Chapter 4

Naïve Learning in Social Networks with Structural Variability: From Consensus to Truth

4.1 Introduction

One of the subjects that attracts attention of social scientists is how individual beliefs evolve in social networks. A well established line of research studies how to extend the rational learning theory into social network settings where individual decision making is based on payoff maximization using observations from the neighbors, see e.g. Bala and Goyal (1998) and Gale and Kariv (2003). Since communication in a social network often involves repeated transfers of knowledge among a large number of agents, the theory based on rational learning soon becomes infeasible even for small numbers of agents. Individuals may use fairly simple updating rules for their beliefs and still arrive at outcomes like those achievable under fully rational learning. Such a naïve learning process is introduced in DeGroot (1974) with a fixed size network and further extended by Golub and Jackson (2010) into a growing sequence of networks. In this approach, each agent is influenced by the agents connected to her, and her belief is the weighted average of beliefs of her neighbors. Under some conditions this updating process converges to some common belief in the time limit with the network size fixed, which is called reaching a consensus, and also possible to reach the truth in the double limit of time and population size, which is called wisdom of crowds. Golub and Jackson (2010) provide sufficient conditions for a network structure to achieve this wisdom of crowds phenomenon. This is very good news as it supports the idea that social networks may eventually lead the society to the truth. However, as we will show in this chapter, the appealing wisdom of crowds phenomenon heavily relies on the - unfortunately - rather un-
realistic assumption that agents always update their beliefs in exactly the same way. Allowing some flexibility in the belief updates, i.e., agent $i$ only every now and then includes the belief of agent $j$ in her belief updating, may destroy wisdom of crowds. The unsettling observation is that consensus is still reached, leading to the question whether consensus can be distinguished from wisdom of crowds. The answer we give is negative. We claim and support by mathematical rigor that wisdom of crowds is indeterminable if structural variability is allowed. For this we use a surprisingly simple argument: some wise networks (defined in Section 4.5) can be written as average of two non-wise networks, while the randomization (described later) of the two non-wise networks is obviously non-wise.

Why should a communication network be variable? Unlike physical networks such as electrical grid or telephone networks, social connections are difficult to measure and are only stable over a short period which may be long enough for reaching a consensus but too short to achieve wisdom of crowds. One may keep in touch with his/her friends on a monthly basis, but need not to talk to all friends on a daily basis. Even at work, people do not talk to every colleague every day especially when they are in a large workplace. This fact may lead to different research results according to different approaches. A survey based research tends to find long term relationship whereas an observation based research usually emphasizes frequent communications in short periods. Therefore it is important to take into account that social networks are possibly variable.

In this chapter a social network is the communication structure in a society. Social networks are represented by non-negative matrices, being referred to influence matrices, where a positive $i,j$ entry indicates that agent $i$ puts positive weight on trusting agent $j$. Let $X$ and $Y$ be two influence matrices. Our idea can be captured by the expression $P = \alpha X + (1 - \alpha) Y$ with $\alpha \in (0,1)$. This expression has two possible interpretations:

- A natural interpretation is that $P$ is the weighted averaged communication structure of the two basic structures $X$ and $Y$. In this case, $X$ and $Y$ can be seen as networks over some short periods and $P$ can be seen as a long term network where the time span covers both periods of $X$ and $Y$. We say that $P$ is the superposition of $X$ and $Y$.

- We can consider a random process $\{\tilde{P}^{(t)} : t = 1,2,\ldots\}$ of networks, where $\tilde{P}^{(t)}$ is randomly chosen from $(X,Y)$ with probability $\{\alpha, 1 - \alpha\}$ independently for each $t$. $\tilde{P}^{(t)}$ is then a random variable of communication structure whose expected value is equal to $P$, i.e., $E[\tilde{P}^{(t)}] = \alpha X + (1 - \alpha) Y = P$. We call $\tilde{P}^{(t)}$ the randomization of $X$ and $Y$. $P$ is the expected structure of the randomized process $\{\tilde{P}^{(t)}\}$.

Indeed, the two interpretations reflect two ways of modeling. Superposition averages information before performance measures being evaluated, thus simplifies analysis. On the other
hand, randomization includes more information in performance measures but also makes the model more complicated. Although using which approach is a freedom of choice, problems may occur if results of different approaches conflict. Unfortunately, this is what we will see in modeling wisdom of crowds.

We examine the validity of wisdom of crowds under both randomization and superposition. As we will show, under randomization consensus can be reached per sample path under certain conditions. Wisdom of crowds is achieved under sufficient conditions from the superposition model. Moreover, we provide examples that a wise network can be expressed as the superposition of two non-wise networks. The corresponding randomization model of the two non-wise networks is non-wise, but for some sample paths it appears that the truth can be reached. This might lead to a mistake in determining whether there is a wisdom of crowds.

This chapter is structured as follows. In Section 4.2 we introduce the DeGroot model of social learning. Network variability is described in Section 4.3. In the subsequent section, conditions of reaching a consensus are provided for both the variable and fixed networks approaches. In Section 4.5 we define wisdom of crowds formally and illustrate it with an example. Numerical simulations are shown in Section 4.6 to demonstrate consensus and wisdom of crowds. Section 4.7 suggests a variant of the randomization model. Discussions about the theory and real world applications are given in Section 4.8.

