What makes an opinion leader: expertise versus popularity

Job Market Paper

For the latest version, please click here.

Theodoros Rapanos*

November 20, 2015

Abstract

This paper studies learning based on information obtained through social or professional networks. Building on the framework first proposed by DeGroot (1974), agents repeatedly update their beliefs by weighting the information acquired from their peers. The innovation lies in the introduction of dynamically updated weights. This allows agents to weight a contact with poor information little at first, but more later on, if that contact has in the meantime gathered better information from other, more knowledgeable agents. The main finding is that individuals' social influence will depend on both their popularity (as captured by eigenvector centrality) and their expertise (as captured by information precision) in a simple and intuitively appealing way. It is moreover shown that even completely uninformed agents can contribute to social learning, and that under some network structures, providing certain agents with better information could actually lead society to worse assessments. The paper also discusses how the relationship between expertise and popularity in a network affects the learning process.

JEL classification codes: D71, D83, D85, Z13

*Dept. of Economics, Stockholm University, 10691 Stockholm, Sweden. email: theodoros.rapanos@ne.su.se.
I would like to thank my thesis advisor, Yves Zenou, for his very helpful guidance and comments. I am also thankful to Antonio Cabrales, Ana Maria Ceh, Yannis Ioannides, Paul Klein, Jordi Massó, Tomàs Rodríguez Barraquer, David Seim, Marc Sommer, Dimitri Vayanos, Jörgen Weibull, Karl Wärneryd, as well as the participants at CRETE 2015 conference and SOFI Netwoks Workshop, and the seminar audiences at Stockholm University, the Stockholm School of Economics, and Universitat Autònoma de Barcelona for their useful feedback that helped improve this paper. Financial support from Svenska Handelsbanken and the Jan Wallander and Tom Hedelius Foundation through grant H2012-0361:1 is gratefully acknowledged.

1 Introduction

Acquisition and aggregation of information is a critical part of the decision making process of individuals, firms, organisations, and governments. As a rule, however, obtaining access to the primary sources of information may be quite difficult, if it is feasible at all. Consequently, most information reaches agents through secondary sources or contacts, who may have themselves acquired it indirectly. Social and professional networks arise thus as a major channel of information diffusion in a society.

The present paper focuses on identifying some determinants of social influence. What are the characteristics of the individuals that get to lead public opinion? Are the beliefs of the experts, or those of the most popular agents that have a greater influence in a society? Is it the most popular individual who should be entrusted with passing on information or raising awareness about an issue in the public?

In order to provide some answers to the above questions, this paper builds on a benchmark model of social influence introduced by DeGroot (1974). The simple but compelling idea behind this model is that agents update their beliefs by repeatedly communicating with their neighbours, and weighting their information. There is, however, an important innovation introduced here: Agents are no longer assumed to assign fixed weights to their peers, as in the standard model. Instead, they update these weights every period, adapting them to reflect the information their peers gain access to. Using this richer setup, it is shown that each agent's social influence stems from two components: a position-driven, *popularity*, and an information-driven one, *expertise*. This new approach enables a social planner to design targeted policy interventions based on the above characteristics, and evaluate their impact. The rest of the introduction provides a short overview of the relevant literature, and discusses the aforementioned points more thoroughly.

Individuals use information acquired through their networks in various facets of their lives. The key role that social contacts can play in job search was documented in the benchmark work by Granovetter (1973), and more recent evidence from empirical studies seems to corroborate, if not strengthen this finding.[1] Consumers often seek the advice of friends who have a deeper knowledge or a better understanding of the relevant area before deciding to buy a new computer or car. Even in everyday consumption decisions, information transmitted through social networks can be crucial (see, for example, Moretti, 2011). The importance of social networks as information transmission mechanisms has grown

[1] See Ioannides and Datcher Loury (2004) for a comprehensive survey.

over the last years due to the emergence of the digital social media. As recent evidence suggests, people use online networks to exchange information on a broad variety of topics, ranging from health issues and the use of medication (Lefebvre and Bornkessel, 2013) to immigration decisions and life in a new country (Dekker and Engbersen, 2014).

Information aggregation —that is, how people combine information collected from various sources— is also an important stage of the decision making process, and has occupied researchers and scholars at least since the time of Condorcet (1785). Although in general a distinct process, information aggregation is ineluctably tied to that of information acquisition: unless people are in a position to accurately track the pieces of information communicated to them back to their original sources (which, as recent research suggests, it is not the case; see for example Choi, Gale, and Kariv, 2008; Chandrasekhar, Larreguy, and Xandri, 2015), they are bound to treat already aggregated information that is passed on to them as new.

There are two main paradigms in the literature, often referred to as *fully rational* or *Bayesian learning*, and *boundedly rational* or *naïve learning* respectively. In practice though this distinction is not always straightforward, since several models encompass elements of both approaches. The main idea behind the Bayesian benchmark is that agents are fully rational, in the sense that they interpret and use in an optimal way any information that becomes available to them, either through communication or through the observation of the actions and the payoffs of their peers. Hence fully Bayesian learning not only entails the use of Bayes' rule by the agents when updating their beliefs and forming posteriors, but also encompasses the idea that agents can optimally extract information from others' actions. As (Bayesian) consistency would suggest, under some regularity conditions agents in large networks will converge to the same beliefs and/or actions, presumably the optimal ones.[2]

The Bayesian approach hinges on the assumption that agents possess the mental capacity to optimally extract and aggregate information in the aforementioned way, or at least along similar lines. Although in some cases this assumption may be plausible, empirical evidence suggests that even in very small and simple network structures this may not be true when repeated interaction is involved (Choi et al., 2008). In fact, observations from a recent field experiment (Chandrasekhar et al., 2015) are compatible with the assumption that virtually all subjects exhibit a non-Bayesian behaviour. In some cases, it may be actually challenging

[2] Nevertheless, factors such as the existence of disconnected or weakly connected network components (Gale and Kariv, 2003), excessively influential agents (Acemoğlu, Dahleh, Lobel, and Ozdaglar, 2010) or highly unbalanced network structures (Mossel, Sly, and Tamuz, 2015) can lead society astray.

even for the modeler to apply fully Bayesian analysis, especially if the assumption of common knowledge of the network structure is relaxed.

The present work adheres to the paradigm of boundedly rational learning. In particular, agents are assumed to update their beliefs through a so-called *average-based updating process*, first introduced by DeGroot (1974). In his seminal study, the author proposes a method for how a group of individuals, such as a committee, exchange opinions and update their beliefs about the value of some unknown parameter they wish to estimate. The process is simple and intuitively appealing: Each member of the group assigns some weight (a degree of "trust") to each other member, and in every period they revise their opinion by taking a weighted average of the beliefs of their peers. The weights can be subjective, and thus not all members need to agree on them. Trust is not necessarily reciprocal, and some members may even disregard completely the opinion of some other members of the committee. In modern network theory terminology, this communication structure can be seen as a directed and weighted network. As DeGroot shows, under some regularity assumptions (for example, any piece of information should be able to flow through the entire network, irrespectively of its origin), a consensus among the members of the group will be attained.

Based on DeGroot's model, DeMarzo, Vayanos, and Zwiebel (2003) study the formation of (political) opinions in the presence of what they refer to as *persuasion bias*. People, either because they are not aware of the structure of the network, or because they lack the cognitive ability or the time required to fully track the path that information has followed in the network, fail to account for the repetition of the information they receive. Hence, due to the influence of some prominently positioned individuals, their beliefs may be driven away from both the true value of the parameter, and the society's initial average beliefs .

Golub and Jackson (2010) maintain the same framework, but introduce a rigorous network-theoretical framework, and a more standard networks-based approach. Their main finding is that persuasion bias will be present even in arbitrarily large societies. Only if the influence of prominent individuals goes to zero as the network grows can society learn efficiently. This result shows that persuasion bias is not, in general, remedied by large numbers, suggesting thus that the intuition behind Condorcet's auspicious finding may no longer apply if some members of the society receive disproportionately high attention.

The present work can be seen as an addition to the literature on average-based social learning, since it retains the spirit underlying the DeGroot model as well as its main idea: agents revise their beliefs by simply weighting the beliefs of their peers. It differs, however, from the existing literature in a key aspect: the introduction of dynamically updated weights, that enable agents to adjust the degree of trust they assign to their contacts to the flow of information. Dynamic weights capture the intuitive idea that individuals may initially assign a low weight to the belief of a peer who is uninformed or possesses low-quality information, but which they can subsequently increase, if that peer acquires information from contacts who are considered experts in the field, or simply have access to more accurate information.

The main findings are the following. First, it is shown that in the present model of learning, as under DeGroot learning, agents' beliefs will over time converge to a stable consensus in strongly connected networks.

Second, again similarly to the DeGroot model, the influence of each agent's initial beliefs in the formation of the consensus belief can be obtained explicitly. Unlike, however, DeGroot learning, the present aggregation process is no longer a "black box", since the determinants of each agent's influence can be explicitly calculated. In particular, it is shown that it can be attributed to three components: the agent's *popularity* (expressed as his or her eigenvector centrality in the network), the agent's *expertise* (expressed as the precision of the information he or she possesses), and a parameter that captures how information is distorted by the network, and which is hence common for all agents in a given network. Interestingly, and perhaps surprisingly enough, not only the relaxation of the assumption of fixed weights does not increase the informational requirements for the calculation of social influences on behalf of the modeler or the social planner, but in fact it makes it easier in cases where the latter is not fully aware of the entire network structure.

Third, it is shown that agents with low initial expertise (i.e. low-quality information), and even agents possessing no information at all initially, can play a crucial role in the learning process. This is a compelling feature that facilitates the study of networks where the majority of information originates from a minority of individuals. This is a common empirical observation in network analysis (see, for example, Galeotti and Goyal, 2010, and the references therein). Moreover, it is a prediction that is in line with findings from the Bayesian strand of literature (see, for example, Mueller-Frank, 2013), but perhaps also with common intuition: individuals without any information or knowledge of their own are often in a position to affect public opinion by propagating information or opinions that they have acquired from more knowledgable contacts. Hence, although such agents merely act as conduits for the transmission of information, their contribution can be significant, especially if they are centrally located or have direct access to expert agents. In contrast, under DeGroot learning, agents without any information would either be completely ignored, since they would be given a zero initial weight that they would carry over forever thereon, or would be given a positive but largely arbitrary weight, based on their neighbours' assessment of their future access to information. In the present model such issues are overcome by allowing for weights to vary over time.

Finally, the breakdown of social influence into an information-driven, a popularity-driven, and a (common) network-driven component allows room for the evaluation of policy interventions.

The rest of the paper is structured as follows: Section 2 introduces the general framework of the model, and then provides a brief overview of some basic network-related concepts that are used later on; readers familiar with economics of networks or graph theory may harmlessly skip the latter part. The main part of the paper begins at Section 3 with the introduction of a model of *dynamic* average-based learning; the belief-updating process is presented and motivated. Section 4 studies the dynamics of the new process, establishes convergence of beliefs, and presents the main theorem of the paper, which is used in Section 5 to study the efficiency of target policy interventions. Section 6 concludes. Section A of the Appendix provides the mathematical tools that are used throughout the paper. The proofs of the propositions and theorems have been deferred to Section B of the Appendix.

2 The setup

2.1 The agents

Consider a society consisting of a finite number of individuals who would like to gather more information about a parameter of interest or form an opinion regarding an issue they need to make a decision on. The pattern of communication among the agent is captured by a network $\mathcal{G}$.

As in the rest of the literature on average-based updating, it will be assumed that the agents are only interested in estimating the true value of the unknown parameter, and stick to the stipulated updating process; they do not seek to maximize their social influence, nor do they have something to gain from propagating a particular belief.

Before proceeding with the rest of the technicalities, it would be useful to provide some guidelines about the notation used in this paper. Matrices shall be denoted with bold capital letters, for example $\mathbf{X}$, and vectors with bold lowercase letters, for example $\mathbf{y}$. The (i,j)-th element of matrix $\mathbf{X}$ shall be denoted with x_{ij}, and the i-th element of vector $\mathbf{y}$ with y_i. All vectors are defined as column vectors, and thus transposed vectors, e.g. $\mathbf{y}^\top$, will be row vectors. Finally, if $\mathbf{y}$ in an n-dimensional vector, $\mathbf{D_y}$ shall denote the $n \times n$ diagonal matrix with the elements of vector $\mathbf{y}$ on its main diagonal (and the rest of its elements equal to zero).

2.2 The network

A network is modelled as a —potentially directed— graph $\mathcal{G} = \langle \mathcal{N}, \mathcal{E} \rangle$, where the set of *vertices* or *nodes*, $\mathcal{N} := \{1, 2, \ldots, n\}$, represents a set of agents who can potentially interact with each other, and the set of *edges* $\mathcal{E} \subseteq \mathcal{N}^2$ represents the links among them. In many applications of network theory it is more convenient to represent a network using a matrix $\mathbf{G} := [g_{ij}]_{(i,j) \in \mathcal{N}^2} \in \{0,1\}^{n \times n}$, where $g_{ij} := 1$ if there is a directed edge from node i to node j (i.e. agent i is linked to agent j), and $g_{ij} := 0$ otherwise. In network theory terminology, matrix $\mathbf{G}$ is referred to as the *adjacency matrix* of network $\mathcal{G}$.

In the analysis in the following sections of this paper, the adjacency matrix $\mathbf{G}$ will represent the pattern of communication and transmission of information across the network. Consider any two agents $i, j \in \mathcal{N}$. A link from agent i to agent j, $g_{ij} = 1$, has the interpretation that agent i has access to agent j's belief. It shall be then said that agent i *observes*, *pays attention to*, or *listens to* agent j, or in network theory terminology, agent i is an *in-neighbour* of agent j. Equivalently, it can be said that agent j *receives attention* from or is an *out-neighbour* of agent i.[3]

Two important remarks are in order at this point. First, as the above discussion suggests, attention may not be reciprocal: the fact that agent i can observe agent j's belief does not necessarily imply that j is in a position to observe i's belief. Hence the adjacency matrix $\mathbf{G}$ will be, in general, non-symmetric. Second, it is reasonable to assume that every agent can observe himself.[4] The diagonal of $\mathbf{G}$ will thus consist of ones, that is, $g_{ii} = 1$ for all $i \in \mathcal{N}$. This assumption will be maintained throughout this paper and will not be stated explicitly again.

The set of all agents that agent i pays attention to (that is, all the out-neighbours of agent i) in network $\mathcal{G}$ constitutes the *out-neighbourhood* of agent i, and is denoted with $\mathcal{D}_\mathcal{G}(i)$. Using

[3] This terminology has its roots in the drawing of networks as graphs (see, for example, networks A and B in Example 1). A directed link *from* agent i *to* agent j means that i *gives* attention to j; hence j is an out-neighbour of i. This of course implies that j receives attention from i; hence i is an in-neighbour of j.

[4] In the terminology introduced above, this implies that every agent is an out-neighbour of himself.

mathematical notation, for any $i \in \mathcal{N}$

$$\mathcal{D}_{\mathcal{G}}(i) := \{j \in \mathcal{N} | g_{ij} = 1\}.$$

Notice that the out-neighbourhood of any agent is a non-empty set, since $i \in \mathcal{D}_{\mathcal{G}}(i)$ for every $i \in \mathcal{N}$.

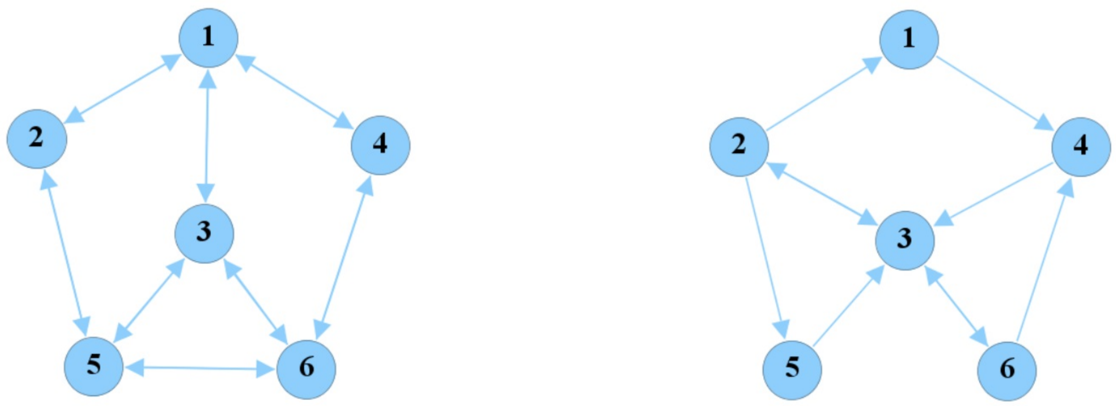
The following terms are common in the networks literature, and will be used throughout the analysis that follows.

