Mathematical Social Sciences 60 (2010) 69–78

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Trust, influence, and convergence of behavior in social networks ☆

Zhengzheng Pan *

Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:
Received 11 November 2008
Received in revised form
4 March 2010
Accepted 31 March 2010
Available online 7 April 2010

JEL classification:
D83
D85
C62
C63

Keywords:
Social networks
Learning
Consensus
Simulation

ABSTRACT

I propose a social learning framework where agents repeatedly take the weighted average of all agents' current opinions in forming their own for the next period. They also update the influence weights that they place on each other. It is proven that both opinions and the influence weights are convergent. In the steady state, opinions reach consensus and influence weights are distributed evenly. Convergence occurs with an extended model as well, which indicates the tremendous influential power possessed by a minority group. Computer simulations of the updating processes provide supportive evidence.

1. Introduction

We do not live like Robinson Crusoe, and interaction with others has an important role in our day to day lives (Barabási, 2003). The big successes of Facebook, MySpace and other similar online community utilities are clear evidence of how people want to be connected. Moreover, a social network acts as a platform and backbone of social activities which assist and enable dissemination of information that influences individuals' decision making process as well as the development of economies. Examples of how social networks affect individuals and communities arise in the diffusion of information and knowledge (Granovetter, 1973), stock market (Shiller, 1995), marketing (Chan and Misra, 1990; Vernette, 2004), and politics (Roch, 2005), just to name a few.

I propose a social learning model where agents continuously update their opinions by taking weighted averages of those of others. Social influence during the learning process is modeled by a row-stochastic matrix, i.e. the sum of the elements in each row equals 1. This kind of naive learning mechanism was seminally proposed by French (1956) and subsequently formalized by DeGroot (1974), which has been employed in multiple models since its introduction and recently adopted by Golub and Jackson (2007). DeMarzo et al. (2003) justified this simplified framework as a reflection of persuasion bias, which conforms to empirical findings. My variation of the model can be distinguished from others with three basic assumptions.

First, one novel feature of my model is a time-varying influence weight matrix. The influence indicator, or variables that play similar roles in social learning, is in most cases assumed to be constant over time (French, 1956; Bala and Goyal, 1998; DeMarzo et al., 2003; Friedkin and Johnsen, 1997; Golub and Jackson, 2007). My main hypothesis is that agents redistribute the influence weights they place on others and the influence matrix continuously gets updated over time. This assumption reflects the changes in attitude that people tend to make during social interaction. There has been studies on time-varying weight matrices, such as Hegselmann and Krause (2002) model and Weisbuch et al. (2002) model. The difference between their models and mine is that agents in their models only place positive weights on those with close opinions; whereas agents in my framework place positive weights even on the most remote opinion holders.

Second, in many social learning models, evolution of opinions is examined according to accuracy of estimation of a predesignated

☆ This work was mostly done when I was at the Economics Department of Virginia Tech. I thank Dr. Robert P. Gilles for his instruction and comments throughout this study. I have benefited greatly from discussions with Dr. Sudipta Sarangi. Also, I appreciate Dr. Hans Haller's suggestions on theoretical analysis. Special thanks are due to Benjamin Golub and an anonymous reviewer for their insightful comments on earlier drafts, especially the reviewer's helpful inputs on some of the proofs. Mistakes are all mine.

* Tel.: +1 540 998 9623.
E-mail address: zzpan@vt.edu.

0165-4896/$ – see front matter
doi:10.1016/j.mathsocsci.2010.03.007

value or state (Bala and Goyal, 1998; DeMarzo et al., 2003; Golub and Jackson, 2007). Thus, in those models agents' opinions are referred to as *beliefs* in the sense that they are essentially estimates of the true value. In this paper, there is no "true value" and agents are not aiming at getting close to such a "true" state. The reason is that a true value does not always exist in real life. For instance, when asked about opinions on food and music, there is no "right" answer. Even for some seemingly less subjective topics such as justifications of a war or credibility of a presidential candidate, the "truth" may never be known. Also, this assumption has made the convergence outcome more interesting, in the sense that individuals come to an agreement with a wider diversity of initial opinions.

Third, a lot of previous studies focused on observational learning. That is, agents observe others' payoff-maximizing decisions and update their own opinions accordingly (Bala and Goyal, 2001; Banerjee and Fudenberg, 2004; Gale and Kariv, 2003). The examination of convergence is thus based on optimality with regard to agents' utility functions. My model differs from them in that there is no preference or utility function involved. The assumption is made in views that individuals do not always share information for a specific payoff-related purpose. Rather, communication with others stems from our needs for social interaction. Consequently, the main finding is that the convergence of my model crucially relies on the details of the interaction dynamics.

Social learning studies have focused on the final outcome of the learning process. Especially, whether the opinions and behavioral decisions of members in a society conform (Bala and Goyal, 1998; Banerjee and Fudenberg, 2004; Ellison and Fudenberg, 1993, 1995; Golub and Jackson, 2007). The core of this paper is the convergence of both opinions and the influence distribution. There are two main questions. First, what are the conditions for convergence? Second, what are the characteristics of the final outcomes?

With a time-invariant matrix, the matrix is defined to be convergent if and only if the opinions are (Golub and Jackson, 2007). In contrast to that, I prove that convergence of either factor does not guarantee that of the other when both change over time. On the other hand, convergence of both occurs when the details of updating rules of the proposed model in this paper are specified. The patterns of final opinions show conformism. That is, not only do opinions converge, all agents' opinions converge to the same value. With the given interaction dynamics, the influence weights converges to be equally distributed among agents.

With a time-varying matrix, Lorenz (2005) provided a general framework and pointed out three sufficient conditions on the sequence of weight matrices such that the weighted averaging process converges to a consensus status. The three sufficient conditions are: positive diagonal elements, symmetric zero elements, and minimum weights. The comparison between Lorenz's work and my findings is quite interesting. First, he shows the existence of pairwise disjoint classes of communication, which suggests isolated islands; my model leads to an opposite scenario in which all agents place positive weights on all others. Second, Lorenz's Proposition 5 appears to be very similar to the principle of my main theorem (Theorem 2), in that we both use minimal weights.

The key difference is that the matrix in my model actually has all zeros alone the diagonal at the steady state, which violates both the first and third conditions in Lorenz's paper. Note that such zero diagonal pattern appears as the limit of my model, whereas Lorenz's conditions should be applied to the intermediate steps. However, even if a zero diagonal matrix is used during every period of the updating, consensus is still reached given all other elements is larger than a positive number δ_c.[1]

[1] One may refer to the proof of Theorem 2 of this paper in the Appendix to check this argument.

The critical point here is that these seemingly discrepant findings do not contradict or disapprove Lorenz's theorems. First, the focus of his theorems is the consensus pattern of the weight matrix, not the opinions. Such weight pattern is one of many ways to reach opinion conformity, i.e., sufficient but not necessary. Secondly, Lorenz stated clearly that the conditions for consensus pattern of weights are also sufficient, not necessary. So it does not affect the validity of his mathematical induction if one finds other sufficient conditions that result in the same outcomes.

An extended model was examined to better understand the interaction effects. The idea is to introduce a group of so-called "persistent" agents that only perform limited interactions with others. It is shown that persistence on opinion affects more on the convergence path than persistence on influence weights does. Moreover, persistent agents can significantly alter the final opinion of the whole society; while it is assumed that they only constitute a minority group of the whole population (less than 10%). For that reason, my model shows cascade-like phenomena and support the *opinion leaders* argument of Katz and Lazarsfeld (1955).

I have simulated the complete updating processes for both the basic and extended models. The outcomes greatly support the theoretical prediction. The convergence behavior is very robust, consistent, and independent of initial conditions.

The paper is organized as follows. The basic model and updating rules are developed in Section 2. Section 3 formalizes the patterns and conditions of equilibrium and convergence. In Section 4, the persistent agent experiment is introduced. Section 5 illustrates simulation results. Section 6 concludes. Proofs and discussions are presented in Appendix.

