Social Learning and Media

Maksat Abasov

In partial fulfillment of the requirements
for the degree of Master of Arts

Supervisor: Professor Rosario Nunzio Mantegna

Economics Department
Central European University
Hungary

08 June 2016

Acknowledgments

I would like to express my sincere gratitude to Professor Rosario Mantegna for his constant guidance and frequent feedback, and to Luca Marotta, PhD for his help during the coursework and thesis discussions. Also, I would like to thank my friends and relatives for supporting me during my two years in Budapest.

Abstract

This paper builds theoretical models of social learning under assumption that agents are subject to different biases. The first part of the paper examines the effect of such biases on the consensus and convergence speed. I find that consensus must be reached in such societies, whereas convergence speed, is distorted significantly by these biases. In the second part, I build a model of dynamic media communication and find that correlation in beliefs in the society leads to a long-term slanting behavior of media outlets.

Contents

1 Introduction

Decision making of any kind requires individuals and societies gather information. People we communicate with and the media we pay attention to are two major sources of such information. It has been shown recently in a wide range of literature that biases related to human limited cognitive abilities are inevitable in processing this information (Rabin and Schrag (1999), Klayman (1995)). The goal of this paper is to develop a tractable model to study the implications of these biases on social communication and information aggregation dynamics. It particularly assumes heuristically biased agents connected in the social network and media outlets that use this bias in a strategic way to attract maximum attention. The focus of the paper is on the dynamics of belief formation and how the media affects it.

As a motivating example, consider a group of people who need to assess the state of the economy in order to make savings decisions. Each of them has a prior belief about the current state of economy. These beliefs might have been formed from the signals that people observe at work, e.g. the number of people hired or fired, or in the streets, e.g. the number of expensive cars or the number of homeless people. Agents communicate with their friends, neighbors, and relatives about their views.

The crucial assumption here is that the communication is not strategic, as it might complicate the analysis, and the truthful communication assumption fits well in the examples provided throughout the paper. In addition, agents can pay attention to media outlets' messages, e.g. TV channels, newspapers, or bloggers. In the rest of the paper, for simplicity, I will use TV channels as a leading example, but the results extend to other media sources as well. Channels are assumed to be private independent enterprises, aiming at maximizing profits by attracting people's attention. This can be interpreted as channels getting profits by selling advertisement time, for which they can charge a price proportionate to their audience and the amount of time this audience devotes for a given channel. I do not consider cases, although it might be quite interesting, when media is used to achieve other goals, e.g. influencing political views in authoritarian regimes. An agent's preferences for media are assumed to be irrational, particularly, agents pay attention to the news which are in line with or closest to their own beliefs. I will refer to this phenomenon as confirmation bias, although with a slight abuse of terminology as this irrationality need not be based only on psychology of human nature. People may underweight the messages not in line with their own beliefs simply due to their educational level or experience, and may just fail to understand argumentation for the message. For example, it might be easier for one to believe that the rich or migrants are guilty for a bad economic situation than to understand a Central Bank's report on the state of the economy. The next assumption is that agents are biased only to media messages, but not to their friends and neighbors. This is partly done for the purposes of tractability, and partly captures the intuition that it is easier for a person to be convinced by another person whom she knows, rather than by a TV news.

Finally, media is able to observe the objective news and manipulate, or slant it in order to attract more attention and increase profits. Following the example used in Mullainathan and Shleifer (2005), I consider media receiving information on annual inflation rate at the level of 6.3%. Slanting this number is quite easy for any purposes, and the exact message can have a wide range. One extreme could be a message about an economic disaster due to a persistently high inflation rate, and the other could tell us about a recovering economy due to slowing down of inflation growth compared to a year before.

The first strand of literature this paper relates to is the theoretical models of learning and information aggregation in social networks. There are two broad groups of learning processes described in the literature: sophisticated and naive learning. Sophisticated learning includes different settings of Bayesian learning (Banerjee (1992), Bala and Goyal (2001), Gale and Kariv (2003)), "streams" model (Acemoglu et al. (2014)), models of strategic communication (Acemoglu et al. (2010)). These models assume rationality of the agents which requires strong cognitive abilities from them. For instance, the benchmark model of Bayesian learning assumes people are capable of Bayesian updating, and Acemoglu et al. (2014) assume that agents are capable of memorizing beliefs of all the agents in a social network. On the contrary, naive learning assumes that agents use some simple heuristic rule to update their own beliefs. In DeGroot (1974) they update beliefs sequentially by taking averages of their neighbors' opinions. DeMarzo et al. (2001) indicate that the DeGroot process creates so called persuasion bias in communication, meaning that agents do not account for repetitions of the same messages disclosed earlier. With slight modification of the original model, they also show convergence results for the quite general setting of social network, followed by Golub and Jackson (2010) showing that naive communication asymptotically leads to efficient aggregation in the absence of influential agents. This result is contrary to that of the sophisticated learning process as influential agents in Bayesian setting intensify efficient aggregation (Acemoglu et al. (2010)).

The described theoretical models have been tested and there is convincing evidence that social networks in fact propagate learning. The evidence comes from many kinds of sources, mainly from the lab and field experiments. Papers compare the baseline DeGroot model with the Bayesian learning model (Chandrasekhar et al. (2015), Grimm and Mengel (2014)) and with "streams" model (Mobius et al. (2015)) in which information can be tagged to its origin. Although results are opposing each other, the consensus belief is that in environments where "tagging" and Bayesian inference is difficult due to cognitive requirements, agents may exhibit DeGroot-type learning.

The other distinct strand of literature I refer to is the literature on media. DellaVigna and Gentzkow (2009) provide an overview of empirical evidence on effects of persuasive communication, including the effect of media on beliefs of population. Most of the works they cite show the significant effect of media news on the opinions of the population. Using a field experiment, Gerber et al. (2007) find that television advertisements had a strong effect on the favorability for different candidates during the gubernatorial campaign in Texas. DellaVigna and Kaplan (2006) find that in 2000

US presidential elections, when Fox News TV channel was not yet introduced in all the cities, exposure to Fox News had a significant positive effect for Republican votes. Using a similar method, Enikolopov et al. (2010) find that the areas with exposure to opposition channels during the 1999 parliamentary election in Russia showed significantly larger support for the anti-government party. From the evidence provided it is clear that the media's message has an impact in political debate and there are few reasons to believe that the media has no effect on other issues.

Apart from this evidence, it has been shown, both in academic literature and general publications, that media sources themselves follow the beliefs distribution, implying that changes in beliefs affect the messages media present. Gentzkow and Shapiro (2010) find that US newspaper readers have a significant preference for "like-minded news", and firms respond strongly to these preferences slanting news towards consumer beliefs. Mullainathan and Shleifer (2005) overviews other evidence of such effect and explains it using human irrational tendency not to give credit to beliefs inconsistent with their own (Rabin and Schrag (1999)) and that people seek information that confirm their beliefs (Klayman (1995)). They build a theoretical model which studies the implications of such behavior on equilibrium media strategy and find that heterogeneous beliefs among the population lead the media to distort the information, providing biased news, even in competitive markets.

This paper is different from the previous literature in that it combines two topics on communication which enables to extending the range of questions. First, it extends the naive learning literature by endogenizing beliefs of the media hubs and allowing them to act strategically. Second, the current literature on media, and Mullainathan and Shleifer (2005) particularly, study temporary equilibria with fixed beliefs distribution and focus on how media slants. Since there is evidence that the slanting affects the beliefs distribution, this paper relaxes the assumption of fixed distribution and puts beliefs dynamics into perspective. The most natural way to do it, in my view, is to locate agents into the social network, which enables tracking beliefs through time and additionally study the effects of network structure and beliefs distribution across a network on social learning. The contribution of this paper is that it makes the first (as far as I know) effort to bring together models of social learning and media behavior. Also, the paper looks at dynamic rather than static consequences of media slanting the news. Finally, the paper shows the effect of social network topology and correlated beliefs on the social learning process and media behavior.

2 Social learning with biases

In this chapter I present models of social learning described by different biases, and analyze their effect on convergence properties of the network.