Definition 1: Some Networks Terminology

- A *directed walk* in a network $\mathcal{G}$ is a sequence of (potentially repeated) nodes that are sequentially connected via directed links. Formally, it is a sequence of nodes $\langle i_1, i_2, \ldots, i_{H-1}, i_H \rangle$ in $\mathcal{G}$ such that $g_{i_h i_{h+1}} = 1$ for all $h \in \{1, 2, \ldots, H-1\}$.
- A *directed path* from node $i_1 \in \mathcal{N}$ to another node $i_H \in \mathcal{N}$ in a network $\mathcal{G}$ is a directed walk consisting of *distinct* nodes. Formally, it is a sequence of nodes $\langle i_1, i_2, \ldots, i_{H-1}, i_H \rangle$ in $\mathcal{G}$, with $i_k \neq i_l$ for $k \neq l,\ k, l \in \{1, 2, \ldots, H\}$, such that $g_{i_h i_{h+1}} = 1$ for all $h \in \{1, 2, \ldots, H-1\}$.
- A *simple cycle* of length H in a network $\mathcal{G}$ is a closed walk consisting of H distinct nodes. Formally, it is a sequence of nodes $\langle i_1, i_2, \ldots, i_{H-1}, i_H \rangle$ in $\mathcal{G}$ such that $g_{i_h i_{h+1}} = 1$ for $h \in \{1, 2, \ldots, H\}$, with $i_1 = i_H$ and $i_k \neq i_l$ for all other $k, l \in \{1, 2, \ldots, H\}$ with $k \neq l$.
- A network $\mathcal{G}$ is said to be *strongly connected* if there exists a directed path from any node to any other node in $\mathcal{G}$.
- The *period* of a strongly connected network $\mathcal{G}$ is defined as the greatest common divisor of the lengths of its simple cycles.
- A strongly connected network is called *aperiodic* if its period is equal to 1, otherwise it is called *periodic*.

Example 1. *The networks shown in Figure 1 are strongly connected. A directed link (arrow) indicates that the agent at its origin observes the information of the agent at its target; hence information flows opposite to the direction of the arrows. To keep graphs as simple as possible, and since all agents are assumed to observe themselves, self-loops have not been drawn. Observe that while attention can be reciprocal, as it is in Network A, may not necessarily be the case. In Network B, for example, agent 3 receives attention form almost everyone in the network, but gives attention only to agents 2 and 6.*

Example 1: Strongly connected networks



Network A　　　　Network B

2.3 Popularity: the eigenvector centrality measure

Eigenvector centrality was first proposed by Bonacich (1972) as a measure of influence, prestige, or popularity in a network. It captures the idea that *what makes an agent important in a network is how well-connected this agent is to other important agents*. More specifically, each agent's eigenvector centrality is a weighted average of the eigenvector centralities of his or her *in-neighbours*. That is, an individual is considered more influential if he or she *receives* attention from influential individuals. A formal definition is provided below.

Definition 2: Eigenvector Centrality

Consider a strongly connected network $\mathcal{G} = \langle \mathcal{N}, \mathcal{E} \rangle$ with adjacency matrix $\mathbf{G}$. The *eigenvector centrality profile* of network $\mathcal{G}$ is defined as the positive left eigenvector of $\mathbf{G}$, that is, as a vector $\mathbf{c} := [c_i]_{i\in\mathcal{N}} > \mathbf{0}$ satisfying

$$\mathbf{c}^\top \mathbf{G} = \rho_{_G} \mathbf{c}^\top \tag{1}$$

normalised so that

$$\|\mathbf{c}\|_1 := \sum_{i=1}^{n} |c_i| = 1, \tag{2}$$

where $\rho_{_G}$ is the spectral radius of adjacency matrix $\mathbf{G}$, and $\|\cdot\|_1$ denotes the vector 1-norm. The *eigenvector centrality* of agent $i \in \mathcal{N}$ is given by the element $c_i \in [0,1]$.

The eigenvector centralities of the agents in Networks A and B introduced in Example 1 are given below. In the following figures, the size of the nodes has been adjusted to represent their eigenvector centrality.

Example 2. *In network A, agents 1, 3, 5, and 6 are each given attention by three peers. Agent 1 is nevertheless less important than agent 5 under the eigenvector centrality measure. The reason is that, although agents 1 and 5 have in common two peers that pay attention to them, namely, agents 2 and 3, the third in-neighbor of agent 5 (i.e. agent 6) is more important than the third in-neighbour of agent 1 (i.e. agent 4). By the same token, agent 3 is more important than agent 5 because agent 1, who listens to agent 3, is more important than agent 2, who listens to agent 5.*

In network B, notice that agents 1 and 5 are equally important since the only peer that pays attention to them is agent 2, and hence they both derive all their prestige *or* popularity *from this agent. The same holds true for agents 2 and 6, who are given attention only by agent 3.*

Example 2: Eigenvector centralities

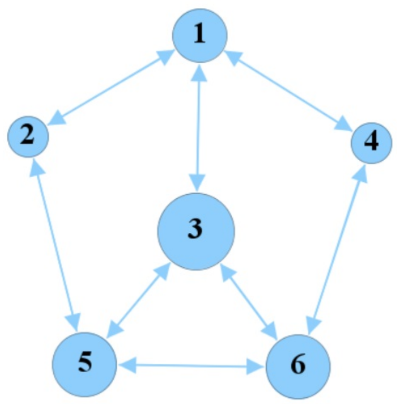


Figure 2.A

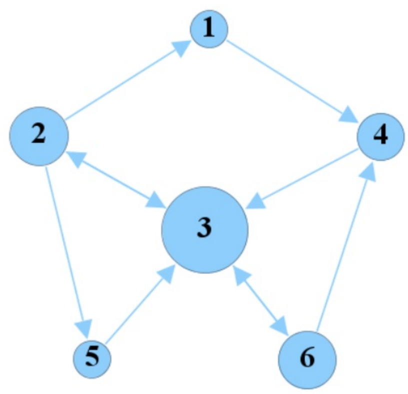


Figure 2.B

i	c_i
1	0.167
2, 4	0.129
3	0.198
5, 6	0.188

Table 2.A

i	c_i
1, 5	0.094
2, 6	0.173
3	0.319
4	0.146

Table 2.B

Eigenvector centrality can be potentially problematic as a measure of influence since it is self-referential, and as such, it may not be always well-defined.[5] Nevertheless, the assumption of strong connectedness is sufficient to guarantee that there is one and only one positive eigenvector associated with matrix $\mathbf{G}$ (see Section B.1 in the Appendix for a proof).[6] Bonacich and Lloyd (2001) and Jackson (2008) provide a motivation for the use of eigenvector centrality as a measure of influence or prestige, and propose alternative measures that can be used in the cases that the former is not well-defined. An algorithm based on a variant of eigenvector centrality that does not presuppose strong connectedness, known as *PageRank*, was used in the first versions of Google search engine to determine the order of appearance of the search results (Page, Brin, Motwani, and Winograd, 1999, sections 2.4, 2.5 and 6).

2.4 Beliefs

The first part of the present paper studies the evolution of the beliefs of the agents in a network through communication. The term *belief* shall be used to refer to an agent's opinion or accumulated information at a period of reference rather than to some (Bayesian) posterior. This constitutes an abuse of terminology, since what will be referred to as "belief" in this paper is technically a *statistic* for the accumulated information. This term, however, has been extensively used, and has become standard in the average-based updating literature in the last decade (see, for example, DeMarzo et al., 2003; Jackson, 2008; Golub and Jackson, 2010, 2012; Acemoğlu, Como, Fagnani, and Ozdaglar, 2013). Thus for conformity reasons, and in order not to cause confusion, the (ab)use of the term *belief* is maintained in the present work as well.

The choice to follow the path set by the existing literature, and not to involve priors in the analysis, should not be taken as a direct or indirect statement that priors are unimportant or that they should not be a part of a statistical updating process. This is done because the purpose of the present work is to study how information is transmitted and accounted for through a network, rather how this information is incorporated into the existing, prior beliefs of the agents. These initial beliefs may have been formed based on past observations,

[5] To see this, notice that if $\rho_G^{-1}\mathbf{G}$ is interpreted as a linear mapping, then $\mathbf{c}$ can be seen as a fixed point. There are mappings with no real (non-zero) fixed points, mappings with real but non-positive fixed points, and mappings with more than one non-negative fixed points. Any of the above would be problematic as a measure of centrality.

[6] Recall that eigenvectors are unique up to the relative magnitude of their entries. Since $\mathbf{c}$ is an eigenvector of matrix $\mathbf{G}$, any positive multiple of $\mathbf{c}$ is also an eigenvector of $\mathbf{G}$, and contains exactly the same information about $\mathcal{G}$ as $\mathbf{c}$ does; hence it could be also considered a vector of eigenvector centralities. Normalisation (2) serves only to uniquely pin down agents' centralities, and to facilitate the definition of some measures introduced in the sections that follow. Normalising eigenvector $\mathbf{c}$ with respect to its 2-norm, so that $\|\mathbf{c}\|_2 := \sqrt{\sum_{i=1}^{n} |c_i|^2} = 1$, is also quite common. The results in this paper are not affected by that particular choice.

information from other sources, accumulated knowledge or experience, or even be arbitrary. Furthermore, nothing prevents agents in an average-based updating model from using the information eventually accumulated through this process to update their existing priors in some boundedly Bayesian or semi-Bayesian way.

The reader may have noticed that no rigorous definition of beliefs has been given so far, nor has it been specified how they are represented mathematically. Depending on the context, beliefs can be represented as a percentage, a value (expressed as a number), a set of values (expressed as a vector) or even as more general objects, such as functions or probability distributions. In fact, the model discussed here can admit as beliefs any objects that are members of some convex subset $\mathcal{B}$ of a linear space.[7]

Golub and Jackson (2010) examine the relationship between the influence of individuals or small groups in large societies and the efficiency of the learning process. Since beliefs *per se* are not directly the focus of the paper, they assume for simplicity, and without loss of generality, that beliefs can be represented by a number in the unit interval, $\mathcal{B} = [0, 1]$. For the purpose of the paper by DeMarzo et al. (2003), who study the phenomenon of uni-dimensional opinions, it makes better sense for agents' beliefs to be expressed as vectors whose elements represent their view or opinion on a series of m issues; in that case $\mathcal{B} = \mathbb{R}^m$ for some $m \in \mathbb{N}^*$. In the seminal work by DeGroot (1974) the objects being updated are subjective probability distributions for the value of some parameter of interest, and hence $\mathcal{B}$ is a space of probability distributions. It becomes hence apparent then that the type of updating process discussed in this paper is quite versatile, and can be adapted to various setups. Note also that the findings of the above papers (consensus, speed of convergence, wisdom of the crowds) do not depend on the representation of the beliefs, apart of course from those that directly concern the structure of the beliefs *per se*.

2.5 Expertise

It will be also assumed that agents assign a degree of certainty, or *precision*, to their beliefs, which will be referred to as *expertise*. This is a non-negative number that expresses how much they trust the information they possess. Agents are assumed to start with some

[7] A constructive approach can be helpful in understanding why this requirement is sufficient for the process to be well-defined. Firstly, the space of beliefs $\mathcal{B}$ must have a basic structure, in order for the process of aggregation of beliefs to be both well-defined, and conceptually meaningful. This structure can be imposed by assuming that this space is abelian under addition. Yet, since the updating process to be discussed below will be based on the averaging of beliefs, with the weights being real numbers, the operation of scalar multiplication needs to be defined; hence the linear space. The convexity assumption simply guarantees that the object resulting from the updating process will still be some valid belief.

initial expertise, which changes as they communicate with their neighbours and exchange opinions. Hence, after a round of communication, the expertise of individuals who observe experts should be expected to increase more than that of individuals who do not have access to experts.

Initial expertise on the topic of interest may differ across agents for a variety of reasons, such as access to more trustworthy first-hand information, experience, or education, to mention some. In modelling terms, expertise can be captured by some appropriate parameter or statistic, depending on the application. Consider the case in which the initial belief of each agent i is equal to some noisy signal s_i that he or she receives about the true value of the parameter in question. Then the precision associated with it can be defined as some sufficient statistic for the variance of agent's signal-generating distribution (see Example 3 below).

3 The model

Assume that there is some unknown state of the world, say θ, that agents would like to estimate, in order to take a decision or a make a choice in an efficient way. It is assumed that agents have some initial beliefs, to be denoted with $b_i(0)$ for each agent $i \in \mathcal{N}$. In the simplest case, these initial beliefs are equal to some unbiased signal s_i about the true state θ. These signals will be assumed to be independent, but not identically distributed.

Agents update their initial beliefs through communication with their neighbours. In each round, they ask their out-neighbours for their beliefs, as well as an assessment of how precise or accurate these beliefs are. Then they update their own beliefs by weighting the information they receive based on their peers' expertise. The belief of agent i after t rounds of communication, where $t \in \{0, 1, 2, \ldots\}$, will be denoted with $b_i(t) \in \mathcal{B}$.

Example 3. *Consider a group of prospective investors who would like to predict the future price movement of the stock of an import company, say,* Harry Lime & Co. *This will depend on the company's quarterly earnings report, to be announced in the near future. Assume that the company's profits were in fact v^*, but investors do not have access to this information yet. Instead, they each observe some noisy signal v_i about profits, with $v_i \sim N\left(v^*, \sigma_i^2\right)$. These signals are independent from each other ($v_i \perp v_j$), and potentially heterogeneous ($\sigma_i \neq \sigma_j$ for $i \neq j$). Smaller variances reflect that some investors may follow closer the developments in the imports sector, or may have access to inside information.*

In this example, investors' initial beliefs can be assumed to be equal to the signals they observe,

$b_i(0) := v_i$, *and their expertise can be defined as the inverse of the variance of their signal,* $\pi_i(0) := \frac{1}{\sigma_i^2}$.

3.1 The canonical average-based updating process

It is useful to begin by presenting the canonical average-based updating process, due to DeGroot (1974). As described above, agents start with some initial beliefs which they update by consulting with their out-neighbours. Before the communication process begins, each agent $i \in \mathcal{N}$ assigns a *weight* $\bar{\gamma}_{ij} \in [0,1]$ to each out-neighbour $j \in \mathcal{D}_{\mathcal{G}}(i)$, including himself, so that $\sum_{j=1}^{n} \bar{\gamma}_{ij} = 1$. These weights reflect the relevance or trustworthiness of the opinion of each neighbour; they may be derived from an objective formula or simply be some —potentially subjective— assessment of the informational value contained in each belief. In DeMarzo et al. (2003) the weight $\bar{\gamma}_{ij}$ is referred to as the *direct influence* of agent j on agent i. If $j \notin \mathcal{D}_{\mathcal{G}}(i)$, meaning that agent i cannot observe agent j, the corresponding weight is set equal to zero: $\bar{\gamma}_{ij} := 0$.

