Prejudice, group cohesion and the dynamics of disagreement

Daniel C. Opolot[a,*]

[a] *The African Institute of Financial Markets and Risk Management (AIFMRM), University of Cape Town*

Abstract

This paper studies how individual prejudice towards others' attributes and behaviour or government policies and programs generates everlasting public disagreement. We study an endogenous model of opinion formation where agents compromise between respecting their own personal prejudice and conforming their opinions to those held by others with whom they share close ties. We characterize equilibrium opinions, showing that their distribution and average depends on Bonacich centralities. An agent's Bonacich centrality measures the extent to which her prejudice is incorporated into others' equilibrium opinions. We find that the magnitude of disagreement positively depends on three reinforcing factors: the extent and intensity of prejudice, and intensities of subgroup cohesion. We also examine the rate of decay of disagreement and demonstrate how it can be used to empirically estimate the intensity of prejudice of a social group.

Keywords: Prejudice, social learning, networks, group cohesion, disagreement.

JEL: D83, D85, Z13, J15

1. Introduction

Disagreement is a common phenomenon in society and has significant economic implications. For example, heterogeneity in opinions is the cause of speculative trade in financial markets; discrepancies in ideological views between policy makers may lead to public debt accumulation in inefficient ways; and divergent opinions among leaders of an organization can affect the organization's success.[1] Given its costs to society, disagreement – its causes and how to quantify it – deserves attention. There is a small but growing literature on disagreement (Friedkin and

*Corresponding author: University of Cape Town, Private Bag X3, Rondebosch, 7701 Cape Town
Email address: daniel.opolot@uct.ac.za (Daniel C. Opolot)

[1] See Haruvy et al. (2007) and Scheinkman and Xiong (2004) for the effects of heterogeneous opinions on speculative trade; Alesina and Tabellini (1990) on how divergent opinions may affect efficient policy making; and Voss et al. (2006) on how divergent opinions may affect organizational success.

Johnsen, 1999; Krause, 2000; Hegselmann et al., 2002; Acemoğlu et al., 2013; Melguizo, 2016).[2] However, existing models make no mention of how prejudice – generally understood as a set of preconceived and inflexible opinions – might be a driving factor behind public disagreement.

This paper incorporates individual prejudice into the process of opinion formation. The models of opinion formation, which date back to the seminal work of DeGroot (1974), argue that people's opinions change through social learning. Although this may generally be true, decades of research in psychology has demonstrated that some attitudes/beliefs are resistant to change. This component of attitudes may in part be heritable and inculcated through culturalization (Tesser, 1993); it is stored in memory and activated automatically with little conscious control (Fazio, 1986; Houston and Fazio, 1989). We refer to the component of attitudes that is resistant to change as prejudice, which is in-line with how individual prejudice is generally understood.

Although prejudices may change in the long-run, by definition, they change more slowly compared to the rate at which overall opinions change through social learning. Intuitively, if individuals possess heterogeneous levels of prejudices, exchanging information through social learning need not drive the group to a consensus. This is because once the process of social learning stops and all available information has been shared among members of the group, heterogeneity in opinions will persist due to heterogeneity in prejudices. We develop a model of opinion formation that includes prejudice as a primary driver of public disagreement.

A key feature of our model is that when adjusting their opinions, agents compromise between respecting their own personal prejudice and forming beliefs in line with those held by others with whom they share close ties. There is empirical and experimental evidence in support of this assumption. Kahan et al. (2012) and Kahan et al. (2017) for example show that the conflict of interest between following ones' prejudice or conforming to public opinion drives divergence of opinions regarding climate change.[3] We suppose that agents interact through a social network, whereby close ties capture relationships with family members, friends, colleagues and acquaintances. Although prejudices may generally change over time, we assume that they

[2] Krause (2000) and Hegselmann et al. (2002) examines how bounded confidence generates disagreement; Melguizo (2016) studies how homophily drives disagreement; Acemoğlu et al. (2013) shows that stubborn agents can be a source of disagreement, and in a similar vein, Friedkin and Johnsen (1999) show that partially stubborn agents can generate permanent disagreement.

[3] Kahan et al. (2012) and Kahan et al. (2017) show that disagreement arises from the conflict of interest between conforming ones' beliefs to their political outlook, and to the beliefs held by others with whom they share close ties. There is also empirical evidence from psychology literature confirming such conflicts of interests in the process of opinion formation, see for example Wilson et al. (2000) and the references therein.

change more slowly compared to changes in overall opinions due to learning from others. These assumptions together give rise to a model of opinion adjustment where agents take weighted averages of their neighbours' opinions and add this average to their personal prejudice. Recent laboratory and field experiments support the idea that individuals follow naïve rules of learning by averaging (Corazzini et al., 2012; Grimm and Mengel, 2014; Chandrasekhar et al., 2015). With this model at hand, our subsequent analysis is guided by three specific questions.

1. How do prejudices and network structure shape the composition of equilibrium opinions?
2. How do prejudices and network structure influence the magnitude of disagreement?
3. What is the rate of convergence of the opinion formation process?

Our first set of results is a characterization of the composition of equilibrium opinions. We find that the distribution of and average equilibrium opinions are functions of *Bonacich centralities* (Bonacich, 1987). The Bonacich centrality of an agent depends on her position in the network and the *intensity of prejudice*. Intensity of prejudice is the intensity with which individuals conform to their prejudice. When the intensity of prejudice is small, then having a large number of direct neighbours and being connected to other highly influential agents is associated with higher Bonacich centrality. When the intensity of prejudice is sufficiently large, then Bonacich centrality depends only on the number of immediate neighbours. This is because when intensity of prejudice is high, agents become impatient and do not wait to incorporate opinions of other agents from distant positions in the network. Our results thus demonstrate how the network structure and prejudice interactively shape the composition of equilibrium opinions.

Our second set of results is a quantification of disagreement. Since prejudices are the primary cause of disagreement in our model, we define the magnitude of disagreement as the difference between the distribution of opinions in the model with prejudices and the distribution of opinions in the model without prejudices. We show that disagreement increases with both the extent and intensity of prejudice. Disagreement also increases with *intensities of group cohesion*. A subgroup of agents is cohesive if every member of the group has more than half of her interactions with other members of the group. The intensity of group cohesion is the relative total weight of interactions within the group, to the total weight of all interactions of group members. We discuss the implications of these findings for government and organizational policies aimed at reducing disagreement.

Finally, we characterize the rate of decay of disagreement. We show that the rate of decay is proportional to the intensity of prejudice and is independent of the network structure. The direct implication of this result is that the parameter of intensity of prejudice can be estimated by tracking the step-by-step evolution of opinions. This result also highlights the difference between the framework studied in this paper and the canonical DeGroot model of opinion formation.[4] In the DeGroot model and its variations, the speed of learning depends only on the properties of the network. For example, Golub and Jackson (2010) and Golub and Jackson (2012) show that the speed of learning in the DeGroot model depends only on the second largest eigenvalue of the matrix capturing interactions among agents; and Jadbabaie et al. (2013) show that in the variation of the DeGroot model, according to which agents linearly combine their personal experiences with the opinions of their neighbours, the speed of learning depends on agents' centralities in the network. Our results thus highlight how different models of opinion formation can be distinguished through laboratory and field experiments by tracking the rates of decay of disagreement.

Our work builds on the existing literature on disagreement. Friedkin and Johnsen (1999) study a model of opinion formation where agents are partially stubborn. That is, with a positive probability they keep their initial opinions, and with a complementary probability they update to a weighted average of their neighbours' opinions.[5] In Acemoğlu et al. (2013), regular agents, who update their opinions by taking weighted averages of their neighbours' opinions, are randomly matched with fully stubborn agents who never change their opinions. Both models generate long-run disagreement. But they do not provide an explicit motivation for prejudice as the primary cause of disagreement. They also do not quantify disagreement or examine how prejudice interacts with the interaction structure to drive disagreement.

Using a different framework, Krause (2000) and Hegselmann et al. (2002) show that disagreement emerges in a model of bounded confidence in which agents only incorporate other's opinions when they are sufficiently close to their own. Melguizo (2016) studies a related model where agents rewire their links based on how close their opinions and types are. We do not focus on the role of bounded confidence or on dynamic network formation.

The remainder of the paper is organized as follows. Section 2 outlines the model of opinion formation with prejudices. Section 3 examines how the network affects the composition of equilibrium opinions. In Sections 4 and 5, we quantify disagreement and examine its relationship

[4]In the DeGroot model of opinion formation, the intensity of prejudice is zero so that agents only take weighted averages of their neighbours opinions (DeGroot, 1974). The DeGroot framework of learning is known to converge to a consensus.

[5]Bindel et al. (2015) studies the price of disagreement in Friedkin and Johnsen (1999) model, which is the ratio of the total payoff in equilibrium to the total payoff at the social social optimum.

to prejudice and group cohesion. Section 6 characterizes the speed of learning and Section 7 offers concluding remarks. Technical proofs are relegated to Appendix B.

2. A model of opinion formation with prejudices

2.1. Agents and interactions

We consider a society of finite size denoted by a set $N = \{1, 2, \cdots, n\}$, where individuals interact through a social network. For each individual i, a *neighbourhood* is the set of other agents (e.g. family, friends, colleagues, e.t.c) they interact with. The interactions are summarized by an $n \times n$ non-negative interaction matrix W, whereby $w_{ij} > 0$ indicates the weight that i attaches to j' opinions. Interactions are directed, so that $w_{ij} > 0$ need not imply $w_{ji} > 0$, or equality between w_{ij} and w_{ji}. We assume that W is row-stochastic, which means that an agent's interactions with her neighbours are normalized to sum to one. Relaxing this assumption does not impact the results qualitatively but affects whether the learning process converges. We write N_i for the set of neighbours of i and d_i for its cardinality.

2.2. Prejudice and the cost of miscoordination

We model opinion formation as a learning process where agents repeatedly minimize the cost of miscoordination. When interacting with others, an agent's behaviour is driven by two competing motives. While an agent wants to agree with her personal *long-held opinions*, her utility depends on the degree to which her opinion coordinates with those held by others with whom she shares close ties. We interpret long-held opinions as *prejudice*, defined as a set of preconceived and *inflexible* opinions or beliefs about individual attributes, group behaviour or government policies and public programs. We interpret inflexibility of prejudices to imply that although they may change, they change more slowly compared to the rate at which overall opinions change due to learning from others. There is sufficient evidence in psychology literature showing that some attitudes/beliefs are resistant to change, and may in part be heritable (Tesser, 1993). That is, inculcated through culturalization. This component of attitudes is stored in memory and activated automatically with little conscious control (Fazio, 1986; Houston and Fazio, 1989). As individuals adjust their attitudes through learning, the inflexible beliefs are not completely erased (Wilson et al., 2000).

The desire to coordinate ones' opinion and behaviour with that of ones' neighbours is driven by individual desire for social conformism. Since the work of Asch and Guetzkow (1951) on social pressure, individual desire for social conformism is by now a well-studied phenomenon. More recently, Salganik et al. (2006) find evidence of conformity in individual taste in music, reflecting individual opinions about what constitutes good music. In economics, social conformism has been observed in work habits and effort exerted (e.g. Falk and Ichino (2006), Chen et al. (2010), Zafar (2011) and Abeler et al. (2011)), and participation in public good provision (Carpenter, 2004).

Our model is particularly in line with the empirical evidence in Kahan et al. (2012) and Kahan et al. (2017), showing that the observed patterns of disagreement regarding the role of human activity on climate change, results from the conflict of interest between conforming ones' beliefs to their political outlook (which is representative of ones' prejudices), and to the beliefs held by others with whom they share close ties. We then suppose that when adjusting their opinions, agents compromise between respecting their own personal prejudice, and conforming their opinions to those held by others with whom they share close ties.

To formalize these ideas, let p_i and $\bar{p}_i$ be the opinion and the extent of prejudice of agent i respectively. We assume that $p_i \in \mathbb{R}_+$, that is, *unidimensional* and represented by a positive real number. This assumption is consistent with empirical evidence suggesting that while individuals have opinions on many issues, spanning domains such as politics, lifestyle and the economy, an individual's opinions on all dimensions can be described using a *unidimensional spectrum*. Poole and Daniels (1985) and Ansolabehere et al. (2008) find that the voting behaviour of both legislators and individual voters can be explained by a single liberal-conservative dimension. Following in a similar line of argument, we assume that personal prejudices can also be described using a unidimensional spectrum, and more specifically, it takes values in the range $[0, 1]$. For any issue, say individual attribute, group behaviour, or government policy and public program, a value of $\bar{p}_i = 1$ means that agent i is fully prejudiced towards an issue, and a value of $\bar{p}_i = 0$ means i is not prejudiced. A value between zero and one then means that an agent is partially prejudiced. In situations where agents hold prejudices on many related issues, $\bar{p}_i$ is then a unidimensional parameter capturing the overall or average prejudice of an agent. The assumption regarding the range of values of $\bar{p}_i$ does not affect our results in a qualitative sense. Thus, in situations where it is suitable, it is feasible to assume $\bar{p}_i \in [-1, 1]$ with -1 representing a fully negatively prejudiced agent, 1 for a fully positively prejudiced agent, zero for no prejudice, and any other value in-between means an agent is only partial prejudiced.

Let $\mathbf{p}$ and $\bar{\mathbf{p}}$ denote the vectors of opinions and prejudices respectively, and write $\mathbf{p}_{-i}$ as the vector of opinions with i's opinion excluded. Each agent minimizes the cost $C_i(p_i, \mathbf{p}_{-i}, \bar{p}_i)$ of mis-coordinating her opinion against her personal prejudice and the opinions of her neighbours. That is

$$C_i(p_i, \mathbf{p}_{-i}, \bar{p}_i) = -\lambda_i \sum_{j=1}^{n} w_{ij}(p_j - p_i)^2 - (1 - \lambda_i)(p_i - \bar{p}_i)^2 \tag{1}$$

where λ_i is the intensity with which i conforms to the opinions of her neighbours; as such, we interpret $1 - \lambda_i$ as the intensity with which i conforms to her prejudice, or simply the *intensity of prejudice*. The first order condition to (1) is

$$p_i = -\lambda_i \sum_{j=1}^{n} w_{ij} p_j + (1 - \lambda_i)\bar{p}_i \tag{2}$$

Although we focus on a model of opinion formation, the cost function in (1) admits a game theoretic interpretation where p_i is i's action or effort and $\bar{p}_i$ is i's preference. Under this interpretation, an agent minimizes the cost of mis-coordinating her actions against her personal preference and that against the actions of her neighbours. This interpretation is used by Kuran and Sandholm (2008) to model cultural integration. Our model and results can thus be interpreted as a model of evolution of behaviour in the presence of relatively static preferences.