Model. n agents are connected through social network $G(N)$ which is characterized by a directed unweighted adjacency matrix N. Each of them is endowed with initial belief $b_i^{(0)}$, where i indexes agents, and the upper script denotes the period. Thus, initial beliefs form a vector $b^{(0)} \in \mathbb{R}^n$, beliefs in period t are denoted as $b^{(t)}$. Dynamics of the learning process is as follows: in each period agents communicate with their neighbors and update their own beliefs by taking some weighted average of the beliefs of their neighbors. Each agent assigns weights to her neighbors, and each neighbor discloses her beliefs truthfully. Let us denote the matrix of weights as $T^{(t)}$. Then, the system can be described as the following iterative process:

$$b^{(t+1)} = T^{(t)} b^{(t)}$$

Each element of the vector $b^{(t+1)}$ is a weighted average of beliefs in the vector $b^{(t+1)}$. Therefore, the rows of matrix $T^{(t)}$ should sum up to 1, and matrix $T^{(t)}$ should be a (row) stochastic matrix. I will also refer to this matrix as a listening matrix, as it conveys the structure of the information flow. In order to define the process one only needs to define the rule, according to which the weighting stochastic matrix is going to be determined. For example, in the DeGroot model, the weighting matrix is constant. In what follows I will discuss two processes with different matrix generation rules. In the first model, the weights are determined endogenously and they depend on the vector of beliefs. Every agent will puts more weight on her own beliefs and beliefs close to her own. I will refer to this model as process of confirmatory biased learning. In the second model, I assume that agents in every period interact only with a random subset of their neighbors. Particularly, any link is be activated with probability $p < 1$, creating frictions in communication. I will refer to this model as a model of imperfect diffusion learning process.

In the following sections I describe the two models, and analyze their convergence under conditions similar to those of the DeGroot process[1]. First, I consider a simple two agents network, then using intuition obtained in small model I analyze the more general case of n agents. Finally, I examine the effect of different network characteristics on the speed of convergence. The exact value of the consensus and the speed of convergence go beyond the scope. However, some properties of speed of convergence are shown using simulations method.

Now, let me first introduce some definitions. I use interchangeably the concepts of convergence and consensus. Both mean that all beliefs in the network converge to some unique value: $b_i^t \to b^*$ for all i. The diameter of the network is the length of the longest directed path. It connects the most far distant agents in the network. I define the set of links, or link structure, as $E = \{(i, j) : i \text{ listens to } j\}$. Note that the link structure does not contain information on particular weights, but weights depend on the link structure: $T_{ij} > 0$ if and only if $(i, j) \in E$. The listening structure can be also represented by a directed matrix Q

2.1 Confirmatory bias

In the confirmatory bias model I assume that the agents put higher weight on messages closer to their own beliefs than on those who are farther. To formalize this, I use the

[1]DeGroot process reaches consensus if it is applied to a strongly connected directed network

following weighting function:

$$w_{ij} = \frac{\exp(-c(b_i - b_j)^2)}{1 + \sum_{k \in N_i} \exp(-c(b_i - b_k)^2)}$$
$$w_{ii} = \frac{1}{1 + \sum_{k \in N_i} \exp(-c(b_i - b_k)^2)}$$

where N_i denotes the set of agent i's neighbors, and c is a confirmatory bias parameter. The higher the parameter is, the less weight the agent puts on the opinions which are farther. The function $f(x) = \exp(-cx^2)$ attains its maximal value 1, when $x = 0$, thus agents put the highest weights on their own beliefs.

The numerator of the function is bell shaped, so that it is convex in some neighborhood of agent's belief, concave farther away and relatively flat even farther. This function conveys the intuition behind the confirmatory bias. The denominator is used only to normalize weights to 1. Another assumption is that agents put a certain weight on themselves. I assume that agents put the highest weight on their own beliefs, and model it by setting the numerator in selfweight to be $1 \geq \exp(-c * x^2)$ for any $x \in [-1, 1]$.

Combined, these weights form a matrix $W(b)$, which depends on the belief vector. The model then can be summarized by the following iterative nonlinear dynamical process process:

$$b^{(t+1)} = W(b^{(t)})b^{(t)}$$

Note that the process is deterministic, so, given the initial beliefs vector, the solution to this process should be unique[2], and can be obtained by iteratively applying the formula above.

Two agents case

First, let me consider the simplest case where the system consists of two nodes. This simple case illustrates the main logic of the proof for the general case.

[2]Solution to a dynamic system is a sequence of values of elements of the process. In this case, th solution is a sequence of agents' beliefs

The model is based on the network of two agents. In order for this system to be strongly connected, both agents should have directed links to each other, so Q is a 2×2 matrix of ones. Let $\{b_1^{(t)}\}_1^\infty$ and $\{b_2^{(t)}\}_1^\infty$ be the sequences of first and second agents' beliefs. Initial beliefs are given by $b_1^{(0)}$ and $b_2^{(0)}$. Then, the system evolves according to a rule:

$$b_1^{(t+1)} = \frac{1}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_1^{(t)} + \frac{\exp(-c(b_1^{(t)}-b_2^{(t)})^2)}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_2^{(t)}$$
$$b_2^{(t+1)} = \frac{\exp(-c(b_1^{(t)}-b_2^{(t)})^2)}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_1^{(t)} + \frac{1}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_2^{(t)}$$

Proposition 1. *Two-agents confirmatory bias model converges to consensus, moreover, consensus belief is* $\frac{b_1^{(0)}+b_2^{(0)}}{2}$

Proof. Without loss of generality let $b_1^{(0)} \geq b_2^{(0)}$. Then, for any t the relation $b_1^{(t)} \geq b_2^{(t)}$ will hold as well. This is true because both are the weighted averages of each other. Let $\Delta_{(t)} = b_1^{(t)} - b_2^{(t)}$ be the difference between the two beliefs in period t. Then, the following expression describes the updating rule for the distances in each period:

$$\begin{aligned} d_{(t+1)} &= \frac{1-\exp(-c(b_1^{(t)}-b_2^{(t)})^2)}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_1^{(t)} \\ &+ \frac{\exp(-c(b_1^{(t)}-b_2^{(t)})^2)-1}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)} b_2^{(t)} \\ &= \frac{1-\exp(-c\Delta_{(t)}^2)}{1+\exp(-c\Delta_{(t)}^2)} \Delta_{(t)} \end{aligned}$$

Note that:

$$\frac{1-\exp(-c\Delta_{(t)}^2)}{1+\exp(-c\Delta_{(t)}^2)} < 1$$

Thus, distances decrease for each t: $\Delta_{(t+1)} < \Delta_{(t)}$, so that $\exp(-c\Delta_{(t+1)}^2) < \exp(-c\Delta_{(t)}^2)$. This implies that distances decrease even faster in each consecutive period:

$$\frac{1-\exp(-c(b_1^{(t+1)}-b_2^{(t+1)})^2)}{1+\exp(-c(b_1^{(t+1)}-b_2^{(t+1)})^2)} < \frac{1-\exp(-c(b_1^{(t)}-b_2^{(t)})^2)}{1+\exp(-c(b_1^{(t)}-b_2^{(t)})^2)}$$

Then, for any t:

$$1 > a \equiv \frac{1 - \exp(-c\Delta^2_{(0)})}{1 + \exp(-c\Delta^2_{(0)})} > \frac{1 - \exp(-c\Delta^2_{(t)})}{1 + \exp(-c\Delta^2_{(t)})}$$

So, the sequence of distances is bounded from above by a linear process:

$$\tilde{\Delta}_{(t+1)} = a\tilde{\Delta}_{(t)}; \qquad\qquad 0 < a < 1$$

which converges to 0. Thus:

$$\lim_{t\to\infty} \Delta_{(t)} = 0$$

Then, the original system of beliefs converges to consensus as the distances vanish.

Note also that

$$|b_1^{(t+1)} - b_1^{(t)}| = |b_2^{(t+1)} - b_2^{(t)}| = \frac{\exp(-c(b_1^{(t)} - b_2^{(t)})^2)}{1 + \exp(-c(b_1^{(t)} - b_2^{(t)})^2)} d_{(t)}$$

meaning that in each period beliefs move towards each other by the same distance, so the point $\frac{b_1^{(0)}+b_2^{(0)}}{2}$ belongs to $[b_2^{(t)} + b_1^{(t)}]$ for any t. Thus, by nested intervals theorem, $b_2^{(t)}, b_1^{(t)} \to \frac{b_1^{(0)}+b_2^{(0)}}{2}$

□

General case

Let us now turn to more general case of directed strongly connected network of size n. Let d be the network's diameter. The process evolves according to a rule $b^{(t+1)} = W(b^{(t)})b^{(t)}$, where $W(b) : [0,1]^n \to M^{n\times n}$ is a weighting function, returning a stochastic matrix of size $n \times n$. As each row of the resulting matrix contains nonnegative elements summing up to one, one iteration of the process maps some n-cube $[b_{min}, b_{max}]^n$ into itself: $W(b)b \subset [b_{min}, b_{max}]^n$.

Let me first recall some useful results from previous research without proofs.