The DeGroot model can generically admit any weights $\bar{\gamma}_{ij} \in [0,1]$. Of particular interest, though, is the case in which the weights that agents assign to their out-neighbours are *consistent*, in the sense that they are equal to relative expertise of each neighbour.[8] It is also a quite plausible choice for the agents in the absence of information about the network structure beyond their out-neighbours. In that case, the direct influence of agent j on agent i (the weight that agent i assigns to agent j) will be given by

$$\bar{\gamma}_{ij} = \frac{g_{ij}\pi_j(0)}{\sum_{k=1}^{n} g_{ik}\pi_k(0)}. \tag{3}$$

Once the weights have been set, the communication process begins. In every period $t \in \{1,2,\ldots\}$ each agent i observes the beliefs of his or her out-neighbours $j \in \mathcal{D}_{\mathcal{G}}(i)$, and revises his or her beliefs accordingly. In particular, the new belief of agent i will be a weighted average of the previous-period beliefs of his or her out-neighbours

$$b_i(t) = \sum_{j=1}^{n} \bar{\gamma}_{ij} b_j(t-1)$$

or, using matrix notation,

$$\mathbf{b}(t) = \bar{\mathbf{\Gamma}}\mathbf{b}(t-1) \tag{4}$$

[8] Notice that under an appropriate definition of expertise, the first-period weights will be the *Bayesian* or "objective" ones, as in Examples 3 and 4.

where $\bar{\mathbf{\Gamma}} := [\bar{\gamma}_{ij}]_{(i,j)\in\mathcal{N}^2}$ is the matrix of weights, and $\mathbf{b}(t) := [b_i(t)]_{i\in\mathcal{N}}$ is the *belief profile* in period t. Observe that $\bar{\mathbf{\Gamma}}$ is, by definition, a row stochastic matrix.[9] Notice that in the context of Example 3, the belief of each agent i after the first round of communication, $b_i(1)$, will be a sufficient statistic for v^*, given the information that agent i has access to through his or her neighbours.

By iterating on process (4) we can express the belief profile in period t as a function of the initial beliefs

$$\mathbf{b}(t) = \bar{\mathbf{\Gamma}}^t \mathbf{b}(0).$$

It follows that the *cumulative weight* assigned by agent i to agent j following t rounds of communication, will be given by the (i,j)-th element of matrix $\bar{\mathbf{\Gamma}}^t$, denoted with $\bar{\gamma}_{ij}(t)$. The limiting belief profile can be calculated then as a function of the matrix of weights, $\bar{\mathbf{\Gamma}}$, and the initial belief profile, $\mathbf{b}(0)$, as

$$\lim_{t\to+\infty} \mathbf{b}(t) = \lim_{t\to+\infty} \bar{\mathbf{\Gamma}}^t \mathbf{b}(0). \tag{5}$$

The limiting belief of agent i will be therefore given by

$$\lim_{t\to+\infty} b_i(t) = \sum_{j=1}^{n} \lim_{t\to+\infty} \bar{\gamma}_{ij}(t)\, b_j(0). \tag{6}$$

A version of the main result in DeGroot (1974), adapted to the context of the present paper, is given below.

Proposition 1: Reaching a Consensus (DeGroot 1974)

Assume that $\mathcal{G}$ is strongly connected and aperiodic, and agents follow the average-based updating process described by expression (4). Then, for all $i, j \in \mathcal{N}$, the cumulative weight assigned by agent i to agent j in the limit is given by

$$\bar{\gamma}_j^{(\infty)} = \lim_{t\to+\infty} \bar{\gamma}_{ij}(t) \tag{7}$$

where $\bar{\gamma}_j^{(\infty)}$ is the j-th element of the left eigenvector $\bar{\boldsymbol{\gamma}}^{(\infty)}$ of matrix $\bar{\mathbf{\Gamma}}$. In DeMarzo et al. (2003), this limiting weight is referred to as the *social influence* of agent j.

The above proposition readily gives rise to three remarks. First, if the stipulated condi-

[9] A non-negative square matrix $\mathbf{A} \in \mathbb{R}_+^{n\times n}$ is said to be *row stochastic* if the elements of each of its rows sum up to 1, that is, if $\mathbf{A}\mathbb{1}_n = \mathbb{1}_n$, where $\mathbb{1}_n$ is an n-dimensional vector of ones. This is why such matrices are often called *right stochastic*. Similarly, a non-negative square matrix $\mathbf{A}$ is called *column stochastic* or *left stochastic* if its columns sum up to 1, that is, if $\mathbb{1}_n^\top \mathbf{A} = \mathbb{1}_n^\top$.

tions are met, each agent i will have the same limiting influence on *all* other agents in the network, and hence the term *social influence* of agent i. Indeed, as expression (7) suggests

$$\lim_{t\to\infty}\bar{\gamma}_{ij}(t) = \lim_{t\to\infty}\bar{\gamma}_{hj}(t) \qquad \text{for all } i,j,h \in \mathcal{N}. \tag{8}$$

Notice, of course, that different agents will have in general different social influences, that is, $\bar{\gamma}_j^{(\infty)} \neq \bar{\gamma}_k^{(\infty)}$ for $j \neq k$. As (7) implies, all rows of matrix $\bar{\mathbf{\Gamma}}^t$ will be identical in the limit, and more specifically

$$\lim_{t\to+\infty}\bar{\mathbf{\Gamma}}^t = \mathbb{1}_{\mathrm{n}}\left(\bar{\boldsymbol{\gamma}}^{(\infty)}\right)^{\top}.$$

Second, it follows from expressions (6) and (7) that all agents will have the same limiting beliefs, or as it shall be said hence forth, they will reach a *consensus*.[10] In particular, for any $i \in \mathcal{N}$ it will hold

$$\lim_{t\to+\infty} b_i(t) = \sum_{j=1}^{n}\bar{\gamma}_j^{(\infty)} b_j(0).$$

Third, observe the vector of social influences, $\bar{\boldsymbol{\gamma}}^{(\infty)}$, can be seen as a vector of *weighted* eigenvector centralities. From a technical (although not a conceptual) point of view, it is essentially the stationary distribution of a homogeneous and aperiodic Markov chain with transition matrix $\bar{\mathbf{\Gamma}}$.[11]

3.2 The dynamic average-based updating process

3.2.1 Motivation and preliminaries

In some cases though it would be more reasonable to assume that the weights agents' assign to their neighbours are not constant but rather change based on how reliable the second-hand information the latter have access to is. Consider, for example, a person who has to decide whether to accept or turn down a job offer, and asks the opinions of his friends and colleagues. It may be the case that one of them used to work in the past for the firm making the offer, and she has thus some partial but perhaps outdated information about it. Yet she may still be in contact with her former colleagues at the firm, whom she may contact. It would make therefore sense for the prospective employee to place some rather moderate

[10]For a formal definition of *consensus*, see Definition 4.

[11]This is a standard result from Markov chains theory, and shall not be discussed further since it is out of the scope of this paper. For a more detailed discussion, see, for example, Section 5 in DeGroot (1974) and the references therein. Moreover, a similar, but more general approach is used to analyse the *dynamic average-based updating process* introduced in this paper, so the reader is referred to Section 4 for a motivation and a more technical analysis.

weight on the opinion of his friend, but increase after the latter has consulted with her more informed contacts. Similar arguments could apply, among other, in the case of a prospective buyer of a house in a neighbourhood he has never lived in, a student who has to decide whether to pursue post-graduate education or work in the industry, and a first-time traveller to a holiday destination.

More generally, assume that an agent $i \in \mathcal{N}$ has to decide how to weight the opinions of his out-neighbours $j \in \mathcal{D}_{\mathcal{G}}(i)$. It could be the case that one of them, agent j, is less well-informed compared to other out-neighbours of agent i, but she is able to observe a third agent, k, who is an expert in the issue in question, and whom agent i cannot directly observe. In that case it would make sense for agent i to assign a small weight to agent j in the first round of communication, but a larger one in the subsequent rounds, since by then agent j's belief will have incorporated information from her better-informed friend, agent k. Analysing this problem using the canonical average-based updating process discussed above is not possible since matrix $\bar{\mathbf{\Gamma}}$ has been assumed to be fixed.

Another implication of the fixed-weights assumption is that agents with no reliable first-hand information (that is, zero initial expertise) will be completely ignored, and thus will play no role in shaping public opinion. It would be reasonable though to consider that such agents can have a significant, albeit indirect, influence by simply spreading information they acquired from their out-neighbours. This case is of particular interest when it comes to agents who occupy a prominent position in the network, but rely on the their peers for information on a topic.

The aforementioned issues could in principle be addressed within the canonical average-based updating model, for example by letting agents weight their neighbours based on the precision of the information the latter are expected to receive in future periods. Such an approach though could be highly problematic. This not only would increase distortion due to inappropriate weighting of information, but also the weights used would have to be quite arbitrary; using some "correct" or "objective" weights would require advance knowledge of the information precision of ones's neighbours, and that of the neighbours of their neighbours, and so on, which would effectively translate into a requirement for full knowledge of the network structure. The updating process to be introduced in this section, instead, addresses the above problems by allowing for weights that vary over time.

Another important question that the model proposed here can help us answer, is what makes an agent influential in a network. In the framework used in our analysis there can be two sources of influence: network position, or *popularity*, and information precision, or *expertise*. It is not straightforward however how these attributes combine to determine the social influence of each agent. Under what conditions would a relatively badly informed or non-expert, yet centrally positioned agent, be more influential than an expert who is not in the spotlight? How much more influential would the former be? Up to what extent can people with a high degree of knowledge or specialisation in an area rely on their expertise to stir public opinion, disregarding social networking? Although the canonical model does not provide direct answers to the above questions, its dynamic counterpart introduced in this section provides a more suitable framework to study these issues.

In order for this to be achieved, we need to take a step backwards, and study how the weights assigned to neighbours' opinions, or the the *direct influences* $\bar{\gamma}_{ij}$ as called above, are determined in the first place. That is, we need to decompose the matrix of direct influences $\bar{\mathbf{\Gamma}}$ into a part depending only on the information or knowledge of the agents, and a part depending only on the position of the agents in the network. The following assumption simplifies the analysis to follow.

Assumption 1. For every agent $i \in \mathcal{N}$ there exists some agent $j \in \mathcal{D}_{\mathcal{G}}(i)$ such that $\pi_j(0) > 0$.

Assumption 1 states that every agent i has at least one out-neighbour j (who could potentially be himself) who receives an informative signal ($\pi_j(0) > 0$). The purpose it serves is to keep technicalities and notation at a minimum, and does not qualitatively affect our results. From a practical point of view it is not a very restrictive assumption, since in most applications it would be reasonable to assume that agents have some direct or indirect access to some information, even arbitrarily inaccurate, about the value of the unknown parameter. A sufficient, although not necessary, condition for this to hold is that every agent places positive precision to his initial belief, that is, $\boldsymbol{\pi}(0) > \mathbf{0}_n$. This is satisfied if it is assumed that every agent receives some signal with finite variance about the true value of the unknown parameter. Assumption 1 will be assumed to hold for the remainder of this paper, and it will not be explicitly reiterated in the theorems and propositions to follow.

3.2.2 The process

With the technicalities in order, the model can be now introduced. At the beginning of each period $t \in \{1, 2, \ldots\}$, every agent $i \in \mathcal{N}$ collects from each out-neighbour $j \in \mathcal{D}_{\mathcal{G}}(i)$ a report $\left(b_j(t-1), \pi_j(t-1)\right)$ consisting of that neighbour's previous-period *belief* $b_j(t-1)$ and the corresponding *accumulated expertise* $\pi_j(t-1)$. Then, based of these reports, agent i updates his or her belief and expertise (i.e. the precision assigned to that belief). The updating process, and how these beliefs and precisions are formed, are described below.

As discussed above in this section, in the initial period, $t = 0$, agents hold beliefs $\mathbf{b}(0)$, with the corresponding precisions being $\boldsymbol{\pi}(0)$.In the first round of communication, agents exchange information of the form $\left(b_j(0), \pi_j(0)\right)$ in the manner described above. Agent i's updated belief in period 1, $b_i(1)$, will then be a weighted average of the beliefs reported by his neighbours, with the weight $\gamma_{ij}(1)$ assigned to each belief being equal to its relative initial precision. In the notation introduced above

$$b_i(1) = \sum_{j=0}^{n} \gamma_{ij}(1)\, b_j(0) = \sum_{j=0}^{n} \frac{g_{ij}\pi_j(0)}{\sum_{k=0}^{n} g_{ik}\pi_k(0)} b_j(0).$$

Up to this point the process is almost identical to the standard average-based updating process à la DeGroot presented in Section 3.1. The difference lies in the assumption that, under the current process, agents update not just their beliefs *per se*, but also the corresponding precisions. The precision $\pi_i(1)$ that agent i places to his or her updated belief after the first round of communication, $b_i(1)$, will be assumed to be simply the sum of the initial expertise of all his or her out-neighbours, including agent i's own initial expertise, $\pi_i(0)$:

$$\pi_i(1) = \sum_{j=0}^{n} g_{ij}\pi_j(0).$$

This updating process is repeated *ad infinitum*. In the second round of communication, agent i inquires with his or her out-neighbours $j \in \mathcal{D}_{\mathcal{G}}(i)$ about their new beliefs and expertise, $\left(b_j(1), \pi_j(1)\right)$, and based on these reports revises his or her belief once more. The new weights assigned to each belief reported are calculated based on neighbours' updated precisions, $\pi_j(1)$; hence the precision given to the new belief $b_i(2)$ will be the sum of these precisions.

In general, the belief of any agent $i \in \mathcal{N}$ in period $t \in \{1, 2, \ldots\}$ (that is, after t rounds of communication) will be given by

$$b_i(t) = \sum_{j=0}^{n} \gamma_{ij}(t)\, b_j(t-1), \tag{9}$$

where

$$\gamma_{ij}(t) := \frac{g_{ij}\pi_j(t-1)}{\sum_{k=0}^{n} g_{ik}\pi_k(t-1)} \tag{10}$$

denotes the relative weight that agent $i \in \mathcal{N}$ places on the belief reported by agent $j \in \mathcal{N}$ in the t-th round of communication. Following the terminology introduced by DeMarzo et al. (2003), $\gamma_{ij}(t)$ will be referred to as the *direct influence* of agent j on agent i in period t. Unlike, however, the standard average-based updating process, it should be apparent from expression (10) that in the model introduced here the direct influences will not be constant over time, and hence the time index t. Notice of course that if agent i does not observe agent j, then $\gamma_{ij}(t) = 0$ for every $t \in \mathbb{N}$ since $g_{ij} := 0$.

Dynamic weights $\gamma_{ij}(t)$ are a result of precisions being updated every period. The *aggregate precision* attached by each agent i to his new belief in period t will be assumed to be the sum of the precisions of the beliefs reported by his neighbours that period

$$\pi_i(t) = \sum_{j=0}^{n} g_{ij}\pi_j(t-1). \tag{11}$$

Updating rule (9) can be expressed in matrix form as

$$\mathbf{b}(t) = \mathbf{\Gamma}(t)\,\mathbf{b}(t-1),$$

where $\mathbf{\Gamma}(t) := \left[\gamma_{ij}(t)\right]_{(i,j)\in\mathcal{N}^2}$ is the matrix of direct influences in period t.

Similarly, agents' aggregate precisions in period t can be written in vector form as follows

$$\boldsymbol{\pi}(t) = \mathbf{G}\,\boldsymbol{\pi}(t-1)$$

or as a vector-valued function $\boldsymbol{\pi}(t)$, with $\boldsymbol{\pi} : \mathbb{N} \to \mathbb{R}^n_+$

$$\boldsymbol{\pi}(t) = \mathbf{G}^t\boldsymbol{\pi}(0)$$

for any given vector of initial precisions $\boldsymbol{\pi}(0)$.

Now the dynamic updating procees introduced in this paper can be defined formally.