2. The model

2.1. Agents, opinions, and influence

A finite set $N = \{1, 2, \ldots, n\}$ of agents interact and share *opinions*, which are represented by an $n \times 1$ vector $\mathbf{p}^t = (p_1^t, p_2^t, \ldots, p_n^t)^{\mathrm{T}} \in \mathbb{R}^n$, where p_i^t is agent i's opinion at time t.[2]

An $n \times n$ non-negative matrix $\mathbf{T}$ is referred to as the *influence* matrix. $\mathbf{T}^t$ captures the interaction patterns at time t, i.e., for all $i, j \in N, \mathbf{T}_{ij}^t \in [0, 1]$ indicates the influence weight that agent i places on agent j's opinion at time t. Also, the influence matrix is row-stochastic, i.e.

$$\sum_{j=1}^{n} \mathbf{T}_{ij}^t = 1 \quad \text{and} \quad \mathbf{T}_{ij}^t \geq 0, \quad \text{for all } i, j \in N, \text{ for all } t. \tag{1}$$

Moreover, $\mathbf{T}^t$ may be asymmetric, so that $\mathbf{T}_{ij}^t \neq \mathbf{T}_{ji}^t$ for some i, j.

At this stage, the basic framework settings of agents, opinions and influence are on the same page as Golub and Jackson (2007), except that $\overline{\mathbf{T}}$ in their paper, referred to as the *interaction* matrix, is constant over time.[3]

2.2. Updating processes

At $t = 0$, both opinions and influence weights are arbitrary. That is, each agent i is endowed with an arbitrary initial opinion p_i^0 drawn from a given range. Unlike other social learning models, such as Golub and Jackson (2007), opinions in this model are not correlated with a predesignated "true value" or "real state". Also, each agent receives an arbitrary influence assignment, which is a

[2] Here p_i^t is scalar; whereas in DeMarzo et al. (2003), each agent has a vector of opinions and does multiple estimates, which are interpreted as multidimensional opinions. Their paper also shows that the multidimensional opinions are convergent and thus can be represented on an unidimensional scale.

[3] I use $\overline{\mathbf{T}}$ to denote a time-invariant influence matrix and $\mathbf{T}$ as an influence matrix in general.

row vector with all elements adding up to be 1. The initialization can be formalized as below:

$$\mathbf{p}^0 = (p_1^0, \ldots, p_n^0)^{\mathrm{T}}, \quad \text{where } p_i^0 \in [\underline{M}, \overline{M}] \text{ for all } i \in N; \tag{2}$$

$$\sum_{j=1}^{n} \mathbf{T}_{ij}^0 = 1, \quad \mathbf{T}_{ij}^0 \in [0, 1] \text{ for all } i, j \in N. \tag{3}$$

Updating influence weights

For $t > 0$, the key assumption in redistribution of influence is that an agent places more weights on others who share similar opinions with him. This idea is represented by the concept of distance, which is a simple way to measure the closeness of two agents' opinions. At time t, the distance between agents i and j is defined as:

$d_{ij}^t = |p_i^t - p_j^t|$.

And the basic idea of redistributing influence is:

$$\mathbf{T}_{ij}^{t+1} \propto \frac{1}{d_{ij}^t}, \quad \text{for } j \neq i. \tag{4}$$

Note that d_{ii}^t is always zero. Also, for agents i, j with the same opinion at time t, $d_{ij}^t = 0$. Consequently, applying Eq. (4) is problematic in these cases. It is further assumed that an agent i does not directly change the influence he places on himself. Instead, he only updates the weights on other agents according to the distances. Then agent i normalizes all the weights (including that he puts on his own) to satisfy Eq. (1), which might get $\mathbf{T}_{ii}^t$ modified. As for the cases where agents have very similar opinions, distances less than a small positive number $\underline{d}$ are taken as the same as $\underline{d}$. Thus, the redistribution rule can be written as:

$$\mathbf{T}_{ii}^{t+1} = \frac{\mathbf{T}_{ii}^t}{\mathbf{T}_{ii}^t + \sum_{k \in N_{-i}} w_{ik}^t}, \quad \text{for all } i \tag{5}$$

$$\mathbf{T}_{ij}^{t+1} = \frac{w_{ij}^t}{\mathbf{T}_{ii}^t + \sum_{k \in N_{-i}} w_{ik}^t}, \quad \text{for all } i \text{ and } j \neq i, \tag{6}$$

where N_{-i} denotes the set of agents other than i, $w_{ij}^t = \frac{1}{d_{ij}^t} = \frac{1}{\max(\underline{d}, |p_i^t - p_j^t|)}$, $\underline{d}$ is a very small positive number that prevents taking inversion of 0 when $p_i^t = p_j^t$.

Updating opinions

Then, for $t > 0$, each agent takes a weighted average of others' current opinions in forming his own for the next period. That is, we have the opinions updating rule as:

$$\mathbf{p}^t = \mathbf{T}^t \mathbf{p}^{t-1} \quad \text{for } t > 0. \tag{7}$$

The updating rule in Golub and Jackson (2007) takes a similar form, which is:

$\mathbf{p}^t = \overline{\mathbf{T}}\mathbf{p}^{t-1} = \overline{\mathbf{T}}^t \mathbf{p}^0 \quad \text{for } t > 0.$

However, note that in their model, $\overline{\mathbf{T}}^t$ refers to the t-th power of $\overline{\mathbf{T}}$, i.e. $\overline{\mathbf{T}}^t = \underbrace{\overline{\mathbf{T}} \times \overline{\mathbf{T}} \times \cdots \times \overline{\mathbf{T}}}_{t}$. Here in this paper, $\mathbf{T}^t$ represents the influence matrix at time t, which changes over time and is not computed by raising some initial matrix to powers. The Golub–Jackson rule can be as one of the equivalent forms of my updating rule. That is, by assuming that $\mathbf{T}$ does not change over time, $\mathbf{T}^t = \overline{\mathbf{T}}$, for all t, then $\mathbf{p}^t = \overline{\mathbf{T}}\mathbf{p}^{t-1} = \mathbf{T}^t \mathbf{p}^0$ coincides with (7).

In DeMarzo et al. (2003), the learning process is summarized as

$\mathbf{p}^t = [(1 - \lambda_t)\mathbf{I} + \lambda_t \overline{\mathbf{T}}]\mathbf{p}^{t-1}$,

which can also be transformed to (7) by defining $\mathbf{T}^t = (1 - \lambda_t)\mathbf{I} + \lambda_t \overline{\mathbf{T}}$.

The updating rule in Friedkin and Johnsen (1997) is:

$\mathbf{p}^t = \mathbf{D}\overline{\mathbf{T}}\mathbf{p}^{t-1} + (\mathbf{I} - \mathbf{D})\mathbf{p}^0$,

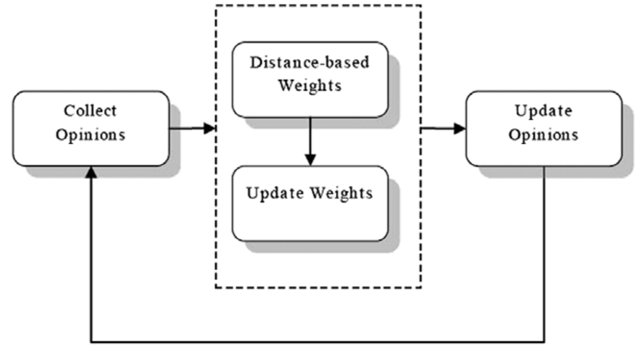


Fig. 1. Updating process of the basic learning model.

where $\mathbf{D}$ is an $n \times n$ matrix with positive diagonal and zero elements elsewhere. The formula can be rewritten as:

$$\mathbf{p}^t = [\mathbf{D}\overline{\mathbf{T}}]^t \mathbf{p}^0 + \sum_{i=0}^{t-1} [\mathbf{D}\overline{\mathbf{T}}]^i (\mathbf{I} - \mathbf{D})\mathbf{p}^0.$$

Note that (7) can be written as:

$\mathbf{p}^t = \mathbf{T}^t \cdot \mathbf{T}^{t-1} \cdots \mathbf{T}^1 \mathbf{p}^0.$

The Friedkin–Johnsen setup can be interpreted as a special case of Eq. (7) with $\mathbf{T}^t = f(t) \times [f(t-1)]^{-1}$, where $f(t) = [\mathbf{D}\overline{\mathbf{T}}]^t + \sum_{i=0}^{t-1} [\mathbf{D}\overline{\mathbf{T}}]^i (\mathbf{I} - \mathbf{D})$.[4]

A period ends after both the opinion vector and influence matrix are properly updated.[5] The updating process repeats during each period, as shown in Fig. 1. So then the question is: does the learning process go anywhere? Or in other words, do opinions and/or the influence matrix converge?