2.3. Evolution of opinions

Following the discussion in Section 2.2 above, each agent i updates her opinion myopically in accordance with the optimal value of p_i in (2). The updating occurs at discrete times $t \in \{0, 1, 2, \cdots\}$. Although prejudices may change over time, we assume (and also by virtue of the definition of prejudice given above) that they change very slowly compared to the rate at which overall opinions are updated. That is, the process of learning by averaging reaches equilibrium before prejudices are updated. We similarly assume that the intensities of prejudices are relatively stable compared to the rate of evolution of overall opinions. Let $p_i(t)$ and $\mathbf{p}(t)$ denote p_i and $\mathbf{p}$ respective at time t. Then from (2), agent i's opinion at t is given by

$$p_i(t) = -\lambda_i \sum_{j=1}^{n} w_{ij} p_j(t-1) + (1 - \lambda_i)\bar{p}_i \quad \text{for } i = 1, \cdots, n. \tag{3}$$

Let Λ be an $n \times n$ diagonal matrix with entries $\lambda_{ii} = \lambda_i$ and zero otherwise. Let also I denote an $n \times n$ identity matrix. Then the system (3) can be expressed in matrix form as

$$\mathbf{p}(t) = \Lambda W \mathbf{p}(t-1) + (I - \Lambda)\bar{\mathbf{p}} \tag{4}$$

The system of equations in (4) describe a simultaneous evolution of individual opinions where agents take weighted averages of neighbours' opinions and add them to their static prejudices. When the intensity of prejudice is zero for all agents, then (4) reduces to the

DeGroot model of opinion formation (DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2010; Acemoğlu et al., 2013). The DeGroot model converges to a consensus provided the network does not consist of cycles that lead agents to alternate in adopting each other's opinions. Disagreement also arises if the society consist of at least two stubborn agents who do not change their opinions (Acemoğlu et al., 2013). In the presence of prejudices, disagreement persists in equilibrium regardless of the network structure. The heterogeneity of λ_i's and $\bar{p}_i$'s is sufficient to generate heterogeneity in equilibrium opinions.

Our subsequent analysis is guided by three objectives. First, to examine how the network combines with prejudices to shape the composition equilibrium opinions. Second, to quantify the disagreement in equilibrium and to establish which network properties and model parameters drive disagreement. Third, to characterize the speed of learning and show how it can be used for parameter estimation in laboratory and field experiments.

3. Convergence and equilibrium opinions

This section examines the relationship between the composition of equilibrium opinions and the network structure. We identify individual network measures that influence the distribution of equilibrium opinions. For the sake of brevity, we focus on the special case where $\lambda_i = \lambda$ for all i; that is, where all agents possess the same intensity of prejudice. Considering this special case enables us to derive concise analytical results regarding the relationship between equilibrium opinions and individual network measures of centrality without affecting the results qualitatively. First, however, we show in the following lemma that the dynamics system (4) converges in a general case where all agents are prejudiced with different intensities.[6]

Lemma 1. *Let $\lambda_i \in (0, 1)$ for all $i \in N$. The dynamics system (4) describing evolution of opinions converges to a well-defined distribution of equilibrium opinions $\mathbf{p}^* = \lim_{t\to\infty} \mathbf{p}(t)$.*

Proof. See Appendix B.2

By a well-defined distribution of equilibrium opinions, we mean that equilibrium opinion p_i^* for each i is finite. Equilibrium opinions will be well-defined only if no single agent has unbounded *influence* on others' opinions. For the dynamics system in (4), the level of influence that each agent commands in equilibrium corresponds to how central they are in the network, and in particular, the *Bonacich centrality* (Bonacich, 1987). The Bonacich centrality of an agent

[6] Convergence conditions for the general case where the network contains a subgroup of agents who are not prejudiced, that is $\lambda_i = 1$ for any i in this subgroup, are provided in Parsegov et al. (2017, Theorem 1).

is a function of how many connections she has, and how many connections her neighbours have, and how many connections her neighbours' neighbours have, and so on. We define two related measures of Bonacich centrality each with a different implication for equilibrium opinions: *Bonacich out-centralities* and *Bonacich in-centralities*.

Let the intensity of prejudice λ be identical for all agents and $\lambda \in (0,1)$. For an interaction matrix W, define an $n \times n$ matrix $B^{[\lambda]}(W)$ as follows.

$$B^{[\lambda]}(W) = (1-\lambda)\sum_{\tau=0}^{+\infty}(\lambda W)^{\tau} \tag{5}$$

The following claim holds (see the proof in Appendix B.1).

Claim 1. *Each element $b_{ij}^{[\lambda]}(W)$ of $B^{[\lambda]}(W)$ is finite, and is equivalent to the expected normalized number of times j's opinion is incorporated into i's opinion in the long-run.*

Claim 1 directly implies that the quantity $b_{ij}^{[\lambda]}(W)$ measures the level of influence that j exerts on i's equilibrium opinion.[7]

Definition 1. *Let $\mathbf{e}$ be a column vector of ones and $B^{[\lambda]}(W)$ as defined above.*

(i) A vector of Bonacich out-centralities $\mathbf{b}^{[\lambda,out]}(W)$ is defined as $\mathbf{b}^{[\lambda,out]}(W) = B^{[\lambda]}(W)\mathbf{e}$.

(ii) A vector of Bonacich in-centralities $\mathbf{b}^{[\lambda,in]}(W)$ is defined as $\mathbf{b}^{[\lambda,in]}(W) = \mathbf{e}^T B^{[\lambda]}(W)$, where $\mathbf{e}^T$ is a transpose of $\mathbf{e}$.

Following the interpretation of $B^{[\lambda]}(W)$ above, the vector of Bonacich out-centralities measures the extent to which an agent is *influenced by* others' opinions. A related measure to Bonacich out-centralities is the $\mathbf{v}$-*adjusted* vector of Bonacich out-centralities $\mathbf{b}^{[\lambda,out]}(W;\mathbf{v})$, defined as $\mathbf{b}^{[\lambda,out]}(W;\mathbf{v}) = B^{[\lambda]}(W)\mathbf{v}$, where $\mathbf{v}$ consists of elements v_i that are not necessarily equal to one. In the case where $\mathbf{v} = \bar{\mathbf{p}}$, then the $\bar{\mathbf{p}}$-*adjusted* vector of Bonacich out-centralities measures the extent to which agents' long-run opinions are influenced by others prejudices.

The vector of Bonacich in-centralities on the other hand measures the extent to which an agent *influences* other's long-run opinions. Recall that $b_{ij}^{[\lambda]}(W)$ is the expected normalized number of times j's opinion is incorporated into i's opinion in the long-run. The Bonacich in-centrality of j is $b_j^{[\lambda,in]}(W) = \sum_{i=1}^{n} b_{ij}^{[\lambda]}(W)$, and measures the total expected normalized number of times j's opinion is incorporated into the opinions of all other agents.

[7]Note that we use "long-run" and "equilibrium" interchangeably. The two are equivalent for the time frame of our model.

A high number of neighbours and being directly connected to other agents with a high number of direct neighbours is on average associated to high Bonacich centrality (Bonacich, 1987). The following proposition relates the vectors of Bonacich out-centralities and Bonacich in-centralities to the distribution of, and the average equilibrium opinions respectively, where average equilibrium opinion is defined as $\text{Ave}[\mathbf{p}^*] = \frac{1}{n}\sum_{i=1}^{n} p_i^*$.

Proposition 1. *Let the intensity of prejudice λ be identical for all agents and $\lambda \in (0,1)$. The distribution of and average equilibrium opinions of a dynamic process described by* (4) *are respectively $\mathbf{p}^* = \mathbf{b}^{[\lambda,out]}(W;\bar{\mathbf{p}})$ and $Ave[\mathbf{p}^*] = \frac{1}{n}\sum_{j=1}^{n} b_j^{[\lambda,in]}(W)\bar{p}_j$.*

Proof. See Appendix B.3

Proposition 1 shows that the distribution of equilibrium opinions is equal to the distribution of $\bar{\mathbf{p}}$-*adjusted* vector of Bonacich out-centralities. The source of disagreement in equilibrium is thus the heterogeneity in the levels of influence that agents command in the network. The levels of influence depend on the topology of the network and intensity of prejudice. If an extreme opinion is defined by how far an agent's equilibrium opinion is from the average, then the agents with extreme opinions are those with the largest and lowest Bonacich out-centralities.

The average equilibrium opinion increases with Bonacich in-centralities. Recall that the Bonacich in-centrality of an agent measures the total expected normalized number of times that agent's opinion is incorporated into the opinions of all other agents. The quantity $b_i^{[\lambda,in]}(W)\bar{p}_i$ is then the overall contribution of i' prejudice into the average opinion. The higher the Bonacich in-centrality of i, the more influence i's prejudice exerts on average equilibrium opinion. The prejudices of agents with the largest Bonacich in-centralities are thus the most adopted by the society. For any government and organizational policies aimed at influencing equilibrium opinions, the Bonacich in-centrality is a suitable measure for identifying "key players" in the society.

Many studies find that Bonacich centrality is the main network measure that influences most dynamic processes on networks. For example, Ballester et al. (2006) show that equilibrium behaviour in network games depends on Bonacich (out-) centralities. Calvó-Armengol et al. (2009) and Liu et al. (2012) show how Bonacich (out-) centrality influence equilibrium behaviour in models of peer effects in education and crime networks respectively. Our results above show that Bonacich centrality also determines the nature of equilibrium opinions in a model of opinion formation with prejudices. In the following section, we develop a measure for disagreement and characterize its relationship to the network topology and intensity of prejudice.

4. Quantifying disagreement

This section quantifies equilibrium disagreement. We define the magnitude of disagreement as the distance between the distribution of opinions in a prejudiced society and the distribution of equilibrium opinions in a non-prejudiced society. When agents are not prejudiced and $\lambda_i = 1$ for all agents, the dynamic system in (4) reduces to the DeGroot model of opinion formation (DeGroot, 1974; DeMarzo et al., 2003; Acemoğlu et al., 2013; Golub and Jackson, 2010). The DeGroot model converges to a consensus in the long-run with equilibrium opinions described by an influence vector π corresponding to the vector of *eigenvector centralities*.[8] A group of agents $C \subset N$ reaches a consensus under W if $p_i^* = p_j^*$ for each i and j in C.

From our analysis of Section 3, a consensus is reached under W in the model with prejudices only if $B^{[\lambda]}(W) = \mathbf{e}\pi^T$, which is a matrix with identical rows equal to π. This equality holds whenever agents are not prejudiced. To see why, note that when $\lambda_i = 1$ for all i, then an equivalent relation to the definition of $B^{[\lambda]}(W)$ in (5) is

$$B(W) = \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t} W^\tau = \mathbf{e}\pi^T \tag{6}$$

Just like $B^{[\lambda]}(W)$, the matrix $B(W)$ also represents the normalized (fractional) number of visits to each state in the long-run. The second equality is because each row of $\frac{1}{t}\sum_{\tau=0}^{t} W^\tau$ converges to an identical vector π, which is equivalent to the stationary distribution of a random walk process on W.[9] We then define the magnitude of disagreement in equilibrium as the distance between the vector of opinions in equilibrium $\mathbf{p}^*$, and the consensus vector $\mathbf{c} = B(W)\bar{\mathbf{p}}$.

Definition 2. *Given the vector of equilibrium opinions $\mathbf{p}^*$ in the model of opinion formation with prejudice, the magnitude of equilibrium disagreement $D(W;\bar{\mathbf{p}})$ under interaction matrix W and vector of prejudices $\bar{\mathbf{p}}$ is defined as*

$$D(W;\bar{\mathbf{p}}) = \|\mathbf{p}^* - \mathbf{c}\|_1$$

where $\| \mathbf{p} \|_1 = \sum_{i=1}^n | p_i |$ is the Euclidean norm of $\mathbf{p}$.

Since $\mathbf{c}$ is the expected consensus vector, and each c_i is the expected consensus value for i. Each $| p_i^* - c_i |$ is then the expected *individual disagreement* of agent i. The magnitude of disagreement in the society is thus the sum of individual disagreements.

[8]The vector of eigenvector centralities corresponds to the vector of Bonacich centralities when $\lambda = 1$.

[9]The matrix $\frac{1}{t}\sum_{\tau=0}^{t} W^\tau$ consists of the normalized number of visits to each state (agent) starting from another. In the long-run, each row of $\lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t} W^\tau$ converges to the stationary distribution π (which is equivalent to the left-eigenvector of W corresponding to the leading eigenvalue)of the random walk process on W; the stationary distribution is the fractional number of visits to each state in the long-run.

Given Definition 2, we proceed by first establishing the relationship between the magnitude of disagreement and prejudices and the network structure. In particular, we show that disagreement depends on the intensity and extent of prejudice, and the spectral properties of the network. Let $1 = \mu_1(W) \geq \mu_2(W) \geq, \cdots, \geq \mu_n(W)$ be the eigenvalues of W, where the first equality is because W is a stochastic matrix. We show that $D(W; \bar{\mathbf{p}})$ is a function of λ, the vector of prejudices $\bar{\mathbf{p}}$, and the eigenvalue $\mu_*(W) = \max\{\mu_2(W), \mid \mu_n(W) \mid\}$.

Since for most connected and strongly connected networks $\mu_2(W) > \mid \mu_n(W) \mid$, we focus on examining the properties of the second largest eigenvalue $\mu_2(W)$.[10] We show that the second largest eigenvalue is a measure of *intensities* of group cohesion. A subgroup of agents is *cohesive* if every member of the subgroup has more than half of her interactions with other members within the subgroup. That is, let $L_k \subset N$ be a subset of N; then L_k is said to be cohesive if for every $i \in L_k$,

$$\sum_{j \in L_k} w_{ij} > \frac{1}{2} \sum_{l \in N} w_{il} \tag{7}$$

The *intensity* of group cohesion is the relative total weight of interactions among group members to the total weight of all interactions of group members. That is, given a cohesive group $L_k \subset N$ and its cardinality n_k, then the intensity of cohesion $\iota(L_k)$ of L_k is

$$\iota(L_k) = \frac{\sum_{i \in L_k} \sum_{j \in L_k} w_{ij}}{\sum_{i \in L_k} \sum_{j \in N} w_{ij}} = \frac{1}{n_k} \sum_{i \in L_k} \sum_{j \in L_k} w_{ij} \tag{8}$$

where the second equality on the right hand side of (8) follows from $\sum_{j \in N} w_{ij} = 1$. We show below that if a network consists of at least two cohesive subgroups, then $\mu_2(W)$ is a measure of intensities of cohesion of subgroups. We first derive the relationship between $D(W; \bar{\mathbf{p}})$ and λ, $\bar{\mathbf{p}}$, $\mu_2(W)$.