Definition 1. *Let T be a stochastic matrix of size n with elements T_{ij}. Then the contraction coefficient is defined as follows:*

$$\tau(T) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |T_{is} - T_{js}|$$

Note that $0 \leq \tau(T) \leq 1$. Equality in the left holds if all the rows of the matrix T are identical[3]. Equality in the right holds if the matrix contains orthogonal rows, or put another way, rows for which the sets of positive entries do not intersect.

The following are Theorem 3.5 and Lemma 4.3 from Seneta (2006):

Theorem 1. *Let x be an arbitrary vector and T a stochastic matrix. If $y = Tx$ then:*

$$\max_{i,i'} |y_i - y_{i'}| \leq \tau(T) \max_{j,j'} |x_j - x_{j'}|$$

Lemma 1. *For any stochastic T_1, T_2:*

$$\tau(T_1 T_2) \leq \tau(T_1)\tau(T_2)$$

Restated version of the main result in DeGroot (1974), Theorem 2:

Theorem 2. *If all the nodes in the network with stochastic adjacency matrix are strongly connected and the network is aperiodic, then a consensus is reached.*

Note that the models presented in this paper assume strong connectedness. Aperiodicity is obtained as I allow selfloops.

Now let us turn to the analysis addressed in this paper.

Definition 2. *For given stochastic matrix T, reweighted matrix is a stochastic matrix T^w such that $T^w_{ij} > 0$ if and only if $T_{ij} > 0$, but $T^w_{ij} = T_{ij}$ does not need to hold.*

It follows that any reweight induces a network with the same link structure as the original one, and changes only the weights of each link. The fact is used in the proof of the following lemma:

[3] Note that this matrix would converge in 1 iteration

Lemma 2. *Let T be a convergent stochastic matrix and d be the diameter of the network induced by T. Let $T^{(1)}, T^{(2)}, \ldots, T^{(d)}$ be the sequence of reweights. Then $T^{(1:d)} \equiv T^{(1)}T^{(2)} \ldots T^{(d)}$ is convergent and has all its entries positive.*

Proof. Recall that stochastic matrix T contains information on link structure of the induced network. This means that agent i listens to all the agents j for whom $T_{ij} > 0$. It follows that in two rounds of communication, agent i listens indirectly to beliefs of her neighbors' neighbors, i.e. those whom each agent $j \in N_i$ listened during first communication. As reweighted matrices preserve the link structure of the original matrix, the same is true for the product of reweights of the matrix T: $(T'T'')_{ij} > 0$ if and only if i listens to some k, who happens to listen to j. In d rounds of communication message from any agent in the network should reach all the other agents, implying that $T_{ij}^{(1:d)} > 0$ for any i and j. By Theorem 2 $T^{(1:d)}$ is convergent.

□

Theorem 3. *Let $G(Q)$ be a directed and strongly connected social network with diameter d. Then given any initial beliefs, it reaches consensus.*

Proof. First, note that $W(b)$ is convergent for any vector b. This is true as the network induced by the matrix is strongly connected and has selfloops. Also, for any b' and b'', $W(b')$ and $W(b'')$ are reweighted matrices of each other and hence diameters are equal. Thus, using Lemma 2 we deduce that $W^{(d-1:0)} = W(b^{(d)})W(b^{d-1}) \ldots W(b^{(1)})W(b^{(0)})$ contains only positive elements, which means that $\tau(W) < 1$.

Using Theorem 1:

$$\Delta_{(d)} \leq \tau(W^{(d-1:0)})\Delta_{(0)}$$

implying $\Delta_d < \Delta_0$

Now note that there is a lower bound on the weight that any agent can put on the belief of any of her neighbors. Consider an agent i and her neighbor j such that $b_i - b_j \leq 1$. Agent i puts minimal weight on j if all the other neighbors also communicate belief b_i: $b_i - b_l = 0$ for all $l \in N_i$. And there can be no smaller weight on j's belief if agent i is connected to everybody in the network, and all $n-2$ other agents hold the same belief as i does. Then, the lower bound on the other agent's belief in the network with $\Delta_{(0)} = \max(b^{(0)}) - \min(b^{(0)}) \leq 1$ is:

$$w^{min}(\Delta_{(0)}) = \frac{\exp(-c\Delta^2_{(0)})}{n-2+exp(-c\Delta^2_{(0)})}$$

Next note that as vector valued function $W(b)$ maps to itself, $\max(b^{(0)})$ should be a lower bound for all the sequence of beliefs. However, using the results of previous paragraph we can narrow this upper bound down. By the period d, due to strong connectedness of the network, any agent i will indirectly receive the message from any other agent j with a strictly positive weight $W^{(d-1:0)}_{ij}$. The lower bound on $W^{(d-1:0)}_{ij}$ is obtained by assuming that agents are separated by the path of length d and that each agent in the path applies lower bound weight $w^{min}(\Delta_{(0)})$. Then, the weight $W^{(d-1:0)}_{ij}$ is bounded by $w^{min}(\Delta_{(0)})^d$. Assume further that only maximal belief changed in period d and only by the amount of lower bound weight. This assumption is in fact too soft. The change in the range of beliefs will certainly be larger, but even in this case we can see that

$$\tau(W^{(d-1:0)}) < \max(b^{(0)})\frac{1-w^{min}(\Delta_{(0)})^d}{1+w^{min}(\Delta_{(0)})^d} - \min(b^{(0)}) < \frac{1-w^{min}(\Delta_{(0)})^d}{1+w^{min}(\Delta_{(0)})^d}\Delta_{(0)}) < 1$$

This is true as we assumed that the agent with belief $\max(b^{(0)})$ put the lower bound weight on agent with belief $\min(b^{(0)})$ only. So that the value $\sum_{s=1}^{n} |T_{is} - T_{js}|$ is maximized when considering these exact agents. Iterating the process further it follows that $\tau(W^{(kd-1:(k-1)d}) < \frac{1-w^{min}(\Delta_{(0)})^d}{1+w^{min}(\Delta_{(0)})^d}\Delta_{((k-1)d)})$ for any positive integer k.

It was shown above that $\Delta_{(d)} < \Delta_{(0)}$, then for any positive integer k, the value of $\tau(W^{(kd-1:(k-1)d})$ will be bounded by $\frac{1-w^{min}(\Delta_{(0)})^d}{1+w^{min}(\Delta_{(0)})^d}\Delta_{(0)}) < 1$.

Now it only remains to define the sequence $\tilde{\Delta}_0 = \tilde{\Delta}_1 = \ldots = \tilde{\Delta}_{d-1}$, $\tilde{\Delta}_t = \frac{1-w^{min}(\Delta_{(0)})^d}{1+w^{min}(\Delta_{(0)})^d}\tilde{\Delta}_{t-d+1}$. The sequence converges to 0 as $\tau(W^{(d-1:0)}) < 1$. It bounds the original sequence $\Delta_{(t)}$, implying that $\Delta_{(t)} \to 0$ as well, proving consensus.

□

Simulations

Now let us turn to the analysis of the speed of convergence and the characteristics of the network that affect the speed. Precise analytic derivations of related issues go beyond the scope of this paper and seem mathematically challenging. Instead, I use simulations method in order to examine different effects on the speed of convergence, but first, I illustrate an example of the process and contrast it with the DeGroot process in Example 1.

Example 1. Figure 1 depicts the exact values of the beliefs dynamics for 30 agents strongly connected in Erdos-Renyi type network with 250 links. I set the confirmatory parameter to be $c = 30$ and iterate the process with the initial beliefs to be uniformly random on $[0, 1]$. In the figure two different processes with the same underlying network and the same initial beliefs are plotted. Every green line represents the belief evolution of a particular agent in confirmatory bias process, while the grey ones plot the beliefs of the DeGroot type agents.

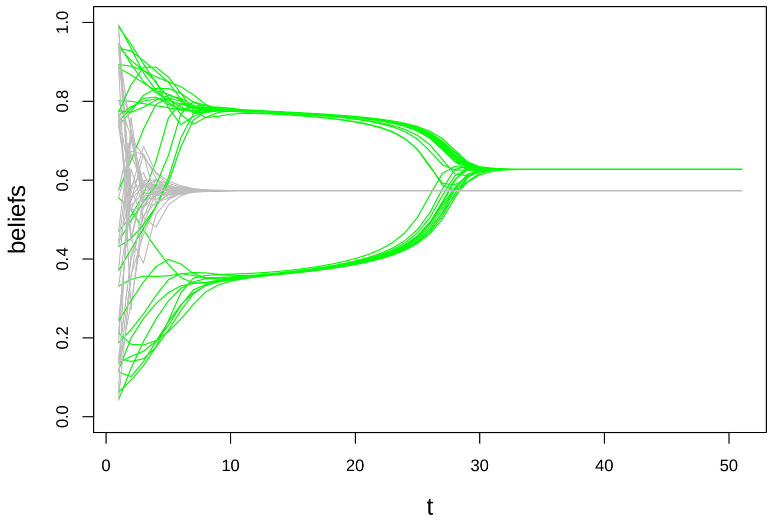


Figure 1: Confirmatory biased and DeGroot processes

The following observations are of a quite interest:

- Learning under confirmatory bias converges much slower.
- Confirmatory bias first forces agents to gather in clusters of similar beliefs, and then the clusters converge between each other.
- Consensus beliefs are different for both processes.