3. Convergence

3.1. Preliminaries

With the basic ideas of influence redistribution, agents tend to place higher weights on similar opinions and those that are relatively different from one's current opinions will have lower weights. So one might guess that diversity would emerge and that it would be possible to sustain islands of different opinions, since an agent does not pay much attention to faraway opinion (Krause, 2000). However, the opposite is true. That is, although there is no "true value" of opinions in this framework, eventually agents will approach to hold approximately the same opinion. First, convergence of opinion and the characteristics of the final outcome are defined as follows.

Definition 1. An opinion vector sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ is *convergent* if $\lim_{t\to\infty} \mathbf{p}^t$ exists. i.e. there exists $\mathbf{p}^* \in \mathbb{R}^n$ s.t. for all $\epsilon > 0$, there exists $t^* > 0$ s.t. $\|\mathbf{p}^t - \mathbf{p}^*\| \le \epsilon$ for $t > t^*$, where for a $1 \times n$ vector x, $\|x\| = \sum_{i=1}^{n} x_i^2$ is defined as its norm.

Definition 2. An opinion vector sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ is *conforming* if there exists $p^* \in \mathbb{R}$ s.t. for all $\epsilon > 0$, there exists $t^* > 0$ s.t. $|p_i^t - p^*| \le \epsilon$ for $t > t^*$, for all $i \in N$.

---

[4] The $\mathbf{T}$ matrices in Golub and Jackson (2007) and DeMarzo et al. (2003) are both stochastic as well. However, it is not necessarily the case in Friedkin and Johnsen (1997).

[5] Initially, following the convention of evolution literature, it was also assumed that mutation actions would be taken. That is, with probability θ, an agent will violate the redistribution rules when updating weights. For example, he might double the influence he places on certain agents, regardless of the actual distances. However, simulation results have shown sound evidence that this mutation process has no significant effects on the final outcomes. Thus, in this paper $\theta = 0$.

Note that $\mathbf{p}^*$ is a vector, whereas p^* is a number. That is, conformism means that in the long run, all agents' opinions converges to the same p^*, which is a stronger argument than convergence. Especially, if at certain time t we have $p_i^t = p^*$ for all i, then the influence distribution will no longer affect opinions, in that the weighted average will always remain to be p^*: the society reaches a steady stage given a consensus of opinions.

Similarly, the convergence of influence matrix is defined below. Based on these definitions, conditions for convergence and characterizations of the final patterns are discussed in the next subsection.

Definition 3. An influence matrix sequence $\{\mathbf{T}^t\}_{t=1}^{\infty}$ is *convergent* if $\lim_{t\to\infty} \mathbf{T}^t$ exists. i.e. there exists $\mathbf{T}^*$ such that for all $\epsilon > 0$, there exists $t^* > 0$ s.t. $\|\mathbf{T}^t - \mathbf{T}^*\| \le \epsilon$ for $t > t^*$, where for an $l \times m$ matrix M, $\|M\| = \sum_{i=1}^{l}\sum_{j=1}^{m}(M_{ij})^2$ is defined as its norm.

3.2. Characterizations and conditions

In the standard Markov theory literature, it is addressed that a transition matrix is convergent if and only if $\lim_{t\to\infty} \overline{\mathbf{T}}^{(t)}\mathbf{p}$ exists for all vectors $\mathbf{p}$.[6] As it has been pointed out in Section 2, $\overline{\mathbf{T}}^{(t)}\mathbf{p}$ can be interpreted as the updated $\mathbf{p}^t$ with a time-invariant $\overline{\mathbf{T}}$ matrix and initial value of $\mathbf{p}$. On the other hand, in my model where $\mathbf{T}^t$ is time-varying, $\mathbf{p}^t = \mathbf{T}^t\mathbf{p}^{t-1} = \mathbf{T}^t \cdot \mathbf{T}^{t-1} \cdots \mathbf{T}^1\mathbf{p}^0$, which makes the setting different and we cannot directly apply the convergent condition in this case.

Theorem 1. *For* $\mathbf{p}^t = \prod_{\tau=t}^{1} \mathbf{T}^\tau \mathbf{p}^0$, *the existence of* $\lim_{t\to\infty}\mathbf{T}^t$ *does not imply that* $\{\mathbf{p}^t\}_{t=1}^{\infty}$ *is convergent and vice versa.*

Theorem 1 states that in general, if both $\mathbf{p}^t$ and $\mathbf{T}^t$ change over time, then we cannot guarantee the equivalence of convergence between the two. Note that Theorem 1 does not impose any restrictions on $\mathbf{T}^t$. On the other hand, the next theorem shows that with the distance-based updating rule, conformism emerges. Please refer to Appendix for proof of this theorem.

Theorem 2. *For* $n > 2$, *with updating rules defined,* $\{\mathbf{p}^t\}_{t=1}^{\infty}$ *is conforming for all row-stochastic* $\mathbf{T}^0$ *and all* $\mathbf{p}^0$.[7]

Theorem 2 applies to societies with more than 2 members. The case where $n = 1$ is trivial. For $n = 2$, the opinion vector $\mathbf{p}^t$ is *not* confirming if $\mathbf{T}^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, because in this case the 2 agents will be switching opinions and flip-flopping between the 2 initial opinions. For the case where $n > 2$, please refer to Appendix for proof of convergence.

With the distance-based updating rule, when conformism occurs, the tentative new weights w_{ij}^t all reach the same upper bound value $\frac{1}{\underline{d}}$. Then after normalization, all $\mathbf{T}_{ij}^{t+1}$ are equal. Theorem 3 below formalizes the argument.

Theorem 3. *For* $n > 2$, *with the updating rules defined,* $\{\mathbf{T}^t\}_{t=1}^{\infty}$ *is convergent for all row-stochastic* $\mathbf{T}^0$ *and all* $\mathbf{p}^0$. *Namely,* $\lim_{t\to\infty}\mathbf{T}^t = \mathbf{T}^*$ *where* $\mathbf{T}_{ii}^* = 0$, $\mathbf{T}_{ij}^* = \frac{1}{n-1}$, *for all i and* $j \ne i$.

Theorem 3 describes the patterns that influence matrix converges to. That is, agents place the same weight on all the other agents and zero weight on himself. Since the elements in the matrix reflect the level of influence that one has on others, in the steady state we have a very equal society in which consensus is reached and influence is evenly distributed.

6 Here $\overline{\mathbf{T}}^{(t)}$ is the power function of $\mathbf{T}$ (Golub and Jackson, 2007), whereas the notation $\mathbf{T}^t$ is used in this paper to indicate $\mathbf{T}$ at time t. The two do not necessarily equal.

7 I appreciate Dr. Hans Haller's contribution to the proof of this Theorem.

4. Interaction effects: an experiment

4.1. Persistent agents

Obviously the updating rules that define the interaction among agents have significant impact on the final outcomes. Thus a minority group of so-called "persistent agents", have been added to the basic setup to examine the interaction effects. Those agents are "persistent" in the sense that they insist on the initial weights they assign to others, or their own opinions, or both. We may see them as individuals who do not have access to others' opinions, or do not want to make any changes. There are three different types of persistent agents.

Type I persistent agents do not update their influence weights. However, type I persistent agents follow the opinions updating rules and take weighted average of others' opinions, which means that their opinions might change over time. On the other hand, type II persistent agents insist on their initial opinions. Although they would redistribute influence, those weights practically have no impact on their opinions. The third type of persistent agents are called "double-persistent" agents, in that they update neither influence nor opinions. Note that different types of persistent agents do not co-exist.

Furthermore, to better illustrate the interaction effects, the society is assumed to be polarized, i.e. at $t = 0$, instead of having arbitrary initial opinions, non-persistent and persistent agents hold 2 different opinions, while the members in each category share the same initial opinion.

4.2. Modified interaction dynamics and convergence

In this extended model, we have a finite set $N = \{1, 2, \ldots, n\}$ of agents, $\sigma \ge 1$ of which are persistent and do not fully interact with others. Since the persistent agents are a minority in the society, we assume that $\frac{\sigma}{n} \le 10\%$. Without loss of generality, we arrange the order of the agents in such a way that number 1 to $n - \sigma$ agents are non-persistent and the rest σ are persistent.

At $t = 0$, $p_{\bar{s}_i}^0 = p_{\bar{s}}^0$ for all $\bar{s}_i \in \bar{S}$;[8] $p_{s_i}^0 = p_s^0 \ne p_{\bar{s}}^0$ for all $s_i \in S$. And an arbitrary positive[9] row-stochastic $n \times n$ influence matrix $\mathbf{T}^0$ shows the initial weight assignment.

For $t > 0$, there are 3 variations of the interaction dynamics in this model, respectively to the 3 different types of persistent agents. We use S_i and σ_i $(i = 1, 2, 3)$ to denote the sets of numbers of the 3 types of persistent agents.