Proposition 2. *For a dynamic process described by (4), the magnitude of disagreement $D(W; \bar{\mathbf{p}})$ at t under W is bounded by*

$$n\bar{p}r \left(\frac{1 - \lambda}{1 - \lambda \mu_*(W)} \right) \leq D(W; \bar{\mathbf{p}}) \leq \frac{n S_{\bar{p}}}{\pi_{\min}} \left(\frac{1 - \lambda}{1 - \lambda \mu_*(W)} \right) \tag{9}$$

where $\pi_{\min} = \min_{i \in N} \pi_i$, $\bar{p} = \min_{i \in N} \bar{p}_i$, $S_{\bar{p}} = \sum_{i \in N} \bar{p}_i$, and $0 \leq r \leq (1 - \lambda)$ is some real number.

Proof. See Appendix B.4

Proposition 2 describes how the network structure and personal prejudices influence the magnitude of equilibrium disagreement. Firstly, the magnitude of disagreement increases with

[10] A network is *connected* if no agent stands in isolation. It is *strongly connected* if a path exists from any one agent to every other agent.

the population size. This is because overall disagreement is defined as the sum of individual disagreements. If the magnitude of disagreement is for example defined as the average of individual disagreements then the former would not directly increase with n, only possibly indirectly. For the definition adopted in this paper, Proposition 2 says that the larger the population size, the harder it is to achieve a consensus in the society.

Secondly, the magnitude of equilibrium disagreement depends positively on the extents and intensities of prejudices. The lower bound increases with $\bar{p}$, which is the extent of prejudice for the least prejudiced agent, and the upper bound increases with $S_{\bar{p}}$, which is the sum of the extents of prejudices. Recall that the extent of prejudice $\bar{p}_i \in [0, 1]$ so that $\bar{p}_i$ is close to one for a highly prejudiced agent. The intensity of prejudice $1 - \lambda$ on the other hand positively influence disagreement through its interaction with the network. In particular, λ influences measures of individual network centralities as discussed in Section 3 above. Proposition 2 then states that the magnitude of disagreement will tend to be higher in societies where agents are highly prejudiced.

Our result suggest that regardless of the network structure, there are two ways through which disagreement can be reduced in the society: by reducing the extent of individual prejudice, and by reducing the intensity of prejudice. Both are individual attributes that are generally inflexible but can be slowly influenced through educational processes. Several papers have documented how diversity education helps reduce prejudices. For example, Hogan and Mallott (2005) show that diversity courses in higher education were effective in improving students' intergroup tolerance (see Astin (1993) and Kulik and Roberson (2008) for a review of the related literature).

The third factor that influences the magnitude of disagreement is the network structure, and it does so in two ways: through the influence vector π and the eigenvalue $\mu_*(W) = \max\{\mu_2(W), \mid \mu_n(W) \mid\}$. Disagreement increases with $\pi_{\min}$, which is the minimum value of the influence vector, hereafter *minimum influence*. The minimum influence is inversely proportional to *network skewness*. The higher network skewness, in the sense that one agent or a subgroup of agents receives disproportionally more attention than others, the smaller is $\pi_{\min}$ (see Example 1 below for a detailed illustration).

For spectral properties of the network, we focus on examining the relationship between the second largest eigenvalue and the structure of the network (for the reasons discussed above). We particularly aim to show that if cohesive subgroups are present in the society, then the second largest eigenvalue describes the intensities of their cohesiveness. When there are no distinct subgroups, then the second largest eigenvalue measures the density of the network. Overall, since W is a stochastic matrix, $\mu_2(W)$ is equal to one if the network is disconnected. It is close to one for sparsely connected networks, and close to zero for densely connected networks. The following example illustrates the relationship between $\mu_2(W)$ and the intensities of group cohesion, and between minimum influence and network skewness.

Example 1: Consider the following $n \times n$ interaction matrix W depicted in (10).

$$W = \begin{bmatrix} 1-\epsilon & \frac{\epsilon}{n-1} & \frac{\epsilon}{n-1} & \cdots & \frac{\epsilon}{n-1} \\ \frac{1}{n-1} & 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \frac{1}{n-1} & \frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1} & \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \end{bmatrix} \tag{10}$$

For $0 < \epsilon < 1$, agent 1 has more direct influence and receives $(1-\epsilon)\frac{n}{n-1}$ more attention than any other agent.[11] Agent 1's influence thus increases as ϵ tends to zero. As ϵ tends to one, W tends to a complete graph where all agents are connected to each other and have the same direct influence. For ϵ sufficiently small, we can partition the society into to cohesive subgroups $L_1 = \{1\}$ and $L_2 = \{2, 3, \cdots, n\}$; both subgroups satisfy the condition for cohesion in (7). The intensities of group cohesion are $\iota(L_1) = 1 - \epsilon$ and $\iota(L_2) = \frac{n-2}{n-1}$. We see that the intensity of cohesion of group L_1 increase as ϵ becomes small, and of L_2 as n increases.

The eigenvalues of W are (Desai and Rao, 1993): $\mu_1(W) = 1$, $\mu_2(W) = \frac{n-2}{n-1} - \epsilon$, $\mu_3(W) = \mu_4(W) = \cdots = \mu_n(W) = \frac{-1}{n-1}$, and hence $\mu_*(W) = \frac{n-2}{n-1} - \epsilon$. Clearly the second largest eigenvalue and hence $\mu_*(W)$ increases with intensities of group cohesion.

The influence vector π of W is (Desai and Rao, 1993): $\pi_1 = \frac{1}{1+(n-1)\epsilon}$ and $\pi_i = \frac{\epsilon}{1+(n-1)\epsilon}$ for $i \geq 2$. As discussed above, π captures the skewness of W so that $\pi_{\min} = \frac{\epsilon}{1+(n-1)\epsilon}$ increases with ϵ. Note that as ϵ increases, agent 1 (subgroup L_1) disproportionately receives more attention and hence making the network more skewed.

The intuition in Example 1 generalizes to all arbitrary networks with two cohesive subgroups. We relegate the discussion for relationship between $\mu_2(W)$ and group cohesion in the case of two subgroups to Section Appendix A. We instead focus on the general case of two or more cohesive subgroups in the next section.

[11] This follows from subtracting the total attention agent 1 receive $= (1-\epsilon) + (n-1) \times \frac{1}{n-1}$ by the attention each of the other agents receives $= \frac{\epsilon}{n-1} + (n-2) \times \frac{1}{n-1}$.

5. The second largest eigenvalue and group cohesion

Several papers point to the role of the second largest eigenvalue in influencing the long-run behaviour and convergence rates of dynamic processes on networks. First, Cavalcanti et al. (2017) relate the second largest eigenvalue to network cohesion. Network cohesion, as opposed to subgroup cohesion in our model, is a measure of how uniform or fragmented a network is. As such, Cavalcanti et al. (2017) equate network cohesion to one minus the second largest eigenvalue. They then show that network cohesion is crucial for conditional convergence in dynamic models of endogenous perpetual growth with network externalities.

Second, Golub and Jackson (2012) show that the second largest eigenvalue determines the speed of convergence in the DeGroot model of opinion formation. They relate the second largest eigenvalue to homophily, the tendency of individuals to interact with others with whom they share attribute (e.g. race, religious beliefs, political orientation). They do so by considering a family of networks formed through a random Bernoulli process so that every neighbour who is listened to is weighted equally. The resulting network is then aggregated into subgroups of agents sharing similar attributes and the interactions between subgroups are symmetric. Homophily is subsequently defined as the second largest eigenvalue of the aggregated interaction matrix. This framework enables Golub and Jackson (2012) to use mean-field approximation techniques to show that as population size tends to infinity, the second largest eigenvalue of the entire network and that of the aggregated network coincide. The relationship between the second largest eigenvalue and homophily is then motivated using the relationship between the former and the conductance of a network (see Lemma 2 for a detailed discussion).

Our analysis in this section is driven by two motivations. Firstly, we aim to provide an alternative motivation as to how and when the second largest eigenvalue measures group cohesion in the society. We do not rely on the relationship between the second largest eigenvalue and the conductance of the network since it does not directly extend to cases with more than two cohesive subgroups. Secondly, although the results in Golub and Jackson (2012) provide insights on the relationship between the second largest eigenvalue and existence of homophilous subgroups, random Bernoulli networks are restrictive and not representative of many geographical and real-world social networks. The problem that arises is that using the second largest eigenvalue as an indication of group interaction is valid only in specific families of networks and conditions. As Golub and Jackson (2012) note in their concluding remarks, going beyond random networks with uniform interactions requires different analytical techniques that do not rely on mean-field approximations. It is therefore necessary to further examine conditions under which the second largest eigenvalue indeed measures group cohesion. This endeavour is not only useful on the theoretical point of view but also for empirical and experimental studies in two ways: (i) Our measure of group cohesion, that is, intensity of group cohesion, is well defined and easily measurable from empirical data; (ii) Our results are for finite networks, which makes them directly applicable to empirical networks.

In the following analysis, we develop methods for examining the relationship between the second largest eigenvalue of the interaction matrix and the intensity of group cohesion. Our results apply to general (both deterministic and random) networks with heterogeneous interactions. In particular, we use the concept of *lumpability* of matrices.

Let $\mathcal{L} = \{L_1, L_2, \cdots, L_k\}$ disjoint subsets be the cohesive subgroups into which N is partitioned, and denote the cardinality of L_l by n_l. Let $\tilde{W}$ be a reduced transition matrix that aggregates subgroups into single points, so that $\tilde{W}_{jl}$ is the overall attention that agents in subgroup L_j attach to agents in subgroup L_l. Given the partition $\mathcal{L}$, let V be an $n \times k$ *collector matrix* with elements $v_{il} = 1$ if $i \in L_l$, and zero otherwise. Then lumpability is defined as follows.[12]

Definition 3. *Let e_i be a basis vector and $\mathbf{e}$ a vector of ones. An interaction matrix W is strongly lumpable under $\mathcal{L}$ into another matrix $\tilde{W}$ if*

$$(e_i - e_j)WV = 0 \quad \textit{for all } i, j \in L_l \in \mathcal{L}. \tag{11}$$

It is near lumpability if for some $\varepsilon > 0$,

$$(e_i - e_j)WV \leq \varepsilon\mathbf{e} \quad \textit{for all } i, j \in L_l \in \mathcal{L}. \tag{12}$$

The definition of strong lumpability in (11) states that for each agent i belonging to subgroup L_m, the quantity $\sum_{j \in L_l} w_{ij}$, which is the total attention that i attaches to members of subgroup L_l, is identical for all i in L_m. The following example helps clarify the relation in (11).

Example 3: Consider an interaction matrix W in (13) with $N = \{1, 2, \cdots, 8\}$. W satisfies

[12] See Buchholz (1994) for a similar definition.

the definition of strong lumpability with $L_1 = \{1,2\}$, $L_2 = \{3,4\}$ and $L_3 = \{5,6,7,8\}$.

$$W = \begin{bmatrix} 0.5 & 0.3 & 0.1 & 0 & 0.06 & 0 & 0.04 & 0 \\ 0.2 & 0.6 & 0 & 0.1 & 0.01 & 0 & 0.09 & 0 \\ 0.1 & 0 & 0.3 & 0.6 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.4 & 0.5 & 0 & 0 & 0 & 0 \\ 0.15 & 0.05 & 0 & 0 & 0.3 & 0.4 & 0.1 & 0 \\ 0.2 & 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0 & 0.2 & 0 & 0 & 0.1 & 0.3 & 0.2 & 0.2 \\ 0 & 0.2 & 0 & 0 & 0 & 0.3 & 0.2 & 0.3 \end{bmatrix} \tag{13}$$

Given the partitioning of N into $\mathcal{L}$, the collector matrix V, and hence the product WV is given by

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad WV = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0 \\ 0.1 & 0.9 & 0 \\ 0.2 & 0 & 0.8 \\ 0.2 & 0 & 0.8 \\ 0.2 & 0 & 0.8 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

The following observation regarding Example 3 generalizes to all interaction matrices that satisfy the definition of strong lumpability. The effect of the collector matrix V on W is to aggregate W into a matrix WV depicting the total weight of interactions each agent attaches to different subgroups. That is, the matrix $\bar{W} = WV$ is an $n \times k$ matrix, with n being the number of agents and k the number of subgroups. Each element $\bar{w}_{il}$ of $\bar{W}$ is given by $\bar{w}_{il} = \sum_{j \in L_l} w_{ij}$, which is the total weight of interactions that agent i attaches to subgroup L_l. The product $e_i WV$ is then the ith row of $\bar{W}$, which is a vector with elements $\bar{w}_{iL_l} = \sum_{j \in L_l} w_{ij}$ for all L_l. The definition for strong lumpability then requires that for every pair of agents i and j belonging to the same subgroup, $\bar{w}_{iL_l} = \bar{w}_{jL_l}$ for each subgroup L_l. But for any pair of distinct subgroups L_l and L_m, $\bar{w}_{iL_l}$ need not equal $\bar{w}_{iL_m}$. This is clearly the case for W in (13) above.

Strong lumpability on its own is a restrictive concept in that it requires all agents belonging to the same subgroup to have identical total interactions with members of another subgroup. It however allows for heterogeneity of interactions between individuals and across subgroups. Consider a society that can be categorized into three subgroups based on political orientation: left-leaning, centre and right-leaning. Strong lumpability then requires that the probability that a left-leaning agent interacts with a centre-leaning agent is identical for all left-leaning agents, and similarly for the interactions between left-leaning and right-leaning agents. But the probabilities of interactions between left-leaning with centre-leaning, and left-leaning with right-leaning agents need not be identical.

To expand on the scope of networks for which our results are applicable, we introduce the concept of near lumpability defined in (12). Near lumpability requires the network to only be close to being strongly lumpable. Note that definition (12) can equivalently be stated as $W = P + \varepsilon R$, where P is strongly lumpable and R is an arbitrary matrix. Despite being restrictive to an extent, the concept of lumpability is useful in empirical and experimental studies because the networks that satisfy these conditions are easier to set up and implement. Moreover, both strong and near lumpability permit heterogeneous interactions among agents and across subgroups, and is not restrictive in terms of the size of the network.