### 4.2 DeGroot model of social learning

We introduce the social learning model first studied by DeGroot (1974). Consider a society of agents $\mathcal{N} = \{1, \ldots, n\}$, where agent $i \in \mathcal{N}$ has a belief $f^{(t)}_i$ at time $t \in \{0, 1, \ldots\}$. Here we assume $f^{(t)}_i \in [a, b] \subset \mathbb{R}$ with $-\infty < a < b < \infty$. The belief vector of all agents is denoted by $f^{(t)} = (f^{(t)}_1, \ldots, f^{(t)}_n)^\top$. At time 0, initial beliefs are given as $f^{(0)}_i = \mu + e_i$ for each $i \in \mathcal{N}$, where constant $\mu$ is said to be the true state of nature and $e_i$ is a random noise sampled from a distribution with bounded support, zero mean and positive variance.

Beliefs are updated based on confidence in other agents. The confidence structure is expressed by a social network, where a link from node $i$ to node $j$ means agent $i$ trusts agent $j$. In other words, agent $i$ is said to be influenced by agent $j$. An agent may put different weights on agents who influence her. The confidence structure and the influence weights are described by an influence matrix $P = [p_{ij}]_{i,j \in \mathcal{N}}$, where $p_{ij} \geq 0$ and $\sum_{j \in \mathcal{N}} p_{ij} = 1$. The weight $p_{ij}$ denotes the trust that agent $i$ puts on agent $j$. The total confidence of each agent is normalized to 1 which makes $P$ a row-stochastic matrix. The belief of agent $i \in \mathcal{N}$ in the next period is defined as the weighted average of beliefs in the current period of the agents.
that receive positive weight from \( i \). More specifically,

\[
f^{(t)}_i = \sum_{j \in \mathcal{N}} p_{ij} f^{(t-1)}_j, \quad i \in \mathcal{N}, t \in \{1, 2, \ldots\},
\]

or in matrix form

\[
f^{(t)} = P f^{(t-1)} = P f^{(0)}, \quad t \in \{1, 2, \ldots\}.
\]

Note that the initial beliefs \( f^{(0)}_i \)'s are randomly drawn at time 0, which means the sample mean \( (\sum_{i \in \mathcal{N}} f^{(0)}_i) / n \) does not need to be \( \mu \). On the other hand, the theoretical mean \( \mathbb{E}[f^{(0)}_i] = \mu + \mathbb{E}[e_i] = \mu \) for all \( i \in \mathcal{N} \), and therefore \( \mathbb{E}[f^{(t)}_i] = \mu \) for \( t \geq 1 \). There are two important questions about the dynamical system (4.1): whether it converges; and if it does, how far the limit is from the true state \( \mu \). DeGroot (1974) studied conditions of convergence of (4.1), and Golub and Jackson (2010) tackled the second question by extending the model into a growing network in size. We modify the DeGroot model by letting the network, i.e. the confidence structure, be variable, and discuss these two questions in the subsequent sections.

### 4.3 Network variability

Everyone maintains long run relations with some good friends. Even though, it is not necessary to exchange opinions with all friends on a daily basis. In the framework of the DeGroot model this means that even if a positive weight is put between two agents, it may not be used in every updating step of beliefs. To incorporate this feature into the model, we consider a random sequence \( \{\hat{\mathcal{P}}^{(t)} : t = 1, 2, \ldots\} \) whose elements are independently drawn from a finite set \( \mathcal{A} \) of row-stochastic matrices according to a probability distribution \( R \) over \( \mathcal{A} \). The corresponding belief process \( \hat{f}^{(t)} \) is defined by

\[
\hat{f}^{(t)} = \hat{\mathcal{P}}^{(t)} \hat{f}^{(t-1)}, \quad \hat{f}^{(0)} = \mu + e_i.
\]

Let \( P = \sum_{A \in \mathcal{A}} R(A) \cdot A \). Clearly, it holds that \( \mathbb{E}[\hat{\mathcal{P}}^{(t)}] = P \) for each \( t \in \{1, 2, \ldots\} \). Intuitively, \( \mathcal{A} \) can be thought of as the collection of communication structures from which one structure is chosen every day. For instance, a middle manager of a company talks to the same or higher level managers when there is a management meeting (a meeting day), otherwise she talks to her subordinates (a normal day). Her daily schedule for the next year can be seen as randomly selected from a meeting day and a normal day, because a meeting is usually arranged a few days in advance but not known some months before. The structure \( P = \sum_{A \in \mathcal{A}} R(A) \cdot A \) is thus the averaged communication structure over a long period, say a month or a year.
When a collection \( \mathcal{A} \) and a distribution \( R \) over \( \mathcal{A} \) are given, we call the corresponding \( P \) the superposition of \( \mathcal{A} \) and \( \{ \hat{P}^{(t)} : t = 1, 2, \ldots \} \) the randomization of \( \mathcal{A} \). For notational simplicity, we use \( \hat{F}^{(t)} \) instead of \( \{ \hat{P}^{(t)} : t = 1, 2, \ldots \} \) when there is no ambiguity. Figure 4.1 illustrates the superposition of two networks.

When it comes to real data such as email communication or on-line social networking data, it is usually possible to detect several special patterns and divide the time span of the data into sub-periods according to those patterns. If the whole period of the data is used to estimate the communication structure, we get \( P \). If sub-periods are made and the structure is estimated for each sub-period, we obtain a collection \( \mathcal{A} \) together with frequency \( R \). Therefore, it is possible to use either the superposition or the randomization version of the model, and hence it is important to know what happens when the randomized version is considered.