Type I *persistent agents*

We have $S_1 \subset N$, $|S_1| = \sigma_1$. For $t > 0$, $\mathbf{T}^t$ and $\mathbf{p}^t$ are updated based on following rules.

(i) For $\bar{s} \in \overline{S_1}$, the updating rules for $\mathbf{T}^t$ take the same form as Eqs. (5) and (6):

$$\mathbf{T}_{\bar{s}\bar{s}}^{t+1} = \frac{\mathbf{T}_{\bar{s}\bar{s}}^t}{\mathbf{T}_{\bar{s}\bar{s}}^t + \sum_{j\in N_{-\bar{s}}} w_{\bar{s}j}^t}, \tag{5'}$$

$$\mathbf{T}_{\bar{s}j}^{t+1} = \frac{w_{\bar{s}j}^t}{\mathbf{T}_{\bar{s}\bar{s}}^t + \sum_{j\in N_{-\bar{s}}} w_{\bar{s}j}^t} \quad \text{for all } j \ne \bar{s}, \tag{6'}$$

where

$$w_{\bar{s}j}^t = \frac{1}{d_{\bar{s}j}^t} = \frac{1}{\max(\underline{d}, |p_{\bar{s}}^t - p_j^t|)}.$$

8 Here for a set S, $\bar{S} = N - S$.

9 Note that for this variation, the initial influence matrix is assumed to be positive instead of non-negative. Because for one special case where $\mathbf{T}_{s\bar{s}}^0 = 0$ for all $s \in S_1$ and $\bar{s} \in \overline{S_1}$, the persistent agents are essentially of type II instead of type I.

(ii) For $s \in S_1$ we have:

$$\mathbf{T}^t_{sj} = \mathbf{T}^0_{sj}, \quad \text{for all} \quad j. \tag{8}$$

And opinion updating is the same as that in the basic model indicated by Eq. (7), i.e. $\mathbf{p}^t = \mathbf{T}^t\mathbf{p}^{t-1}$ for $t > 0$.

Theorem 4. *For societies with type* I *persistent agents,* $\{\mathbf{p}^t\}_{t=1}^{\infty}$ *is conforming, i.e.* $\lim_{t\to\infty}\mathbf{p}^t = (p^*, \dots, p^*)^{\mathrm{T}}$. *Also,* $\{\mathbf{T}^t\}_{t=1}^{\infty}$ *is convergent, for all non-persistent agents* $\bar{s} \in \overline{S_1}$, *for all* $s' \neq \bar{s}$, $\lim_{t\to\infty}\mathbf{T}^t_{\bar{s}\bar{s}} = 0$ *and* $\lim_{t\to\infty}\mathbf{T}^t_{\bar{s}s'} = \frac{1}{n-1}$.

That is, similar to the basic model, the initially polarized opinions in a society with type I persistent agents conform to the same value, which results in equal distribution of influence weights. In addition, simulations show that the final opinion in this case is closer to non-persistent agents' initial opinion. More on this issue is explained in Proposition 1 later in this section and discussions in the simulation section. Proofs can be found in Appendix.

Type II *persistent agents*

We have $S_2 \subset N$, $|S_2| = \sigma_2$. For $t > 0$, $\mathbf{T}^t$ and $\mathbf{p}^t$ are updated based on following rules.

For all $i \in N$, the updating process of $\mathbf{T}^t$ is the same as it is in the basic model, i.e. we apply Eqs. (5) and (6):

$$\mathbf{T}^{t+1}_{ii} = \frac{\mathbf{T}^t_{ii}}{\mathbf{T}^t_{ii} + \sum_{j\in N_{-i}} w^t_{ij}}, \quad \text{for all } i, \tag{5}$$

$$\mathbf{T}^{t+1}_{ij} = \frac{w^t_{ij}}{\mathbf{T}^t_{ii} + \sum_{j\in N_{-i}} w^t_{ij}} \quad \text{for all } j \neq i, \tag{6}$$

where

$$w^t_{ij} = \frac{1}{d^t_{ij}} = \frac{1}{\max(\underline{d}, |p^t_i - p^t_j|)}.$$

Then, agents in S_2 do not update their opinions, i.e.

(i) For $\bar{s} \in \overline{S_2}$, we use Eq. (7):

$$p^t_{\bar{s}} = \sum_{j=1}^{n} \mathbf{T}^t_{\bar{s}j} \cdot p^{t-1}_j. \tag{7$'$}$$

(ii) For $s \in S_2$, persistent agents insist on their initial opinions:

$$p^t_s = p^0_s. \tag{9}$$

Theorem 5. *For societies with type* II *persistent agents,* $\{\mathbf{p}^t\}_{t=1}^{\infty}$ *is conforming to persistent agents' initial opinion, i.e.* $\lim_{t\to\infty}\mathbf{p}^t = (p^0_s, \dots, p^0_s)^{\mathrm{T}}$. *Also,* $\{\mathbf{T}^t\}_{t=1}^{\infty}$ *is convergent,* $\lim_{t\to\infty}\mathbf{T}^t_{ii} = 0$ *and* $\lim_{t\to\infty}\mathbf{T}^t_{ij} = \frac{1}{n-1}$, *for all* $i \in N$, *for all* $j \neq i$.

Theorem 5 states that type II persistent agents are able to drive the whole society to converge to their initial opinions, by not changing their views. This shows the dominant influential power that a minority group have and reflects the interesting fact that when people are constantly exposed to something, they are highly likely to believe that information even though they used to have a completely different point of view. Please refer to Appendix for the proof of this theorem.

Double-persistent Agents

We have $S_3 \subset N$, $|S_3| = \sigma_3$. For $t > 0$, $\mathbf{T}^t$ and $\mathbf{p}^t$ are updated based on following rules.

(i) For $\bar{s} \in \overline{S_3}$, we use Eqs. (5) and (6) to update $\mathbf{T}^t$:

$$\mathbf{T}^{t+1}_{\bar{s}\bar{s}} = \frac{\mathbf{T}^t_{\bar{s}\bar{s}}}{\mathbf{T}^t_{\bar{s}\bar{s}} + \sum_{j\in N_{-\bar{s}}} w^t_{\bar{s}j}}, \tag{5$'$}$$

$$\mathbf{T}^{t+1}_{\bar{s}j} = \frac{w^t_{\bar{s}j}}{\mathbf{T}^t_{\bar{s}\bar{s}} + \sum_{j\in N_{-\bar{s}}} w^t_{\bar{s}j}} \quad \text{for all } j \neq \bar{s}, \tag{6$'$}$$

where

$$w^t_{\bar{s}j} = \frac{1}{d^t_{\bar{s}j}} = \frac{1}{\max(\underline{d}, |p^t_{\bar{s}} - p^t_j|)}.$$

Updating rule of opinions follows Eq. (7):

$$p^t_{\bar{s}} = \sum_{j=1}^{n} \mathbf{T}^t_{\bar{s}j} \cdot p^{t-1}_j. \tag{7$'$}$$

(ii) For $s \in S_3$, double-persistent agents do not update influence weights or opinions, i.e. we apply Eqs. (8) and (9):

$$\mathbf{T}^t_{sj} = \mathbf{T}^0_{sj}, \quad \text{for all } j; \quad \text{and} \quad p^t_s = p^0_s.$$

Theorem 6. *For societies with double-persistent agents,* $\{\mathbf{p}^t\}_{t=1}^{\infty}$ *is conforming to persistent agents' initial opinion, i.e.* $\lim_{t\to\infty}\mathbf{p}^t = (p^0_s, \dots, p^0_s)^{\mathrm{T}}$. *Also,* $\{\mathbf{T}^t\}_{t=1}^{\infty}$ *is convergent,* $\lim_{t\to\infty}\mathbf{T}^t_{\bar{s}\bar{s}} = 0$ *and* $\lim_{t\to\infty}\mathbf{T}^t_{\bar{s}s'} = \frac{1}{n-1}$, *for all* $\bar{s} \in \overline{S_3}$, *for all* $s' \neq \bar{s}$.

This variation seems to be a combination of the other two. Actually the double-persistent and type II persistent agents have the same impact on the updating process of the whole population, since the change in type II persistent agents' influence weights virtually plays no role. The key here is the persistence on initial opinions. If persistent agents insist on their initial opinion, then whether they update their influence weights or not, the final opinion conforms to their initial view.