To relate the interaction matrix W to its aggregated form $\tilde{W}$, we need to define another collector matrix U from V as $U = (V^T V)^{-1} V^T$. By definition, U is a $k \times n$ matrix with elements $u_{li} = \frac{1}{n_l}$ if i belongs to subgroup L_l and zero otherwise. The corresponding aggregated interaction matrix is then defined as $\tilde{W} = UWV$. Similarly, by definition $\tilde{W}$ is a $k \times k$ matrix, where k is the number of subgroups. The diagonal entries $\tilde{w}_{ll}$ for $l = 1, \cdots, k$ of $\tilde{W}$ are equivalent to the intensity of cohesiveness of subgroup l.[13] For Example 3 above, we compute U and $\tilde{W}$ to be

$$U = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad \tilde{W} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0 \\ 0.2 & 0 & 0.8 \end{bmatrix}$$

Given the definitions of strong and near lumpability above, our next objective is twofold. First, we show that the second largest eigenvalue of $\tilde{W}$ measures the intensities of cohesion within subgroups. Consider for example a matrix $\tilde{W}(\alpha)$ derived from $\tilde{W}$ as follows.

$$\tilde{W}(\alpha) = \alpha I + (1-\alpha)\,\tilde{W}$$

with $\alpha \in [0, 1)$, so that $\tilde{W}(\alpha) = \tilde{W}$ when $\alpha = 0$, and as α increases, the intensities of co-

[13]To see why, note that the product matrix $\bar{W} = WV$ is an $n \times k$ matrix, with n being the number of agents and k the number of subgroups. Each element $\bar{w}_{il}$ of $\bar{W}$ is the total weight of interactions that agent i attaches to subgroup L_l, hence $\bar{w}_{il} = \sum_{j \in L_l} w_{ij}$. The collector matrix U on the other hand is a $k \times n$ matrix with elements $u_{li} = \frac{1}{n_l}$ if i belongs to subgroup L_l and zero otherwise. The aggregated matrix $\tilde{W} = UWV$ thus consists of elements $\tilde{w}_{lm} = \frac{1}{n_l} \sum_{i \in L_l} \sum_{j \in L_m} w_{ij}$. So if $l = m$, then $\tilde{w}_{ll}$, the lth diagonal element of $\tilde{W}$ is equivalent to $\iota(L_l)$, the intensity of cohesiveness of subgroups L_l as defined in (8).

hesion within groups (i.e. the diagonal elements $\tilde{W}_{jj}(\alpha)$ for $j = 1, \cdots, k$) also increase. The relationship between the second largest eigenvalue $\mu_2(\tilde{W}(\alpha))$ of the new matrix $\tilde{W}(\alpha)$, and the parameter measuring intensities of group cohesion α is[14]

$$\mu_2(\tilde{W}(\alpha)) = \mu_2(\tilde{W}) + \alpha\left(1 - \mu_2(\tilde{W})\right)$$

Since $\mu_2(\tilde{W}) < 1$, $\mu_2(\tilde{W}(\alpha))$ thus increases with the parameter of intensities of subgroup cohesion α. We generalize this argument in Proposition 3 below.

Our second objective is to show that when the network is nearly lumpable, then the second largest eigenvalue of the aggregated interaction matrix $\tilde{W}$ is close to the second largest eigenvalue of the non-aggregated interaction matrix W. If the interaction matrix is strongly lumpable then the second largest eigenvalues of $\tilde{W}$ and W are identical. The line of argument we follow is that if the second largest eigenvalue of the aggregated interaction matrix measures the intensities of group cohesion, and the second largest eigenvalues of the aggregated and non-aggregated interaction matrices are identical or close to being identical, then from Proposition 2, the magnitude of equilibrium disagreement increases with intensities of group cohesion

Proposition 3. *The second largest eigenvalue $\mu_2(\tilde{W})$ of $\tilde{W}$ measures the intensities of cohesion within subgroups of $\mathcal{L}$. If W is strongly lumpable with respect to a partition $\mathcal{L}$ to $\tilde{W}$, then $\mu_2(\tilde{W}) = \mu_2(W)$.*

Let W be nearly lumpable with respect to a partition $\mathcal{L}$ to $\tilde{W}$ so that $W = P + \varepsilon R$, where P is strongly lumpable and R is some arbitrary matrix. Let $R_{\max} = \max_{i\in N}\sum_{j=1}^{n} r_{ij}$ be the maximum row-sum of R. Then

$$\mid \mu_2(W) - \mu_2(\tilde{W}) \mid \leq \varepsilon R_{\max}\rho(R) \tag{14}$$

where $\rho(R) = 1 + \frac{1}{|\mu_n(R)|}$ and $\mid \mu_n(R) \mid$ is the magnitude of the smallest eigenvalue of R.

Proof. See Appendix B.6

Proposition 3 shows that if W is strongly lumpable then $\mu_2(W) = \mu_2(\tilde{W})$. For the interaction matrix in Example 3 above, the eigenvalues of $\tilde{W}$ are $\mu_1(\tilde{W}) = 1$, $\mu_2(\tilde{W}) = 0.862$ and $\mu_3(\tilde{W}) = 0.638$, which are identical to the first three eigenvalues of W, and hence $\mu_2(W) = \mu_2(\tilde{W}) = 0.862$. Proposition 3 also states that provided W is sufficiently close to being strongly

[14]Let $\mathbf{v}$ be the right eigenvector corresponding to the second largest eigenvalue $\mu_2(\tilde{W})$ of $\tilde{W}$ so that $\tilde{W}\mathbf{v} = \mu_2(\tilde{W})\mathbf{v}$. If $\mathbf{v}$ is also the right eigenvector o $\tilde{W}(\alpha)$, then $\tilde{W}(\alpha)\mathbf{v} = \mu_2(\tilde{W}(\alpha))\mathbf{v} = \alpha\mathbf{v} + (1-\alpha)\tilde{W}\mathbf{v} = (\alpha + (1-\alpha)\mu_2(\tilde{W}))\mathbf{v}$. Hence, $\mu_2(\tilde{W}(\alpha)) = \mu_2(\tilde{W}) + \alpha\left(1 - \mu_2(\tilde{W})\right)$.

lumpable, that is ε is sufficiently small, then $\mu_2(W)$ measures the intensities of group cohesion in W. Overall, we have shown that if a society consists of cohesive subgroups, which may be based on political and religious orientation, income classification, racial lines or simply based on geographical constraints, then the second largest eigenvalue measures the extent of their cohesiveness. The results of Proposition 2 then imply that the existence of cohesive subgroups drives public disagreement, and the more cohesive the subgroups, the larger the magnitude of disagreement.

Decades of research have shown that prejudices are prevalent in societies where individuals identify with specific groups that are distinct in nature: for example, on racial, political, educational and income grounds. The lesser the interaction between subgroups, which is equivalent to higher intensity of cohesiveness, the larger the extent of prejudice. Indeed, Allport (1954), Masson and Verkuyten (1993), Wright et al. (1997) and Levin et al. (2003) show that intergroup contact reduces prejudice. In our analysis above, the positive effect of group cohesion on equilibrium disagreement is a secondary effect. The primary cause of equilibrium disagreement is the presence of prejudice. In the DeGroot model for example, where agents do not hold prejudices and only take weighted average of neighbours' opinions, a consensus emerges even when the society is made up of cohesive subgroups. Our results, together with those on the relationship between intergroup contact and prejudice, then suggest that government policies and public programs aimed at fostering social, economic and cultural integration, should in principal also lead to reduction in public disagreement. In particular, instead of investing in diversity education which is a long-term and expensive process, governments and organizations could instead invest on programs that encourage intergroup contacts.

6. The rate of decay of disagreement

For any model of learning, examining the speed of convergence is just as relevant as examining the properties of equilibrium behaviour. It is very important to understand whether the predicted equilibrium behaviour can be reached at the time scales of economic relevance. In this section, we show that the speed of learning can also be used to distinguish between models of learning by averaging and parameter estimation. Since disagreement in the society about factual issues occurs more often than not, some papers have studied variations of the DeGroot model so as to generate disagreement as an equilibrium behaviour (Friedkin and Johnsen, 1999; Hegselmann et al., 2002; Acemoğlu et al., 2013; Melguizo, 2016). The question that arises however is whether the observed disagreement is indeed an equilibrium behaviour or that agents follow that DeGroot (which in our case means they do not hold prejudices) but the observed disagreement is because reaching a consensus in the DeGroot model takes very long. The following example helps illustrate this point.

Consider the case of public opinion around climate change. By 2009, over 97% of scientists who registered climate science as their area of expertise said that human activity is a significant contributing factor in changing mean global temperature. This value indicates a near consensus within climate science community. In contrast, 82% of the overall scientific community believed that human activity contributed to global temperature changes, and only 58% of the American public believed the same (Doran and Zimmerman, 2009). Empirical evidence suggests that public disagreement over climate change stems not from the public's incomprehension of science but from the process of opinion formation (Kahan et al., 2012).

In the above example, it is not possible to immediately conclude which forces contribute to the disparities in observed disagreement across groups. The observed near-consensus among climate scientist can be because they do not hold prejudices; that is the intensity of prejudice is zero or very low. But it can also be a result of the network structure. Compared to the rest of the society, it could be that the community of climate scientists is small and its members have relatively identical influence (Bonacich centralities). These two explanations both fit the predictions discussed in Section 4. Alternatively, it could be that the community of climate scientists and the rest of the society all follow the DeGroot model but learning in the former converges faster than in the latter. Here, we show that examining the speed of convergence helps to distinguish between models of learning.

We define the speed of convergence as the time it takes the learning process to get close to equilibrium. Define the distance between the vector of opinions $\mathbf{p}(t)$ at t and equilibrium vector of opinions $\mathbf{p}^*$ as $DE(t; W; \bar{\mathbf{p}}) = \parallel \mathbf{p}(t) - \mathbf{p}^* \parallel_1$, where as before, $\parallel \mathbf{p} \parallel_1 = \sum_{i=1}^{n} \mid p_i \mid$ is the Euclidean norm of $\mathbf{p}$. For some small real number $\varepsilon > 0$, the convergence time is the time it takes for the distance $DE(t; W; \bar{\mathbf{p}})$ to get below ε.

Definition 4. *The convergence time $CT(\varepsilon; W; \bar{\mathbf{p}})$ to $\varepsilon > 0$ under interaction matrix W is*

$$CT(\varepsilon; W; \bar{\mathbf{p}}) = \min\{t : \; DE(t; W; \bar{\mathbf{p}}) < \varepsilon\} \tag{15}$$

The convergence time is a function of the vector of the extent of prejudice $\bar{\mathbf{p}}$. This is because the vector of prejudice can be selected in such a way that $\mathbf{p}(1)$, and hence $\mathbf{p}(t)$ for $t \geq 2$, is close to $\mathbf{p}^*$; Under this scenario, learning stops after a few steps.

Proposition 4. *The convergence time in the model of learning by averaging described by* (4) *is bounded by*

$$\frac{\ln\left(\varepsilon/2rn\bar{p}\right)}{\ln\left(\lambda\right)} - 1 \leq CT(\varepsilon; W; \bar{\mathbf{p}}) \leq \frac{\ln\left(\varepsilon/2nS_{\bar{p}}\right)}{\ln\left(\lambda\right)} - 1 \tag{16}$$

where $0 \leq r \leq \frac{\bar{p}_m}{S_{\bar{p}}}$ *and* $\bar{p}_m = \max_i \bar{p}_i$.

Proof. See Appendix B.7

Proposition 4 shows that the convergence time increases with intensity and extent of prejudice, but independent of the network structure. Compared to the results in Section 4, although the intensity of prejudice increases the magnitude of equilibrium disagreement, it ensures that convergence to equilibrium occurs rapidly. The underlying reason for this relationship is that the higher the intensity of prejudice the less attention agents pay to their neighbours' opinions. The extreme case is when $\lambda = 0$; under this scenario, each agent sticks to their prejudice and no information exchange occurs in the society. Learning thus stops after a single step of iteration.

The result of Proposition 4 has relevant implications for empirical analysis. Since the convergence time depends only on the intensity of prejudice, the latter can be empirically estimated through laboratory or field experiments by tracking the rate of decay of disagreement. In particular, λ can be estimated by tracking the change $\Delta\mathbf{p}(t)$ in the vector of opinions in the tth iteration. Define $\Delta\mathbf{p}(t) = \parallel \mathbf{p}(t-1) - \mathbf{p}(t) \parallel_1$ as the distance between the vector of opinions at $t-1$ and t. The following corollary provides the relationship between $\Delta\mathbf{p}(t)$ and λ.

Corollary 1. *The distance* $\Delta\mathbf{p}(t)$ *the vector of opinions at* t *and* $t-1$ *is bounded by*

$$\frac{2rn\bar{p}_m\lambda^t(1-\lambda)}{(1-\lambda^{t+1})} \leq \Delta\mathbf{p}(t) \leq \frac{2nS_{\bar{p}}\lambda^t(1-\lambda)}{(1-\lambda^{t+1})} \tag{17}$$

where $0 \leq r \leq \frac{\bar{p}_m}{S_{\bar{p}}}$.

Proof. See Appendix B.8

In Corollary 1, the change in the vector of opinions at t is independent of the network structure. It decays exponentially with time and generally decreases with intensity of prejudice. The change in the vector of opinions can be directly estimated from any laboratory and field experiment that has individual opinions as output. The parameter λ can then be estimated by fitting the function of the form $C(n, \bar{\mathbf{p}})\frac{(1-\lambda)\lambda^t}{(1-\lambda^{t+1})}$ to the evolution of $\Delta\mathbf{p}(t)$. The function $C(n, \bar{\mathbf{p}})$ depend son the population size and the vector of extents of prejudice. The latter can be estimated separately prior to commencement of an experiment; see for example Schuman et al. (1997) and Tejada et al. (2011) regarding empirical techniques for measuring the extent of prejudice. Although the intensity of prejudice will generally differ across agents, the results in Corollary 1 would estimate the average intensities within groups.

The results in Proposition 4 and Corollary 1 can also be used to distinguish between models of learning by averaging. As highlighted in the example of public opinions on climate change, the literature still lacks robust methods for identifying whether the observed disagreement is a result of slow convergence or an equilibrium outcome. For the DeGroot model and its variations that do not incorporate prejudice, the speed of convergence depends on the network structure. For example, Golub and Jackson (2010) and Golub and Jackson (2012) show that the speed of learning depends only on the second largest eigenvalue of the interaction matrix in the DeGroot model, and Jadbabaie et al. (2013) shows that the speed of learning in a modified DeGroot model, according to which agents linearly combine their personal experiences with the opinions of their neighbours, depends on agents' eigenvector centralities. In the presence of prejudice, we showed above that the speed of learning depends only on the intensity of prejudice.