The next question we might be interested in relates to the speed of convergence. First natural thing to check is the invariance of speed for the initial belief vector as it is for the DeGroot process. For this, I consider the same network as in the Example 1 and simulate the process by changing the vector of initial beliefs. It turns out that the speed of the time of convergence is quite sensitive to the vector of initial beliefs. Figure 2 plots a log-log scaled frequencies of the number of periods it took for the network to converge. Number of simulations was 10000. In every repetition I simulated a random initial belief vector, iterated the system, and fixed the number of periods in which beliefs reached consensus.

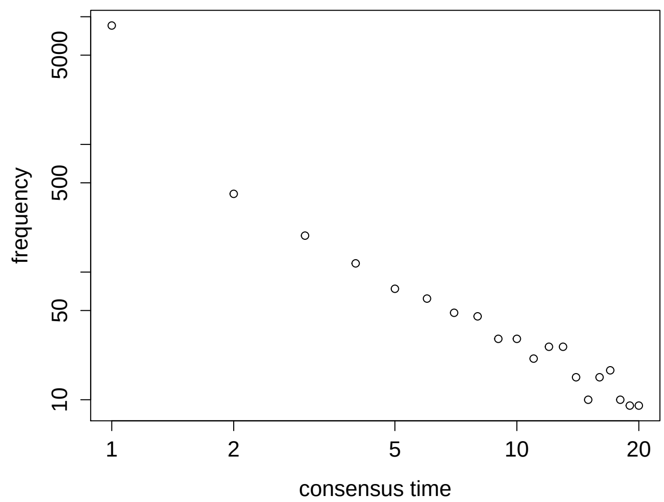


Figure 2: Simulations results

What is interesting here is that the log-frequencies follow a power law. This particularly mean that most of times system converges in some reasonable number of periods, but with a quite low frequency convergence time might become extremely large. This fact shows the sensitivity of the result for small disturbances in the initial belief as everything in the network is fixed except for initial beliefs, which are being changed slightly explode the consensus time.

Next, in order to get more sense of what is happening I run broader simulations of the process. I fix the size of the network $n = 30$, then vary the values of number of links, correlation of beliefs and the confirmatory bias parameter and look at their effect on the speed of convergence, running around 176.5 thousands of simulations. I trace two measures of convergence speed: the percent of observations in which consensus was achieved in 20 iterations and the average number of iterations it took for the process to reach consensus. I present the results of these simulations in the Figure 3.

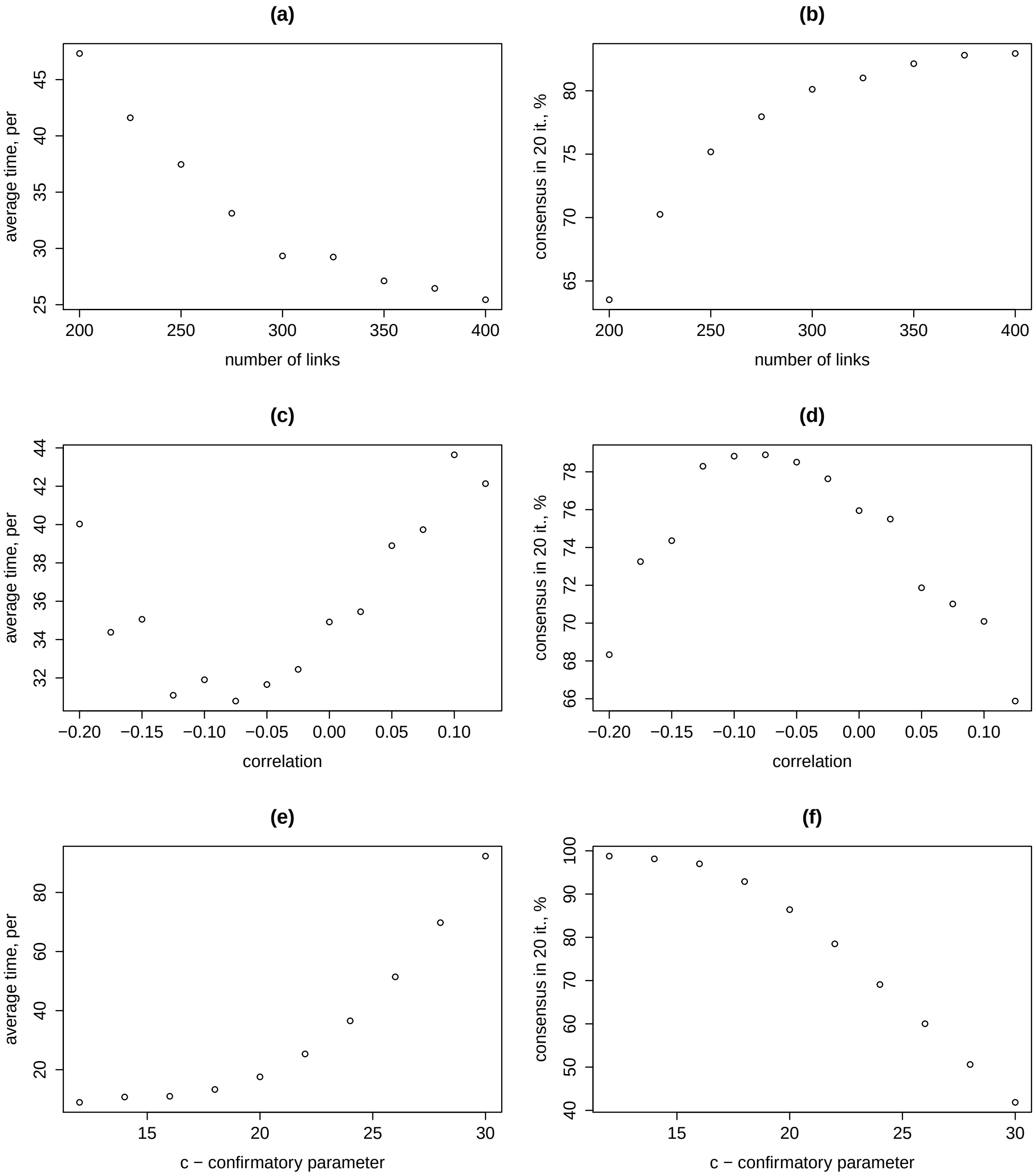


Figure 3: Simulations results

Upper two subfigures (a) and (b) show the effect of number of links in the network on (a) average consensus time, and (b) percent of consensus in first 20 periods. The monotonic effect is evident on both graphs, so that we can say that the number of links has a positive effect on convergence.

Subfigures (c) and (d) in the middle depict the effect of the correlation of initial beliefs beliefs. To measure it I consider the set E of all links. Then fro each $(i, j) \in E$ I create the pair (b_i, b_j) and prescribe values to two different vectors, first number in the pair to the first vector, second number to the second vector. Roughly speaking, the first vector is the vector of beliefs of agents-receivers and the second is the vector of beliefs of the agents, who send their beliefs. Correlation is computed for these two vectors, and we see that the effect on convergence is non-monotonic, generally increasing in absolute value of correlation.

Subfigures (e) and (f) depict the negative effect of confirmatory parameter. The more agents are biased the slower the network converges.

Next, I combine all the effects to assess their significance and quantify them while controlling for other effects I run the following regressions:

	Model 1	Model 2
(Intercept)	-24.21^{***}	1.21^{***}
	(1.24)	(0.01)
# of links	-0.11^{***}	0.00^{***}
	(0.00)	(0.00)
correlation	34.03^{***}	-0.27^{***}
	(5.06)	(0.02)
correlation2	116.40^{*}	-1.03^{***}
	(48.85)	(0.21)
c	4.34^{***}	-0.03^{***}
	(0.03)	(0.00)
R^2	0.09	0.21
Adj. R^2	0.09	0.21
Num. obs.	176411	176411

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Determinants of speed of convergence

Model 1 regresses consensus time and Model 2 - the probability of reaching consensus in 20 iterations. In order to control for non-monotonicity of convergence in correlation, I add the square of its measure, correlation2. Coefficients are of different signs as I expected (the more periods it takes to converge the less percent of times should the network converge in less than 20 periods). Thus, controlling for all the effects, all of them remain statistically significant with Model 2 also having quite high predictive power.