Moreover, given a certain initiation pattern, we can describe the convergence speed in case of persistent agents, as presented in the proposition below. Namely, if non-persistent agents are endowed with the same diagonal weight at the initial state, their opinions change very little during each period of updating. Actually, the change is bounded upwards by the cut-off value that we use to prevent taking inverse of zero-distance.

Recall that contrary to the outcomes with type II and double-persistent agents, type I persistent agents do not change their influence weights but do listen to others and change their opinions accordingly. In that case the final standpoint of an initially polarized society is closer to the opinion that the majority holds in the beginning. In this convex updating process, persistent and non-persistent agents are moving towards a point in between their initial opinions; persistent agents take much bigger steps and therefore they meet near non-persistent agents' starting point.

Proposition 1. *Let the non-persistent agents place equal weight on themselves in the initial period, i.e,* $\mathbf{T}^0_{ii} = \mathbf{T}^0_{jj}$ *for all* $i, j \notin S$. *Then, the change of opinion of non-persistent agents from one period to the next is less than* $\underline{d}$, *i.e.,* $|p^{t+1}_i - p^t_i| < \underline{d}$ *for all* $i \notin S$ *for all* t.

5. Simulation

By Theorem 2, conformity indicates a steady state of the updating process. It also results in the even distribution of influence weights that $\mathbf{T}^t$ converges to. Speed of convergence is not covered in the theoretical analysis and is recorded during simulations, where 5 different society sizes are used: $n \in \{20, 40, 60, 80, 100\}$. The standard deviation of opinions p^t_i is checked every n periods of updating. Simulation is terminated once conformity is detected, represented by a standard deviation of 0. The cut-off point $\underline{d} = 10^{-3}$ for all sizes.

5.1. Basic model

Recall the complete interaction dynamics discussed in Section 2. At $t = 0$ agents are endorsed with arbitrary initial influence

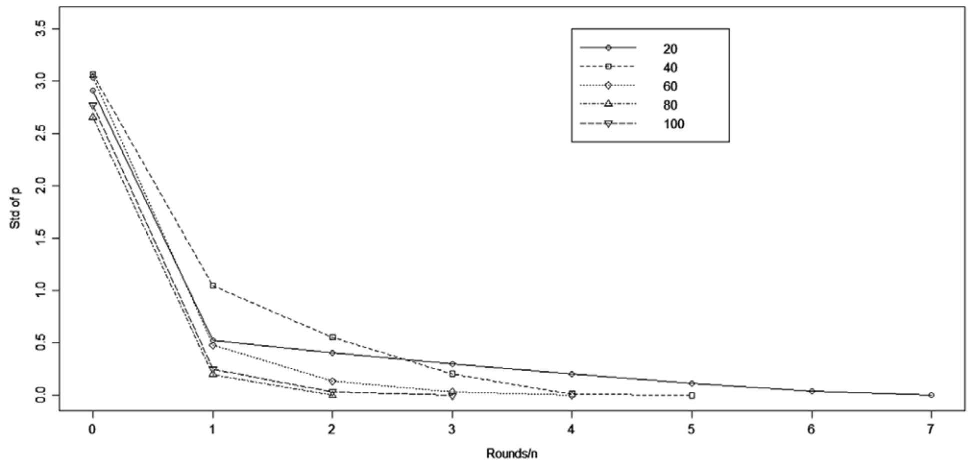


Fig. 2. Convergence paths of the basic model.

assignments and for the simulation, initial opinions are randomly drawn from 0 to 10. Then agents redistribute the weights and update their opinions during each period.

Conforming to Theorem 2, all the simulation results show that convergence of opinions occurs; and all agents hold the same opinion at the steady state. The influence matrix $\mathbf{T}^t$ also converges as Theorem 3 states. Fig. 2 shows the convergence paths of opinions of 5 different society sizes: 20, 40, 60, 80, and 100.[10] The y-axis shows the standard deviation of $p_1^t, \ldots, p_n^t$, which converges to 0. The x-axis shows total rounds adjusted to size, i.e. $x = \frac{t}{n}$. For instance, we see that when $n = 20$, conformity is reached at $x = 7$; whereas for $n = 40$, conformity occurs at $x = 5$. It means that for a 20-agent society, the learning process stabilizes after $7 * 20 = 140$ rounds of updating, and it takes $5 * 40 = 200$ rounds of updating for a 40-agent society to conform. Indeed, it takes about 200 rounds of updating for all 5 sizes. A bigger society does not necessarily imply slower convergence speed.

5.2. Persistent experiment

In this experiment, the initial opinions of normal non-persistent agents are assigned to be 0, while those of persistent agents (all three types) are assigned to be 10; so that it would be easier to see the interaction effects.

During each round of the simulation, a control group is also conducted. We introduce so-called "placebo" agents instead of persistent agents in the control group, and the initial opinions follow the 0–10 assignment rule. However, the placebo agents in a control group behave the same as the normal agents: they update both influence and opinions.

Both opinions and the influence matrix converge as stated in the theorems. The focus of the experiment is the outcome of final opinions. Proposition 1 shows that non-persistent agents' opinions change very little if they have equal initial weights on themselves. Even when the initial weights are completely random and $\mathbf{T}_{ii}^0 \neq \mathbf{T}_{jj}^0$ for some i, j, observations of the simulation indicate that the change of non-persistent agents' opinions is still very small in cases of all 3 types of persistence. Consequently, it takes very long time to reach the steady state when persistent agents are of type II or double-persistent. Besides, it is true that the convergence paths of opinions with type II and double-persistent agents are not different.

Table 1
Summary of final opinions by different agent types, $n = 20$.

Agent type	Convergence speed	Final opinion
Control	29	0.06588
Type I persistent	1	0.0006
Type II persistent	4253	10
Double-persistent	4253	10

Results of simulations with $n = 20$ and 10% persistent agents are shown in Table 1, where convergence speed is denoted by number of checks, which is executed very n rounds.[11] Recall that persistent agents have initial opinions of 10 and normal ones have 0, it clearly shows that persistent agents could "drag" the final opinions to converge closer to 10, if they stick to their initial opinions. On the other hand, keeping the initial influence distribution does not have much impacts. Note that $\underline{d} = 10^{-3}$ and the final opinion with control and type I persistent agents is $6.6 * 10^{-2}$ and $6 * 10^{-4}$, respectively.

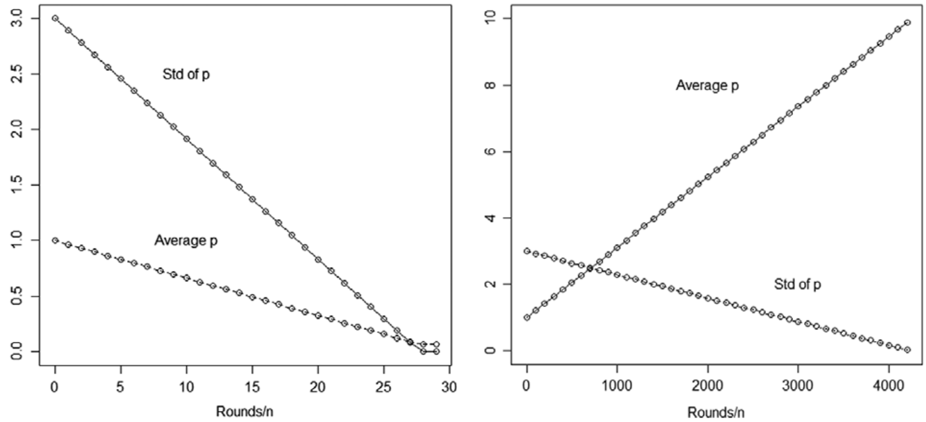
Fig. 3 illustrates the different convergence paths with placebo and type II/double-stubborn persistent agents. We see in that in both cases, the standard deviation of opinions shows a downward slope towards 0 which indicates convergence to conformity. However, average opinion of the control group goes from 1.0 to near 0 in subfigure (a); whereas that with agents persistent on their opinions forms an upward slope line from 1.0 to 10 in subfigure (b). That is, agents learn to conform on the majority's initial opinion (i.e. 0) with the absence of persistence, and the whole group converge to minority's initial opinion (i.e. 10) when the minority do not change their opinions.

6. Conclusion

I introduced a social learning framework where agents update not only their opinions but also the influence weights they place on others. The basic assumptions used for the interaction dynamics are that (1) a non-negative row-stochastic matrix indicates the influence weights that agents place on each other; (2) during each period, an agent takes the weighted average of others' opinions as his updated opinion for the next period; they also redistribute the weights on others, based on the rule that weights are proportional to the closeness of opinions.