7. Concluding remarks

The question of how peoples' attitudes and behaviours evolve is at the centre of social and behavioural sciences. The models of learning by averaging others' opinions or behaviours, commonly known as naïve learning, have been instrumental in explaining the processes of opinion formation and behavioural change. Empirical studies also suggest that people indeed tend to follow naïve rules of learning. The prediction of the canonical models of naïve learning is that the society will converge to a consensus in the long-run. This prediction is however not consistent with the observed persistence of disagreement in the society across factual issues. To reconcile the discrepancies between theoretical predictions and empirical observation, some papers in the literature have come up with modifications of the canonical models of naive learning that generate disagreement in equilibrium.

The objective of this paper has been to provide a behavioural foundation for prejudices as individual attributes that naturally lead to persistent disagreement in the society. We have shown that indeed, persistent prejudices lead to long-lasting disagreement. Although the mere presence of prejudice is sufficient to generate long-run disagreement, what drives the heterogeneity in equilibrium opinions is not the heterogeneity of the actual prejudices per se, but rather the heterogeneity in network centralities of agents. Network centralities measure the level of influence agents exert on each other's equilibrium opinions.

Our analysis subsequently quantifies the extent of disagreement in the society and characterizes its evolution. We show that the intensity of prejudice and of group cohesion interactively drive the extent of disagreement. We also demonstrate how the rate of decay of disagreement can be used as a mechanism for distinguishing different models of naïve learning, and as a method for estimating the parameter of intensity of prejudice. One aspect that is largely missing in the study of opinion and behaviour formation is empirical studies that attempt to establish which among the existing models best describes reality. Our analysis on convergence rates of learning provides a potentially suitable method for distinguishing between existing models through laboratory and field experiments.

Appendix A. Relationship between the second largest eigenvalue and group cohesion for the case of two subgroups

The general relationship between intensities of group cohesion and $\mu_2(W)$ for the case of two cohesive groups can be established by considering the relationship between $\mu_2(W)$ and the *conductance* $\Psi(W)$ of W, defined as follows.

Let $S \subset N$ and define $C(S) = \sum_{i\in S} \pi_i$, the *capacity* of S. Define also $F(S) = \sum\limits_{i\in S}\sum\limits_{j\in N\backslash S} \pi_i w_{ij}$, the *ergodic flow* out of S. These definitions imply that $0 \leq F(S) \leq C(S) \leq 1$. Define the quantity $\Psi(W,S) = F(S)/C(S)$. For both S and $N\backslash S$, the quantities $1-\Psi(W,S)$ and $1-\Psi(W,N\backslash S)$ are the respective measures of intensities of cohesion. The conductance of W is defined as

$$\Psi(W) = \min_{\substack{0<|S|<n \\ C(S)\leq\frac{1}{2}}} \Psi(W,S)$$

so that $1-\Psi(W)$ measures the maximum intensity of group cohesion among the two subgroups. Generally, however, the conductance measures the minimum relative connection strength between subgroups S and $N\backslash S$. The following lemma shows that $\mu_2(W)$ is directly related to $\Psi(W)$.

Lemma 2. *(Diaconis and Stroock, 1991, Proposition 6) The second eigenvalue $\mu_2(W)$ of W is bounded by*

$$1-2\Psi(W) \leq \mu_2(W) \leq 1-\frac{\Psi(W)^2}{2} \tag{A.1}$$

Example 2: Consider an interaction matrix W derived as follows. Let G be an adjacency matrix of a network $G(N,E)$, where E is the set of directed links connecting agents. The adjacency matrix has elements $g_{ij} = 1$ if i pays attention to j, and zero otherwise. Let $d_i = \sum_{j\in N} g_{ij}$ be the degree of i, that is the number of agents that i pays attention to. Write I_D for the diagonal matrix with d_i as its elements, and define the interaction matrix $W = I_D^{-1}G$. Let $\mathcal{E} =\mid E \mid$ be the cardinality of E. Assume that the network is undirected so that existence of relation $i \to j$ implies existence of $j \to i$; we then simply write (i,j) to mean an undirected link exists between i and j.

Claim 2. *The stationary distribution of W is then $\pi = (\frac{d_1}{2\mathcal{E}}, \frac{d_2}{2\mathcal{E}}, \cdots, \frac{d_n}{2\mathcal{E}})^T$.*

Claim 2 follows from the following argument. Let $[\pi^T W]_i$ denote the product of π^T with the ith column of W. Then

$$\pi_i = [\pi^T W]_i = \sum_{j \in N} \pi_j w_{ij} = \sum_{j:(i,j)\in E} \frac{d_j}{2\mathcal{E}} \frac{1}{d_j} = \frac{1}{2\mathcal{E}} \sum_{j:(i,j)\in E} 1 = \frac{d_i}{2\mathcal{E}} = \pi_i$$

Let S be the subset that minimizes $\Psi(W,S)$ and write $\mathcal{E}_{SS'}$ for the number of links between subset S and its complement $S' = N \backslash S$. Then $\Psi(W) = F(S)/C(S) = \mathcal{E}_{SS'}/\sum_{i\in S} d_i$. The quantity $\mathcal{E}_{SS'}/\sum_{i\in S} d_i$ is the likelihood that an agent $i \in S$ is connected to another agent j outside of S, so that $1 - \Psi(W)$ measures the intensity of cohesion within S.

Appendix B. Proofs

Appendix B.1. Proof of Claim 1

To prove the claim, we show that the matrix $B^{[\lambda]}(W)$ measures the normalized expected number of visits to each state for a random walk process on the matrix λW. That is, for a random walk process on λW, each element $b_{ij}^{[\lambda]}(W)$ of $B^{[\lambda]}(W)$ is the expected normalized number of visits to j in the long-run starting from i. Each visit to j starting from i implies that j's opinion is being incorporated into i's opinions.

Let Y_t be the random walk process on λW, and let $I_t = 1$ if the process is in j at period t and zero otherwise, so that the number of visits to j from i in t transitions is $\sum_\tau^t I_\tau$. Let $P(E)$ be the probability of event E. The expected number of visits to j starting from i after t iterations is then

$$\begin{aligned}\mathbb{E}\left[\sum_{\tau=0}^{t} I_\tau \mid Y_0 = i\right] &= \sum_{\tau=0}^{t} \mathbb{E}\big[I_\tau \mid Y_0 = i\big] = \sum_{\tau=0}^{t} \Big[1 \times P(I_\tau \mid Y_0 = i) + 0 \times (1 - P(I_\tau \mid Y_0 = i))\Big] \\ &= \sum_{\tau=0}^{t} P(I_\tau \mid Y_0 = i) = \sum_{\tau=0}^{t} (\lambda w_{ij})^\tau\end{aligned}$$

In the long-run, $\mathbb{E}\left[\sum_{\tau=0}^{\infty} I_\tau \mid Y_0 = i\right] = \sum_{\tau=0}^{\infty} (\lambda w_{ij})^\tau$. Normalizing with a factor of $\sum_{\tau=0}^{\infty} \lambda = \frac{1}{1-\lambda}$ then implies that $b_{ij}^{[\lambda]}(W) = (1-\lambda)\sum_{\tau=0}^{\infty}(\lambda w_{ij})^\tau$ is the expected normalized number of visits to j in the long-run starting from i.

The elements of $B^{[\lambda]}(W)$ are finite since the summation on the right hand side of (5) is convergent. That is, $\sum_{\tau=0}^{+\infty}(\lambda W)^\tau = (I - \lambda W)^{-1}$, which is well-defined whenever $\lambda \in (0,1)$. See Appendix B.2 in the proof of Lemma 1 for details.

Appendix B.2. Proof of Lemma 1

The dynamic system in (4) is convergent whenever the largest eigenvalue $\mu_1(\Lambda W)$ of ΛW is less than one. This condition is observable once the matrix ΛW is expressed in terms of eigenvalue decomposition. Let $\mu_1(\Lambda W) \geq \mu_2(\Lambda W) \geq \cdots \geq \mu_n(\Lambda W)$ be the eigenvalues of ΛW, and let $\mathbf{v}_j$ and $\mathbf{r}_j$ be the respective left and right eigenvectors of eigenvalue $\mu_j(\Lambda W)$. Then $\mathbf{p}^*$ can be re-expressed as

$$\mathbf{p}^* = (I - \Lambda W)^{-1}\, \bar{\mathbf{p}}^{[\Lambda]} = \left[\sum_{\tau=0}^{\infty} (\Lambda W)^\tau\right] \bar{\mathbf{p}}^{[\Lambda]} = \left[\sum_{\tau=0}^{\infty}\sum_{j=1}^{n} \mu_j^\tau(\Lambda W)\mathbf{r}_j\mathbf{v}_j^T\right] \bar{\mathbf{p}}^{[\Lambda]} \tag{B.1}$$

The expression on the right hand side of the second equality of (B.1) converges if and only if $\mu_1(\Lambda W) < 1$. Wang and Xi (1997, Lemma 2) show that for any tow Hermitian matrices A and B, $\mu_1(AB) \leq \mu_1(A)\mu_1(B)$. Applying this result to $\mu_1(\Lambda W)$ then implies that $\mu_1(\Lambda W) \leq \mu_1(\Lambda)\mu_1(W)$. Since W is a stochastic matrix, $\mu_1(W) = 1$, then $\mu_1(\Lambda W) \leq \mu_1(\Lambda)$. Since Λ is a diagonal matrix, $\mu_1(\Lambda) = \lambda_{\max}$, where $\lambda_{\max} = \max_{i\in N} \lambda_i$. If $\lambda_i \in (0,1)$ for all $i \in N$, then $\mu_1(\Lambda W) < 1$. This completes the proof.

Appendix B.3. Proof of Proposition 1

From the proof of Lemma 1 in Appendix Appendix B.2, equilibrium opinions are given by

$$\mathbf{p}^* = (I - \Lambda W)^{-1}\, \bar{\mathbf{p}}^{[\Lambda]} = \left[\sum_{\tau=0}^{\infty} (\Lambda W)^\tau\right] \bar{\mathbf{p}}^{[\Lambda]} \tag{B.2}$$

Substituting for $\lambda_i = \lambda$ for all $i \in N$ yields

$$\mathbf{p}^* = (1-\lambda)\left[\sum_{\tau=0}^{\infty} (\lambda W)^\tau\right] \bar{\mathbf{p}} = B^{[\lambda]}(W)\bar{\mathbf{p}} = \mathbf{b}^{[\lambda,out]}(W;\bar{\mathbf{p}}) \tag{B.3}$$

The average opinion is given

$$\text{Ave}[\mathbf{p}^*] = \frac{1}{n}\sum_{i=1}^{n} \bar{p}_i^* = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} b_{ij}^{[\lambda]}(W)\bar{p}_j\right) = \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{i=1}^{n} b_{ij}^{[\lambda]}(W)\right)\bar{p}_j = \frac{1}{n}\sum_{j=1}^{n} b_j^{[\lambda,in]}(W)\bar{p}_j$$

Appendix B.4. Proof of Proposition 2

Given the long-run vector of opinions $\mathbf{p}^*$, let $M^{[\lambda]}(\infty) = B^{[\lambda]}(W)$ so that $\mathbf{p}^* = M^{[\lambda]}(\infty)\bar{\mathbf{p}}$ that is

$$M^{[\lambda]}(\infty) = \left(1-\lambda\right)\sum_{\tau=0}^{\infty} (\lambda W)^\tau$$

We use $M^{[\lambda]}(\infty)$ instead of $B^{[\lambda]}(W)$ throughout the following sections for notational convenience.

We then proceed by first rewriting $\mathbf{e}\pi^T$ to have the same structure as $\mathbf{M}^{[\lambda]}(t)$. Let $\Pi = \mathbf{e}\pi^T$, and note that since Π is derived by infinitely iterating W^t, then $\Pi^t = \Pi$ for any $t = 1, 2, \cdots$. The following relation holds for any $\tau \geq 0$

$$\Pi = (1-\lambda)\sum_{\tau=0}^{\infty}(\lambda\Pi)^{\tau} = (1-\lambda)\Pi\sum_{\tau=0}^{\infty}(\lambda)^{\tau} = \Pi \tag{B.4}$$

where the last equality if because $\sum_{\tau=0}^{\infty}\lambda^{\tau} = \frac{1}{1-\lambda}$. The magnitude of disagreement can thus be rewritten as

$$\begin{aligned} DP(W;\bar{\mathbf{p}}) &= \left\|\left(M^{[\lambda]}(\infty) - \Pi\right)\bar{\mathbf{p}}\right\|_1 \\ &= \left\|\left[\left(1-\lambda\right)\sum_{\tau=0}^{\infty}(\lambda W)^{\tau} - (1-\lambda)\sum_{\tau=0}^{\infty}(\lambda\Pi)^{\tau}\right]\bar{\mathbf{p}}\right\|_1 \\ &= \left(1-\lambda\right)\left\|\sum_{\tau=0}^{\infty}\lambda^{\tau}\left(W^{\tau}\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\right)\right\|_1 \end{aligned} \tag{B.5}$$

We use the following lemma to complete the proof.

Lemma 3. *Let the underlying network of interactions be connected. Then* $\|W^{\tau}\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\|_1$ *is bounded by*

$$nr'\bar{p}\mu_*^t(W) \leq \|W^{\tau}\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\|_1 \leq nS_{\bar{p}}\frac{\mu_*^t(W)}{\pi_{\min}} \tag{B.6}$$

where $\pi_{\min} = \min_{i\in N}\pi_i$, $\bar{p} = \min_{i\in N}\bar{p}_i$, $S_{\bar{p}} = \sum_{i=1}^{n}\bar{p}_i$, *and* $0 \leq r' \leq 1$.

Proof. See Appendix B.5

By making use of the triangular inequality and substituting the results of Lemma 3 into (B.5), we have

$$\begin{aligned} DP(W;\bar{\mathbf{p}}) &\leq \left(1-\lambda\right)\sum_{\tau=0}^{\infty}\lambda^{\tau}\,\|W^{\tau}\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\|_1 \\ &\leq \left(1-\lambda\right)\frac{nS_{\bar{p}}}{\pi_{\min}}\sum_{\tau=0}^{\infty}\left(\lambda\mu_*(W)\right)^{\tau} \\ &= \frac{nS_{\bar{p}}}{\pi_{\min}}\left(\frac{1-\lambda}{1-\lambda\mu_*(W)}\right) \end{aligned} \tag{B.7}$$

where the last equality if because $\sum_{\tau=0}^{\infty}\left(\lambda\mu_*(W)\right)^{\tau} = \frac{1}{1-\lambda\mu_*(W)}$.