2.2 Imperfect diffusion

For imperfect diffusion process I assume that each period agents communicate with random set of their neighbors. Particularly, each neighbor in every period is listened with the probability p, independently for any agent and period. As in the DeGroot model agent i prescribe weight of T_{ij} to agent j's opinion, but the difference is that it is only done when in a particular period agent i turns out to listen to j. If some neighbors are not listened their weights are added to agent i's selfweight.

In order to describe the process formally, note that with this setup, each link occurance accounts for different combination of links. There are 2^m of such specifications. Let Ω be the set containing them, and $T^{(i)}$ be a random draw from this set. The probability of draw of a particular matrix is $p^a(1-p)^{a'}$, where a is a number of active links and a' is a number of nonactive links. Thus, the process evolves according to the following rule:

$$b^{(t+1)} = T^{(t)}b^{(t)}$$

Due to the randomness of the process, solution is not unique and will depend on the particular realization of listening matrices in each period. The goal for this chapter is to check whether this solution converges to single value for all the agents under conditions similar to those for DeGroot process convergence as well as to examine effects of different network characteristics on convergence speed.

Two agents case

As for two-agents case for confirmatory biased system, here also the listening matrix contains links from agent 1 to agent 2 and from agent 2 to agent 1, meaning that all the entries of T are positive. This ensures that:

$$\tau(T) < 1$$

Then, the iterative (random) solution to a system is:

$$b^{(t+1)} = T^{(t)}T^{(t-1)} \dots T^{(0)}b^{(0)} \equiv \tilde{T}_n^{(t)}b^{(0)}$$

Proposition 2. *Two-agents imperfect diffusion model converges to consensus.*

Proof. According to Theorem 1:

$$\max(b^{(t+1)}) - \min(b^{(t+1)}) \leq \tau(\tilde{T}_n^{(t)})(\max(b^{(0)}) - \min(b^{(0)})) \tag{1}$$

Now consider the product of two random matrices $T'T''$ such that one of them is formed with link from agent 1 to agent 2 invoked and the other one with link from agent 2 to agent 1 active. This product will also provide all-positive-entries matrix with $\tau(T'T'') < 1$.

Now deconstruct $\tilde{T}_n^{(t)}$ into the product of matrices by singling out matrices T, $T'T''$, $T''T'$. As $t \to \infty$ the probability that number of occurances of these particular three matrices is finite is equal to 0, thus there should be infinite number of occurances with probability one. And thus:

$$\tau(\tilde{T}_n^{(t)}) \leq \tau(T^{(t)}) \dots \tau(T) \dots \tau(T'T'') \dots$$

meaning that:

$$\tau(\tilde{T}_n^{(t)}) \to 0$$

which combined with (1) proves convergence of two agent model for any initial beliefs vector.

□

General case

Theorem 4. *A directed strongly connected social network of size n and with m links evolves as imperfect diffusion process. Then given any initial beliefs, it reaches consensus with probability 1.*

Proof. The process can be rewritten as:

$$b^{(t)} = T^{(t)}T^{(n-1)} \dots T^{(1)}T^{(0)}b^{(0)}$$

For a given n let us decompose the matrix product in the right hand side in the following way: $S^{(0)} = T^{(n-1)} \dots T^{(0)}$, $S^{(i)} = T^{(n(i+1)-1)}T^{(n(i+1)-1)} \dots T^{(ni)}$. Then

$$b^{(ni)} = S^{(i-1)} \dots S^{(0)}b^{(0)} \equiv S_0^{i-1}b^{(0)}$$

Note that the probability that a full matrix T is contained in the product $S^{(i)}$ is $(1-(1-p)^m)^n$ (where $(1-p)^m$ is a probability that one given random matrix is not T, and the whole expression is the probability of the converse event for the i.i.d. sequence of such matrices.

Now we are interested in the probability of convergence:

$$Pr(\max(b^{(t)}) - \min(b^{(t)}) \to 0)$$

By the Theorem 1:

$$\max(b^{(ni)}) - \min(b^{(ni)}) \leq \tau(S_0^{i-1})(\max(b^{(0)}) - \min(b^{(0)}))$$

Thus if $\tau(S^{i-1})(\max(b^{(0)}) - \min(b^{(0)})) \to 0$ for $i \to \infty$, it follows that $\max(b^{(t)}) - \min(b^{(t)}) \to 0$, which implies:

$$\begin{aligned}
Pr(\max(b^{(t)}) - \min(b^{(t)}) \to 0) &\geq Pr(\tau(S^{i-1})(\max(b^{(0)}) - \min(b^{(0)})) \to 0) \\
&\geq Pr(\prod \tau(S^{(j+r)} \dots S^{(j+1)}S^{(j)})(\max(b^{(0)}) - \min(b^{(0)})) \to 0) \\
&= Pr(\prod \tau(S^{(j+r)} \dots S^{(j+1)}S^{(j)}) \to 0)
\end{aligned}$$

where the second inequality is due to Lemma 1 $\tau(S_0^{i-1}) \leq \prod_{j=1}^{i-1} \tau(S^{(j)})$.

Now it remains to prove that $\tau(S^{i-1})(\max(b^{(0)}) - \min(b^{(0)})) \to 0$ for $i \to \infty$. This is true if there exists a subset of natural numbers A such that for each i in A: $\tau(S^{(j+r)} \dots S^{(j+1)}S^{(j)}) < 1$, as then it should be the case that $\tau(S^i \leq \prod_{j=1}^{i-1} \tau(S^{(j)}) \to \infty$. Such A exists, to see this recall that $S^{(i)} = T^{(n(i+1))}T^{(n(i+1)-1)} \dots T^{(ni)}$. The probability that matrix T is contained in the product is $A = (1-(1-p)^m)^n$, thus for $r > d$ with positive probability $S^{(j+r)} \ldots S^{(j+1)}S^{(j)}$ can contain at least d matrices T, ensuring that the product has only positive entries, and thus $\tau(S^{(j+r)} \ldots S^{(j+1)}S^{(j)}) < 1$. And as it has positive probability, in the infinite sequence of such products it will appear infinitely many times as well. Thus:

$$\begin{aligned} Pr(\prod \tau(S^{(j+r)} \ldots S^{(j+1)}S^{(j)}) \to 0) \\ &\geq Pr(\cap_{i\in A}\{\tau(S^{(j+r)} \ldots S^{(j+1)}S^{(j)}) < 1\}) \\ &= \prod_{i\in A} Pr(\tau(S^{(j+r)} \ldots S^{(j+1)}S^{(j)}) < 1) = 1 \end{aligned}$$

Then, it implies that

$$Pr(\max(b^{(t)}) - \min(b^{(t)}) \to 0) = 1$$

or that beliefs reach consensus with probability 1.

□

Simulations

In this section I illustrate the process and run the same simulations as in previous sections. The parameters of the model are: $p = 0.2$, $n = 30$, $m = 250$

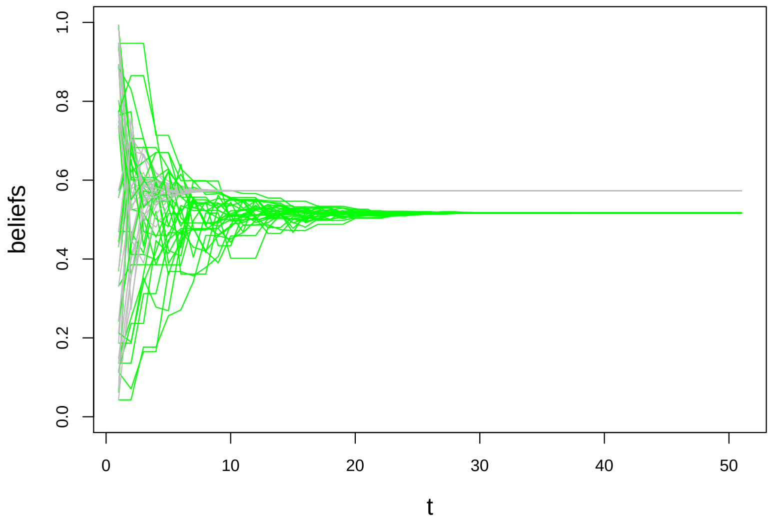


Figure 4: Confirmatory biased and DeGroot processes

Figure illustrates following observations:

- Learning under confirmatory bias converges much slower.
- Confirmatory bias first forces agents to gather in clusters of similar beliefs, and then the clusters converge between each other.
- Consensus beliefs are different for both processes.