Definition 3: The Dynamic Average-based Updating Process

Agents are said to follow the *dynamic average-based updating process* if their updated beliefs after each round of communication equal a weighted average of the beliefs reported by their out-neighbours (including themselves), where the weight assigned to each belief is equal to the relative aggregate precision associated with it. Using mathematical notation, agents' beliefs in period $t \in \{1, 2, \dots\}$ (that is, after t rounds of communication) can be written as

$$\mathbf{b}(t) = \mathbf{\Gamma}(t)\,\mathbf{b}(t-1) \tag{12}$$

where the matrix of agents' direct influences is given by

$$\mathbf{\Gamma}(t) = \left[(\mathbf{G}^t \boldsymbol{\pi}(0) \mathbb{1}_n{}^\top) \circ \mathbb{I}_n\right]^{-1} \left[\mathbf{G} \circ \mathbb{1}_n (\mathbf{G}^{t-1} \boldsymbol{\pi}(0))^\top\right] \tag{13}$$

and $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product of matrices $\mathbf{A}$ and $\mathbf{B}$.[12]

It would be useful at this point to express the belief profile $\mathbf{b}(t)$ in any period $t \in \{1, 2, ...\}$ as a function of the initial belief profile $\mathbf{b}(0)$. By recursive backwards substitutions, expression (12) can be written as

$$\mathbf{b}(t) = \mathbf{\Gamma}(t)\, \mathbf{\Gamma}(t-1) \cdots \mathbf{\Gamma}(1)\, \mathbf{b}(0). \tag{14a}$$

or, in more compact form[13]

$$\mathbf{b}(t) = \prod_{\tau=1}^{t} \mathbf{\Gamma}(\tau)\, \mathbf{b}(0) \tag{14b}$$

The *cumulative influence*, or simply *influence* $w_{ij}(t)$ of agent j on agent i after t rounds of communication is defined as the (i,j)-th element of matrix $\mathbf{W}(t)$

$$w_{ij}(t) := \left[\mathbf{W}(t)\right]_{ij}$$

where

$$\mathbf{W}(t) := \prod_{\tau=1}^{t} \mathbf{\Gamma}(\tau). \tag{15}$$

Hence the belief updating process given by expressions (14a, 14b) can be written as

$$\mathbf{b}(t) = \mathbf{W}(t)\, \mathbf{b}(0) \tag{16}$$

[12] The Hadamard product of two equidimensional matrices is the matrix of the products of their respective elements. For a formal definition as well as some properties that are used in the proofs of the statements in this paper, see Section A.1 in the Appendix.

[13] Since matrix multiplication is generally non-commutative, the order of multiplications in the product prescribed by the product operator ($\mathbf{\Pi}$) is not uniquely defined. In this paper, however, it shall be used to denote the so-called *backwards matrix product*, that is

$$\prod_{\tau=1}^{t} \mathbf{X}(\tau) := \mathbf{X}(t)\, \mathbf{X}(t\text{-}1) \cdots \mathbf{X}(2)\, \mathbf{X}(1)$$

for some $\mathbf{X} \in \mathbb{R}_+^{n \times n}$. A brief discussion of these products, as well as some additional references, are provided in Section 3.3.

It would be useful at this point to summarise the updating process introduced in this section. As described by expression (12), each period agents form their beliefs by weighting the (previous period) beliefs of their out-neighbours. Expression (13) stipulates what these period-specific weights are: The weight that agent i assigns to the belief of each neighbour j in period t is equal to the accumulated expertise or precision of neighbour j in period t, normalised by the sum of the accumulated precisions of all neighbours of agent i that period. In any period t, the element w_{ij} of matrix $\mathbf{W}(t)$ expresses how much agent i's belief has been affected by agent j's initial belief over the course of all past periods.

3.2.3 Timing

At this point it would be useful to summarize the timing of the dynamic average-based updating process. Each agent $i \in \mathcal{N}$ starts with initial belief $b_i(0)$ to which he assigns precision $\pi_i(0)$. The belief profile of the agents at the beginning of each period $t \in \{1, 2, ...\}$ is denoted with $\mathbf{b}(t\text{-}1)$, and the corresponding precisions with $\boldsymbol{\pi}(t\text{-}1)$. The timing of the updating process that takes place every period t is the following:

[1] The t-th round of communication takes place. Agent i collects from each out-neighbour j a report of his or her previous period beliefs and precision (expertise), that is, a pair $\left(b_j(t\text{-}1), \pi_j(t\text{-}1)\right) \in \mathcal{B} \times \mathbb{R}_+$ for every $j \in \mathcal{D}_{\mathcal{G}}(i)$.

[2] Agent i updates the weight he or she assigns to each neighbour j (i.e. the direct influence of agent j on agent i) $\gamma_{ij}(t\text{-}1)$ to $\gamma_{ij}(t)$, according to expression (10)

[3] Agent i updates his or her belief $b_j(t\text{-}1)$ according to (9). The new belief, $b_j(t)$, is the weighted average of the beliefs $b_j(t\text{-}1)$ reported by agent i's out-neighbours, using the new weights $\gamma_{ij}(t)$ calculated in stage [2] above.

[4] Agent i calculates the precision of his or her updated belief as shown in expression (11). The new precision $\pi_i(t)$ is simply the sum of precisions of his or her out-neighbours (including own-precision) reported in stage [1] above.

Hence, "beliefs in period t" or "expertise in period t", $\mathbf{b}(t)$ and $\boldsymbol{\pi}(t)$ respectively, shall refer to the belief and precision profiles of the agents *at the end* updating process, and after all communication has taken place in period t.

3.3 A note on backward matrix products

Note that although the updating rule stipulated by expressions (14) may be reminiscent of an inhomogeneous, or as it is sometimes called, a non-stationary Markov chain, it is in fact a different process. First, from a conceptual point of view, the process described in this paper is very different from an inhomogeneous Markov chain. Observe that, unlike a Markov chain, the dynamic average-based updating process is entirely deterministic. Moreover, the elements of the matrix of direct influences, $\mathbf{\Gamma}(t)$, represent weights, and not transition probabilities, as the elements of a Markov matrix $\mathbf{M}(t)$ do. Consequently, the object being updated is a vector of beliefs $\mathbf{b}(t)$, not a probability distribution $\mathbf{p}(t)$ as in the case of a Markov chain. Hence, although $\mathbf{p}(t)$ is by definition a stochastic vector, this will not be true in general for belief profile $\mathbf{b}(t)$.

Second, from a technical perspective, recall that the dynamics of an inhomogeneous Markov chain are captured by what is often referred to as a *forward product* of stochastic matrices, that is, a product of the form $\mathbf{M}(1)\,\mathbf{M}(2)\,\cdots\,\mathbf{M}(t)$. The distribution in period t will be thus given by

$$\mathbf{p}(t) = \mathbf{p}(0)\,\mathbf{M}(1)\,\mathbf{M}(2)\,\cdots\,\mathbf{M}(t).$$

Recall, however, from expression (14a), that the dynamic average-based updating process is described by a *backwards product*

$$\mathbf{b}(t) = \mathbf{\Gamma}(t)\,\mathbf{\Gamma}(t-1)\,\cdots\,\mathbf{\Gamma}(1)\,\mathbf{b}(0)$$

As non-commutativity of matrix multiplication would suggest, these two processes are different both in dynamics and in asymptotics. Hence, the resulting beliefs (or "marginal distributions", if it is a Markov chain) in any time period will be in general different under each process ($\mathbf{b}(t) \neq \mathbf{p}(t)$ for $t \in \{1, 2, \ldots\}$), even if the starting points and all transition matrices are identical ($\mathbf{b}(0) = \mathbf{p}(0)$ and $\mathbf{\Gamma}(t) = \mathbf{M}(t)$). Backwards products, moreover, depend more heavily on the first matrix in the sequence, $\mathbf{\Gamma}(1)$, than forward products do on $\mathbf{M}(1)$.[14]

[14] Unfortunately, although the literature on forward matrix products is quite rich, there is a dearth of studies on backwards products, perhaps due their more limited applications (namely studying aspects of Markov Decision Processes, and distributed algorithms, aside from DeGroot-type updating). Chatterjee and Seneta (1977), (Seneta, 1981, chapter 4.6), and Leizarowitz (1992) provide some sufficient conditions for convergence; Anthonisse and Tijms (1977) and Federgruen (1981) study the rate of convergence of such sequences. The author of the present paper is not aware of any work providing an explicit formula for the limit of convergent sequences of backward products, analogous to those we have for forward products (see, for example, Isaacson and Madsen, 1976, Theorem V.4.7). The proofs in the present paper are based on results derived in the aforementioned papers, as well as on certain results from the Markov chains literature that do not depend on the direction of the matrix product.

Technically, the canonical average-based updating process à la DeGroot is also described by a backwards product (and should be thought of as such). Since, however, the matrix of direct influences is constant over time ($\mathbf{\Gamma}(t) := \bar{\mathbf{\Gamma}}$), the standard results for homogeneous Markov chains can be used in that case.

4 Information exchange dynamics and convergence of beliefs

4.1 Reaching a consensus in the dynamic model

This section studies the conditions under which a common belief arises in the network. Although the analysis is asymptotic, it may still be a very good approximation of the finite belief and influences dynamics, especially in the cases where convergence is fast.

Definition 4: Consensus

It is said that the agents in a network $\mathcal{G}$ reach a *consensus* if for any initial belief profile $\mathbf{b}(0) \in \mathcal{B}^n$, and any vector of initial precisions $\boldsymbol{\pi}(0) \in \mathbb{R}_+^n$, it holds

$$\lim_{t \to +\infty} \left(b_i(t) - b_j(t)\right) = 0 \qquad \text{for all } (i,\ j) \in \mathcal{N}^2. \tag{17}$$

If moreover there exists some belief $b^{(\infty)} \in \mathcal{B}$ such that

$$\lim_{t \to +\infty} b_i(t) = b^{(\infty)} \qquad \text{for all } i \in \mathcal{N} \tag{18}$$

the consensus shall be called *definitive*, and $b^{(\infty)}$ will be referred to as the *consensus belief*. Otherwise, the consensus will be called *oscillatory*.

As expression (17) suggests, a consensus is reached if after any —potentially arbitrarily large— number of communication rounds, all agents end up holding the same beliefs with each other. Notice that this does not automatically imply that these beliefs will be constant over time; it could be the case that all agents change their beliefs synchronously every period, or more accurately, keep oscillating indefinitely among a number of different beliefs. Interestingly enough though, it turns out that the dynamic average-based updating process discussed here cannot lead to oscillatory consensuses. The following result is an immediate application of Theorem 1 in Chatterjee and Seneta (1977).

Proposition 2: Stable Beliefs in the Limit

Consider a strongly connected network $\mathcal{G}$, and suppose that agents $\mathcal{N}$ reach a consensus by following the dynamic average-based updating process stipulated in Definition 3. Then this consensus must be definitive.

Hence, if beliefs end up being identical across agents, they will also be constant over time. For the remainder of the paper, the qualifier *definitive* will be omitted; since there can be no oscillatory consensuses in the current setup, it should be clear that the term *consensus* will refer to definitive consensuses.

The result below establishes convergence of the dynamic average-based updating process.

Proposition 3: Convergence

Consider a strongly connected network $\mathcal{G} = \langle \mathcal{N}, \mathcal{E} \rangle$, and assume that agents follow the dynamic average-based updating process. Then there exists a unique stochastic vector $\mathbf{w}^{(\infty)} := \left[w_1^{(\infty)}, w_2^{(\infty)}, \ldots, w_n^{(\infty)}\right]^\top$ such that[15]

$$\lim_{t \to +\infty} w_{ij}(t) = w_j^{(\infty)} \tag{19}$$

for all $i, j \in \mathcal{N}$. The limiting weight $w_j^{(\infty)}$ is called the *social influence* of agent j. It follows that the agents in $\mathcal{G}$ will reach a consensus, with the *consensus belief* $b^{(\infty)} \in \mathcal{B}$ given by

$$b^{(\infty)} = \sum_{j=1}^{n} w_j^{(\infty)} b_j(0) \tag{20}$$

Proposition 3 suggests that if agents follow the communication process introduced in Definition 4, they will all end up having the same belief about the state of the world θ^*. It is moreover shown that the consensus belief will be a weighted average of the agents' initial beliefs, with constant weights. This implies that, asymptotically, the impact of each initial opinion $b_i(0)$ will be the same on all agents in the network, irrespectively whether they can directly observe agent I or not.

Although, as discussed above, there is no general formula for the limit of such a sequence, in this case the consensus beliefs can be computed directly using some "direction-free" results

[15] A vector $\mathbf{y}$ is said to be a *stochastic* or probability vector if it is non-negative, and its elements sum up to 1, that is if $\mathbf{y} \in [0, 1]^n$ and $\sum_{i=1}^n y_i = 1$.

from matrix algebra.

Theorem 1: The Determinants of Social Influence

Assume that network $\mathcal{G}$ is strongly connected, and agents follow the dynamic average-based updating process stipulated in Definition 3. Then the social influence of each agent i is equal to the product of their eigenvector centrality and their relative initial precision, adjusted by a network effects multiplier, that is

$$w_i^{(\infty)} = \alpha_{\mathbf{c},\boldsymbol{\pi}} \; c_i \; \widetilde{\pi}_i(0) \tag{21}$$

where

c_i is agent i's *popularity* (eigenvector centrality)

$\widetilde{\pi}_i(0) := \frac{\pi_i(0)}{\sum_{j=1}^n \pi_j(0)}$ is agent i's *relative initial expertise* (precision), and

$\alpha_{\mathbf{c},\boldsymbol{\pi}} := \alpha(\mathbf{c},\, \boldsymbol{\pi}(0)) = \frac{\sum_{j=1}^n \pi_j(0)}{\sum_{j=1}^n c_j \pi_j(0)}$ is a scalar, common for all agents in a given network, that captures the *network effects* or the distortion in agents' influences induced by the network.

This result is quite interesting since it shows that the social influence of each agent under the dynamic average-based updating process depends only on his or her position in the network or *popularity*, as captured by eigenvector centrality, and on his or her *relative initial expertise* (that is, how precise their information is relative to that of the other agents). It also disentangles these two effects in a clear and straightforward way.

Example 4. *A group of economists planning to attend a conference are reviewing their travel options. The most economical option would be to fly the local airline,* Carcosa Air*, but a severe delay could cause them to miss their connection flight. In fact such delays occur with probability p^*, unknown to the prospective travellers. Before making a decision, the economists can communicate with their friends in order to exchange information.*

In this example, agents' initial expertise or precision, $\pi_i(0)$, can be taken to be the number of times they have flied Carcosa Air in the past, while their initial belief about the probability of a severe delay, $b_i(0)$, can be the percentage of their past flights that were delayed.[16].

[16] Notice that the parameter of interest, that is the probability of delay p^*, can be seen as the (unknown) success probability of a draw from a Bernoulli distribution with support $\Omega = \{0 \textit{ (on time)}, 1 \textit{ (delayed)}\}$. Then the dynamic average-based updating process described in Definition 3 would be the optimal information aggregation process in the absence of persuasion bias. In fact, expression (12) can be derived from the

Nodes representing agents with a higher degree of expertise *are depicted in darker colours in* Figure 4.B. *Agents' popularity (c_i), initial relative and absolute expertise ($\widetilde{\pi}_i(0)$ and $\pi_i(0)$), as well as their social influence at the consensus $\left(w_i^{(\infty)}\right)$, are given in* Table 4.

Example 4: Social influence

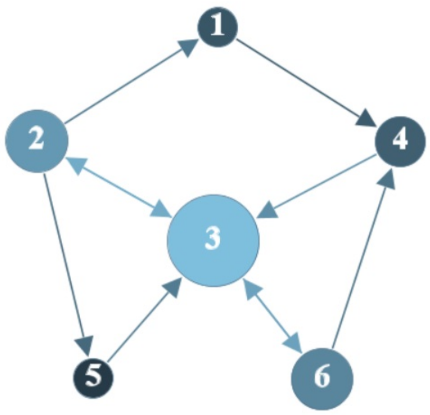


Figure 4.B

i	c_i	$\widetilde{\pi}_i(0)$	$(\pi_i(0))$	$w_i^{(\infty)}$
1	0.094	0.222	(*14*)	0.154
2	0.173	0.111	(*7*)	0.142
3	**0.319**	0.048	(*3*)	0.112
4	0.146	0.206	(*13*)	**0.221**
5	0.094	**0.270**	(**17**)	0.188
6	0.173	0.143	(*9*)	0.183
$\sum$	1	1	(*63*)	1

Table 4.B

In the small society of this example, the individual with the highest expertise is agent 5, while the most popular one is, by far, agent 3. Yet the most influential one is agent 4, who ranks third in expertise, and just fourth in popularity. As this example suggests, in general it is the agent who has the "right" (for that society) mixture of expertise and popularity that gets to influence public opinion the most. Agent 5 is the expert here, but she is rather at the margin of social attention, and this limits her influence. Agent 3, on the other hand, is the most popular agent in the network, yet he is considered to be quite ignorant regarding the topic in question. As a result, his initial belief will be heavily discounted by the other agents.