### 4.4 Reaching a consensus

We first discuss the long run behavior of the belief vectors \( \mathbf{f}^{(t)} \) and \( \hat{\mathbf{f}}^{(t)} \), as defined in (4.2) and (4.3) respectively. In order to do so, we need to introduce some concepts from non-negative matrices theory.\(^1\) All matrices in the rest of this chapter are square matrices. Given a non-negative matrix \( T \), we say there exists a path from \( i \) to \( j \) if there exists some \( k > 0 \) such that the \((i, j)\) element of matrix \( T^k \) is positive. A non-negative matrix \( T \) is said to be irreducible if for every ordered pair of indices \((i, j)\) there exists a path from \( i \) to \( j \). Denote \( \mathbf{1} \) the vector with

\(^1\)Most results introduced here can be found in Seneta (1981).
all elements being equal to 1.

**Proposition 4.1.** If a row-stochastic matrix $P$ is irreducible, then there is a unique solution $\pi$ of

$$
\pi^T P = \pi^T, \quad \pi^T 1 = 1. \tag{4.4}
$$

The solution $\pi$ of (4.4) is called the left Perron-Frobenius eigenvector of $P$. A non-negative matrix $T$ is said to be primitive if there exists some $k > 0$ such that $(T^k)_{ij} > 0$ for all $i, j \in \mathcal{N}$. Note that primitivity implies irreducibility but the converse is not true. The following famous result forms a foundation of the analysis of long run dynamics of (4.1).

**Proposition 4.2.** If a row-stochastic matrix $P$ is primitive, it holds that

$$
\lim_{t \to \infty} P^t = 1 \pi^T, \tag{4.5}
$$

where the convergence is exponentially fast.

The proof of Propositions 4.1 and 4.2 can be found in Seneta (1981).

**Corollary 4.3.** If an influence matrix $P$ is primitive, then the corresponding belief $f^{(t)}_i$ in system (4.1) converges to the limit $\pi^T f^{(0)}$ as $t \to \infty$ for all $i \in \mathcal{N}$.

The above corollary describes a situation that the beliefs of all agents converge to the same limit. DeGroot (1974) refers to this case as consensus.

**Definition 4.1.** We say that a consensus is reached if and only if all $f^{(t)}_i$’s, $i \in \mathcal{N}$, converge to the same limit as $t \to \infty$.

We now turn our attention to the dynamics (4.3) under randomization. We first state a sufficient condition which guarantees the primitivity of the superposition $P$ of $\mathcal{A}$. If a non-negative matrix $T$ is primitive, and another non-negative matrix $\tilde{T}$ has the same dimensions as $T$ and has positive elements in the same positions as $T$, it holds that $\tilde{T}$ is also primitive. We then have the following proposition.

**Proposition 4.4.** If a finite collection $\mathcal{A}$ of influence matrices contains only primitive matrices, then the corresponding superposition $P = \sum_{A \in \mathcal{A}} R(A) \cdot A$ with any distribution $R$ is also primitive.

Recall that under superposition, the process (4.1) becomes deterministic after initial beliefs $f^{(0)}_i$’s being drawn. However, in (4.3) each $\hat{f}^{(t)}_i$, $t = 1, 2, \ldots$, is a random variable with bounded
support provided that \( \hat{f}^{(0)} \) is given. Let \( \Omega \) be the space of all infinite sequences whose members are drawn from set \( \mathcal{A} \), then the following proposition follows from Lebesgue’s dominated convergence theorem, see e.g. Doob (1994).

**Proposition 4.5** (Dominated Convergence). If there exists a random variable \( \hat{f} : \Omega \to \mathbb{R}^n \) such that \( \hat{f}^{(t)} \) converges to \( \hat{f} \) almost surely as \( t \to \infty \), then it holds that

\[
\mathbb{E}[\hat{f} | \hat{f}^{(0)}] = \lim_{t \to \infty} P^t \hat{f}^{(0)} = \mathbf{1} \cdot \pi^T \cdot \hat{f}^{(0)}.
\]

Of more interest is the condition of path wise convergence of process \( \{\hat{f}^{(t)} : t = 1, 2, \ldots\} \) provided \( \hat{f}^{(0)} \) is given. The ultimate reason to look at properties for each sample path is that we have only one history. Though many sampling methods can be used in simulation, the real observation is always unique. Hence any moments or distribution of a random process is not really useful when a particular observation is being tested.

Note that Equation (4.3) can be written as

\[
\hat{f}^{(t)} = \hat{P}^{(t)} \hat{f}^{(t-1)} = \hat{P}^{(t)} \hat{P}^{(t-1)} \ldots \hat{P}^{(1)} \hat{f}^{(0)},
\]

(4.6)
in which the backward product \( \hat{P}^{(t)} \hat{P}^{(t-1)} \ldots \hat{P}^{(1)} \) is used. In other words, whenever a new time period comes, a matrix chosen from \( \mathcal{A} \) is multiplied from the left to the existing product. It has been pointed out in Hajnal (1958) that the product of two different primitive matrices may not be primitive, which means \( \hat{f}^{(t)} \) may not converge path wise even if all matrices in \( \mathcal{A} \) are primitive. Fortunately, using results of Anthonisse and Tijms (1977), we are able to obtain conditions for path wise convergence of \( \{\hat{f}^{(t)} : t = 1, 2, \ldots\} \).

Here we introduce a subclass of primitive matrices which is originally defined by Hajnal (1958). A non-negative matrix \( T \) is said to be *scrambling* if for any two rows \( i \) and \( j \), there exists at least one column, say \( k \), such that both \( t_{i,k} > 0 \) and \( t_{j,k} > 0 \). In terms of network structure, this means that any two agents in the network both receive a link from at least one common agent who can be either of them or someone else.