For the lower bound, we use the following results on reverse triangular inequality from Diaz and Metcalf (1966, Theorem 1). Let $\mathbf{u}$ be a unit vector in the Hilbert Space. Suppose the vectors $\mathbf{x}(1), \cdots, \mathbf{x}(n)$, whenever $\mathbf{x}(\tau) \neq 0$ satisfies $0 \leq r \leq \frac{Re\langle\mathbf{x}(\tau),\mathbf{u}\rangle}{\|\mathbf{x}(\tau)\|}$, where Re stands for the real number and $\langle\mathbf{x},\mathbf{y}\rangle = \sum_{i=1}^{n}x_i y_i$ is the inner product of $\mathbf{x}$ and $\mathbf{y}$. Then

$$\left\|\sum_{\tau=1}^{n}\mathbf{x}(\tau)\right\| \geq r\sum_{\tau=1}^{n}\|\mathbf{x}(\tau)\| \tag{B.8}$$

Applying inequality (B.8) to (B.5) yields

$$\begin{aligned}
DP(W;\bar{\mathbf{p}}) &\geq r\Big(1-\lambda\Big)\sum_{\tau=0}^{\infty}\lambda^{\tau}\left\|W^{\tau}\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}}\right\|_1 \\
&\geq nr'\bar{p}r\Big(1-\lambda\Big)\sum_{\tau=0}^{\infty}\left(\lambda\mu_*(W)\right)^{\tau} \\
&= nr'\bar{p}r\left(\frac{1-\lambda}{1-\lambda\mu_*(W)}\right)
\end{aligned} \tag{B.9}$$

We now derive bounds for r using the definitions above. The summations on the right hand side of (B.5) consists of the vectors $\bar{\mathbf{x}}(\tau)=\bar{\mathbf{p}}(\tau)-\Pi\bar{\mathbf{p}}$ for $\tau=0,1,2,\cdots$, where $\bar{\mathbf{p}}(\tau)=W^{\tau}\bar{\mathbf{p}}$. To derive an upper bound for $\frac{Re\langle\bar{\mathbf{x}}(\tau),\mathbf{u}\rangle}{\|\bar{\mathbf{x}}(\tau)\|}$, we choose a τ so that the corresponding $\bar{\mathbf{x}}(\tau)$ maximizes the former. Since $\bar{\mathbf{x}}(\tau)$ decays uniformly to zero, its largest value occurs when $\tau=0$; denoting it by $\mathbf{v}(0)$, we have

$$\mathbf{v}(0)=(1-\lambda)\lambda^0\left[W^0\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}}\right]=(1-\lambda)(\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}})=(1-\lambda)\bar{\mathbf{x}}(0)$$

Let $\pi_{\bar{p}}=\pi^T\bar{\mathbf{p}}=\sum_{i=1}^{n}\pi_i\bar{p}_i$. Hence $\mathbf{v}(0)=(1-\lambda)(\bar{p}_1-\pi_{\bar{p}},\cdots,\bar{p}_i-\pi_{\bar{p}},\cdots,\bar{p}_n-\pi_{\bar{p}})$. We can choose $\mathbf{u}=e_i$ so that $i\in\underset{j\in N}{\operatorname{argmax}}\ v_j(0)$. Let $v_{\max}(0)=\max_{i\in N}v_i(0)$. Then

$$\|\mathbf{v}(0)\|=\sum_{i\in N}\mid\bar{p}_i-\pi_{\bar{p}}\mid\geq v_{\max}(0) \tag{B.10}$$

$$\frac{Re\langle\mathbf{v}(0),\mathbf{u}\rangle}{\|\mathbf{v}(0)\|}\leq(1-\lambda)\frac{v_{\max}(0)}{v_{\max}(0)}=1$$

Hence, for $0\leq r\leq(1-\lambda)$, and $0\leq r'r\leq(1-\lambda)$. Substituting into B.9, then for some real number $0\leq r\leq(1-\lambda)$

$$DP(W;\bar{\mathbf{p}})\geq n\bar{p}r\left(\frac{1-\lambda}{1-\lambda\mu_*(W)}\right)$$

Appendix B.5. Proof of Lemma 3

The proof uses results from the literature of Markov chains. In particular, we use the following steps for the proof of bounds for mixing times in Levin et al. (2009), where the concept of total variation distance is used. The total variation distance of any two distributions ν and π on a state space X is defined as $\|\nu-\pi\|_{\text{var}}=\frac{1}{2}\sum_{x\in X}\mid\nu(x)-\pi(x)\mid$. This definition implies that $\left\|W^t\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}}\right\|_1=2\left\|W^t\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}}\right\|_{\text{var}}$. The total variation distance can be expanded as follows

$$\begin{aligned}
\left\|W^t\bar{\mathbf{p}}-\Pi\bar{\mathbf{p}}\right\|_{\text{var}} &= \frac{1}{2}\sum_{i\in N}\left|\sum_{j\in N}\left(w_{ij}^t-\pi_j\right)\bar{p}_j\right| \\
&\leq\frac{1}{2}\sum_{i\in N}\left(\sum_{j\in N}\pi_j\bar{p}_j\left|\frac{w_{ij}^t}{\pi_j}-1\right|\right)
\end{aligned} \tag{B.11}$$

Let the network associated to W be connected and that W is aperiodic so that π is well-defined and unique. Under these conditions, W satisfies the conditions in Levin et al. (2009, Lemma 12.1). The proof then follows from Levin et al. (2009, Lemma 12.2) as follows. Let $\{f_i\}_{i=1}^n$ be the real valued eigenfunctions of W corresponding to real eigenvalues $\{\mu_i\}$. In the proof of Levin et al. (2009, Lemma 12.3), it is shown that

$$\left|\frac{w_{ij}^t}{\pi_j} - 1\right| \leq \frac{[\mu_*(W)]^t}{\pi_{\min}}$$

Let $S_{\bar{p}} = \sum_{j\in N} \bar{p}_j$, and noting that $\sum_{j\in N} \pi_j \bar{p}_j \leq \sum_{j\in N} \bar{p}_j$, equation (B.11) then becomes

$$\left\|W^t\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\right\|_{\text{var}} \leq \frac{1}{2}\frac{\mu_*^t(W)}{\pi_{\min}} \sum_{i\in N}\sum_{j\in N} \pi_j \bar{p}_j \leq \frac{n}{2}\frac{\mu_*^t(W)}{\pi_{\min}} S_{\bar{p}}$$

Multiplying by a factor of two yields the desired upper bound.

We again use the reverse triangular inequality to establish the lower bound for $\|W^t\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\|_{\text{var}}$. The term $\sum_{j\in N}\left(w_{ij}^t - \pi_j\right)\bar{p}_j$ on the right hand side of (B.11) is a summation of the vector

$$\mathbf{v}_i(t) = \left(\left(w_{i1}^t - \pi_1\right)\bar{p}_1, \cdots, \left(w_{i1}^t - \pi_n\right)\bar{p}_n\right) \quad \text{for } t \geq 0$$

Since $w_{ij}^t \to \pi_1$ as $t \to \infty$ for all j, $\mathbf{v}_i(t)$ with the largest entries and hence magnitude is $\mathbf{v}_i(0)$ given by

$$\mathbf{v}_i(0) = \left(\left(1 - \pi_1\right)\bar{p}_1, \cdots, \left(1 - \pi_n\right)\bar{p}_n\right)$$

We can choose $\mathbf{u} = e_i$ so that $i \in \underset{j\in N}{\operatorname{argmax}}\ v_j(0)$. Let $v_{\max}(0) = \max_{i\in N} v_i(0)$. Then

$$\|\mathbf{v}(0)\| = \sum_{i\in N} \mid (1-\pi_i)\,\bar{p}_i \mid = \sum_{i\in N} (1-\pi_i)\,\bar{p}_i \geq v_{\max}(0) \tag{B.12}$$

The second equality in (B.12) is because $\pi_i \leq 1$ for all $i \in N$. We then have

$$\frac{Re\langle\mathbf{v}(0), \mathbf{u}\rangle}{\|\mathbf{v}(0)\|} \leq \frac{v_{\max}(0)}{v_{\max}(0)} = 1$$

Hence for $0 < r \leq 1$, the right hand side of (B.11) can be re-written using the reverse triangular inequality as follows

$$\left\|W^t\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\right\|_{\text{var}} \geq \frac{1}{2} r \sum_{i\in N}\left(\sum_{j\in N} \pi_j \bar{p}_j \left|\frac{w_{ij}^t}{\pi_j} - 1\right|\right) \geq \frac{1}{2} r\bar{p} \sum_{i\in N}\sum_{j\in N}\left|w_{ij}^t - \pi_j\right| \tag{B.13}$$

where $\bar{p} = \min p_i$. Let w_i^t be the ith-row of W^t. Then $\|w_i^t - \pi\|_{\text{var}} = \frac{1}{2}\sum_{j\in N}\left|w_{ij}^t - \pi_j\right|$ is the total variation distance between w_i^t and π. Levin et al. (2009, Page 156, Equation (12.13)) show that $\|w_i^t - \pi\|_{\text{var}} \geq \frac{1}{2}\mu_*^t(W)$, and hence $\sum_{j\in N}\left|w_{ij}^t - \pi_j\right| \geq \mu_*^t(W)$. Substituting into (B.13) yields,

$$\left\|W^t\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\right\|_{\text{var}} \geq \frac{n}{2} r\bar{p}\mu_*^t(W) \tag{B.14}$$

And hence $\|W^t\bar{\mathbf{p}} - \Pi\bar{\mathbf{p}}\|_1 \geq nr\bar{p}\mu_*^t(W)$.

Appendix B.6. Proof of Proposition 3

The proof proceeds by proving the first statement of the proposition:

1) *The second largest eigenvalue $\mu_2(\tilde{W})$ of $\tilde{W}$ measures the intensities of cohesion within subgroups of $\mathcal{L}$. If W is strongly lumpable with respect to a partition $\mathcal{L}$ to $\tilde{W}$, then $\mu_2(\tilde{W}) = \mu_2(W)$.*

Let $\mu_1(\tilde{W}) \geq \mu_2(\tilde{W}) \geq, \cdots \geq \mu_k(\tilde{W})$ be the eigenvalues of $\tilde{W}$. We first show that $\tilde{\mu}_2$ measures the intensities of cohesion within subgroup $\mathcal{L} = \{L_1, L_2, \cdots, L_k\}$. By definition, $\tilde{W}_{jj}$ for $j = 1, \cdots, k$, represents the intensity of interactions among agents of subgroup L_j. The higher the value of $\tilde{W}_{jj}$, the more cohesive subgroup L_j. We begin by showing that increasing the values of $\tilde{W}_{jj}$ directly increases $\mu_2(\tilde{W})$, which in turn implies that the second eigenvalue of $\tilde{W}$ measures the intensities of cohesion within subgroups.

Let D_α be a diagonal matrix with entries $\alpha_j \in (0,1)$ for $j = 1, \cdots, k$. Define a matrix $\tilde{W}_\alpha$ derived from $\tilde{W}$ as follows.

$$\tilde{W}_\alpha = (I - D_\alpha) + D_\alpha \tilde{W} \tag{B.15}$$

where I is an identity matrix. Just as $\tilde{W}$, $\tilde{W}_\alpha$ is also row-stochastic, which implies that $\mu_1(\tilde{W}) = \mu_1(\tilde{W}_\alpha) = 1$. Since D_α is a diagonal matrix, reducing α_j increases the intensity of interaction among agents of subgroup L_j and hence its intensity of cohesion. We are interested in establishing the relationship between $\mu_2(\tilde{W})$ and $\mu_2(\tilde{W}_\alpha)$.

Let A and B be $n \times n$ non-negative matrices with eigenvalues $a_1 \geq \cdots \geq a_n$ and $b_1 \geq \cdots \geq b_n$ respectively. Then the eigenvalues $c_1 \geq \cdots \geq c_n$ of $C = A + B$ have the following upper bounds (Bhatia, 2001).

$$c_{i+j-1} \leq a_i + b_j \quad \text{whenever } 0 < i, j, i + j < n \tag{B.16}$$

Letting $i = 1$ and $j = 2$, it follows from (B.15) that

$$\mu_2(\tilde{W}_\alpha) \leq \mu_1 (I - D_\alpha) + \mu_2(D_\alpha \tilde{W}) = (1 - \alpha_k) + \mu_2(D_\alpha \tilde{W}) \tag{B.17}$$

where the second equality follows from the fact that the eigenvalues of $(I - D_\alpha)$ are its diagonal elements, so that the largest eigenvalue is $1 - \alpha_k$. We use the following lemma to place an upper bound on $\mu_2(D_\alpha \tilde{W})$.

Lemma 4. *(Wang and Xi, 1997, Lemma 2) Let $G, H \in \mathbb{C}^{n \times n}$ be positive definite Hermitian, and let $1 \leq i_1, \cdots, i_l \leq n$. Then*

$$\prod_{\tau=1}^{l} \mu_{i_\tau}(GH) \leq \prod_{\tau=1}^{l} \mu_{i_\tau}(G) \mu_{i_\tau}(H) \tag{B.18}$$

From (B.18),

$$\mu_1(D_\alpha\tilde{W})\mu_2(D_\alpha\tilde{W}) \leq \mu_1(D_\alpha)\mu_2(D_\alpha)\mu_1(\tilde{W})\mu_2(\tilde{W})$$

Recall that $\tilde{W}$ is row-stochastic, so that $\mu_1(\tilde{W}) = 1$ and $\mu_1(D_\alpha\tilde{W}) = \alpha_1$. Since $\mu_1(D_\alpha) = \alpha_1$ and $\mu_2(D_\alpha) = \alpha_2$, it follows that $\mu_2(D_\alpha\tilde{W}) \leq \alpha_2\mu_2(\tilde{W})$, and hence (B.17) becomes

$$\mu_2(\tilde{W}_\alpha) \leq (1-\alpha_k) + \alpha_2\mu_2(\tilde{W}) = 1 - \alpha_2\left(\frac{\alpha_k}{\alpha_2} - \mu_2(\tilde{W})\right) \tag{B.19}$$

It then follows from (B.19) that provided the elements of D_α are sufficiently close in magnitude such that $\frac{\alpha_k}{\alpha_2} > \mu_2(\tilde{W})$, then increasing the intensities of within group interactions by decreasing α_is, increases $\mu_2(\tilde{W}_\alpha)$. Hence $\mu_2(\tilde{W}_\alpha)$ measures the intensities of cohesion within subgroups $\mathcal{L}$.