The following remark follows directly from expression (21).

Corollary 1. *The expertise-driven component of social influence is determined by the* relative *initial expertise (that is, precision of the information) of each agent; any changes in absolute expertise matter only inasmuch as they alter relative precisions.*

As the above statement indicates, it is relative, not absolute precision that matters for the

application of Bayes rule under the assumption that every agent i has a $Beta(\beta, \delta)$ prior distribution for the probability of a delay, p^*, where $\beta = b_i(0)\pi_i(0)$ and $\delta = (1 - b_i(0))\pi_i(0)$. Beta distribution is the conjugate prior of the Bernoulli distribution; for a more in-depth discussion, see Pham-Gia (2004) limiting beliefs. Note that this refers to an agent's relative initial expertise with respect to *all* other agents in the network, not only his or her neighbours; any effects due to differences in connections or position in the network are captured by the popularity-driven component, c_i. This is a consequence of the assumption of strong connectedness of the network.

The result in Theorem 1 is interesting from both a theoretical and an empirical point of view. Not only it is straightforward to see whether important agents derive their influence from their position in the network or the information they possess, but it is also easy to see how a small change in the information precision of some agent, or a rewiring of his links, would affect his social influence as well as the consensus beliefs. This could have direct implications on how some social planner could intervene in order to facilitate or disrupt the flow of information in a network.

4.2 Constant versus dynamic weights and the role of uninformed agents

Another interesting observation is that, under the dynamic average-based updating process, even agents without any credible initial information can affect consensus beliefs. This is because although their initial expertise may be zero, they can affect the beliefs of their neighbours in subsequent periods by passing on second-hand information. Hence, despite that their own social influence will be zero, they will affect the social influences of the other agents, possibly unevenly, through their effect on eigenvector centralities $\mathbf{c}$.

This is a compelling finding, since as empirical literature has shown, information often originates from a small number of individuals. In many cases, acquiring first-hand information may be costly in terms of time and effort. Hence, a large number of people prefer to obtain their information indirectly, through a minority of expert or well-informed individuals, something that Galeotti and Goyal (2010) called *the law of the few*. A prominent example are online communities such as network forums. It is therefore important to understand the role that the initially ignorant agents play in the information diffusion process once they have learned from the experts.

This, however, may not be very clear under the canonical average based updating process, since weights have assumed to be constant. Hence if agents with no information receive zero initial weight from their neighbours, this will be carried over *ad infinitum*. As a result they will be completely ignored, and their presence will have no impact on the consensus beliefs or the social influence of the other agents. The following example shows how the predictions of the dynamic model introduced in this paper can differ from those of the baseline DeGroot model.

Example 5. *Let us revisit the group of investors introduced in Example 3. Recall that they are interested in forecasting as accurately as possible the profits of an imports firm,* Harry Lime & Co.*, before the official announcement is made. Towards this, they communicate with their contacts, and exchange information. The communication pattern is captured by Network A.*

Investors 1 and 5 have been following the imports industry closely, so they are the experts in this example. Investor 2, on the contrary, has no information at all about the developments in HLC *or the imports industry in general, and hence relies entirely on second-hand information from her contacts.*

Investors' popularity (c_i), initial relative and absolute expertise ($\widetilde{\pi}_i(0)$ and $\pi_i(0)$), and their social influence at the consensus under the dynamic updating process ($w_i^{(\infty)}$) and the baseline DeGroot model ($\overline{\gamma}_i^{(\infty)}$) are given in Table 5. In Figure 5.A.I node size represents popularity, while a darker colour represents higher expertise.

Under the dynamic updating process, the agents whose initial belief has the highest social influence ($w_i^{(\infty)}$) is agent 5. On the contrary, agent 2 has zero social influence: since she has zero expertise ($\pi_2(0) = 0$), her initial belief (whatever this may be) will have no impact on the consensus belief.[17] *Despite that, agent 2 cannot be ignored, since she has some role to play in the information diffusion process, and hence in the formation of the consensus belief. In a network without agent 2, the most influential agent would be agent 3, ranks just third if agent 2 is present. The reason is that although she is completely uninformed, she can learn from other agents who possess better information (including her neighbour agent 5, who is one of the leading experts in this network). This way she can contribute in propagating their views, increasing thus their influence. Indeed, without agent 2, agents 1 and 5 would be less popular, and hence less influential. Notice that agent 2 was chosen to be one of the least popular agents in the network; the error from ignoring a more popular agent on the grounds of being initially ignorant could be significantly larger.*

[17]Zero expertise in this case would imply that the signal of the "ignorant" agent originates from a normal distribution with infinite variance. Although this would not be a well-defined probability distribution, it can be still used as a "starting point" to model cases where the signal is uninformative, or the agent is missing prior information. In the literature, such distributions are often called *improper priors*. Using them should not be a problem as long as they are not over-interpreted (as, for example, representing total ignorance), and the posterior they give rise to is a proper distribution (see, for example, Robert, 2007, chapters 1.5 and 3.5). Alternatively, zero precision could be thought of as an arbitrarily low positive precision, due to a signal from a normal distribution with very high variance. Although these two interpretations are not equivalent, treating them as such would be harmless given their limited use in the context of this simple example.

Example 5: The role of ignorant agents

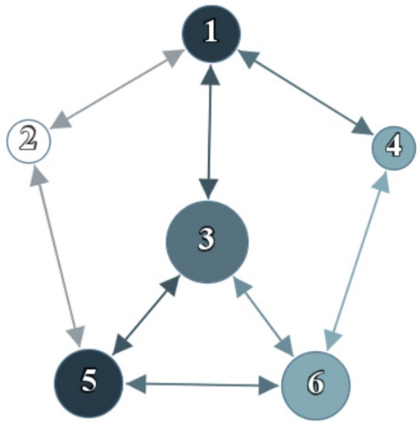


Figure 5.A.I

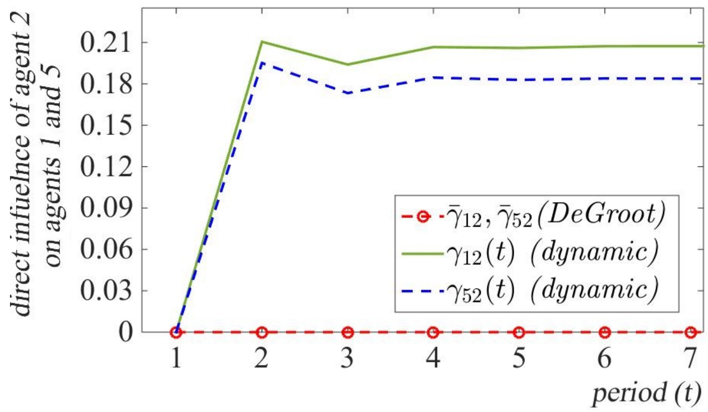


Figure 5.A.II

i	$\widetilde{\pi}_i(0)$	$(\pi_i(0))$	with agent 2 c_i	$w_i^{(\infty)}$ (dynamic)	$\bar{\gamma}_i^{(\infty)}$ (DeGroot)	without agent 2 c_i	$w_i^{(\infty)}$ (dynamic)	$\bar{\gamma}_i^{(\infty)}$ (DeGroot)
1	**0.254**	(**1.8**)	0.167	0.239	0.229	0.162	0.208	0.229
2	0	*(0)*	0.129	0	0	*n/a*	*n/a*	*n/a*
3	0.211	*(1.5)*	**0.198**	0.237	**0.271**	**0.241**	**0.257**	**0.271**
4	0.141	*(1.0)*	0.129	0.103	0.113	0.162	0.116	0.113
5	**0.254**	(**1.8**)	0.188	**0.270**	0.229	0.194	0.248	0.229
6	0.171	*(1.0)*	0.188	0.150	0.157	**0.241**	0.171	0.157
$\sum$	1	*(7.1)*	1	1	1	1	1	1

Table 5.A

This is not the case, however, in the baseline DeGroot model. If the time-constant direct influences $\bar{\gamma}_{ij}$ are assumed to be the optimal first-period weights, as given by expression (3), the direct influence of agent 2 will be zero in all periods. This not only implies that her initial opinion would be completely ignored, but also that she will keep being ignored, even when she would have accumulated knowledge from her neighbours. As shown in Table 5, the presence of uninformed agents such as investor 2 can be completely ignored in the DeGroot model.

Figure 5.A.II shows the evolution of investor 2's direct influence on her neighbours over time. In the beginning, her relative expertise is zero, and hence receives no attention from investors 1 and 3 ($\bar{\gamma}_{12}(1) = \bar{\gamma}_{52}(1) = 0$). Yet, in the next period, investor 1 becomes interested in her belief since it potentially contains information from agents that investor 1 cannot directly observe (in this case, investor 3). The same applies to investor 5. Hence, in the dynamic model, agents

adjust the weights they assign to investor 2's belief in order to account for the second-hand information she possesses in periods $t > 1$.

There are two remarks worth bringing forth before concluding the discussion of the above example. First, an alternative approach using the DeGroot model would be to let agents account for possible information that an initially uninformed neighbour may obtain later on, and assign a positive fixed weight to him. This would nevertheless have to be quite arbitrary, and would require at least some partial knowledge about the uninformed agent's neighbours, or even about the neighbours of his neighbours, and so on. An educated guess would work in this case, but if agents are assumed to trust their neighbours reported expertise, as they trust their reported beliefs, the dynamic-weights process could be a more intuitive approach for them to follow.

Second, the above example shows that under DeGroot updating, decomposing an agent's social influence into a popularity-driven and an expertise-driven part may not be as straightforward, at least with respect to the measures used in the dynamic approach (eigenvector centrality and information precision). Agents 1 and 5 have the same expertise, and despite the fact that agent 5 is more popular than agent 1, they end up having the same social influence. This suggests that the determinants of social influence in the DeGroot model could be more difficult to pin down.

4.3 Network-induced distortion

Recall that according to Theorem 1, agent i's social influence will be given by

$$w_i^{(\infty)} = \alpha_{\mathbf{c},\boldsymbol{\pi}}\, c_i \widetilde{\pi}_i(0). \tag{21}$$

It can be easily seen that in the benchmark case of a complete network it will hold that $\alpha_{\mathbf{c},\boldsymbol{\pi}} = n$ and $c_i = \frac{1}{n}$ for all agents $i \in \mathcal{N}$. It follows then that every agent's social influence will be equal to his or her true relative expertise, that is, $w_i^{(\infty)} = \widetilde{\pi}_i(0)$. However, the communication pattern dictated by an incomplete network, in combination with the failure of agents to account for repetitions of information, induces a distortion on the distribution of social influences. In expression (21), this network-induced distortion of agent i's influence is captured by

$$d_i := \alpha_{\mathbf{c},\boldsymbol{\pi}} c_i.$$

Agents with $d_i > 1$ enjoy greater social influence than the one justified by their expertise, while agents with $d_i < 1$ see their influence weighted down due to the network.

Notice that (21) can be rewritten as

$$w_i^{(\infty)} = \widehat{c_i}\ \pi_i(0)$$

where $\widehat{c_i} := \frac{1}{\sum_{i=1}^{n} c_i \pi_i(0)} c_i$ is an alternative normalisation of the vector of eigenvector centralities. Nevertheless, the formulation in (21) is preferable for two reasons. First, if both precision and centrality measures are normalised to sum up to 1, it is much clearer to see how they interact and how much each factor contributes to the agents' social influence. Second, the scalar $\alpha_{\mathbf{c},\boldsymbol{\pi}}$ has a nice intuitive interpretation: it captures the distortion in network $\mathcal{G}$, that is, it can be readily seen that

$$\alpha_{\mathbf{c},\boldsymbol{\pi}} = \sum_{i \in \mathcal{N}} d_i.$$

Yet, since this measure depends on the size of the network, it would be more meaningful to scale it by the size n of the population in the network:

$$\bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} := \frac{1}{n} \alpha_{\mathbf{c},\boldsymbol{\pi}}. \tag{22}$$

The following statement follows directly form the definition of $\bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0}$.

Corollary 2. *The following relationship exists between scaled distortion in $\mathcal{G}$, $\bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0}$, and the covariance between popularity and expertise in $\mathcal{G}$,* $\mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)]$*:*

$$\begin{aligned} \mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)] > 0 &\Longleftrightarrow \bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} < 1 \\ \mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)] = 0 &\Longleftrightarrow \bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} = 1 \\ \mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)] < 0 &\Longleftrightarrow \bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} > 1.\ \diamond \end{aligned} \tag{23}$$

Based on the above result, three different patterns of expertise/popularity allocation emerge:
Pattern A $\bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} < 1$: An $\bar{\alpha}$ smaller than 1 suggests that more centrally positioned agents will on average possess more precise information. An example is a star network with the central agent having higher expertise than the peripheral agents. In that sense, $\bar{\alpha}$ can be a measure of how the network affects "inequality" (in terms of influence): networks with smaller $\bar{\alpha}$ reinforce the influence of agents who would anyway be influential due to their high precision.

Pattern B $\bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} = 1$: There is zero correlation between agents' position in the network and the precision of their signals. This will be the case if all agents have the same popularity, such as for example, in a circle, a line, or a complete network. Another case that induces

$\bar{\alpha} = 1$ is that in which all agents have the same initial expertise, irrespectively of the structure of the network. Note that $\bar{\alpha} = 1$ can arise even in cases in which agents differ from each other both in popularity and expertise.

Pattern C $\bar{\alpha}_{\mathbf{c},\pi^0} > 1$: In this case it is the less central agents who possess on average more precise information. A star network with the agent in the centre having less precise information than the average precision in the network is such an example.

Notice that, similarly to the canonical DeGroot model, learning in the present model will be suboptimal in general. It can be shown that in the DeGroot model, the consensus belief will be correct only if the matrix of direct influences. $\bar{\mathbf{\Gamma}}$, is balanced, that is, if and only if

$$\sum_{j=1}^{n} g_{ij}\bar{\gamma}_{jj} = 1$$

for all agents $i \in \mathcal{N}$ (see DeMarzo et al. (2003), Theorem 2). Under the dynamic model introduced above, it suffices if all agents in the network are equally popular.

5 Efficiency of learning and some policy implications

The analysis in the above section establishes a straightforward relationship between the characteristics of the individuals in a network, in terms of expertise and popularity, and their social influence. It moreover quantifies the degree to which the network distorts information as the latter flows through it: the opinions of popular individuals get to be heard more, and hence receive more attention than what their informational value would justify. Conversely, the opinions of some less popular but better informed agents are underweighted compared to the optimum.

A question that arises naturally following the preceding discussion is whether some network configurations, or some information allocation patterns, favour learning more than others. The answer to this has direct policy implications. Assume, for example, that providing information to individuals is costly. What are the characteristics of the individual or the organisation that a policy maker should target in order to better inform society about the benefits of a new technology, or the right measures to help prevent a disease? Conversely, if the purpose is to disrupt the flow of information, and create confusion in a criminal organisation, which member should be given false information?

In order to be able to provide answers to such questions, some measure of efficiency of the learning process should be introduced. The obvious candidate would be the expected deviation of the consensus belief from the true value of the parameter of interest, θ^*. Assume that, as discussed in Section 3, the initial belief $b_i(0)$ of each agent i is equal to realisation of a signal s_i. All agents' signals can be stacked into a vector $\mathbf{s} := [s_i]_{i\in\mathcal{N}}$. Then the *consensus bias* in network $\mathcal{G}$ is defined as

$$Bias_{\mathcal{G}}^{(\infty)}(\theta^*,\ \mathbf{s}) := \mathsf{E}\left[b^{(\infty)} - \theta^*\right] . \tag{24}$$

Hence, any norm of $Bias_{\mathcal{G}}^{(\infty)}(\theta^*,\ \mathbf{s})$, or simply its absolute value, in case that θ^* is a scalar, could potentially serve as an efficiency measure. It turns out, however, that if the signals that agents receive before the beginning of the communication process are unbiased, so will be the consensus belief as well.