For any \( \omega \in \Omega \), i.e. any infinite sequence of matrices chosen from \( \mathcal{A} \), let

\[
B^{(t)}_{\omega} = \hat{P}^{(t)}_{\omega} \hat{P}^{(t-1)}_{\omega} \ldots \hat{P}^{(1)}_{\omega}
\]

be a partial backward product up to time \( t \) of sequence \( \{\hat{P}^{(1)}_{\omega}, \hat{P}^{(2)}_{\omega}, \ldots\} \). The following condition \( [C] \) is crucial for path wise convergence.

\( [C] \) There is an integer \( k \geq 1 \) such that for each \( t \geq k \) the matrix \( B^{(t)}_{\omega} \) is scrambling.
**Proposition 4.6** (Anthonisse and Tijms (1977)). *Condition C is the necessary and sufficient condition of the following assertion: there is an integer $z \geq 1$, a number $\beta$ with $0 < \beta < 1$ and for any $\omega \in \Omega$ there is a vector $v_\omega$ such that $v_\omega^T 1 = 1$ and for all $j \in N$ and every element $(B^{(t)}_\omega)_{ij}$ in row $j$,

$$|(B^{(t)}_\omega)_{ij} - (v_\omega)_j| \leq \beta^{[i/z]} \text{ for all } t \geq 1,$$

where $[x]$ is the largest integer less than or equal to $x$.

Proposition 4.6 says that $B^{(t)}_\omega$ converges to $1v_\omega^T$ as $t \to \infty$ exponentially fast. By translating [C] to property of members of $\mathcal{A}$, one has the following proposition.

**Proposition 4.7** (Anthonisse and Tijms (1977)). *If each $A \in \mathcal{A}$ is scrambling, then for any $\omega \in \Omega$ there is a vector $v_\omega$ such that $v_\omega^T 1 = 1$ and

$$\lim_{t \to \infty} B^{(t)}_\omega = 1v_\omega^T,$$

where the convergence is exponentially fast.

**Corollary 4.8.** *If each $A \in \mathcal{A}$ is scrambling, then for any $\omega \in \Omega$ there is a vector $v_\omega$ such that $v_\omega^T 1 = 1$ and

$$\lim_{t \to \infty} (\hat{f}^{(t)}_\omega)_i = v_\omega^T \hat{f}^{(0)} \text{ for all } i \in N.$$ 

Intuitively, Corollary 4.8 states that if the collection of confidence structure contains only networks such that for any two agents in it there exists at least one agent who influences both of them, then any realization of a sequence of networks will lead to consensus. However, the consensus level depends on the realization of the sequence.

### 4.5 Wisdom of crowds: from consensus to truth

It is pointed out that in many situations group decision based on aggregated information outperforms individual decision, e.g. in Surowiecki (2004) where this phenomenon is called wisdom of crowds. In Golub and Jackson (2010) this notion is applied to the DeGroot model. They redefine the concept in precise mathematics to describe a society that not only reaches a consensus but that also discovers the truth. Their idea is simple: let the set of agents $\mathcal{N} = \{1, 2, \ldots, n\}$ grow in size, and take the double limit of time $t$ and network size $n$. Here we first introduce their model with deterministic networks, and then extend it with randomization. By explicitly indicating the influence matrix $P$ with size $n \times n$ by $P(n)$, we consider a sequence
of networks \( \{P(n)\}_{n=n_0}^{\infty} \) with \( n_0 > 0 \). Suppose for each \( n \geq n_0 \), the dynamics (4.1) has a limit \( f_i^{(\infty)}(n) := \lim_{t \to \infty} f_i^{(t)}(n) \) for each \( i = 1, \ldots, n \).

**Definition 4.2.** The sequence of networks \( \{P(n)\}_{n=n_0}^{\infty} \) is said to be wise if

\[
\lim_{n \to \infty} \Pr \left[ \max_{i \leq n} \left| f_i^{(\infty)}(n) - \mu \right| > \varepsilon \right] = 0 \tag{4.10}
\]

for any \( \varepsilon > 0 \).

Note that, although for each \( n \) the convergence of \( f_i^{(t)}(n) \) as \( t \to \infty \) is required, consensus is not a necessary assumption. On the other hand, even if a consensus is reached for each \( n \), the corresponding sequence of networks does not need to be wise.

The following proposition is a rephrase of a result in Golub and Jackson (2010). It gives a sufficient and necessary condition of being wise in terms of network characteristics.

**Proposition 4.9** (Golub and Jackson (2010)). If \( \{P(n)\}_{n=n_0}^{\infty} \) is a sequence of primitive stochastic matrices, then it is wise if and only if the associated left Perron-Frobenious eigenvectors \( \{\pi(n)\}_{n=n_0}^{\infty} \) are such that

\[
\max_{i \leq n} \pi_i(n) \to 0
\]

as \( n \to \infty \).

The assumption in Proposition 4.9 is slightly stronger in the sense that it guarantees consensus for each \( n \geq n_0 \). According to Corollary 4.3, the consensus level with network size \( n \) is \( \pi(n)^\top \cdot f^{(0)}(n) \). Hence \( \pi_i(n) \) is the averaging weight of initial belief of agent \( i \) in the consensus, or the influence power of agent \( i \) in the society. The proposition states that a growing society is wise if and only if it grows in a way that for any member the influence power becomes insignificantly small if \( n \to \infty \).