[10] Multiple simulations were executed for each size and with a different initial condition every time. Convergence of both opinions and the influence matrix shows significant robustness. The figure shows the convergence paths of one simulation (initial condition).

[11] Different sizes are used and the outcomes do not show any significantly different patterns. For comparison purposes, for a given society size and persistent percentage, the same random initial weights are assigned. Since the initial opinions are also identical in this case, we have the same initial conditions for each combination of numbers of persistent and non-persistent agents.



(a) Control. (b) Type II/Double-stubborn.

Fig. 3. Average and standard deviation of opinions, $n = 20$.

It is shown that with the basic learning model, we observe convergence of both opinions and the influence matrix. Moreover, opinions are conforming, and influence weights show an even distribution. Especially, although one pays the least attention to the farthest opinion holders, we do not observe islands of different opinions. Rather, in the steady state we have consensus: everyone holds the same opinion.

An experiment with three types of persistent agents is conducted to concentrate on the different effects that opinion updating and influence updating have on the final outcomes. Initial opinions in this experiment are assumed to be polarized between persistent and non-persistent agents in order to better illustrate the effects of interaction by observation of the final opinions. Non-persistent agents' opinions change very little over time. If persistent agents insist on their weights but update their opinions, consensus is reached near non-persistent agents' initial opinion. Otherwise, the society slowly converges to the opinion of persistent agents, even if there is only one agent who is persistent on his initial opinion.

The persistent experiment shows the tremendous influence possessed by a minority group. It is argued in Watts and Dodds (2007) that the key is not the power of a minority to be influential, but rather how influenceable the majority is. Watts and Dodds' argument suggests an interesting angle to look into the results where we have persistent agents. That is, the non-persistent agents (majority) always assign a (small) positive weight on persistent agents, and take the weighted average of all opinions to update their own. This indeed makes normal agents very influenceable.

I believe these findings can be used for meaningful applications, such as the field experiment on consumer behavior (Leider et al., 2007) and politics (DeMarzo et al., 2003). Network structure and social positions are not fully discussed in this paper; whereas they also impose greatly on the path of social learning (Bala and Goyal, 2001; Friedkin and Johnsen, 1997). Also, interaction network and information network are not always the same and should be separated (Durieu and Solal, 2003; Baron et al., 2002). In later papers, I incorporate network structures and game theory analysis into my model (Pan and Gilles, 2009; Pan, 2009). Also, in Pan and Gilles (2009), unlike the uniform persistent agents in this model, persistent agents may have different opinions among themselves.

Appendix

Proof of Theorem 1

For the first part, suppose we have a sequence $\{\mathbf{T}^t\}_{t=1}^{\infty}$ that

$$\mathbf{T}^t = \bar{\mathbf{T}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{for all } t.$$

Then $\{\mathbf{T}^t\}_{t=1}^{\infty}$ is obviously convergent since $\mathbf{T}^t$ is time-invariant.

However, $\prod_{\tau=t}^{1} \mathbf{T}^{\tau} = \bar{\mathbf{T}}^{(t)} = \begin{cases} \bar{\mathbf{T}} & \text{if } t \text{ is odd,} \\ \mathbf{I} & \text{if } t \text{ is even.} \end{cases}$

That is, $\mathbf{p}^t = \prod_{\tau=t}^{1} \mathbf{T}^{\tau}\mathbf{p}^0$ is not convergent, although $\mathbf{T}^t$ is.
Next, another counter example that shows the second part of the Theorem. Suppose we have a sequence $\{\mathbf{T}^t\}_{t=1}^{\infty}$ that $\mathbf{T}^t = \begin{cases} \bar{\mathbf{T}} & \text{if } t \text{ is odd,} \\ \mathbf{I} & \text{if } t \text{ is even.} \end{cases}$ where $\bar{\mathbf{T}} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$. $\mathbf{T}^t$ is not convergent as it is switching between two matrices.

On the other hand, we see that after the first round of updating,

$$p_1^1 = p_2^1 = p_3^1 = \frac{1}{3}\sum_{i=1}^{3} p_i^0 = p^*.$$

Consequently, regardless of the influence weights, the opinions of all three agents remain the same after that period. Because they are taking weighted averages of the constant p^* and always end up with the opinion of p^*.

In other words, in this case, the sequence $\{\mathbf{p}^t\}_{t=1}^{\infty}$ is convergent although $\{\mathbf{T}^t\}_{t=1}^{\infty}$ is not:

$$\lim_{t\to\infty} \mathbf{p}^t = \lim_{t\to\infty} \prod_{\tau=t}^{1} \mathbf{T}^{\tau}\mathbf{p}^0 = (p^*, p^*, p^*)^{\mathrm{T}},$$

where $p^* = \frac{1}{3}\sum_{i=1}^{3} p_i^0$.

Proof of Theorem 2

Theorem 3.1 in Seneta (1981) states that

$$\max_{i,j} |p_i^{t+1} - p_j^{t+1}| \le \mu_t(\mathbf{T})\{\max_{i,j} |p_i^t - p_j^t|\},$$

where

$$\mu_t(\mathbf{T}) = \frac{1}{2}\max_{i,j}\sum_{k=1}^{n} |\mathbf{T}_{ik}^t - \mathbf{T}_{jk}^t|. \tag{10}$$

By induction it follows that

$$\max_{i,j} |p_i^{t+1} - p_j^{t+1}| \le \left[\prod_{\tau=t}^{1} \mu_\tau(\mathbf{T})\right] \max_{i,j} |p_i^0 - p_j^0|.$$

Since at $t = 0$, all the opinions p_i^0 are arbitrarily drawn from a bounded set, $\max_{i,j} |p_i^0 - p_j^0|$ has an upper bound. Therefore, if can be shown that $\mu_\tau(\mathbf{T})$ is less than and bounded away from 1 for all

τ, then it follows that $\lim_{t\to\infty}\prod_{\tau=t}^{1}\mu_\tau(\mathbf{T})=0$.[12] That implies that $\max_{i,j}|p_i^{t+1}-p_j^{t+1}|\to 0$ as $t\to\infty$, i.e., $|p_i^{t+1}-p_j^{t+1}|$ converges to 0 and the opinion sequence is conforming.

Since $\mathbf{T}^t$ is row-stochastic, we have $\sum_{k=1}^n \mathbf{T}_{ik}^t = 1$, and $\sum_{k=1}^n \mathbf{T}_{jk}^t = 1$. Denote $\delta_{ijk}^{t+1} = \mathbf{T}_{ik}^{t+1} - \mathbf{T}_{jk}^{t+1}$. It follows that

$$\sum_{k=1}^n \delta_{ijk}^t = \sum_{k=1}^n \mathbf{T}_{ik}^t - \sum_{k=1}^n \mathbf{T}_{jk}^t = 0. \tag{11}$$

One can divide the set of indices k into two categories. Define

$$D_+^{ijt} = \{k\in N|\delta_{ijk}^t \ge 0\}, \qquad D_-^{ijt} = \{k\in N|\delta_{ijk}^t \le 0\}.$$

From (11), we have

$$\sum_{k\in D_+^{ijt}} \delta_{ijk}^t = -\sum_{k\in D_-^{ijt}} \delta_{ijk}^t.$$

Hence, (10) can be rewritten as

$$\begin{aligned}\mu_t(\mathbf{T}) &= \frac{1}{2}\max_{i,j}\sum_{k=1}^n |\mathbf{T}_{ik}^t - \mathbf{T}_{jk}^t| \\ &= \frac{1}{2}\max_{i,j}\sum_{k=1}^n |\delta_{ijk}^t| \\ &= \frac{1}{2}\cdot 2\max_{i,j}\sum_{k\in D_+^{ijt}} \delta_{ijk}^t \\ &= \max_{i,j}\sum_{k\in D_+^{ijt}} \delta_{ijk}^t \\ &= \max_{i,j}\sum_{k\in D_+^{ijt}} (\mathbf{T}_{ik}^t - \mathbf{T}_{jk}^t).\end{aligned}$$

Define

$$\nu_{ij}^t = \sum_{k\in D_+^{ijt}} \delta_{ijk}^t = \sum_{k\in D_+^{ijt}} \mathbf{T}_{ik}^t - \sum_{k\in D_+^{ijt}} \mathbf{T}_{jk}^t.$$

Since $\mu_t(\mathbf{T}) = \max_{i,j}\nu_{ij}^t$, if we show ν_{ij}^t is less than and bounded away from 1 for all i,j,t, the proof is complete. Next, we show that for an $l\in D_+^{ijt}$, $\sum_{k\in D_+^{ijt}}\mathbf{T}_{jk}^t \ge \mathbf{T}_{jl}^t > \delta_c$ where $0<\delta_c<1$. This combined with $\sum_{k\in D_+^{ijt}}\mathbf{T}_{ik}^t \le 1$, we have $\sum_{k\in D_+^{ijt}}\mathbf{T}_{ik}^t - \sum_{k\in D_+^{ijt}}\mathbf{T}_{jk}^t < 1-\delta_c$. That is, ν_{ij}^t is less than and bounded away from 1.