Lemma 1: Unbiasedness of the Updating Process

Let agents form their initial beliefs $\mathbf{b}(0)$ as the realisations of some independent signals, stacked into a vector $\mathbf{s}$. Assume that these signals are unbiased, so that $\mathsf{E}[s_i] = \theta^*$ for all agents $i \in \mathcal{N}$. Then, the consensus belief under the dynamic average-based updating process will be unbiased

$$Bias_{\mathcal{G}}^{(\infty)}(\theta^*,\ \mathbf{s}) = \mathbf{0},$$

that is,

$$\mathsf{E}\left[b^{(\infty)}\right] = \theta^*.$$

Notice that the above lemma applies to the baseline DeGroot as well as the dynamic average-based updating process. An important remark is that the consensus belief will be equal to the true value of the parameter in question *only in expectation*. In fact, if the state space for that parameter is infinite, as it is the case in Examples 3 and 4, the probability of the consensus belief being equal to the true state of the world will be zero.

Interestingly enough, a utility-based approach can help us motivate a more meaningful measure for the efficiency of the learning process. Assume that agents in the network intend to use the information they acquired through this communication process to take a decision. Examples could be, among other, the decision to pursue university education, to buy a product or a service (Example 4), or invest in a financial asset (Examples 3 and 5). Such choices will in general depend on each agent's preferences or constraints (income, time, credit constraints, etc). For the purpose of the present analysis, though, it will be simpler to focus on a very basic case: individuals wish to estimate the unknown parameter θ^* as accurately as possible, since any deviations would be costly. The payoff of each agent i can be expressed then by a quadratic loss function

$$u_i(\theta^*,\ b^{(\infty)}) = -(b^{(\infty)} - \theta^*)^2. \tag{25}$$

it follows then that the expected utility of agent i before the realisation of the signals will be captured by a widely used statistical measure: (the negative of) the mean squared error of $b^{(\infty)}$,

$$\mathsf{E}\Big[u_i(\theta^*,\ b^{(\infty)})\Big] = -\mathsf{E}\Big[b^{(\infty)} - \theta^*\Big]^2 = -\mathsf{MSE}\Big[b^{(\infty)}|\theta^*\Big]\ .$$

It is well known though that if an estimator is unbiased, its mean squared error collapses to its variance (see, for example, Greene, 2008, Definition C.4). Hence it holds that

$$\mathsf{E}\Big[u_i(\theta^*,\ b^{(\infty)})\Big] = -\mathsf{Var}\Big[b^{(\infty)}\Big]\ .$$

Maximising, therefore, expected social welfare in that case would amount to minimising the variance of (the estimator of) the consensus belief.

Definition 5: Quality of Assessments Efficiency of Learning

Consider two disjoint sets of agents, $\mathcal{N}$ and $\mathcal{N}'$, with the communication pattern within each population described by a strongly connected network, $\mathcal{G}$ and $\mathcal{G}'$ respectively. Assume that before communication begins, agents receive some independent, noisy, but unbiased signals about a common true state of the world, θ^*. These signals follow distributions f_i for all $i \in \mathcal{N}$, and f_j' for all $j \in \mathcal{N}'$. Denote the consensus belief in each network by $b^{(\infty)}$ and $b'^{(\infty)}$ respectively. Then $b^{(\infty)}$ will be said to be *a better assessment* than $b'^{(\infty)}$ (or equivalently $b'^{(\infty)}$ to be *a worse assessment* than $b^{(\infty)}$) if

$$\mathsf{Var}\Big[b^{(\infty)}\Big] < \mathsf{Var}\Big[b'^{(\infty)}\Big]$$

If, moreover, it holds that

$$\sum_{i \in \mathcal{N}} \pi_i(0) \leq \sum_{j \in \mathcal{N}'} \pi_j(0),$$

then the learning process in network $\mathcal{G}$ with initial expertise distribution $\boldsymbol{\pi}(0)$, will be relatively *more efficient* than the learning process in network $\mathcal{G}'$ with initial expertise distribution $\boldsymbol{\pi}'(0)$.

The first part of the above definition states that a consensus belief is preferable to another if its *ex ante* (before the realisation of the signals) variance around the mean is smaller. The second part defines the learning process in a communication network as relatively more efficient than that in another network, if the agents in the former can reach a better assessment of the unknown parameter, having in aggregate access to the same or more noisy information. This may depend on different network configurations, different distribution of initial expertise across the agents in the network, or both.

Having introduced the appropriate framework, we can now study some of the issues raised above. As standard statistical analysis suggests, more precise information —a less noisy signal— unambiguously decreases the mean squared error of an unbiased estimator. In the present setup, though this is not always the case. In fact, providing an agent with better information can lead to a worse assessments, and decrease social welfare. This is demonstrated through an example.

Example 6. *Let us return to Example 4, where the agents in Network B would like to estimate the probability of a delay that would cause them to miss their connection flight. In this setup, the precision of an agent's signal is captured by the number of times this agent has drawn an observation from the true distribution. Hence, agents who have flown* Carcosa Air *more times in the past have better initial information, or a less noisy signal about the true probability of such delays.*

Air Carcosa believe that people have overestimated the probability of delay. In order to remedy that before the upcoming conference, they decide to provide more information to some agent in the network (for example, offer him or her a free flight). Agent 3 is the most popular economist in that network, which may, intuitively, make him the most suitable candidate for that purpose.

As shown, however, in Table 6.B.I, the learning process will become less efficient, since it will lead to a worse assessment, despite the improvement in aggregate initial expertise. That is, the ex ante *expected error will increase although,* ceteris paribus*, one of the agents gets access to better information.*

The above finding may seem quite surprising at a first glance. Some intuition can be provided perhaps based on Corollary 1, and the fact that an increase in an agent's initial expertise implies a decrease in all other agents' relative initial expertise, and hence their social influence. More specifically, improving agent i's information gives rise to two effects. First, it decreases the variance of that agent's signal, and hence the aggregate variance in the network. It follows from Definition 5 that this effect tends to improve the assessment. There is, however, a second-order effect: as expression (21) implies, increasing an agent's expertise

Example 6: More information can hurt

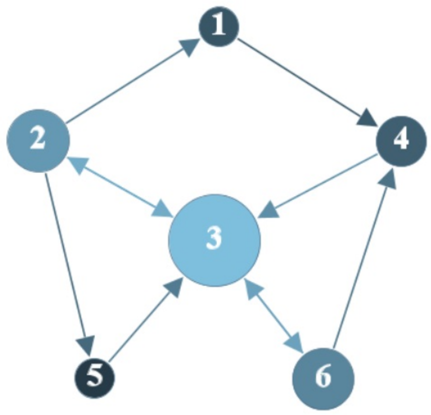


Figure 6.B

measure	*before additional info*	*after additional info*
distortion ($\alpha_{\mathbf{c},\pi}$)	4.914	4.812
scaled distortion ($\bar{\alpha}_{\mathbf{c},\pi^0}$)	1.203	1.229
$\mathrm{MSE}\left[b^{(\infty)}\vert p^*\right]$	0.0183099	0.0183093
$\mathrm{MSE}_{after} - \mathrm{MSE}_{before}$	$6.632 * 10^{-7}$	

Table 6.B.I

i	c_i	*before additional info* $\widetilde{\pi}_i(0)$	$(\pi_i(0))$	$w_i^{(\infty)}$	*after additional info* $\widetilde{\pi}'_i(0)$	$(\pi'_i(0))$	$w_i'^{(\infty)}$
1	0.094	0.222	(14)	0.154	0.219	(15)	0.149
2	0.173	0.111	(7)	0.142	0.109	(7)	0.137
3	**0.319**	**0.048**	(3)	**0.112**	**0.063**	(4)	**0.144**
4	0.146	0.206	(13)	0.221	0.203	(13)	0.213
5	0.094	0.270	(17)	0.188	0.266	(17)	0.181
6	0.173	0.143	(9)	0.183	0.141	(9)	0.176
$\sum$	1	1	(63)	1	1	(**64**)	1

Table 6.B.II

makes him or her *ceteris paribus* more influential. If agent i was already more influential than it would be justified by his or her initial expertise, providing that agent with better information could increase network distortion. This is because other agents, who may possess better information than agent i even after his or her signal improves, will see their social influence decrease further below the optimum. This was the case in Example 6 above with agent 3.

A more general result is given below.

Proposition 4: Conditions for Welfare-improving Policy Interventions

Consider a strongly connected network $\mathcal{G}$, where agents update their beliefs according to dynamic average-based updating process. Assume that agents' signals are independent and follow a normal distribution, $s_i \sim N\left(\theta^*,\ \sigma_i^2\right)$ where $\pi_i(0) := \frac{1}{\sigma_i^2}$, or a binomial distribution, $s_i \sim Bin\ (m_i,\ \theta^*)$ where $\pi_i(0) := m_i$.[18] Then a small increase in agent i's expertise leads to a better assessment if and only if

$$c_i < 2\ \frac{\sum_{j=1}^{n}\ c_j^2 \widetilde{\pi}_j(0)}{\sum_{j=1}^{n}\ c_j \widetilde{\pi}_j(0)} \tag{26}$$

The above proposition shows that better information leads to better assessments only if it is not given to excessively popular agents.

6 Conclusions

The present work contributes to the literature on boundedly rational social learning by proposing a variant of the DeGroot model that accounts for the determinant of social influence. Under the canonical average-based updating process, agents revise their beliefs by weighting the opinions of their peers; the framework introduced in this paper enables them to revise the weights too. Although the introduction of this new element constitutes a step towards a more rational, Bayesian approach, the updating process remains unambiguously naïve: agents' updated beliefs are still just weighted averages of those of their neighbours, and do not account for possible repetitions of information. As a result, the simple and intuitively appealing mechanism behind the standard DeGroot model is retained. Moreover, empirical observations, such as agents' inability to account for the repetition of information, and the subsequent emergence of persuasion bias, emerge in the model as well.

At the same time, however, the richer structure introduced in this paper provides us with a deeper insight into the determinants of social influence. In particular, as shown in Theorem 1, each agent's social influence is driven by two components (apart from a network-specific effect, common for all agents): their popularity, as captured by their eigenvector centrality, and their expertise, as captured by the relative precision of their initial beliefs.

Apart from providing an better understanding of the origins of social influence, the above result has further important implications. As it suggests, even agents with very little or no expertise at all can contribute to social learning: although their direct influence will be ini-

[18] In Examples 3 and 5 agents signals are drawn from normal distribution, while in Examples 4 and 6 signals are observations from a binomial distribution.

tially zero, they may paly an important part in the information diffusion process, and even end up being the king-makers. Ignoring the presence of such agents would lead to miscalculation of the other agents' influence, since the popularity of the uninformed agents' peers would be underestimated. Hence it may be a more suitable tool to analyse network where the majority of information originates from a small number of experts.

Furthermore, the breakdown of social influence into its primary constituents has significant policy implications. First, the amount of information that a social planner needs to have in order to estimate the agents' influences is lower than under the DeGroot model. An assessment of an agent's relative expertise (information precision) and popularity (eigenvector centrality) suffices to get a rough measure of their social influence. Although these informational requirements are still quite strong, they are much milder than the corresponding requirement in the DeGroot model (complete knowledge of the network structure). Second, social influence is described by a mathematically simple formula, expressed in terms of agents' popularity and expertise. This enables the use of comparative statics, and facilitates the design and evaluation of targeted policy interventions. Interestingly enough, it turns out some interventions may have an adverse effect even if they increase the aggregate information that is available in the network. A mistargeted information campaign, for example, could lead society to more inaccurate estimates of the true state of the world.

The present paper studies only the asymptotic behaviour of the learning process. In cases where convergence is fast, this can be a good approximation of the evolution of the short-term dynamics of the model. If, however, convergence of beliefs is slow, as it could be under the presence of homophily (Golub and Jackson, 2012), the finite dynamics of the model become more relevant. In the DeGroot model, the speed of convergence is captured by the subdominant eigenvalue of the matrix of direct influences. In the case with dynamically updated weights though, as Federgruen (1981) suggests, this is not as straightforward.

Finally, it would be interesting to bring the model proposed by this paper to the data, and test its predictions against those of the standard DeGroot model in different setups. Given, however the high informational requirement of such a project, the relevant data would ideally be collected through a field lab experiment similar to the one conducted by Chandrasekhar et al. (2015).

References

ACEMOĞLU, D., G. COMO, F. FAGNANI, AND A. OZDAGLAR (2013): "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, 38, 1–27 🌐.

ACEMOĞLU, D., M. A. DAHLEH, H. LOBEL, AND A. OZDAGLAR (2010): "Bayesian learning in social networks," *The Review of Economic Studies*, 78, 1201–1236 🌐.

ANTHONISSE, J. M. AND H. TIJMS (1977): "Exponential convergence of products of stochastic matrices," *Journal of Mathematical Analysis and Applications*, 59, 360–364 🌐.

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

BONACICH, P. AND P. LLOYD (2001): "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, 23, 191–201 🌐.

CHANDRASEKHAR, A. G., H. LARREGUY, AND J. P. XANDRI (2015): "Testing models of social learning on networks: Evidence from a lab experiment in the field," NBER working paper no. 21468 🌐.

CHATTERJEE, S. AND E. SENETA (1977): "Towards consensus: Some convergence theorems on repeated averaging," *Journal of Applied Probability*, 14, 89–97 🌐.

CHOI, S., D. GALE, AND S. KARIV (2008): "Sequential equilibrium in monotone games: A theory-based analysis of experimental data," *Journal of Economic Theory*, 143, 302–330 🌐.

CONDORCET, M. J. A. N. C. (1785): *Essai sur l'application de l'analyze à la probabilité des décisions rendues à la pluralité des voix*, Paris, France: Imprimérie Royale 🌐.

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

DEKKER, R. AND G. ENGBERSEN (2014): "How social media transform migrant networks and facilitate migration," *Global Networks*, 14, 401–418 🌐.

DEMARZO, P. M., D. VAYANOS, AND J. ZWIEBEL (2003): "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly Journal of Economics*, 118, 909–968 🌐.

FEDERGRUEN, A. (1981): "The rate of convergence for backwards products of a convergent sequence of finite Markov matrices," *Stochastic Processes and their Applications*, 11, 187–192 🌐.

GALE, D. AND S. KARIV (2003): "Bayesian learning in social networks," *Games and Economic Behavior*, 45, 329–246 🌐.

GALEOTTI, A. AND S. GOYAL (2010): "The law of the few," *American Economic Review*, 100, 1468–1492 🌐.

GOLUB, B. AND M. O. JACKSON (2010): "Naïve learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2, 112–149 🌐.

——— (2012): "How homophily affects the speed of learning and best-response dynamics," *The Quarterly Journal of Economics*, 127, 1287–1338 🌐.

GRANOVETTER, M. S. (1973): "The strength of weak ties," *American Journal of Sociology*, 78, 1360–1381 🌐.

GREENE, W. H. (2008): *Econometric analysis*, Upper Saddle River, NJ: Pearson Prentice hall, 6th ed.

HORN, R. A. AND C. R. JOHNSON (1991): *Topics in matrix analysis*, New York, NY: Cambridge University Press.

IOANNIDES, Y. M. AND L. DATCHER LOURY (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42, 1056–1093 🌐.

ISAACSON, D. L. AND R. W. MADSEN (1976): *Markov chains: Theory and applications*, New York, NY: Wiley.

JACKSON, M. O. (2008): *Social and economic networks*, Princeton, NJ: Princeton University Press.

LEFEBVRE, R. C. AND A. S. BORNKESSEL (2013): "Social media as a tool in medicine: digital social networks and health," *Circulation*, 127, 1829–1836 🌐.

LEIZAROWITZ, A. (1992): "On infinite products of stochastic matrices," *Linear Algebra and its Applications*, 168, 189–219 🌐.

MEYER, C. D. (2001): *Matrix analysis and applied linear algebra*, Philadelphia, PA: American Mathematical Society.

MORETTI, E. (2011): "Social learning and peer effects in consumption: Evidence from movie sales," *The Review of Economic Studies*, 78, 356–393 🌐.

MOSSEL, E., A. SLY, AND O. TAMUZ (2015): "Strategic learning and the topology of social networks," *Econometrica*, 83, 1755–1794 🌐.