**Example 4.1.** Consider the network expressed by the following matrix:

\[
P(n) = \begin{bmatrix}
0 & 1/4 & 1/4 & 0 & \cdots & 0 & 1/4 & 1/4 \\
1/4 & 0 & 1/4 & 1/4 & 0 & \cdots & 0 & 1/4 \\
1/4 & 1/4 & 0 & 1/4 & 0 & \cdots & 0 & \vdots \\
& & & \ddots & \ddots & \ddots & \ddots & \vdots \\
& & & & \ddots & \ddots & \ddots & \ddots \\
& & & & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & 1/4 & 1/4 & 0 & 1/4 & 1/4 \\
1/4 & 0 & \cdots & 0 & 1/4 & 0 & 1/4 & 1/4 \\
1/4 & 1/4 & 0 & \cdots & 0 & 1/4 & 1/4 & 0 \\
\end{bmatrix}
\]
It is easy to see that the corresponding left Perron-Frobenius eigenvector \( \mathbf{\pi}(n) \) is \( (1/n, \ldots, 1/n)^T \), whose maximum element converges to 0 as \( n \to \infty \). Therefore, the sequence \( \{P(n)\}_{n=n_0}^{\infty} \) is wise.

Let \( \mathcal{A}(n) \) denote a finite collection of influence matrices of size \( n \times n \). As an analogy to deterministic networks, the collection of networks \( \{\hat{P}^{(t)}(n) : t = 1, 2, \ldots\}_{n=n_0}^{\infty} \) is said to be a randomization of \( \{\mathcal{A}(n)\}_{n=n_0}^{\infty} \) if for each \( n \geq n_0 \), \( \hat{P}^{(t)}(n) \) is independently chosen from \( \mathcal{A}(n) \) for \( t = 1, 2, \ldots \) according to some probability distribution \( R_n \) over \( \mathcal{A}(n) \). We denote by \( \Omega(n) \) the collection of all infinite sequences whose members are chosen from \( \mathcal{A}(n) \). The corresponding superposition \( \{P(n)\}_{n=n_0}^{\infty} \) is given as \( P(n) = \sum_{A \in \mathcal{A}(n)} R_n(A) \cdot A \) for \( n \geq n_0 \). The wisdom of crowds under randomization is then defined as the follows.

**Definition 4.3.** The sequence of randomized networks \( \{\hat{P}^{(t)}(n) : t = 1, 2, \ldots\}_{n=n_0}^{\infty} \) is said to be wise if

1. for every sample path \( \omega \in \Omega(n) \) the limit \( \left( \hat{f}_{\omega}^{(\infty)}(n) \right)_i \) exists for \( i = 1, \ldots, n \), \( n \geq n_0 \), and for any given initial beliefs;
2. for any sequence of sample paths \( \{\omega(n)\}_{n=n_0}^{\infty} \)

\[
\lim_{n \to \infty} \Pr \left( \max_{i \leq n} \left| \left( \hat{f}_{\omega(n)}^{(\infty)}(n) \right)_i - \mu \right| > \epsilon \right) = 0 \quad (4.11)
\]

for any \( \epsilon > 0 \).

If condition [C] holds for every \( \omega \in \Omega(n) \) for all \( n \geq n_0 \), then path wise convergence of beliefs is guaranteed. However, there is no easy way to check wisdom of crowds by characteristics of matrices in \( \mathcal{A}(n) \). This is because unlike the left Perron-Frobenius eigenvectors in Proposition 4.9, it seems not possible to compress the information of the behavior of a sample path into a single function in the randomization model.

### 4.6 Comparing the two models: superposition versus randomization

In Section 4.4 we discussed conditions for consensus under superposition and randomization. The superposition is only meaningful in combination with the corresponding randomization model. Mathematically, superposition essentially implies a fixed network and hence primitivity is enough for consensus to be reached. In randomization models primitivity is not sufficient and a stronger condition that all matrices in the collection are scrambling is used to assure consensus.
In this section we aim at testing wisdom of crowds under randomization and its corresponding superposition model. We focus on a representative case where a wise network is the superposition of two non-wise networks. On the other hand our simulations show that the randomization of the same two non-wise networks is also non-wise. However, in some realizations it is not possible to distinguish the randomization model from its corresponding superposition model, as we will see. It means even if the randomization model is appropriate in modeling the learning process and it suggests that the society is not wise, a mistake in determining wisdom of crowds might occur.

**Example 4.2.** Consider the networks depicted in Figure 4.2. Every agent has a link to herself which is omitted in the figure. The influence matrix $P(n)$ of size $n$ (both odd and even) is given below:

\[
P(n) = \\
\begin{bmatrix}
1/2 & 1/4 & 1/4 \\
1/4 & 1/2 & 0 & 1/4 \\
1/4 & 0 & 1/2 & 0 & 1/4 \\
& & & & \\
& & & & \\
& & & & \\
1/4 & 0 & 1/2 & 0 & 1/4 \\
1/4 & 0 & 1/2 & 1/4 \\
1/4 & 1/4 & 1/2
\end{bmatrix}
\]
This matrix is the superposition of two other influence matrices $X(n)$ and $Y(n)$ such that $P(n) = \alpha X(n) + (1 - \alpha) Y(n)$ for some $\alpha \in (0, 3/4)$, where for $n = 2m + 1$, $X(n)$ and $Y(n)$ are given by

$$
X(2m + 1) = \begin{bmatrix}
1/2 & 1/4 & 1/4 \\
1/3 & 1/2 & 0 & 1/6 \\
1/3 & 0 & 1/2 & 0 & 1/6 \\
& & & & & \\
& & & & & \\
1/3 & 0 & 1/2 & 0 & 1/6 \\
1/3 & 0 & 1/2 & 1/6 \\
& & & & & \\
1/3 & 1/6 & 1/2
\end{bmatrix}
$$