Now I ask the same questions for the imperfect diffusion learning and run simulations as previously. First, I check invariance of speed for the initial beliefs vector. For this I consider the same network as in the Example 2 and simulate process by changing the vector of initial beliefs. It turns out that the speed or the time is not very sensitive to the vector of initial beliefs. To see this I plot the graph of frequencies for the described simulation. Figure 5 show the graph. Figure 6, however shows how randomness of the process itself with fixed initial belief can affect the consensus beliefs. In both cases I present distributions measured in frequencies.

(10 000 repetitions)

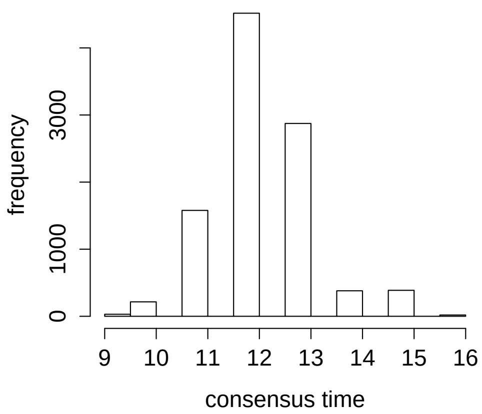


Figure 5: Cons. time distribution

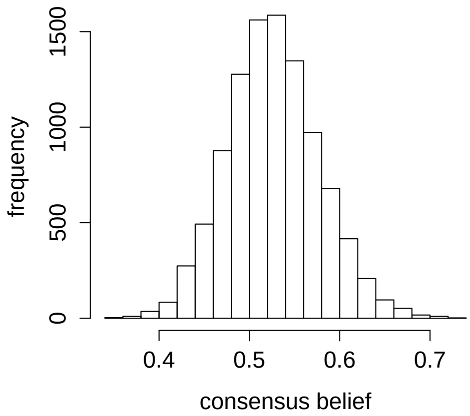


Figure 6: Cons. beliefs distribution

Next for the same parameters as previously, I run simulations for imperfect diffusion case

The following simulations and regression do not provide new, worthwile results in comparison to the confirmatory bias, except for having better predictive power.

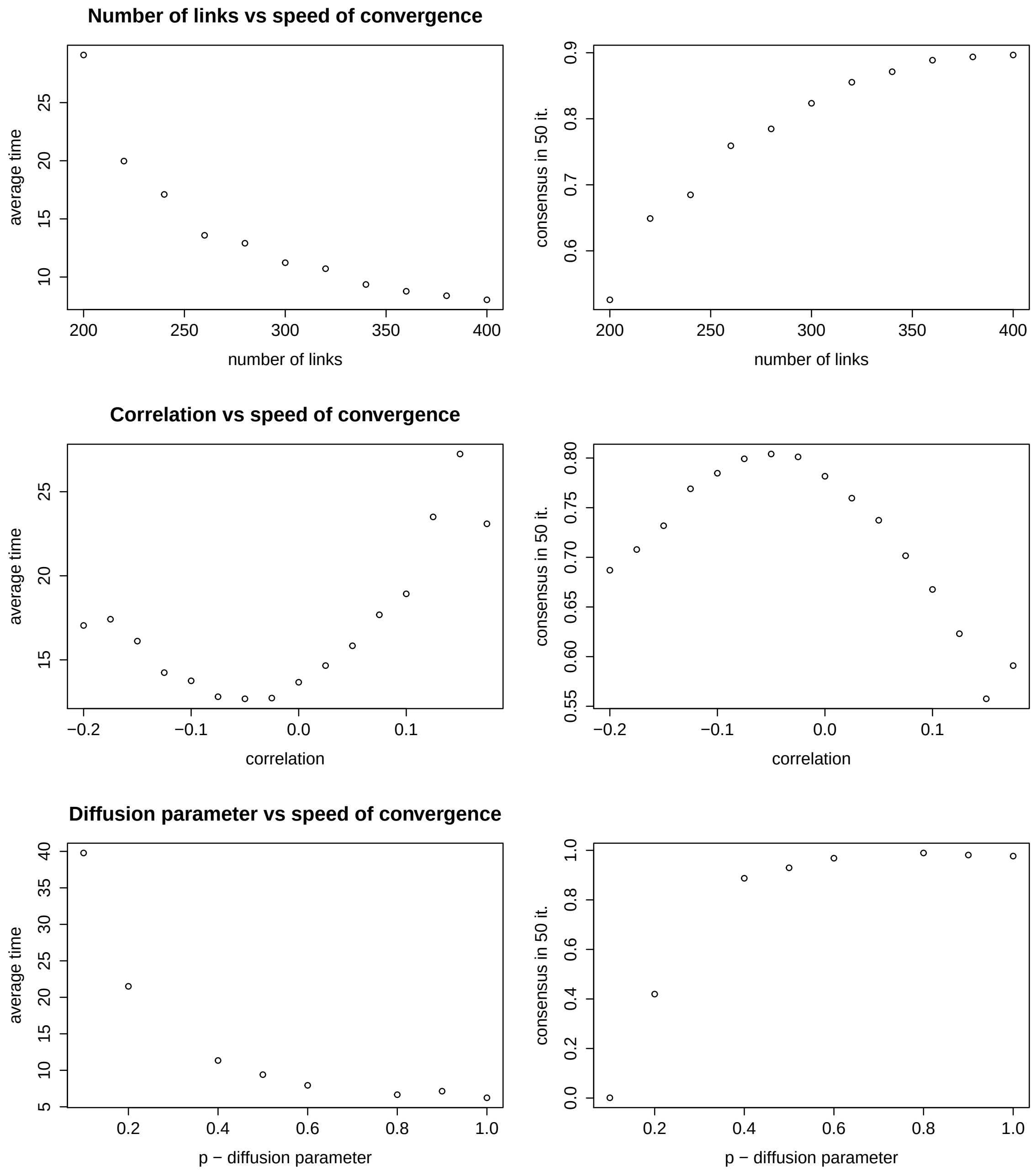


Figure 7: Simulations results

	Model 1	Model 2
(Intercept)	53.97^{***}	-0.19^{***}
	(0.24)	(0.00)
# of links	-0.08^{***}	0.00^{***}
	(0.00)	(0.00)
correlation	3.00^{**}	-0.08^{***}
	(1.14)	(0.02)
corr2	1.51	-0.35^{*}
	(11.08)	(0.16)
p	-27.01^{***}	0.86^{***}
	(0.16)	(0.00)
R^2	0.17	0.43
Adj. R^2	0.17	0.43
Num. obs.	219891	219891

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Determinants of speed of convergence

3 Media communication

As in previous chapter there is a set of n agents endowed with initial noisy beliefs x_i^0 for $i = 1, ..., n$ about underlying state of the world θ. Their goal is to accurately assess the state of the world. Agents are connected through a directed network $G(T)$ described by a (row) stochastic adjacency matrix T, where $T_{ij} > 0$ if and only if there is a directed link from agent i to agent j, meaning that agent i can pass information to agent j. The network is used to share the messages. The weight of each link (i, j) means the attention agent i puts on the message received from agent j. As the matrix T is stochastic, each agents attention to her neighbors sums up to one and is used to weight their beliefs.

Additionally to standard model, there is also a set M of TV channels, whom agents listen as well. Channels are profit maximizers with a profit function of aggregate attention they get from agents. This can be thought of as profits from advertisements. I assume that at any period they can observe the distribution of agents' beliefs and that they can adjust, or "slant" news distorting it without any costs.

The weights that agents put on the messages of both media and neighbors depend on the number of neighbors as well as the composition of messages the media provide. I assume that agents are confirmatory biased towards channels' messages. This means that an agent first chooses channel whose announced message is closest to her belief, and then the rest of the attention she prescribes to each neighbor according to their initial weights. If agent i chooses to listen to channel j, then she puts the following weight to the message of j

$$w_{i,j} = u(x_i, m_j) = \begin{cases} \bar{u} - c(x_i - m_j)^2 & \text{if } \bar{u} - c(x_i - m_j)^2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

where x_i^t is agent i's belief and m_j^t is channel j's message, and parameter c captures the how prone the agents are to confirmatory bias. The higher is c, the more selective are agents towards the messages farther from their own beliefs. The intuition of putting a weight on media is that attention is costly for agents, be it time, cable fees, or news paper price. The remaining weight of $1 - \bar{u} + c(x_i^t - m_j^t)^2$ is distributed among the agent's neighbors. Thus, if agent i communicates with agent k, then the weight she puts on agent k's belief is $[1 - \bar{u} + c(x_i^t - m_j^t)^2]T_{i,k}$.

