößner (2015). "Opinion Dynamics and Wisdom Under Conformity". *Journal of Economic Dynamics and Control* 52, pp. 240–257 (pp. 3, 15).

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber (2015). "Partisan Bias in Factual Beliefs about Politics". *Quarterly Journal of Political Science* 10, pp. 519–578 (p. 29).

Cerreia-Vioglio, Simone, Roberto Corrao, and Giacomo Lanzani (2024). "Dynamic Opinion Aggregation: Long-Run Stability and Disagreement". *Review of Economic Studies* 91.3, pp. 1406–1447 (p. 3).

Chandrasekhar, Arun G., Horacio Larreguy, and Juan Pablo Xandri (2020). "Testing Models of Social Learning on Networks: Evidence From Two Experiments". *Econometrica* 88.1, pp. 1–32 (p. 2).

Cheng, Chen, Xiao Han, Xin Tong, Yusheng Wu, and Yiqing Xing (2026). "Degree-Weighted Social Learning". *American Economic Journal: Microeconomics* Forthcoming (pp. 16–18).

Choi, Syngjoo, Sanjeev Goyal, Frederic Moisan, and Yu Yang Tony To (2023). "Learning in Networks: An Experiment on Large Networks with Real-World Features". *Management Science* 69.5, pp. 2778–2787 (p. 2).

Cookson, J. Anthony and Marina Niessner (2020). "Why Don't We Agree? Evidence from a Social Network of Investors". *The Journal of Finance* LXXV.1, pp. 173–228 (p. 28).

DeGroot, Morris H. (1974). "Reaching a Consensus". *Journal of the American Statistical Association* 69.345, pp. 118–121 (p. 1).

Della Lena, Sebastiano, Luca Paolo Merlino, and Yves Zenou (2026). "Affective Polarization, Media Outlets, and Opinion Dynamics". *Review of Economic Studies*, pp. 1–44 (pp. 2, 19, 28).

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel (2003). "Persuasion Bias, Social Influence, and Unidimensional Opinions". *The Quarterly Journal of Economics* 118.3, pp. 909–968 (pp. 3, 27, 28).

Dvorak, Fabian, Urs Fischbacher, and Katrin Schmelz (2025). "Strategic Conformity or Anti-Conformity to Avoid Punishment and Attract Reward". *The Economic Journal* 135.666, pp. 556–583 (p. 15).

Foerster, Manuel (2019). "Dynamics of Strategic Information Transmission in Social Networks". *Theoretical Economics* 14.1, pp. 253–295 (p. 3).

Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani (2013). "Strategic information transmission networks". *Journal of Economic Theory* 148.5, pp. 1751–1769 (p. 3).

Golub, Benjamin and Matthew O. Jackson (2010). "Naïve Learning in Social Networks and the Wisdom of Crowds". *American Economic Journal: Microeconomics* 2.1, pp. 112–149 (pp. 14, 41–43).

— (2012). "How Homophily Affects the Speed of Learning and Best-Response Dynamics". *The Quarterly Journal of Economics* 127.3, pp. 1287–1338 (pp. 2, 14, 16–18, 34, 35).

Grabisch, Michel, Antoine Mandel, and Agnieszka Rusinowska (2023). "On the Design of Public Debate in Social Networks". *Operations Research* 71.2, pp. 626–648 (p. 28).

Grimm, Veronika and Friederike Mengel (2020). "Experiments on Belief Formation in Networks". *Journal of the European Economic Association* 18.1, pp. 49–82 (p. 2).

Grindrod, Peter, Desmond J. Higham, and Vanni Noferini (2018). "The Deformed Graph Laplacian and its Applications to Network Centrality Analysis". *SIAM Journal on Matrix Analysis and Applications* 39.1, pp. 310–341 (p. 6).

Hegselmann, Rainer and Ulrich Krause (2002). "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation". *Journal of Artificial Societies and Social Simulation (JASSS)* 5.3 (p. 28).

Higgins, E. Tory (1999). ""Saying is Believing" Effects: When Sharing Reality about Something Biases Knowledge and Evaluations". *Shared Cognition in Organizations: The Management of Knowledge*. Ed. by L. L. Thompson, J. M. Levine, and D. M. Messick. Lawrence Erlbaum Associates, pp. 33–48 (p. 15).

Higgins, E. Tory and William S. Rholes (1978). ""Saying is Believing": Effects of Message Modification on Memory and Liking for the Person Described". *Journal of Experimental Social Psychology* 14.4, pp. 363–378 (p. 15).

Kempton, Mark (2016). "Non-Backtracking Random Walks and a Weighted Ihara's Theorem". *Open Journal of Discrete Mathematics* 06.04, pp. 207–226 (p. 33).

Krzakala, Florent, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang (2013). "Spectral Redemption in Clustering Sparse Networks". *Proceedings of the National Academy of Sciences of the United States of America* 110.52, pp. 20935–20940 (p. 6).

Levy, Gilat and Ronny Razin (2019). "Echo Chambers and Their Effects on Economic and Political Outcomes". *Annual Review of Economics* 11, pp. 303–328 (p. 29).

Melguizo, Isabel (2019). "Homophily and the Persistence of Disagreement". *Economic Journal* 129.619, pp. 1400–1424 (p. 28).

Meyer, Carl D. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM (p. 33).

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie (2018). "A Theory of Non-Bayesian Social Learning". *Econometrica* 86.2, pp. 445–490 (pp. 3, 20, 42).

Polanski, Arnold and Fernando Vega-Redondo (2023). "Homophily and Influence". *Journal of Economic Theory* 207 (p. 2).

Robin, Stéphane, Agnieszka Rusinowska, and Marie Claire Villeval (2014). "Ingratiation: Experimental evidence". *European Economic Review* 66, pp. 16–38 (p. 15).

Seneta, Eugene (2006). *Non-negative Matrices and Markov Chains*. 2nd Ed. Springer Science and Business Media (p. 30).

Zafar, Basit (2011). "An Experimental Investigation of why Individuals Conform". *European Economic Review* 55.6, pp. 774–798 (p. 15).

Zou, Wenbo and Xue Xu (2023). "Ingroup Bias in a Social Learning Experiment". *Experimental Economics* 26.1, pp. 27–54 (p. 15).