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

PAGE, L., S. BRIN, R. MOTWANI, AND T. WINOGRAD (1999): "The PageRank citation ranking: Bringing order to the Web," Tech. Rep. 1999-66, Stanford InfoLab 🌐.

PERKINS, P. (1961): "A theorem on regular matrices," *Pacific Journal of Mathematics*, 11, 1529–1533 🌐.

PHAM-GIA, T. (2004): "Bayesian inference," in *Handbook of beta distribution and its applications*, ed. by A. K. Gupta and S. Nadarajah, New York, NY: Marcel Dekker, Inc., 361–422.

ROBERT, C. P. (2007): *The Bayesian choice: From decision-theoretic foundations to computational implementation*, New York, NY: Springer Science+Business Media, 2nd ed.

SENETA, E. (1981): *Non-negative matrices and Markov chains*, New York, NY: Springer Science+Business Media, 2nd ed.

TAM, S. M. (1985): "On covariance in finite population sampling," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 34, 429–433 🌐.

Appendix

A Mathematical appendix

A.1 The Hadamard product

This section discusses shortly the Hadamard product matrix operation and some basic results related to it that are used in the present analysis.

Definition 6: The Hadamard Product

Consider matrices $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{m\times n}$ and $\mathbf{B} = [b_{ij}] \in \mathbb{C}^{m\times n}$ with $m,\ n \in \mathbb{N}^*$. The *Hadamard product*[19] of $\mathbf{A}$ and $\mathbf{B}$, denoted with $\mathbf{A} \circ \mathbf{B}$, is defined as the matrix of the scalar products of their corresponding elements

$$\mathbf{A} \circ \mathbf{B} := \left[a_{ij} b_{ij}\right]_{(i,j)\in\mathcal{M}\times\mathcal{N}} \in \mathbb{C}^{m\times n}$$

where $\mathcal{M} := \{1, \ldots, m\}$ and $\mathcal{N} := \{1, \ldots, n\}$.

Let $\mathbf{A},\ \mathbf{B},\ \mathbf{C} \in \mathbb{C}^{m\times n}$, and consider a conformable matrix of ones, $\mathbb{J}_{m\times n} := \mathbb{1}_m \mathbb{1}_n^\top$, and a scalar $\kappa \in \mathbb{C}$. The Hadamard product possesses the following properties:

[H.1]	*Commutativity*:	$\mathbf{A} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{A}$
[H.2]	*Associativity*:	$\mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C}$
[H.3]	*Distributivity*:	$\mathbf{A} \circ (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) + (\mathbf{A} \circ \mathbf{C})$
[H.4]	*Identity element* $\mathbb{J}$:	$\mathbf{A} \circ \mathbb{J}_{m\times n} = \mathbf{A}$
[H.5]	*Distributive transposition*:	$(\mathbf{A} \circ \mathbf{B})^\top = \mathbf{A}^\top \circ \mathbf{B}^\top$
[H.6]	*Compatibility with scalar multiplication*:	$\kappa(\mathbf{A} \circ \mathbf{B}) = (\kappa\mathbf{A}) \circ \mathbf{B} = \mathbf{A} \circ (\kappa\mathbf{B})$

In addition, consider vectors $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{y} \in \mathbb{C}^n$, and define the diagonal matrices $\mathbf{D_x} := \mathrm{diag}(x_1, \ldots, x_m)$ and $\mathbf{D_y} := \mathrm{diag}(y_1, \ldots, y_n)$. The following observations will be particularly useful in our analysis:

---

[19] Named after French mathematician Jacques Salomon Hadamard (1865-1963). The terms *elemntwise*, *entrywise*, or *Schur product* are also encountered in the literature.

- *Pre-multiplying* matrix $\mathbf{A}$ by a conformable diagonal matrix $\mathbf{D_x}$ multiplies every element of each *row* i of $\mathbf{A}$ by the corresponding element x_i of vector $\mathbf{x}$, that is

$$\mathbf{D_x}\,\mathbf{A} = \left[a_{ij}x_i\right]_{i\in\mathcal{M},j\in\mathcal{N}} \in \mathbb{C}^{m\times n} \tag{27}$$

- *Post-multiplying* matrix $\mathbf{A}$ by a conformable diagonal matrix $\mathbf{D_y}$ multiplies every element of each *column* j of $\mathbf{A}$ by the corresponding element y_j of vector $\mathbf{y}$, that is

$$\mathbf{A}\,\mathbf{D_y} = \left[a_{ij}y_j\right]_{i\in\mathcal{M},j\in\mathcal{N}} \in \mathbb{C}^{m\times n} \tag{28}$$

The above observations give rise to the following properties:

[H.7] [*Multiply row* i *by* x_i] $\mathbf{D_x}\,\mathbf{A} = \mathbf{A}\circ\left(\mathbf{x}\mathbb{1}_{\mathrm{m}}^{\top}\right) = \left[\left(\mathbf{x}\mathbb{1}_{\mathrm{m}}^{\top}\right)\circ\mathbb{I}_m\right]\mathbf{A}$

[H.8] [*Multiply colm.* j *by* y_j] $\mathbf{A}\,\mathbf{D_y} = \mathbf{A}\circ\left(\mathbb{1}_{\mathrm{n}}\mathbf{y}^{\top}\right) = \mathbf{A}\left[\left(\mathbb{1}_{\mathrm{n}}\mathbf{y}^{\top}\right)\circ\mathbb{I}_{\mathrm{n}}\right]$

Finally, the results below are used in the proofs in part B of this appendix.

[H.9] $\mathbf{D_x}(\mathbf{A}\circ\mathbf{B})\mathbf{D_y} = (\mathbf{D_x}\,\mathbf{A})\circ(\mathbf{B}\,\mathbf{D_y}) = (\mathbf{A}\,\mathbf{D_y})\circ(\mathbf{D_x}\,\mathbf{B}) = \mathbf{A}\circ(\mathbf{D_x}\mathbf{B}\,\mathbf{D_y})$

[H.10] $\left[(\mathbf{A}\circ\mathbf{B})\,\mathbf{y}\mathbb{1}_{\mathrm{m}}^{\top}\right]\circ\mathbb{I}_m = \left(\mathbf{A}\,\mathbf{D_y}\mathbf{B}^{\top}\right)\circ\mathbb{I}_m$

A proof of the last two statements can be found in Horn and Johnson (1991). More specifically, property [H.9] is Lemma 5.1.2 in Chapter 5, while property [H.10] follows immediately from Lemma 5.1.3 and the definition of Hadamard product.

A.2 Non-negative matrices and networks

For the sake of convenience, some standard elements of linear algebra theory are presented below.

Definition 7: Irreducible and Primitive Matrices

- A matrix $\mathbf{P}\in\{0,1\}^{n\times n}$ is called a *permutation matrix* if in each row and in each column there exists exactly one entry equal to 1, with all other entries being equal to 0.

- A matrix $\mathbf{A}\in\mathbb{C}^{n\times n}$ is said to be a *reducible* matrix if there exists a permutation matrix $\mathbf{P}$ such that

$$\mathbf{P}^{\top}\mathbf{A}\,\mathbf{P} = \begin{bmatrix} \mathbf{X}_{k\times k} & \mathbf{Y}_{k\times n\text{-}k} \\ \mathbb{O}_{n\text{-}k\times k} & \mathbf{Z}_{n\text{-}k\times n\text{-}k} \end{bmatrix} \tag{29}$$

where $\mathbf{X}, \mathbf{Z}$ are square matrices, and $\mathbb{O}$ is a matrix of zeros.

- A square matrix is called *irreducible* if it is not reducible.
- A non-negative irreducible matrix is said to be a *primitive* matrix if only one of its eigenvalues lies on its spectral circle.

Note that in some older papers, the terms *indecomposable* and *regular* are used respectively instead of irreducible and primitive. The following lemma establishes a relationship between the properties of an adjacency matrix and the structure of the network it represents.

Lemma 2: Matrix Properties and Network Structure

- A network $\mathcal{G}$ is strongly connected if and only if its adjacency matrix $\mathbf{G}$ is irreducible.
- A strongly connected network $\mathcal{G}$ is aperiodic if and only if its adjacency matrix $\mathbf{G}$ is primitive.

Proof. To prove the first statement, it suffices to show that its contrapositive holds true. In other words, it is equivalent to proving the following:

Statement [CP]. *A network $\mathcal{G} = \langle \mathcal{N}, \mathcal{E} \rangle$ is not strongly connected if and only if its adjacency matrix* $\mathbf{G}$ *is reducible.*

First, notice that pre- and post-multiplying a square matrix respectively by a permutation matrix $\mathbf{P}$ and its transpose $\mathbf{P}^\top$, reorders the rows and the columns of matrix in the same way. Hence the transformation

$$\widetilde{\mathbf{G}} := \mathbf{P}^\top \mathbf{G}\, \mathbf{P} \tag{30}$$

simply relabels the agents in $\mathcal{N}$ without essentially changing the structure of the network; matrix $\widetilde{\mathbf{G}}$ represents the same network as $\mathbf{G}$, but with the agents relabelled. More formally, transformation (30) can be seen as applying a bijection $f_N : \mathcal{N} \to \mathcal{N}$ from the set of nodes to itself, and a corresponding bijection $f_E : \mathcal{N}^2 \to \mathcal{N}^2$ with $f_E(i,j) := \big(f_N(i),\ f_N(j)\big)$ from the set of edges of $\mathcal{G}$ to itself. Except for the node labels, network $\widetilde{\mathcal{G}} = \langle f_N(\mathcal{N}),\ f_E(\mathcal{E}) \rangle$, represented by adjacency matrix $\widetilde{\mathbf{G}}$, will be identical to $\mathcal{G}$. An equivalent, therefore, to Statement [CP] is that $\widetilde{\mathbf{G}}$ is not strongly connected if and only if $\mathbf{G}$ is reducible. For convenience, denote the new labels of the nodes of the transformed network with l_i, so that $f_N(\mathcal{N}) = \{l_1, l_2, \ldots, ł_n\}$.

Notice that if **G** is reducible, then by (29) matrix **P** can chosen so that

$$\widetilde{\mathbf{G}} = \begin{bmatrix} \mathbf{X}_{k \times k} & \mathbf{Y}_{k \times n\text{-}k} \\ \mathbf{O}_{n\text{-}k \times k} & \mathbf{Z}_{n\text{-}k \times n\text{-}k} \end{bmatrix}$$

It can now easily be seen from $\widetilde{\mathbf{G}}$ that no directed path exists from any of the nodes in $\mathcal{N}_{\text{isol}} := \{l_{k+1}, l_{k+2}, \ldots, l_n\}$ to any of the nodes in $\mathcal{N}_{\text{main}} := \{l_1, l_2, \ldots, l_k\}$, since the former group pay attention only to agents within $\mathcal{N}_{\text{isol}}$. This suggests that network $\widetilde{\mathcal{G}}$, and hence network $\mathcal{G}$, are *not* strongly connected. This proves the *if* part of Statement [CP].

To prove the *only if* part of Statement [CP], assume that network $\mathcal{G}$ is not strongly connected. Then there must exist (at least) two nodes $i, j \in \mathcal{N}$ such that there are no directed paths from node i to node j. Consider the set of nodes $\mathcal{M}(j) \subset \mathcal{N}$ consisting of those and only those nodes $h \in \mathcal{N}$ such that there is a directed path from node h to node j, and denote its cardinality with m, where $1 \leq m \leq n\text{-}1$.[20] Since there is no directed path from node i to node j, it must be that $i \in \mathcal{N} \setminus \mathcal{M}$. Now consider a transformation similar to (30) that assigns labels from l_1 to l_m to the nodes in $\mathcal{M}(j)$. This can be implemented through a bijection f_N such as the one described above, with $f_N(\mathcal{M}) = \{l_1, \ldots, l_m\}$ and $f_N(\mathcal{N} \setminus \mathcal{M}) = \{l_{m+1}, \ldots, l_n\}$, together with the corresponding bijection f_E. Then the transformed matrix can be written as

$$\widetilde{\mathbf{G}} = \mathbf{P}^\top \mathbf{G}\, \mathbf{P} = \begin{bmatrix} \widetilde{\mathbf{X}}_{m \times m} & \widetilde{\mathbf{Y}}_{m \times n\text{-}m} \\ \widetilde{\mathbf{W}}_{n\text{-}m \times m} & \widetilde{\mathbf{Z}}_{n\text{-}m \times n\text{-}m} \end{bmatrix}$$

where block $\widetilde{\mathbf{X}}$ captures the edges among the nodes in $\mathcal{M}$, block $\widetilde{\mathbf{Z}}$ the edges among the nodes in $\mathcal{N} \setminus \mathcal{M}$, block $\widetilde{\mathbf{Y}}$ the edges from nodes in $\mathcal{M}$ to nodes in $\mathcal{N} \setminus \mathcal{M}$, and block $\widetilde{\mathbf{W}}$ the edges from nodes in $\mathcal{N} \setminus \mathcal{M}$ to nodes in $\mathcal{M}$. Observe, however, that there should not exist any edges from nodes outside $\mathcal{M}$ towards nodes in $\mathcal{M}$. Suppose towards a contradiction that there existed such an edge, emanating from node $q \in \mathcal{N} \setminus \mathcal{M}$. This would imply that there is a directed path from node q to node j, and hence $q \in \mathcal{M}$ by the definition of $\mathcal{M}$. Since no such edges exists, it must be that $\widetilde{\mathbf{W}} = \mathbf{O}_{n\text{-}m \times m}$, suggesting that **G** is a reducible matrix. This completes the proof of the first statement in Lemma 2.

The second statement follows directly from Theorem 1 in Perkins (1961) and the definition of an aperiodic network (see Definition 1). The statement is presented in the same form as Lemma 2 in Golub and Jackson (2010). ◇

[20]Notice that $\mathcal{M}(j)$ will be non-empty since $j \in \mathcal{M}(j)$.

A.3 The Perron-Frobenius theorem

The following statement of the Perron-Frobenius theorem,[21] is based on Meyer (2001), Chapter 8.3, and has been adapted to the context of the present paper.

Proposition 5: The Perron-Frobenius Theorem

Let $\mathbf{G} \in \mathbb{R}_+^{n \times n}$ be a non-negative, irreducible square matrix, and denote its spectral radius with ρ_G. The following statements hold true.

[PF.1] There exists a simple eigenvalue of $\mathbf{G}$ equal to ρ_G.

[PF.2] There exists a positive stochastic eigenvector corresponding to ρ_G, that is, a vector $\mathbf{p} > \mathbf{0}_n$ such that $\mathbf{G}\,\mathbf{p} = \rho_G\,\mathbf{p}$ and $\|\mathbf{p}\|_1 = 1$. This is called the *Perron vector* of $\mathbf{G}$.

[PF.3] The Perron vector is the only non-negative eigenvector of $\mathbf{G}$ up to a positive multiple.

B Proofs

B.1 Existence and uniqueness of eigenvector centrality

We begin by establishing that adjacency matrix $\mathbf{G}$, and hence its transpose, $\mathbf{G}^\top$, is non-negative and irreducible; thus the Perron-Frobenius theorem applies (see Section A.3). Non-negativity holds true by definition, since $\mathbf{G} \in \{0,1\}^{n \times n}$, while irreducibility follows from Lemma 2 and the assumption that $\mathcal{G}$ is strongly connected.

It can now be readily shown that eigenvector centrality is a well defined measure, that is, it exists and it is unique in any strongly connected network $\mathcal{G}$. To establish existence, notice that [PF.1] suggests that ρ_G will be an eigenvalue of $\mathbf{G}^\top$, and hence, by [PF.2], $\mathbf{c}$ will be the Perron vector of $\mathbf{G}^\top$. It will therefore be a positive vector, and thus meaningful as a measure of centrality, since it will not contain any negative or non-real entries. Uniqueness follows from the fact that the Perron vector is the only positive eigenvector of $\mathbf{G}$ (see [PF.3]).

[21] A first version of the theorem, applying to positive matrices, was proved by German mathematician Oskar Perron in 1907. Five years later his colleague Ferdinand Georg Frobenius showed that most of Perron's results carry over to non-negative matrices, provided that they are irreducible.