and

$$
Y(2m + 1) = \begin{bmatrix}
1/2 & 1/4 & 1/4 \\
c & 1/2 & 0 & d \\
c & 0 & 1/2 & 0 & d \\
& & & & & \\
& & & & & \\
c & 0 & 1/2 & 0 & d \\
c & 0 & 1/2 & d \\
& & & & & \\
c & d & 1/2
\end{bmatrix}
$$

with $c = (3 - 4\alpha)/(12 - 12\alpha)$ and $d = 1/2 - c = (3 - 2\alpha)/(12 - 12\alpha)$. For $n = 2(m + 1)$, $X(n)$ and $Y(n)$ can be defined in a similar way, with slight differences in the last row. Both $X(n)$ and $Y(n)$ are primitive. The corresponding left Perron-Frobenious eigenvectors $\pi_X$ and $\pi_Y$ have the following entries:

$$
\pi_{X;1}(2m + 1) = \frac{2^m}{2^{m+2} - 3}, \quad \text{(4.12)}
$$

$$
\pi_{X;2i}(2m + 1) = \pi_{X;2i+1}(2m + 1) = \frac{3 \cdot 2^{m-i-1}}{2^{m+2} - 3} \quad \text{for} \quad i = 1, \ldots, m
$$

and

$$
\pi_{Y;1}(2m + 1) = \frac{(1 - 4c)(2c)^m}{(1 - 2c)^m - 2(2c)^{m+1}},
$$

$$
\pi_{Y;2i}(2m + 1) = \pi_{Y;2i+1}(2m + 1) = \frac{(1 - 4c)(1 - 2c)^{i-1}(2c)^{m-i}/2}{(1 - 2c)^m - 2(2c)^{m+1}} \quad \text{for} \quad i = 1, \ldots, m \quad \text{(4.13)}
$$
for odd \( n = 2m + 1 \), and

\[
\pi_{X:1}(2(m + 1)) = \frac{2^m}{2^{m+2} - 2}, \quad (4.14)
\]

\[
\pi_{X:i}(2(m + 1)) = \pi_{X:i+1}(2(m + 1)) = \frac{3 \cdot 2^{m-i-1}}{2^{m+2} - 2} \quad \text{for} \quad i = 1, \ldots, m,
\]

\[
\pi_{X:2(m+1)}(2(m + 1)) = \frac{1}{2^{m+2} - 2}
\]

and

\[
\pi_{Y:1}(2(m + 1)) = \frac{(1 - 4c)(2c)^m}{2(1 - 2c)^{m+1} - 2(2c)^{m+1}},
\]

\[
\pi_{Y:i}(2(m + 1)) = \pi_{Y:i+1}(2(m + 1)) = \frac{(1 - 4c)(1 - 2c)^{i-1}(2c)^{m-i}/2}{2(1 - 2c)^{m+1} - 2(2c)^{m+1}} \quad \text{for} \quad i = 1, \ldots, m,
\]

\[
\pi_{Y:2(m+1)}(2(m + 1)) = \frac{(1 - 4c)(1 - 2c)^m}{2(1 - 2c)^{m+1} - 2(2c)^{m+1}} \quad (4.15)
\]

for even \( n = 2(m + 1) \).

It is easy to see that \((P(n))_{n=n_0}^{\infty}\) given in Example 4.2 is wise. After some algebra, one has

\[
\lim_{m \to \infty} \pi_{X:1}(2m + 1) = \frac{1}{4}, \quad \lim_{m \to \infty} \pi_{X:1}(2(m + 1)) = \frac{1}{4},
\]

and

\[
\lim_{m \to \infty} \pi_{Y:2m}(2m + 1) = \lim_{m \to \infty} \pi_{Y:2m+1}(2m + 1) = \frac{1 - 4c}{2 - 4c} = \frac{\alpha}{3 - 2\alpha} > 0,
\]

\[
\lim_{m \to \infty} \pi_{Y:2(m+1)}(2(m + 1)) = \frac{1 - 4c}{2 - 4c} = \frac{\alpha}{3 - 2\alpha} > 0,
\]

which implies that \((X(n))_{n=n_0}^{\infty}\) and \((Y(n))_{n=n_0}^{\infty}\) are both non-wise (see Proposition 4.9). We conclude this as follows.

**Proposition 4.10.** It is possible that the superposition of two non-wise sequence of networks is wise.

It is difficult to determine whether the corresponding randomization of \((X(n), Y(n))_{n=n_0}^{\infty}\) is wise, because even if consensus exists for each finite \( n \geq n_0 \) under randomization, the lack of a closed form solution hinders an analytical discussion of wisdom of crowds. Intuitively, it is tempting to jump to the conclusion that the randomization of two non-wise network is also non-wise, rather than to believe that there might be some surprising way of mixing up non-wise networks that leads the society to the truth. It is easy to see that for any \( t \geq \lfloor n/2 \rfloor \),
the backward product $B^{(t)}_\omega(n)$ has at least one column (e.g. column 1) where all elements of it are positive for any $\omega \in \Omega(n)$, where $\Omega(n)$ here is the space of all infinite sequences chosen from $\{X(n), Y(n)\}$. Hence $B^{(t)}_\omega(n)$ is scrambling and condition $[C]$ holds, which implies that consensus is reached path wise. The wisdom of crowds under the randomized model will be tested by simulation.