The channel's preferences are formed by aggregate attention that it gets from the agents, i.e. the sum of the weights that agents put on the message from that channel:

$$\pi_j = \sum_{ilistenstoj} w_{i,j}$$

This can be interpreted as media getting more viewers' time and can show more adds, increasing revenues. The channel's set of actions is any signal $m_k \in R$, that is a channel can say whatever it wants to say. Alternatively, it can be understood as if media gets a continuous set of signals and shows only those that they want.

Timing of the model is as follows. In the beginning of each period channels announce their messages. Agents listen both to their neighbors and one of channels, so first they decide on which channel to listen to and how much weight to put on its message. Next they reweight their attention to neighbors according to initial weights. For any agent each neighbor she communicates to truthfully discloses her belief. At the end of the period agents update their beliefs and media receive their profits.

I consider three distinct features that might be interesting in analysis of social learning. First, most papers in the field assume initial seed of information to be exogenous, which might itself be a wrong assumption to start with, as most probably, the information seeds are correlated across network. I depart from these complications and seek for equilibrium dynamics for cases of exogenous and endogenous initial seed. Further I call them correlated and uncorrelated initial beliefs. Second, following Mobius et al. (2015) and Acemoglu et al. (2010), I assume that communication with neighbors is imperfect, meaning that at each period agent listens to a random segment of neighbors. Every agent listens to any of her neighbors with probability $p \in (0, 1)$. Thus the number of neighbors whom an agent listens follows binomial distribution. Agents also listen to exactly one of the channels. Third, I consider different competition structure in media market. I let the market to be monopolistic and duopolistic and seek for effect of competition on learning.

In order for the model to be tractable and provide intuition for discussed cases I consider a simplified version of general model described earlier. I maintain two major assumptions: specific topology and agents' attention parameters.

The network of size n consists of two equally sized strongly connected cliques. Cliques are connected to each other through m links, half of which are directed from first clique to the second, and the other half is directed opposite way. Strong connectedness of the cliques ensures that left on their own cliques converge to consensus through the time. Weak connectedness in both direction between cliques ensures global consensus, although convergence can be quite slow for low values of m. The division into cliques is done so that the correlation in beliefs across agents would be intuitive. If the initial seed of beliefs will be distributed with different ranges for different cliques, the correlation in beliefs between neighbors will be significantly positive. A stylized example of such network is is depicted in Figure 8. Cliques are colored differently and connected with few links.

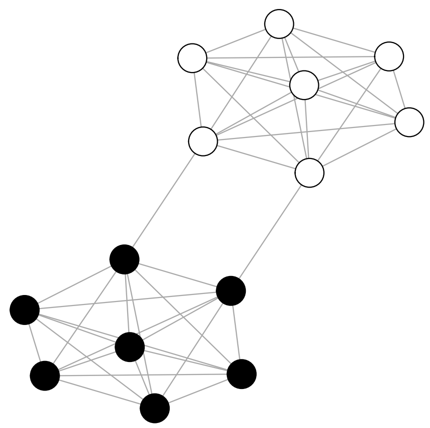


Figure 8: Stylized network

The next assumption concerns the parameters of agents utility function for media messages. I assume that $\bar{u} \geq \frac{c(1-2\epsilon)^2}{16}$. This ensures that first, media can cover the whole range of one clique beliefs, and second, if ϵ is large enough, the media will not have incentives to cover news for both cliques at a time.

The way competition work is ... media is listened every period

I further present the results for considered cases: uncorrelated and correlated beliefs. For each case I consider different media market structure. Also, I present the results for imperfect diffusion model for all the cases discussed in one section. Finally, I first present the simulations of the described process, and than the derivations to support the simulated results.

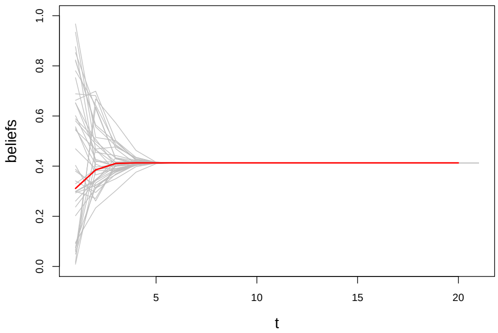
Simulations

All the simulation presented further are modeled on the network of size $n = 40$, with each clique of size 20 Erdos-Renyi type topology subgraphs. Cliques are connected by 40 links, connecting random agents from both cliques. Belief dynamics is traced for 20 periods in perfect diffusion cases and for 30 periods due to slower convergence speed.

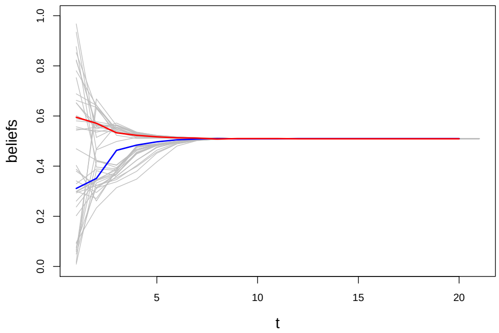
Case 1. Uncorrelated beliefs

To model uncorrelated beliefs I assume that the initial beliefs are iid across all the agents. In the simulations I model these beliefs to be uniformly distributed on $[0, 1]$.

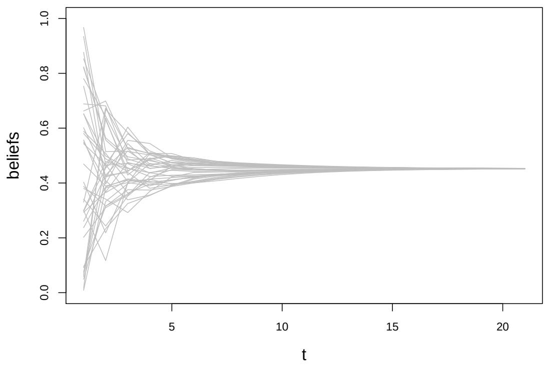
Figure 9 presents simulations for three media market structure cases. In figure 9a monopolist media maximized its aggregate attention in each period by choosing respective message. Beliefs of agents are depicted in grey and media message is in red. In Figure 9b the same is done for duopolistic media, with a competing channel being plotted in blue. Figure 9c is used a reference communication dynamics in the case of no media.



(a) monopolistic media



(b) duopolistic media



(c) without media

Figure 9: Simulations for uncorrelated beliefs

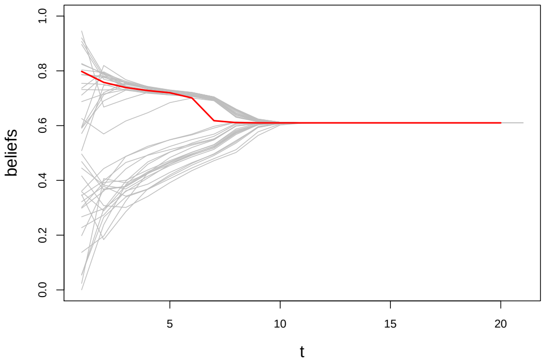
There are three notable observations in these simulations:

- convergence with any type of media market structure is faster than without media
- consensus bias is small relative to no media consensus
- no significant difference between monopoly and duopoly cases

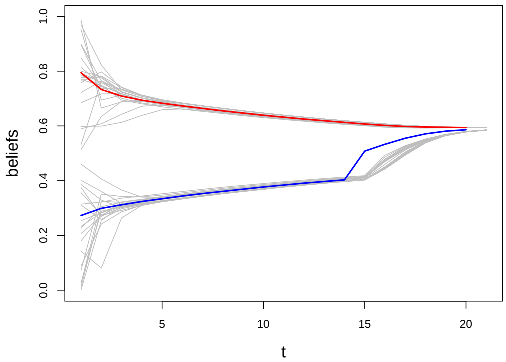
Case 2. Correlated beliefs

TO model correlated beliefs I assume, without loss of generality, that the the beliefs are distributed uniformly among each clique, but the range of beliefs in the first clique is $[0, 1/2 - \epsilon)$ and $(1/2 + \epsilon, 1]$ in the second clique. This ensures the correlation in beliefs across agents and can be a stylized model of geographic differences in beliefs across real life populations.