Denote for each t,

$$\underline{p}^t = \min\{p_1^t,\ldots,p_n^t\}, \qquad \overline{p}^t = \max\{p_1^t,\ldots,p_n^t\}.$$

Thus $p_i^t \in [\underline{p}^t,\overline{p}^t]$ for all i. Recall that in the initialization period $t=0$, $p_i^0\in[\underline{M},\overline{M}]$ for all i, therefore $\overline{p}^0$ and $\underline{p}^0$ exists. Note that since p_i^{t+1} is a convex combination of p_i^t, $p_i^{t+1}\in[\underline{p}^t,\overline{p}^t]$ for all i. That is, $\{\mathbf{p}^t\}_{t=0}^\infty$ is a sequence in a compact set $[\underline{M},\overline{M}]^n$.

Moreover, $\overline{p}^0-\underline{p}^0 \le \overline{M}-\underline{M}$. Denote $\check{p}=\overline{M}-\underline{M}$, then we have $|p_i^t-p_j^t|<\check{p}$, for all i,j,t. For $n>2$, it must be true that either $|D_+^{ijt}|\ge 2$ or $|D_-^{ijt}|\ge 2$. Without loss of generality, assume that $|D_-^{ijt}|\ge 2$. Thus, there exists $l\in D_-^{ijt}$, such that $l\ne j$. Choose such an l. We now have $d_{jl}^t\in[\underline{d},\check{p}]$, for all j,t. So the interim weight $w_{jl}^t\in[\frac{1}{\check{p}},\frac{1}{\underline{d}}]$. Also, $\mathbf{T}_{jj}^t\le 1$ for all j,t. Therefore,

$$\mathbf{T}_{jl}^{t+1} = \frac{w_{jl}^t}{\mathbf{T}_{jj}^t + \sum_{i\in N_{-j}} w_{ji}^t} > \frac{\frac{1}{\check{p}}}{1+\frac{n-1}{\underline{d}}} \gg 0.$$

Denote $\frac{\frac{1}{\check{p}}}{1+\frac{n-1}{\underline{d}}}$ as δ_c. Then for all i,j,t,

$$\nu_{ij}^t = \sum_{k\in D_+^{ijt}}\delta_{ijk}^t = \sum_{k\in D_+^{ijt}}\mathbf{T}_{ik}^t - \sum_{k\in D_+^{ijt}}\mathbf{T}_{jk}^t < 1-\mathbf{T}_{jl}^t < 1-\delta_c.$$

Therefore, $\mu_t(\mathbf{T}) = \max_{i,j}\nu_{ij}^t < 1-\delta_c$. That is, μ_t is less than and bounded away from 1. From the discussion above, we have $|p_i^{t+1}-p_j^{t+1}|\to 0$ since $\max_{i,j}|p_i^{t+1}-p_j^{t+1}| \le \prod_{\tau=t}^{1}\mu_\tau(\mathbf{T})\max_{i,j}|p_i^0-p_j^0|$.

Next, show that $\mathbf{p}^t$ is conforming. Recall that $\{\mathbf{p}^t\}_{t=0}^\infty$ is a sequence in a compact set $[\underline{M},\overline{M}]^n$. Therefore, $\{\mathbf{p}^t\}_{t=0}^\infty$ has a subsequence that is convergent. Suppose that the subsequence is convergent to $\mathbf{p}^*$. If $\mathbf{p}^t$ does not converge to $\mathbf{p}^*$, then there exists another subsequence of $\{\mathbf{p}^t\}_{t=0}^\infty$ that is convergent to $\mathbf{p}^{**}\ne\mathbf{p}^*$.

We have shown that $p_i^t-p_j^t\to 0$ for all i,j, so $\mathbf{p}^*=(p',\ldots,p')$ and $\mathbf{p}^{**}=(p'',\ldots,p'')$. Also, $\overline{p}^t-\underline{p}^t\to 0$. For $\epsilon=|p'-p''|$, by definition, there exists t', s.t. for $t\ge t'$, $\overline{p}^t-\underline{p}^t<\epsilon$, i.e. $|p'-p''|<\epsilon$ since $p',p''\in[\underline{p}^t,\overline{p}^t]$. Therefore, $p'=p''$. That is, $\mathbf{p}^t$ is conforming, $\lim_{t\to\infty}\mathbf{p}^t=(p^*,\ldots,p^*)^{\mathrm{T}}$.

Proof of Theorem 3

By Theorem 2, there exists p^* such that for $\epsilon=\frac{\underline{d}}{2}$, there exists $t^*>0$ such that $|p_i^t-p^*|\le\epsilon$ for $t>t^*$, for all $i\in N$.

Hence for $t>t^*$, $|p_i^t-p_j^t|\le 2\epsilon=\underline{d}$, for all i,j. By the given updating rule, for all $j\ne i$, $d_{ij}^t=\underline{d}$. That is, $w_{ij}^t=\frac{1}{\underline{d}}$, for all $j\ne i$. Thus for all i and $j,k\ne i$,

$$\mathbf{T}_{ij}^{t+1}=\mathbf{T}_{ik}^{t+1}=\frac{\frac{1}{\underline{d}}}{\mathbf{T}_{ii}^t+\frac{n-1}{\underline{d}}}, \qquad \mathbf{T}_{ii}^{t+1}=\frac{\mathbf{T}_{ii}^t}{\mathbf{T}_{ii}^t+\frac{n-1}{\underline{d}}}.$$

Then for all $t>t^*$ we have

$$\frac{\mathbf{T}_{ii}^{t+1}}{\mathbf{T}_{ii}^t}=\frac{1}{\mathbf{T}_{ii}^t+\frac{n-1}{\underline{d}}}.$$

Since $\mathbf{T}_{ii}^t\ge 0$ for all i,t, we have $\frac{1}{\mathbf{T}_{ii}^t+\frac{n-1}{\underline{d}}}\le\frac{1}{0+\frac{n-1}{\underline{d}}}=\frac{\underline{d}}{n-1}$. Recall that $\underline{d}$ is a small positive number much less than 1 that we use to prevent taking reverse of 0, $\frac{\underline{d}}{n-1}$ is less than and bounded away from 1. That is,

$$\frac{\mathbf{T}_{ii}^{t+1}}{\mathbf{T}_{ii}^t}=\frac{1}{\mathbf{T}_{ii}^t+\frac{n-1}{\underline{d}}}<\frac{\underline{d}}{n-1}\ll 1.$$

$$\Rightarrow \mathbf{T}_{ii}^t\to 0 \Rightarrow \mathbf{T}_{ij}^t\to\frac{1}{n-1} \quad \text{for all } i \text{ and } j\ne i.$$

That is, $\lim_{t\to\infty}\mathbf{T}^t=\mathbf{T}^*$ where $\mathbf{T}_{ii}^*=0$, $\mathbf{T}_{ij}^*=\frac{1}{n-1}$, for all i and $j\ne i$.

Proof of Theorem 4

Following the same notation and logic of the proof of Theorem 2, we have

$$\mu_t(\mathbf{T})=\frac{1}{2}\max_{i,j}\sum_{k=1}^n|\mathbf{T}_{ik}^t-\mathbf{T}_{jk}^t|=\max_{i,j}\sum_{k\in D_+^{ijt}}(\mathbf{T}_{ik}^t-\mathbf{T}_{jk}^t).$$

Now for an $l\in D_+^{ijt}$, $\sum_{k\in D_+^{ijt}}\mathbf{T}_{jk}^t\ge\mathbf{T}_{jl}^t>\delta_c$ where $0<\delta_c<1$ if $j\in\overline{S_1}$, or $\sum_{k\in D_+^{ijt}}\mathbf{T}_{jk}^t\ge\mathbf{T}_{jl}^t>\min_{s\in S,j\in N}\mathbf{T}_{sj}^0$ if $j\in S_1$. We know $0<\min_{s\in S,j\in N}\mathbf{T}_{sj}^0<1$ since the arbitrary initial influence matrix

[12] Note that $0<\mu_\tau<1$ is not sufficient for $\prod_{\tau=t}^{1}\mu_\tau(\mathbf{T})$ to converge to 0. For example, define a sequence $\{a_i\}_{i=1}^\infty$ with $a_i=\frac{1}{2}\sqrt{\underbrace{2+\sqrt{2+\sqrt{2+\cdots\sqrt{2}}}}_{i}}$. We see that $0<a_i<1$ for all i but not bounded away from 1. The product of a_i is the Viète product $\prod_{i=1}^\infty a_i\to\frac{2}{\pi}\ne 0$.

is assumed to be positive, as explained in previous section. Besides, $\sum_{k\in D_+^{ijt}} \mathbf{T}_{ik}^t \leq 1$.