### 4.6.1 A simulation to test wisdom of crowds

In what follows we revisit Example 4.2 under the randomization model. Due to the lack of a closed form solution of the limit values we will demonstrate our findings by simulation. As a first step, we show that consensus is reached path wise by generating some realizations of $\{P(n)\}_{n=n_0}^{\infty}$. Let network size $n = 25$, and the probability of choosing $X(n)$ be $0.3$, i.e. $\alpha = 0.3$. We generate 5 sample paths:

- Sample path 1: $X X Y Y Y Y X X X X X Y Y X Y Y Y \ldots$,
- Sample path 2: $X Y X X Y Y Y Y X Y X Y Y Y Y Y X \ldots$,
- Sample path 3: $Y Y Y Y Y Y X X Y Y Y Y Y Y Y Y Y Y Y \ldots$,
- Sample path 4: $X X X X Y Y Y Y X Y Y Y Y Y Y Y Y \ldots$,
- Sample path 5: $Y X Y Y X Y Y Y Y Y Y Y Y Y Y Y Y \ldots$.

The initial belief vector $\hat{f}^{(0)}(n)$ is fixed to

$$\hat{f}^{(0)}_i(n) = 1 - 2(i - 1)/(n - 1)$$

for $i \in \mathcal{N}$ for all sample paths. Under this construction, the initial beliefs are equally distributed on $[-1, 1]$ with mean 0.

Firstly, convergence to consensus under $X(n)$, $Y(n)$, and the corresponding superposition $P(n)$ are demonstrated, where $f^{(0)}(n) = \hat{f}^{(0)}(n)$ is considered. Figure 4.3 depicts the distributions of individual beliefs at some points in time using box plots. It is clear that a consensus is reached for each network, but the consensus levels are different.

Figure 4.4 (a) and (b) illustrate belief distributions under sample paths 1 and 3 of the randomization model. The convergence to consensus is verified as expected. The consensus levels, on the other hand, are path dependent. Average beliefs of five sample paths are plotted on Figure 4.4 (c). One can see that for $t < 200$, the average beliefs behave in a very random way. This is because influence matrices $X(n)$ and $Y(n)$ push individual opinions into different directions. At the same time, the variances of beliefs is reaching zero, which results in the convergence of average beliefs for $t > 200$. 
Figure 4.3: Convergence of beliefs under fixed influence matrices with \( n = 25 \) and \( \alpha = 0.3 \).

(a) Update by \( X(25) \)  
(b) Update by \( Y(25) \)  
(c) Update by \( P(25) \)

Figure 4.4: Path wise convergence of beliefs under randomization with \( n = 25 \) and \( \alpha = 0.3 \).

(a) Sample path 1  
(b) Sample path 3  
(c) Average beliefs

Figure 4.5 are the histograms of consensus level of 5000 sample paths with the randomization model, where \( n = 25 \) and \( \alpha \) is 0.02 for subfigure (a) and 0.73 for subfigure (b). A small \( \alpha \) means that network \( X(n) \) is relatively rare. Since the beliefs are influenced by \( Y(n) \) almost all the time, the consensus level is more likely to be negative which coincides with the findings of Figure 4.3 (b) that the consensus level under \( Y(n) \) is negative. A large \( \alpha \) affects the consensus distribution in the opposite way. Nevertheless, the sample means of both histograms are very close to zero, verifying the assertion in Proposition 4.5.

Now we are ready to look at wisdom of crowds under growing randomized networks. Let \( \mathcal{X}(n) = (X(n), Y(n)) \) where \( X(n) \) and \( Y(n) \) are given in Example 4.2. The probability \( \alpha = \Pr[B^{(i)}(n) = X(n)] = 0.3 \). We vary \( n \) among \( 25, 50, 75, \ldots, 500 \). For each \( n \), we start the belief updating process from \( \tilde{f}^{(0)}(n) \) where initial individual beliefs are sampled from the standard normal distribution and then sorted by descending order. Thus the true state of the world is zero. Each simulation run stops at time \( \tau(n) \) so that \( \left| \max_i[f_i^{(\tau(n))}(n)] - \min_i[f_i^{(\tau(n))}(n)] \right| < \varepsilon \).
Figure 4.5: Histograms of consensus level under randomization with different values of $\alpha$.

The stopping criterion $\epsilon$ is taken to be $10^{-3}$. Roughly speaking, a consensus is reached at time $\tau(n)$. In Figure 4.6, those consensus levels are plotted against growing network size. As a comparison, the consensus levels under the corresponding superposed networks are also shown in the figure. There are two samples in this figure showing different dynamics of consensus. In the upper sample, clearly the randomization model exhibits a constant larger deviation of consensus from the true state zero than the superposition model, meaning that this growing network under randomization is not wise. On the other hand, the sample on the bottom sends us a confusing signal. One can observe that for $n > 400$, the consensus level under the randomization model seems to converge to zero, and is almost undistinguishable from the superposition model. We know that under the superposition model the network is wise. Hence this sample provides evidence for the randomized model being wise as well, which conflicts with the other sample and the model suggestion.

We have seen that on sample path level, the randomization model of non-wise networks can lead to similar results as the corresponding superposition model, which suggests that the network is wise. We conclude that using the superposition model as an approximation of the corresponding randomization model is dangerous as the two models can come to different conclusions with respect to wisdom of crowds. Therefore a modeler shall be very careful with selecting the right model.

4.7 An implementation with local randomization

In the previous section we have shown an example of the randomization model where the entire network is selected from two alternatives. This is restrictive, though our model is defined in a more general way. Here we provide an implementation with local randomization.
Figure 4.6: Two samples of consensus levels under growing networks.