The relationship between the eigenvalues of W and $\tilde{W}$ follows from Horn and Johnson (1990, Theorem 1.3.20); that is, Suppose that $A \in \mathbb{M}^{m,n}$ and $B \in \mathbb{M}^{n,m}$, with $m \leq n$. Then the n eigenvalues of BA are the m eigenvalues of AB together with $n-m$ zeroes. This result directly implies that the n eigenvalues of $W = DK$ are the k eigenvalues of $\tilde{W} = KD$ together with $n-k$ zeroes; and hence $\mu_2(W) = \mu_2(\tilde{W})$.

We now prove the second statement of the proposition:

2) *Let W be nearly lumpable with respect to a partition $\mathcal{L}$ to $\tilde{W}$ so that $W = P + \varepsilon R$, where P is strongly lumpable and R is some arbitrary matrix. Let $R_{\max} = \max_{i\in N}\sum_{j=1}^{n} r_{ij}$ be the maximum row-sum of R. Then*

$$\mid \mu_2(W) - \mu_2(\tilde{W}) \mid \leq \varepsilon R_{\max}\rho(R) \tag{B.20}$$

where $\rho(R) = 1 + \frac{1}{|\mu_n(R)|}$ and $\mid \mu_n(R) \mid$ is the smallest eigenvalue of R.

The definition of near lumpability in (12) implies that W can be rewritten as $W = P + \varepsilon R$, where P is strongly lumpable and R is some arbitrary transition matrix. Applying the collector matrix V and its corresponding normalization U, then

$$\tilde{W} = UWV = UPV + \varepsilon URV = \tilde{P} + \varepsilon URV \tag{B.21}$$

We apply the following inequalities on eigenvalues to establish the relationship between $\mu_2(W)$ and $\mu_2(\tilde{W})$. Let A and B be $n \times n$ non-negative matrices with eigenvalues $a_1 \geq \cdots \geq a_n$ and $b_1 \geq \cdots \geq b_n$ respectively. Then the eigenvalues $c_1 \geq \cdots \geq c_n$ of $C = A + B$ have the following bounds (Bhatia, 2001).

$$c_{i+j-1} \leq a_i + b_j \quad \text{whenever } 0 < i, j, i+j < n, \text{ and} \quad c_i \geq a_i + b_n \quad \text{for } 0 \leq i \leq n. \tag{B.22}$$

Letting $i = 1$ and $j = 2$, it follows from the first inequality of (B.22) that

$$\mu_2(W) \leq \mu_2(P) + \varepsilon\mu_1(R) \tag{B.23}$$

and from the second inequality

$$\mu_2(\tilde{W}) \geq \mu_2(\tilde{P}) + \varepsilon\mu_k(URV) \tag{B.24}$$

Since $\mu_2(P) = \mu_2(\tilde{P})$ by virtue of lumpability of P, it follows that

$$\mid \mu_2(W) - \mu_2(\tilde{W}) \mid \leq \mid \varepsilon\mu_1(R) - \varepsilon\mu_k(URV) \mid \leq \varepsilon \mid \mu_1(R) \mid + \varepsilon \mid \mu_k(URV) \mid \tag{B.25}$$

To establish the relationship between $\mu_1(R)$ and $\mu_k(URV)$, we first examine the eigenvalues of VUR, and then later apply the relation $\mu_k(URV) = \mu_k(VUR)$. This relation follows from Horn and Johnson (1990, Theorem 1.3.20); that is, Suppose that $A \in \mathbb{M}^{m,n}$ and $B \in \mathbb{M}^{n,m}$, with $m \leq n$. Then the n eigenvalues of BA are the m eigenvalues of AB together with $n - m$ zeroes. We then use the following lemma for bounding $\mu_k(VUR)$.

Lemma 5. *(Wang and Xi, 1997, Lemma 2) Let $G, H \in \mathbb{C}^{n\times n}$ be positive definite Hermitian, and let $1 \leq i_1, \cdots, i_l \leq n$. Then*

$$\prod_{\tau=1}^{l} \mu_{i_\tau}(GH) \leq \prod_{\tau=1}^{l} \mu_{i_\tau}(G)\mu_{i_\tau}(H) \tag{B.26}$$

and

$$\prod_{\tau=1}^{l} \mu_{i_\tau}(GH) \geq \prod_{\tau=1}^{l} \mu_{i_\tau}(G)\mu_{n-\tau+1}(H) \tag{B.27}$$

The inequality (B.26) implies that

$$\mu_1(VUR).\mu_2(VUR).\cdots.\mu_k(VUR) \leq \Big[\mu_1(VU).\mu_2(VU).\cdots.\mu_k(VU)\Big]\Big[\mu_1(R).\mu_2(R).\cdots.\mu_k(R)\Big] \tag{B.28}$$

The matrix VU is of size n and consists of k diagonal block matrices; it thus consists of the first k eigenvalues equal to one. That is $\mu_1(VU) = \mu_2(VU) = \cdots = \mu_k(VU) = 1$, so that

$$\mu_1(VUR).\mu_2(VUR).\cdots.\mu_k(VUR) \leq \mu_1(R).\mu_2(R).\cdots.\mu_k(R) \tag{B.29}$$

and that

$$\mu_k(VUR) \leq \frac{\mu_1(R).\mu_2(R).\cdots.\mu_k(R)}{\mu_1(VUR).\mu_2(VUR).\cdots.\mu_{k-1}(VUR)} \tag{B.30}$$

Using (B.27), $\mu_{k-1}(VUR)$ is bounded from below by

$$\begin{aligned}
&\mu_1(VUR).\mu_2(VUR).\cdots.\mu_{k-1}(VUR) \\
&\qquad \geq \Big[\mu_1(VU).\mu_2(VU).\cdots.\mu_{k-1}(VU)\Big]\Big[\mu_{n-k+2}(R).\mu_{n-k+3}(R).\cdots.\mu_n(R)\Big] \\
&\qquad = \mu_{n-k+2}(R).\mu_{n-k+3}(R).\cdots.\mu_n(R)
\end{aligned}$$

so that

$$\mu_{k-1}(VUR) \geq \frac{\mu_{n-k+2}(R).\mu_{n-k+3}(R).\cdots.\mu_n(R)}{\mu_1(VUR).\mu_2(VUR).\cdots.\mu_{k-2}(VUR)}$$

Substituting into (B.30) yields

$$\mu_k(VUR) \leq \frac{\mu_1(R).\mu_2(R).\cdots.\mu_k(R)}{\mu_{n-k+2}(R).\mu_{n-k+3}(R).\cdots.\mu_n(R)} \leq \frac{k\mu_1(R)}{(k-1)\mu_n(R)} \leq \frac{\mu_1(R)}{\mu_n(R)} \tag{B.31}$$

And hence

$$\mid \mu_2(W) - \mu_2(\tilde{W}) \mid \leq \varepsilon \mid \mu_1(R) \mid \left(1 + \frac{1}{\mid \mu_n(R) \mid}\right) \tag{B.32}$$

Since R is non-negative, it follows from Gershgorin circle theorem that

$$\mid \mu_1(R) \mid \leq R_{\max} = \max_{i \in N} \sum_{j=1}^{n} r_{ij}$$

Appendix B.7. Proof of Proposition 4

We start by showing that $\mathbf{p}(t) = M^{[\lambda]}(t)\bar{\mathbf{p}}$, where

$$M^{[\lambda]}(t) = \frac{1-\lambda}{1-\lambda^{t+1}} \left[\sum_{\tau=0}^{t} (\lambda W)^{\tau}\right] \tag{B.33}$$

To show that $\mathbf{p}(t) = M^{[\lambda]}(t)\bar{\mathbf{p}}$, we start by setting $\mathbf{p}(0) = \bar{\mathbf{p}}$ without loss of generality; recall that $\mathbf{p}^*$ is independent of $\mathbf{p}(0)$. Moreover, prejudices, by definition, are preconceived beliefs. So it is correct to assume that when forming an opinion on a new non-familiar topic or event, an agent starts with a preconceived opinion that is influenced by the environment that she lives in.

Now, note that for a pair of agents i and j, $m_{ij}^{[\lambda]}(\infty)$, the element in the ith-row and jth-column of $M^{[\lambda]}(\infty)$, is the expected normalized number of times j's opinion is incorporated in i's opinion. This interpretation is supported by the following claim.

Claim 3. *For a random walk process on λW, $\frac{1-\lambda^{t+1}}{1-\lambda} \sum_{\tau=1}^{t} (\lambda w_{ij})^{\tau}$ is the expected normalized number of visits to j starting from i after t iterations*

Claim 3 follows from the following argument. Let Y_t be the random walk process on λW, and let $I_t = 1$ if the process is in j at period t and zero otherwise, so that the number of visits to j from i in t transitions is $\sum_{\tau}^{t} I_{\tau}$. Let $P(E)$ be the probability of event E. The expected number of visits to j starting from i after t transitions is then

$$\begin{aligned}
\mathbb{E}\left[\sum_{\tau}^{t} I_{\tau} \mid Y_0 = i\right] &= \sum_{\tau}^{t} \mathbb{E}\Big[I_{\tau} \mid Y_0 = i\Big] = \sum_{\tau}^{t} \Big[1 \times P(I_{\tau} \mid Y_0 = i) + 0 \times (1 - P(I_{\tau} \mid Y_0 = i))\Big] \\
&= \sum_{\tau}^{t} P(I_{\tau} \mid Y_0 = i) = \sum_{\tau=1}^{t} (\lambda w_{ij})^{\tau}
\end{aligned}$$

Normalizing by a factor of $\sum_{\tau=0}^{t}\lambda^\tau = \frac{1-\lambda^{t+1}}{1-\lambda}$ then implies that $b_{ij}^{[\lambda]}(W) = \frac{1-\lambda}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}(\lambda w_{ij})^\tau$ is the expected normalized number of visits to j after t iterations starting from i. This value is equivalent to the normalized number of times j's opinion is incorporated into i's opinion after t iterations. The matrix $M^{[\lambda]}(t)$ defined in B.33 thus describes the normalized number of times agents' opinions are incorporated in others' opinions after t iterations. As such, it measures the level of influence each agent commands after t iterations. In the long-run,

$$\lim_{t\to\infty} M^{[\lambda]}(t) = \lim_{t\to\infty}\frac{1-\lambda}{1-\lambda^{t+1}}\left[\sum_{\tau=0}^{t}(\lambda W)^\tau\right] = \Big(1-\lambda\Big)\sum_{\tau=0}^{\infty}(\lambda W)^\tau = M^{[\lambda]}(\infty)$$

where the second equality follows from $\lim_{t\to\infty}\frac{1-\lambda}{1-\lambda^{t+1}} = \Big(1-\lambda\Big)$.

The long-run vector of opinions is then $\mathbf{p}^* = \lim_{t\to\infty}\mathbf{p}(t) = \lim_{t\to\infty} M^{[\lambda]}(t)\bar{\mathbf{p}} = M^{[\lambda]}(\infty)\bar{\mathbf{p}}$. Working backwards, the vector of opinions at t is then $\mathbf{p}(t) = M^{[\lambda]}(t)\bar{\mathbf{p}}$. We proceed by first providing bounds for $DE(t;W)$.

Lemma 6. *The distance $DE(t;W;\bar{\mathbf{p}})$ to the long-run distribution after t iterations is bounded by*

$$2rn\bar{p}\lambda^{t+1} \le DE(t;W;\bar{\mathbf{p}}) \le 2nS_{\bar{p}}\lambda^{t+1} \tag{B.34}$$

where $0 \le r \le \frac{\bar{p}_m}{S_{\bar{p}}}$ and $\bar{p}_m = \max_i \bar{p}_i$.

PROOF. Given the definitions of $M^{[\lambda]}(t)$ and $M^{[\lambda]}(\infty)$, $DE(t;W;\bar{\mathbf{p}})$ is given by

$$\begin{aligned} DE(t;W;\bar{\mathbf{p}}) &= \left\|\Big[M^{[\lambda]}(t) - M^{[\lambda]}(\infty)\Big]\bar{\mathbf{p}}\right\|_1 \\ &= \left\|\left[\frac{1-\lambda}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}(\lambda W)^\tau - (1-\lambda)\sum_{\tau=0}^{\infty}(\lambda W)^\tau\right]\bar{\mathbf{p}}\right\|_1 \\ &= (1-\lambda)\left\|\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\lambda^\tau W^\tau\bar{\mathbf{p}} - \sum_{\tau=t+1}^{\infty}\lambda^\tau W^\tau\bar{\mathbf{p}}\right\|_1 \end{aligned} \tag{B.35}$$

The quantity $\|W^\tau\bar{\mathbf{p}}\|_1$ has the following lower and upper bounds.

$$\|W^\tau\bar{\mathbf{p}}\|_1 = \sum_{i=1}^{n}\left|\sum_{j=1}^{n} w_{ij}\bar{p}_j\right| = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\bar{p}_j \le \sum_{i=1}^{n}\sum_{j=1}^{n}\bar{p}_j = nS_{\bar{p}} \tag{B.36}$$

where the second equality is because w_{ij} and $\bar{p}_j$ are positive for all i and j; the third inequality in (B.36) is because $w_{ij}\bar{p}_j \le \bar{p}_j$ for all j and i. As before, $S_{\bar{p}} = \sum_{j=1}^{n}\bar{p}_j$. For the lower bound,

$$\|W^\tau\bar{\mathbf{p}}\|_1 = \sum_{i=1}^{n}\left|\sum_{j=1}^{n} w_{ij}\bar{p}_j\right| = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\bar{p}_j \ge \sum_{i=1}^{n}\bar{p}\sum_{j=1}^{n} w_{ij} = n\bar{p} \tag{B.37}$$

where the fourth equality is because $\sum_{j=1}^{n} w_{ij} = 1$. As before, $\bar{p} = \min_i \bar{p}_i$. Using the triangular inequality and substituting the upper bound into (B.35) yields

$$\begin{aligned}
DE(t; W; \bar{\mathbf{p}}) &\leq (1-\lambda)\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\|\lambda^{\tau}W^{\tau}\bar{\mathbf{p}}\|_1 + \sum_{\tau=t+1}^{\infty}\|\lambda^{\tau}W^{\tau}\bar{\mathbf{p}}\|_1\right] \\
&\leq (1-\lambda)nS_{\bar{p}}\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\lambda^{\tau} + \sum_{\tau=t+1}^{\infty}\lambda^{\tau}\right] \\
&= (1-\lambda)nS_{\bar{p}}\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}.\frac{1-\lambda^{t+1}}{1-\lambda} + \frac{\lambda^{t+1}}{1-\lambda}\right] \\
&= 2nS_{\bar{p}}\lambda^{t+1} \qquad \text{(B.38)}
\end{aligned}$$