B.2 Proof of Proposition 3

Notice that network $\mathcal{G}$ is aperiodic, since it has been assumed that agents are out-neighbours of themselves (Perkins, 1961, Theorem 1). Then Lemma 2 implies that matrix $\mathbf{G}$ will be primitive. Given that, existence of the limiting vector $\mathbf{w}^{(\infty)}$, and hence existence of a (definitive) consensus can established using Proposition 2 in the present paper and the corollary of Theorem 5 in Chatterjee and Seneta (1977). Uniqueness of the consensus follows directly from Theorem 3 in the latter paper.

B.3 Proof of Theorem 1

$$
\begin{aligned}
\mathbf{W}(t) &= \prod_{\kappa=1}^{t} \boldsymbol{\Gamma}(\kappa) \\
&= \prod_{\kappa=1}^{t} \left[\left(\mathbf{G}^{\kappa}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1} \left[\mathbf{G}\circ \mathbb{1}_n\left(\mathbf{G}^{\kappa-1}\boldsymbol{\pi}(0)\right)^\top\right] \\
&= \prod_{\kappa=1}^{t} \left[\left(\mathbf{G}^{\kappa}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1} \mathbf{G} \left[\left(\mathbf{G}^{\kappa-1}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] && \text{by [H.8]} \\
&= \left[\left(\mathbf{G}^{t}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1}\mathbf{G} \left[\left(\mathbf{G}^{t-1}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] \left[\left(\mathbf{G}^{t-1}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1}\mathbf{G} \left[\left(\mathbf{G}^{t-2}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]\cdots \\
&\qquad \cdots \left[\left(\mathbf{G}^{2}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1}\mathbf{G} \left[\left(\mathbf{G}\,\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] \left[\left(\mathbf{G}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1}\mathbf{G} \left[\left(\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] \\
&= \left[\left(\mathbf{G}^{t}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1} \underbrace{\mathbf{G}\ldots\mathbf{G}}_{t-1\text{ terms}}\ \mathbf{G} \left[\left(\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] \\
&= \left[\left(\mathbf{G}^{t}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1} \mathbf{G}^t \left[\left(\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right] \\
&= \left[\left(\mathbf{G}^{t}\boldsymbol{\pi}(0)\mathbb{1}_n^\top\right)\circ \mathbb{I}_n\right]^{-1} \mathbf{G}^t\, \mathbf{D}_{\boldsymbol{\pi}(0)} && (31)
\end{aligned}
$$

where we have used the properties of Hadamard product discussed in Section A.1 of the Appendix, and $\mathbf{D}_{\boldsymbol{\pi}(0)} := \text{diag}(\pi_1(0), \pi_2(0), \ldots, \pi_n(0))$ is a diagonal matrix with the elements of vector $\boldsymbol{\pi}(0)$ on its main diagonal.

Proposition 6

Let $\mathcal{G}$ be a strongly connected, aperiodic network with adjacency matrix $\mathbf{G}$. Denote its spectral radius by ρ_G, and its Perron vector by $\mathbf{p}$. Then

$$\lim_{t \to +\infty} \left(\frac{\mathbf{G}}{\rho_G} \right)^t = \frac{\mathbf{p}\,\mathbf{c}^\top}{\mathbf{c}^\top \mathbf{p}} \tag{32}$$

Proof. We know that a network $\mathcal{G}$ is aperiodic, and hence strongly connected, if and only if its adjacency matrix $\mathbf{G}$ is primitive, and hence irreducible (Lemma 2). Recall that eigenvector centrality is defined as the left eigenvector of $\mathbf{G}$ (Definition 2). Then (32) follows directly from the theorem on primitive matrices and expression (8.3.10) in Meyer (2001). ◇

Now we can use Proposition 6 to obtain an expression for the social influence of the agents. From (32) it follows that

$$\begin{aligned}
\lim_{t \to +\infty} \mathbf{W}(t) &= \lim_{t \to +\infty} \left\{ \left[\left(\left(\frac{\mathbf{G}}{\rho_G} \right)^t \boldsymbol{\pi}(0)\, \mathbb{1}_n^\top \right) \circ \mathbb{I}_n \right]^{-1} \left(\frac{\mathbf{G}}{\rho_G} \right)^t \mathbf{D}_{\boldsymbol{\pi}(0)} \right\} \\
&= \left[\left(\lim_{t \to +\infty} \left(\frac{\mathbf{G}}{\rho_G} \right)^t \boldsymbol{\pi}(0)\, \mathbb{1}_n^\top \right) \circ \mathbb{I}_n \right]^{-1} \lim_{t \to +\infty} \left(\frac{\mathbf{G}}{\rho_G} \right)^t \mathbf{D}_{\boldsymbol{\pi}(0)} \\
&= \left[\left(\frac{\mathbf{p}\,\mathbf{c}^\top}{\mathbf{c}^\top \mathbf{p}} \boldsymbol{\pi}(0)\, \mathbb{1}_n^\top \right) \circ \mathbb{I}_n \right]^{-1} \frac{\mathbf{p}\,\mathbf{c}^\top}{\mathbf{c}^\top \mathbf{p}} \mathbf{D}_{\boldsymbol{\pi}(0)} \\
&= \left[\left(\mathbf{p}\,\mathbf{c}^\top \boldsymbol{\pi}(0)\, \mathbb{1}_n^\top \right) \circ \mathbb{I}_n \right]^{-1} \mathbf{p}\,\mathbf{c}^\top \mathbf{D}_{\boldsymbol{\pi}(0)} && \text{by [H.6]} \\
&= \left(\mathbf{c}^\top \boldsymbol{\pi}(0) \right)^{-1} \left[\left(\mathbf{p} \mathbb{1}_n^\top \right) \circ \mathbb{I}_n \right]^{-1} \mathbf{p}\,\mathbf{c}^\top \mathbf{D}_{\boldsymbol{\pi}(0)} && \text{by [H.6]} \\
&= \left(\mathbf{c}^\top \boldsymbol{\pi}(0) \right)^{-1} \mathbf{D}_{\mathbf{p}}^{-1} \mathbf{p}\,\mathbf{c}^\top \mathbf{D}_{\boldsymbol{\pi}(0)} \\
&= \left(\mathbf{c}^\top \boldsymbol{\pi}(0) \right)^{-1} \mathbb{1}_n \mathbf{c}^\top \mathbf{D}_{\boldsymbol{\pi}(0)} \\
&= \left(\mathbf{c}^\top \boldsymbol{\pi}(0) \right)^{-1} \mathbb{1}_n \left(\mathbf{c}^\top \circ \boldsymbol{\pi}(0)^\top \right) && \text{by [H.8]} \\
&= \mathbb{1}_n \left(\mathbf{c}^\top \boldsymbol{\pi}(0) \right)^{-1} \left(\mathbf{c} \circ \boldsymbol{\pi}(0) \right)^\top && \text{by [H.5]} \\
&= \mathbb{1}_n \left(\frac{\mathbb{1}_n^\top \boldsymbol{\pi}(0)}{\mathbf{c}^\top \boldsymbol{\pi}(0)} \right) \left(\mathbf{c} \circ \frac{\boldsymbol{\pi}(0)}{\mathbb{1}_n^\top \boldsymbol{\pi}(0)} \right)^\top && \text{by [H.6]} \\
&= \mathbb{1}_n\, \alpha_{\mathbf{c},\boldsymbol{\pi}} \left(\mathbf{c} \circ \widetilde{\boldsymbol{\pi}}(0) \right)^\top
\end{aligned}$$

where $\mathbf{D}_{\pi(0)} := \mathrm{diag}(\boldsymbol{\pi}(0))$, $\alpha_{\mathbf{c},\boldsymbol{\pi}} := \frac{\mathbb{1}^\top \boldsymbol{\pi}(0)}{\mathbf{c}^\top \boldsymbol{\pi}(0)} = \frac{\sum_{i=1}^n \pi_i(0)}{\sum_{i=1}^n c_i \pi_i(0)}$ is a scalar that captures the effects of the network on social influence, and $\widetilde{\boldsymbol{\pi}}(0) := \frac{\boldsymbol{\pi}(0)}{\mathbb{1}^\top \boldsymbol{\pi}(0)}$ is the vector of relative initial precisions of the agents in network $\mathcal{G}$.

B.4 Proof of Corollary 1

Consider any change in the agents' initial expertise, from $\boldsymbol{\pi}(0)$ to $\boldsymbol{\pi}'(0)$ that does not affect relative precisions $\widetilde{\boldsymbol{\pi}}(0)$. Then we can write $\boldsymbol{\pi}'(0) = \kappa \boldsymbol{\pi}(0)$ for some $\kappa > 0$. From expression (21), the new social influence of agent i is

$$w_i' = \alpha_{\mathbf{c},\kappa\boldsymbol{\pi}} c_i \widetilde{\pi}_i(0) = \frac{\sum_{j=1}^n \kappa \pi_j(0)}{\sum_{j=1}^n c_j \kappa \pi_j(0)} c_i \widetilde{\pi}_i(0) = \frac{\sum_{j=1}^n \pi_j(0)}{\sum_{j=1}^n c_j \pi_j(0)} c_i \widetilde{\pi}_i(0) = \alpha_{\mathbf{c},\boldsymbol{\pi}} c_i \widetilde{\pi}_i(0) = w_i.$$

B.5 Proof of Corollary 2

Notice that in this case the sampling fraction relevant for calculating covariance is the entire network population. Then the standard formula for covariance (see, for example, Tam, 1985) yields

$$\begin{aligned}
\mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)] &= \frac{1}{n} \sum_{i=1}^n \left(c_i - \frac{1}{n} \sum_{j=1}^n c_j \right) \left(\pi_i(0) - \sum_{j=1}^n \pi_j(0) \right) \\
&= \frac{1}{n} \left[\sum_{i=1}^n c_i \pi_i - \frac{1}{n} \sum_{i=1}^n \pi_i - \frac{1}{n} \sum_{i=1}^n \pi_i(0) + n \frac{1}{n^2} \sum_{i=1}^n \pi_i(0) \right] \\
&= \frac{1}{n} \left[\sum_{i=1}^n c_i \pi_i - \frac{1}{n} \sum_{i=1}^n \pi_i \right], \qquad (33)
\end{aligned}$$

where the second equation holds true since $\sum_{i=1}^n c_i = 1$. Then from (33) it follows that

$$\mathrm{Cov}[\mathbf{c},\ \boldsymbol{\pi}(0)] > 0 \ \Leftrightarrow \ \sum_{i=1}^n c_i \pi_i > \frac{1}{n} \sum_{i=1}^n \pi_i > 0 \ \Leftrightarrow \ \frac{1}{n} \frac{\sum_{i=1}^n \pi_i}{\sum_{i=1}^n c_i \pi_i} < 1 \ \Leftrightarrow \ \bar{\alpha}_{\mathbf{c},\boldsymbol{\pi}^0} < 1.$$

B.6 Proof of Lemma 1

Recall from expression (20) that the consensus belief will be

$$b^{(\infty)} = \sum_{i=1}^n w_i^{(\infty)} b_i(0)$$

The *ex ante* expectation of the consensus belief (that is, before the signals are realised, and the prior beliefs are formed) will be

$$
\begin{aligned}
\mathsf{E}\Big[b^{(\infty)}\Big] &= \mathsf{E}\left[\sum_{i=1}^{n} w_i^{(\infty)} b_i(0)\right] \\
&= \sum_{i=1}^{n} w_i^{(\infty)} \mathsf{E}[b_i(0)] \\
&= \sum_{i=1}^{n} w_i^{(\infty)} b^{(\infty)} \\
&= b^{(\infty)}.
\end{aligned}
$$

B.7 Proof of Proposition 4

From Theorem 1, the *ex ante* variance of the consensus belief $b^{(\infty)}$ is given by

$$
\begin{aligned}
\mathsf{Var}\Big[b^{(\infty)}\Big] &= \mathsf{Var}\left[\sum_{j=1}^{n} w_j^{(\infty)} b_j(0)\right] \\
&= \alpha_{\mathbf{c},\boldsymbol{\pi}}^2 \sum_{j=1}^{n} c_j^2 (\widetilde{\pi}_j(0))^2 \mathsf{Var}\Big[b_j(0)\Big] \\
&= \left(\frac{\sum_{j=1}^{n} \pi_j(0)}{\sum_{j=1}^{n} c_j \pi_j(0)}\right)^2 \sum_{j=1}^{n} c_j^2 \left(\frac{\pi_j(0)}{\sum_{j=1}^{n} \pi_j(0)}\right)^2 \mathsf{Var}\Big[b_j(0)\Big] \\
\mathsf{Var}\Big[b^{(\infty)}\Big] &= \frac{1}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2} \sum_{j=1}^{n} c_j^2 [\pi_j(0)]^2 \mathsf{Var}\Big[b_j(0)\Big] .
\end{aligned}
\tag{34}
$$

Let us consider first the case where $s_i \sim N\left(\theta^*, \; \pi_i(0)^{-1}\right)$. Then for each initial belief it holds that $\mathsf{Var}\Big[b^{(\infty)}\Big] = \frac{1}{\pi(0)}$, and hence it follows from expression (34) above that

$$
\begin{aligned}
\mathsf{Var}\Big[b^{(\infty)}\Big] &= \frac{1}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2} \sum_{j=1}^{n} c_j^2 [\widetilde{\pi}_j(0)]^2 \frac{1}{\pi(0)} \\
\mathsf{Var}\Big[b^{(\infty)}\Big] &= \frac{\sum_{j=1}^{n} c_j^2 \pi_j(0)}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2}.
\end{aligned}
\tag{35}
$$

We can now calculate how $\mathsf{Var}\Big[b^{(\infty)}\Big]$ changes with $\pi_i(0)$:

$$
\begin{aligned}
\frac{d\mathsf{Var}\Big[b^{(\infty)}\Big]}{d\pi_i(0)} &= \frac{d}{d\pi_i(0)} \left[\frac{\sum_{j=1}^{n} c_j^2 \pi_j(0)}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2}\right] \\
&= \frac{1}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^4} \left[c_i^2 \left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2 - 2c_i \sum_{j=1}^{n} c_j \pi_j(0) \sum_{j=1}^{n} c_j^2 \pi_j(0)\right]
\end{aligned}
$$

$$= \frac{1}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^3} \left[c_i \sum_{j=1}^{n} c_j \pi_j(0) - 2 \sum_{j=1}^{n} c_j^2 \pi_j(0) \right] \tag{36}$$

Then using (36) we can show that

$$\frac{d\mathsf{Var}\left[b^{(\infty)}\right]}{d\pi_i(0)} > 0 \Leftrightarrow c_i > 2 \frac{\sum_{j=1}^{n} c_j^2 \pi_j(0)}{\sum_{j=1}^{n} c_j \pi_j(0)},$$

which gives condition (26) if the signals are normally distributed.

Consider now the case with $s_i \sim Bin(\pi_i(0),\ \theta^*)$. It will then hold that $\pi_i(0) := q_i + r_i$, where q_i are the successes and r_i the failures in the observations drawn. Assume that based on this signal, the initial beliefs are formed as

$$b_i(0) := \frac{q_i}{q_i + r_i} = \frac{q_i}{\pi_i(0)}.$$

If the expertise $\pi_i(0)$ of the agents (the "sample size") is known, then it follows that

$$\mathsf{Var}[b_i(0)] := \mathsf{Var}\left[\frac{q_i}{q_i + r_i}\right] = \frac{1}{\pi_i(0)} \mathsf{Var}[q_i] = \frac{\theta^*(1-\theta^*)}{\pi_i(0)}.$$

Substituting this into expression (34) gives

$$\mathsf{Var}\left[b^{(\infty)}\right] = \theta^*(1-\theta^*) \frac{\sum_{j=1}^{n} c_j^2 \pi_j(0)}{\left(\sum_{j=1}^{n} c_j \pi_j(0)\right)^2}. \tag{37}$$

The rest of the proof is similar to the case with $s_i \sim N\left(\theta^*,\ \frac{1}{\pi_i(0)}\right)$, since expression (37) is proportional to (35).