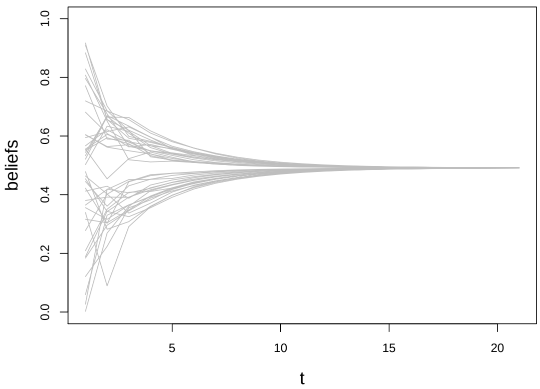
Figure 10 presents the learning simulations for the same model as previously. The main observation we can make looking at the figures is that in contrast to previous examples the learning is much slower in the model with competitive media. Shortly, competition decreases speed of convergence and increases slant throughout the whole process. We can also observe, that monopolistic setting converges much faster than reference model without media.



(a) monopolistic media



(b) duopolistic media

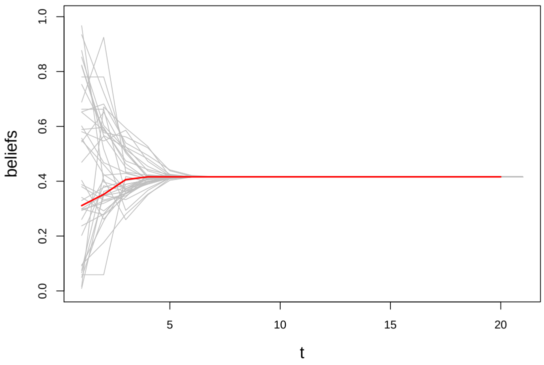


(c) without media

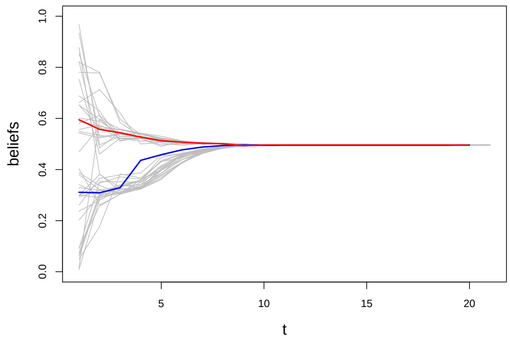
Figure 10: Simulations for correlated beliefs

Case 3. Imperfect diffusion

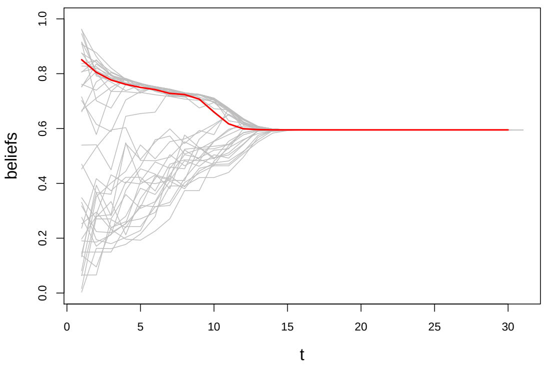
Figure 11 presents the simulations of belief dynamics for a setting where diffusion is not perfect with p, probability of communicating with a specific neighbor equal to 0.5 for the same network described previously, but traced for 30 periods. The main observation is that apart from slower convergence due to diffusion imperfection, all the results are similar to perfect diffusion case, meaning that diffusion only affects the speed of convergence and do not affect slanting strategies.



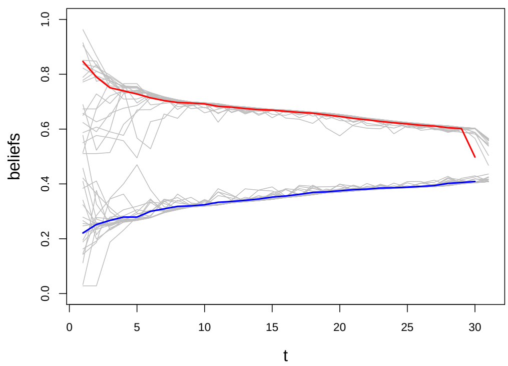
(a) monopolistic media, uncorrelated beliefs



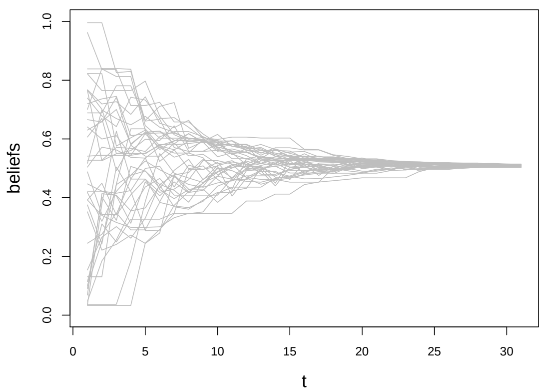
(b) duopolistic media, uncorrelated beliefs



(c) monopolistic media, correlated beliefs



(d) duopolistic media, correlated beliefs



(e) without media

Figure 11: Simulations for correlated beliefs

4 Conclusions

This paper analyzes social learning processes under assumption of significant biases in human behavior. The analysis is motivated by the previous literature which provides evidence of such biases. I examined the effects of those biases on social learning process and addressed the issues related to reaching consensus and the speed of convergence to find how learning under these assumptions is different from previous results.

I presented three major specifications of models with biases. First specification is the model where agents are confirmatory biased, irrationally paying more attention to opinions which are more in line with their own. Second model assumes that people do not communicate to the whole set of their neighbors each period, but rather they listen to a random share. In the third specification I endogenize the behavior of the influential agents, media outlets, and let them use agents' confirmatory bias for news to extract profit.

The first set of results of this paper is the generality of the convergence in strongly connected networks. Although biases cause strong frictions to the learning process, the analysis shows that networks reach consensus. Particular pattern of convergence, however, depends on the assumptions we apply. Also, different determinants of the speed of convergence were examined. The part of results are obtained by simulations method. They particularly show that the speed of convergence is significantly and strongly affected by the number of links, correlation of beliefs among population, by the level confirmatory biasedness, and by the frequency of interpersonal communication. Thus, the learning converges faster for highly connected networks which do not show evidence of any kind of homophily (which leads to correlations in beliefs).

The second set of results is aimed to provide intuition on real world communication. It assumes that people learn both from their friends and colleagues as well as from media news. Key assumption is that the agents are confirmatory biased towards media news and try to cherry pick channels which communicate the news more in line with their own beliefs. Recent literature on this topic viewed the model as static. And the accepted result is that media slants when it faces heterogeneity in population's beliefs. My contribution is that by combining the network structure and the media interaction, I study the dynamics of this process. It particularly shows the importance of correlation of beliefs. Simulations has shown that media bias is not as serious issue for a society with uncorrelated beliefs, as they converge quite fast, on the other hand, if correlation is high, system is most probably will suffer from long-term media bias and high inattention of large parts of society to each other.

Finally, my approach attempts to bring network arguments into main body of economic research. Certainly, there are many drawbacks in the layout, assumptions and derivations in these approach. I plan to investigate this issues more broadly in my future research.

References

Acemoglu, D., Bimpikis, K., and Ozdaglar, A. (2014). Dynamics of information exchange in endogenous social networks. *Theoretical Economics*, 9(1):41–97.

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

Bala, V. and Goyal, S. (2001). Conformism and diversity under social learning. *Economic theory*, 17(1):101–120.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817.

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

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

DellaVigna, S. and Gentzkow, M. (2009). Persuasion: empirical evidence. Technical report, National Bureau of Economic Research.

DellaVigna, S. and Kaplan, E. (2006). The fox news effect: Media bias and voting. Technical report, National Bureau of Economic Research.

DeMarzo, P. M., Zwiebel, J., and Vayanos, D. (2001). Persuasion bias, social influence, and uni-dimensional opinions. *Social Influence, and Uni-Dimensional Opinions (November 2001). MIT Sloan Working Paper*, (4339-01).

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2010). Media and political persuasion: Evidence from russia. *American Economic Review, forthcoming.*

Gale, D. and Kariv, S. (2003). Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Gerber, A., Gimpel, J. G., Green, D. P., and Shaw, D. R. (2007). The influence of television and radio advertising on candidate evaluations: Results from a large scale randomized experiment. *Unpublished paper, Yale University*.

Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, pages 112–149.

Grimm, V. and Mengel, F. (2014). An experiment on belief formation in networks. *Available at SSRN 2361007*.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, 32:385–418.

Mobius, M., Phan, T., and Szeidl, A. (2015). Treasure hunt: Social learning in the field. Technical report, National Bureau of Economic Research.

Mullainathan, S. and Shleifer, A. (2005). The market for news. *American Economic Review*, pages 1031–1053.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *Quarterly journal of Economics*, pages 37–82.

Seneta, E. (2006). *Non-negative matrices and Markov chains*. Springer Science & Business Media.