Suppose now each individual in the society can independently choose her own neighborhood to trust, from a set of two alternatives. Let $x_i$ and $y_i$ be two influence vectors of agent $i \in \mathcal{N}$, where the $j$-th entries $x_{ij}$ and $y_{ij}$ represent the trust weights put on agent $j \in \mathcal{N}$ by agent $i$. Consider that in each period agent $i \in \mathcal{N}$ select $x_i$ with probability $\alpha_i$ and $y_i$ with probability $1 - \alpha_i$. Since each agent select her own influence vector independently, there are $2^n$ different network structures. Let $\mathcal{A} = \{A_1, A_2, \ldots, A_{2^n}\}$ where $A_k, k \in \{1, \ldots, 2^n\}$, represents one network structure where the $i$-th row of $A_k$ is either $x^T_i$ or $y^T_i$. Let $\delta_k$ denote the probability that $A_k$ is formed, where $\delta_k$ can be expressed as a product of $n$ numbers such that the $i$-th multiplier is $\alpha_i$ if the $i$-th row of $A_k$ is $x^T_i$ and $(1 - \alpha_i)$ otherwise. The randomized network $\hat{P}(t)$ at time $t \geq 1$ is then randomly selected from $\mathcal{A}$ according to probability distribution $(\delta_1, \ldots, \delta_{2^n})$. The corresponding superposed network $P$ is the weighed averaged
network structure, i.e.,

\[ P = \sum_{k=1}^{2^n} \delta_k A_k \]

\[ = \prod_{i=1}^{n} \alpha_i \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_n^\top \end{bmatrix} + (1 - \alpha_1) \prod_{i=2}^{n} \alpha_i \begin{bmatrix} y_1^\top \\ x_2^\top \\ \vdots \\ x_n^\top \end{bmatrix} + \cdots + (1 - \alpha_i) \prod_{i=1}^{n} \alpha_i \begin{bmatrix} y_1^\top \\ y_2^\top \\ \vdots \\ y_n^\top \end{bmatrix} \]

\[ = D_\alpha \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_n^\top \end{bmatrix} + (I - D_\alpha) \begin{bmatrix} y_1^\top \\ y_2^\top \\ \vdots \\ y_n^\top \end{bmatrix}, \]

where \( D_\alpha \) denotes the diagonal matrix whose \( i \)-th diagonal element is \( \alpha_i \), and \( I \) is the identity matrix.

As shown above, our framework can be easily applied to local randomization such that individuals selects whom to trust independently. Although the randomization procedure is complicated, the corresponding superposed network has a surprisingly simple structure, allowing further analysis with ease.

### 4.8 Discussion

In this chapter we have studied consensus and wisdom of crowds under variable networks. In our consideration the variability is achieved by randomly selecting a network structure from some alternatives. Within this framework, it is shown that consensus can be reached under certain sufficient conditions. However, when it comes to wisdom of crowds, not only is there no analytical way to check it based on the characteristics of the alternative networks, but also a conflict can occur between intuition and aggregated information in terms of superposed network structure. From our simulation results, it turns out that a non-wise variable network can even behave very similarly to a wise one based on the corresponding deterministic superposed network.

Our finding raises a fundamental question: how should we model the world? If we have two models where one is simple and fits our observation quite well but the other suggests a different conclusion but still have some chance to reproduce the observed results, which model should we believe? In most cases the former is thought to be appropriate as a first hand approximation, however, this conclusion is dangerous in absence of a quality discus-
sion of this approximation. Since the world itself is a highly complex system, simplicity cannot be a criterion of model selection, and to support our judgement using solely goodness of fit may cause serious consequences when we make a mistake.

An example from social networks study is the so called majority illusion paradox reported in Lerman, Yan, and Wu (2015). It is a phenomenon that even though a majority of agents in the network observe that more than half of their neighbors choose to take a new action, the real fraction of people who take the new action can be far less than 50%. This situation happens when the new action adopters have high degrees and are linked with many non-adopters with low degrees, especially when the network has a power law degree distribution. If agents do not know the entire network structure but make their decision based on local observations, then most of them will reach the conclusion that the majority is new action adopters whereas it is not.

Interestingly, our findings could help explain the effect of economic bubbles. What if most people in the society think the economy grows healthily but it actually is booming into a bubble? In fact, in human history, bubbles had never been recognized as such until they collapsed. Research on detecting bubbles are still on going (see e.g. Jarrow, Kchia, and Protter (2011) on financial markets and for more literature). Our study contributes to an understanding of why it is so difficult to detect economic bubbles. People update their opinion via a social network and consensus might erroneously be mistaken as wisdom of crowds. We have shown, however, that even if there are smallest fluctuations in the network structure while updating, the consensus is path dependent and the proximity to truth is mere chance. The message from our study is that it is difficult to determine whether a society is wise, and we tend to overestimate its wisdom.

In this chapter we considered a learning model where all agents update their beliefs. With a similar approach, this learning model can also be interpreted as a “random chatter” model, where an imaginary chatter from outside of the society is considered to chat with a random person in each period in the society and aggregates information from whom she has met. In a new period, a neighbor of the one that the chatter has chatted with in the previous period is selected to chat with the probability given by the influence rate. Therefore, the influence matrix $P$ becomes a probability matrix and the model becomes a Markov chain. The corresponding left Perron-Frobenious vector $\pi$ (or stationary distribution in Markov chain theory) then represents the proportions of time the chatter spends with each individual in the long run. If individuals in the society keep their initial beliefs unchanged, the chatter will find an aggregated belief of the whole society $\pi^T f^{(0)}$ which is equivalent to the consensus in our learning model. This random chatter model is related to the famous Google PageRank (see Brin and Page (1998) and Page et al. (1999)). By applying perturbation analysis of
Markov chains, it is possible to study the behavior of the random chatter model under randomization of network structure where the probability $\alpha$ is close to zero or even as a variable of time. This is a direction of our future research.