We again use the reverse triangular inequality for the lower bound (see Appendix B.2) and the results in (B.37). That is, given r defined below,

$$\begin{aligned}
DE(t; W; \bar{\mathbf{p}}) &= (1-\lambda)\left\|\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\lambda^{\tau}W^{\tau}\bar{\mathbf{p}} - \sum_{\tau=t+1}^{\infty}\lambda^{\tau}W^{\tau}\bar{\mathbf{p}}\right\|_1 \\
&\geq r(1-\lambda)\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\|\lambda^{\tau}W^{\tau}\bar{\mathbf{p}}\|_1 + \sum_{\tau=t+1}^{\infty}\|\lambda^{\tau}W^{\tau}\bar{\mathbf{p}}\|_1\right] \\
&= r(1-\lambda)n\bar{p}\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}\lambda^{\tau} + \sum_{\tau=t+1}^{\infty}\lambda^{\tau}\right] \\
&= r(1-\lambda)n\bar{p}\left[\frac{\lambda^{t+1}}{1-\lambda^{t+1}}.\frac{1-\lambda^{t+1}}{1-\lambda} + \frac{\lambda^{t+1}}{1-\lambda}\right] \\
&= 2rn\bar{p}\lambda^{t+1} \qquad \text{(B.39)}
\end{aligned}$$

The largest vector in the summation of the second equality of (B.39) is when $\tau = t = 0$; denoting it by $\mathbf{v}(0)$, we have

$$\mathbf{v}(0) = \frac{1-\lambda}{1-\lambda}(\lambda W)^0\bar{\mathbf{p}} - (1-\lambda)(\lambda W)^0\bar{\mathbf{p}} = \bar{\mathbf{p}} - (1-\lambda)\bar{\mathbf{p}} = \lambda\bar{\mathbf{p}}$$

We can choose $\mathbf{u} = e_i$ so that $i \in \underset{j\in N}{\text{argmax}}\ v_j(0)$. Let $\bar{p}_m = \max_i \bar{p}_i$ so that $v_{\max}(0) = \max_{i\in N} v_i(0) = \lambda\bar{p}_m$; then

$$\frac{Re\langle\mathbf{v},\mathbf{u}\rangle}{\|\mathbf{v}\|} \leq \frac{\lambda\bar{p}_m}{\lambda S_{\bar{p}}} = \frac{\bar{p}_m}{S_{\bar{p}}}$$

And hence $0 \leq r \leq \frac{\bar{p}_m}{S_{\bar{p}}}$. We use the following relations to complete the proof: $\ln(\varepsilon) - \ln(2nS_{\bar{p}}) = \ln(\varepsilon/2nS_{\bar{p}})$ and $\ln(\varepsilon) - \ln(2rn\bar{p}) = \ln(\varepsilon/2rn\bar{p})$

For $\lambda \neq 0$, if $t \geq \frac{\ln(\varepsilon/2nS_{\bar{p}})}{\ln\left(\lambda\right)} - 1$, then $DE(t; W; \bar{\mathbf{p}}) \leq \varepsilon$. The upper bound for $CT(\varepsilon; W)$ then follows from Definition 4.

Similarly, if $t \leq \frac{\ln(\varepsilon/2rn\bar{p})}{\ln\left(\lambda\right)} - 1$, then $DE(t; W; \bar{\mathbf{p}}) \geq \varepsilon$. And the lower bound for $CT(\varepsilon; W)$ follows from Definition 4.

Appendix B.8. Proof of Corollary 1

Following from the proofs of Propositions 2 and 4 above, $\Delta\mathbf{p}(t)$ is given by

$$\begin{aligned}
\Delta\mathbf{p}(t) &= \left\|\left[M^{[\lambda]}(t-1) - M^{[\lambda]}(t)\right]\bar{\mathbf{p}}\right\|_1 \\
&= \left\|\left[\frac{1-\lambda}{1-\lambda^t}\sum_{\tau=0}^{t-1}(\lambda W)^\tau - \frac{1-\lambda}{1-\lambda^{t+1}}\sum_{\tau=0}^{t}(\lambda W)^\tau\right]\bar{\mathbf{p}}\right\|_1 \\
&= (1-\lambda)\left\|\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\lambda^\tau W^\tau\bar{\mathbf{p}} - \frac{\lambda^t}{(1-\lambda^{t+1})}W^t\bar{\mathbf{p}}\right\|_1 \\
&\leq (1-\lambda)\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\|\lambda^\tau W^\tau\bar{\mathbf{p}}\|_1 + \frac{\lambda^t}{(1-\lambda^{t+1})}\left\|W^t\bar{\mathbf{p}}\right\|_1\right] \\
&= (1-\lambda)nS_{\bar{p}}\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\lambda^\tau + \frac{\lambda^t}{(1-\lambda^{t+1})}\right] \\
&= (1-\lambda)nS_{\bar{p}}\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}.\frac{1-\lambda^t}{1-\lambda} + \frac{\lambda^t}{(1-\lambda^{t+1})}\right] \\
&= \frac{2nS_{\bar{p}}\lambda^t(1-\lambda)}{(1-\lambda^{t+1})}
\end{aligned} \tag{B.40}$$

We again use the reverse triangular inequality for the lower bound (see Appendix B.2). That is, given r defined below,

$$\begin{aligned}
\Delta\mathbf{p}(t) &= (1-\lambda)\left\|\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\lambda^\tau W^\tau\bar{\mathbf{p}} - \frac{\lambda^t}{(1-\lambda^{t+1})}W^t\bar{\mathbf{p}}\right\|_1 \\
&\geq r(1-\lambda)\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\|\lambda^\tau W^\tau\bar{\mathbf{p}}\|_1 + \frac{\lambda^t}{(1-\lambda^{t+1})}\left\|W^t\bar{\mathbf{p}}\right\|_1\right] \\
&= rn\bar{p}_m(1-\lambda)\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}\sum_{\tau=0}^{t-1}\lambda^\tau + \frac{\lambda^t}{(1-\lambda^{t+1})}\right] \\
&= rn\bar{p}_m(1-\lambda)\left[\frac{\lambda^t(1-\lambda)}{(1-\lambda^{t+1})(1-\lambda^t)}.\frac{1-\lambda^t}{1-\lambda} + \frac{\lambda^t}{(1-\lambda^{t+1})}\right] \\
&= \frac{2rn\bar{p}_m\lambda^t(1-\lambda)}{(1-\lambda^{t+1})}
\end{aligned} \tag{B.41}$$

The steps for bounding r are identical to those in Appendix B.7 above; hence, $0 \leq r \leq \frac{\bar{p}_m}{S_{\bar{p}}}$.

Acknowledgements

I gratefully acknowledge the URC Postdoctoral Research Fellowship funding from the University of Cape Town. This paper has benefited from comments by Suraj Shekhar, Co-Pierre Georg, Stefania Innocenti and Michael Rose. I also acknowledge comments from the seminar participants at the School of Economics and AIFMRM – University of Cape Town. For all the help and advice that I have received from others, the usual disclaimer applies: that all errors, omissions, and shortcomings are mine alone.

References

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. American Economic Review 101 (2), 470–92.

Acemoğlu, D., Como, G., Fagnani, F., Ozdaglar, A., 2013. Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research 38 (1), 1–27.

Alesina, A., Tabellini, G., 1990. A positive theory of fiscal deficits and government debt. The Review of Economic Studies 57 (3), 403–414.

Allport, G. W., 1954. The nature of prejudice.

Ansolabehere, S., Rodden, J., Snyder, J. M., 2008. The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting. American Political Science Review 102 (2), 215–232.

Asch, S. E., Guetzkow, H., 1951. Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men, 222–236.

Astin, A. W., 1993. Diversity and multiculturalism on the campus: How are students affected? Change: The Magazine of Higher Learning 25 (2), 44–49.

Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's who in networks. wanted: The key player. Econometrica 74 (5), 1403–1417.

Bhatia, R., 2001. Linear algebra to quantum cohomology: the story of alfred horn's inequalities. The American Mathematical Monthly 108 (4), 289–318.

Bindel, D., Kleinberg, J., Oren, S., 2015. How bad is forming your own opinion? Games and Economic Behavior 92, 248–265.

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

Buchholz, P., 1994. Exact and ordinary lumpability in finite markov chains. Journal of applied probability 31 (1), 59–75.

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. The Review of Economic Studies 76 (4), 1239–1267.

Carpenter, J. P., 2004. When in rome: conformity and the provision of public goods. The Journal of Socio-Economics 33 (4), 395–408.

Cavalcanti, T. V., Giannitsarou, C., Johnson, C. R., 2017. Network cohesion. Economic Theory 64 (1), 1–21.

Chandrasekhar, A. G., Larreguy, H., Xandri, J. P., 2015. Testing models of social learning on networks: Evidence from a lab experiment in the field. Tech. rep., National Bureau of Economic Research.

Chen, Y., Harper, F. M., Konstan, J., Xin Li, S., 2010. Social comparisons and contributions to online communities: A field experiment on movielens. The American economic review 100 (4), 1358–1398.

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

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

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

Desai, M. P., Rao, V. B., 1993. On the convergence of reversible markov chains. SIAM journal on matrix analysis and applications 14 (4), 950–966.

Diaconis, P., Stroock, D., 1991. Geometric bounds for eigenvalues of markov chains. The Annals of Applied Probability, 36–61.

Diaz, J., Metcalf, F., 1966. A complementary triangle inequality in hilbert and banach spaces. Proceedings of the American Mathematical Society 17 (1), 88–97.

Doran, P. T., Zimmerman, M. K., 2009. Examining the scientific consensus on climate change. Eos, Transactions American Geophysical Union 90 (3), 22–23.

Falk, A., Ichino, A., 2006. Clean evidence on peer effects. Journal of labor economics 24 (1), 39–57.

Fazio, R. H., 1986. How do attitudes guide behavior. Handbook of motivation and cognition: Foundations of social behavior 1, 204–243.

Friedkin, N. E., Johnsen, E. C., 1999. Social influence networks and opinion change. Advances in Group Processes 16, 1–29.

Golub, B., Jackson, M. O., February 2010. Naïve learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics 2 (1), 112–49.
URL http://www.aeaweb.org/articles?id=10.1257/mic.2.1.112

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

Grimm, V., Mengel, F., 2014. An experiment on belief formation in networks. Working Paper.

Haruvy, E., Lahav, Y., Noussair, C. N., 2007. Traders' expectations in asset markets: experimental evidence. American Economic Review 97 (5), 1901–1920.

Hegselmann, R., Krause, U., et al., 2002. Opinion dynamics and bounded confidence models, analysis, and simulation. Journal of artificial societies and social simulation 5 (3).

Hogan, D. E., Mallott, M., 2005. Changing racial prejudice through diversity education. Journal of College Student Development 46 (2), 115–125.

Horn, R. A., Johnson, C. R., 1990. Matrix analysis. Cambridge university press.

Houston, D. A., Fazio, R. H., 1989. Biased processing as a function of attitude accessibility: Making objective judgments subjectively. Social cognition 7 (1), 51–66.

Jadbabaie, A., Molavi, P., Tahbaz-Salehi, A., 2013. Information heterogeneity and the speed of learning in social networks. Columbia Business School Research Paper 13-28.

Kahan, D. M., Peters, E., Dawson, E. C., Slovic, P., 2017. Motivated numeracy and enlightened self-government. Behavioural Public Policy 1 (1), 54–86.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., Mandel, G., 2012. The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature climate change 2 (10), 732–735.

Krause, U., 2000. A discrete nonlinear and non-autonomous model of consensus formation. Communications in difference equations, 227–236.

Kulik, C. T., Roberson, L., 2008. Common goals and golden opportunities: Evaluations of diversity education in academic and organizational settings. Academy of Management Learning & Education 7 (3), 309–331.

Kuran, T., Sandholm, W. H., 2008. Cultural integration and its discontents. The Review of Economic Studies 75 (1), 201–228.

Levin, D. A., Peres, Y., Wilmer, E. L., 2009. Markov chains and mixing times. American Mathematical Soc.

Levin, S., Van Laar, C., Sidanius, J., 2003. The effects of ingroup and outgroup friendships on ethnic attitudes in college: A longitudinal study. Group Processes & Intergroup Relations 6 (1), 76–92.

Liu, X., Patacchini, E., Zenou, Y., Lee, L.-F., 2012. Criminal networks: Who is the key player?

Masson, C. N., Verkuyten, M., 1993. Prejudice, ethnic identity, contact and ethnic group preferences among dutch young adolescents. Journal of Applied Social Psychology 23 (2), 156–168.

Melguizo, I., 2016. Endogenous homophily and the persistence of disagreement. Mimeo.

Parsegov, S. E., Proskurnikov, A. V., Tempo, R., Friedkin, N. E., 2017. Novel multidimensional models of opinion dynamics in social networks. IEEE Transactions on Automatic Control 62 (5), 2270–2285.

Poole, K. T., Daniels, R. S., 1985. Ideology, party, and voting in the us congress, 1959–1980. American Political Science Review 79 (2), 373–399.

Salganik, M. J., Dodds, P. S., Watts, D. J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. science 311 (5762), 854–856.

Scheinkman, J., Xiong, W., 2004. Heterogeneous beliefs, speculation and trading in financial markets. In: Paris-Princeton Lectures on Mathematical Finance 2003. Springer, pp. 217–250.

Schuman, H., Steeh, C., Bobo, L., Krysan, M., 1997. Racial attitudes in america: Trends and interpretations, rev.

Tejada, A. J. R., Luque, M. N., Rojas, O. M. L., Moreno, P. J. P., 2011. Prejudiced attitude measurement using the rasch rating scale model. Psychological reports 109 (2), 553–572.

Tesser, A., 1993. The importance of heritability in psychological research: The case of attitudes. PSYCHOLOGICAL REVIEW-NEW YORK- 100, 129–129.

Voss, Z. G., Cable, D. M., Voss, G. B., 2006. Organizational identity and firm performance: What happens when leaders disagree about “who we are?”. Organization Science 17 (6), 741–755.

Wang, B.-Y., Xi, B.-Y., 1997. Some inequalities for singular values of matrix products. Linear algebra and its applications 264, 109–115.

Wilson, T. D., Lindsey, S., Schooler, T. Y., 2000. A model of dual attitudes. Psychological review 107 (1), 101.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., Ropp, S. A., 1997. The extended contact effect: Knowledge of cross-group friendships and prejudice. Journal of Personality and Social psychology 73 (1), 73.

Zafar, B., 2011. An experimental investigation of why individuals conform. European Economic Review 55 (6), 774–798.