Therefore, define $\delta_s = \min_{s\in S, j\in N}\{\delta_c, \mathbf{T}_{sj}^0\}$. Then $\mu_t(\mathbf{T}) < 1-\delta_s$, i.e., $\mu_t(\mathbf{T})$ is less than and bounded away from 1. Thus follow the same proof as we have for Theorems 2 and 3, we have that $\lim_{t\to\infty} p_i^t = p^*$ for all i and for $\bar{s} \in \overline{S_1}$, for all $s' \neq \bar{s}$, $\lim_{t\to\infty} \mathbf{T}_{\bar{s}\bar{s}}^t = 0$ and $\lim_{t\to\infty} \mathbf{T}_{\bar{s}s'}^t = \frac{1}{n-1}$.

Proof of Theorem 5

With type II persistent agents, the updating process can be rewritten as

$$\mathbf{p}^t = \widetilde{\mathbf{T}}^t \mathbf{p}^{t-1}, \tag{12}$$

where

$$\widetilde{\mathbf{T}}^t = \begin{pmatrix} \mathbf{T}_{11}^t & \dots & \mathbf{T}_{1j}^t & \dots & \dots & \mathbf{T}_{1n}^t \\ \vdots & & \vdots & & & \vdots \\ \mathbf{T}_{n-\sigma_2,1}^t & \dots & \mathbf{T}_{n-\sigma_2,j}^t & \dots & \dots & \mathbf{T}_{n-\sigma_2,n}^t \\ 0 & \dots & 0 & \dfrac{1}{\sigma_2} & \dots & \dfrac{1}{\sigma_2} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \dfrac{1}{\sigma_2} & \dots & \dfrac{1}{\sigma_2} \end{pmatrix}. \tag{13}$$

Then following the notation in proof of Theorem 2, for $i \in \overline{S_2}$ and $j \in S_2$, the set of non-persistent agents $\{l | l \in \overline{S_2}\} \subseteq D_+^{ijt}$. If no persistent agent is in the set D_+^{ijt}, which means that $\{l | l \in \overline{S_2}\} = D_+^{ijt}$ then

$$\max_{i\in\overline{S_2}, j\in S_2} \sum_{k\in D_+^{ijt}} (\widetilde{\mathbf{T}}_{ik}^t - \widetilde{\mathbf{T}}_{jk}^t) \leq 1-(n-\sigma_2)\delta_c,$$

since there are $n-\sigma_2$ non-persistent agents not in D_+^{ijt} and each has a minimum value of δ_c. Otherwise, if $\{l | l \in \overline{S_2}\} \subset D_+^{ijt}$, i.e., there is at least one persistent agent s_x in the set D_+^{ijt}, then

$$\max_{i\in\overline{S_2}, j\in S_2} \sum_{k\in D_+^{ijt}} (\widetilde{\mathbf{T}}_{ik}^t - \widetilde{\mathbf{T}}_{jk}^t) \leq 1-\frac{1}{\sigma_2},$$

since $\sum_{k\in D_+^{ijt}} \widetilde{\mathbf{T}}_{jk}^t \geq \widetilde{\mathbf{T}}_{js_x}^t = \frac{1}{\sigma_2}$. In conclusion, denote $\delta_s = \min((n-\sigma_2)\delta_c, \frac{1}{\sigma_2})$ then

$$\max_{i\in\overline{S_2}, j\in S_2} \sum_{k\in D_+^{ijt}} (\widetilde{\mathbf{T}}_{ik}^t - \widetilde{\mathbf{T}}_{jk}^t) \leq 1-\delta_s.$$

So similar to the proof of Theorem 4 above, denote $\delta = \min(\delta_c, \delta_s)$,

$$\mu_t(\widetilde{\mathbf{T}}) = \frac{1}{2}\max_{i,j} \sum_{k=1}^n |\widetilde{\mathbf{T}}_{ik}^t - \widetilde{\mathbf{T}}_{jk}^t| = \max_{i,j} \sum_{k\in D_+^{ijt}} (\widetilde{\mathbf{T}}_{ik}^t - \widetilde{\mathbf{T}}_{jk}^t) \leq 1-\delta.$$

So μ_t is less than and bounded away from 1. Thus the difference of opinions converges to zero: $|p_i^t - p_j^t| \to 0$.

Next, we need to prove that $p_i^t \to p_s^0$. Suppose that $p_i^t \to p^*$ and $p^* \neq p_s^0$. Then for $\epsilon^* = \frac{1}{2}|p_s^0 - p^*|$, since for all $s \in S$, $p_s^t = p_s^0$, $\forall t$, $|p_s^0 - p^*| > \epsilon^*$ for all t. However, it is assumed that $p_i^t \to p^*$, which implies that $\forall \epsilon > 0$, there exists $t > 0$, s.t. $|p_i^t - p^*| < \epsilon$ for all i. Thus we have a contradiction, which means that $p^* = p_s^0$. In other words, $\lim_{t\to\infty} p_i^t = p_s^0$, $\forall i$.

Then similar to the basic model, with conforming opinions, $\lim_{t\to\infty} \mathbf{T}_{ii}^t = 0$ and $\lim_{t\to\infty} \mathbf{T}_{ij}^t = \frac{1}{n-1}$, for all $i \in N$, for all $j \neq i$.

Proof of Theorem 6

In this case we can apply the same opinion updating process as that with type II persistent agents, as shown in Eqs. (12) and (13). The statement as regard to the convergence of both opinions and the influence matrix follows if we mimic the proof of Theorem 5 (omitted).

Proof of Proposition 1

Given that $p_i^0 = p_j^0 = p_{\bar{s}}^0$ for all non-persistent $i, j \notin S$, $p_{s_i}^0 = p_{s_j}^0 = p_s^0$ for all persistent $s_i \in S$, and that $|S| = \sigma$, we have that after the first updating of weights:

$$\mathbf{T}_{ii}^1 = \frac{\mathbf{T}_{ii}^0}{\mathbf{T}_{ii}^0 + (n-\sigma-1)(1/\underline{d}) + \sigma(1/|p_{\bar{s}}^0 - p_s^0|)}, \quad \text{for all } i \notin S,$$

$$\mathbf{T}_{ik}^1 = \frac{1/\underline{d}}{\mathbf{T}_{ii}^0 + (n-\sigma-1)(1/\underline{d}) + \sigma(1/|p_{\bar{s}}^0 - p_s^0|)}, \quad \text{for all } k \neq i, k \notin S, \text{ and}$$

$$\mathbf{T}_{is_j}^1 = \frac{1/|p_{\bar{s}}^0 - p_s^0|}{\mathbf{T}_{ii}^0 + (n-\sigma-1)(1/\underline{d}) + \sigma(1/|p_{\bar{s}}^0 - p_s^0|)}, \quad \text{for all } s_j \in S.$$

Since $\mathbf{T}_{ii}^0 = \mathbf{T}_{jj}^0$ for all $i, j \notin S$, we have $\mathbf{T}_{ik}^1 = \mathbf{T}_{jk}^1$ for all $i, j \notin S$ and all $k \in N$ that ik, jk are not on the diagonal. The diagonal elements for non-persistent agents are also equal: $\mathbf{T}_{ii}^1 = \mathbf{T}_{jj}^1$. Given that non-persistent agents have identical weight distribution on any initial opinion, the weighted average opinion is the same also, i.e., $p_i^1 = p_j^1$ for all $i, j \notin S$. By induction, we can extend the argument to all periods t. That is, $p_i^t = p_j^t = p_{\bar{s}}^t$ for all $i, j \notin S$, for all t.

As for the persistent agents, if they are of type II or double-stubborn, then by definition their opinions remain the same over time. If we have type I persistent agents, then since their initial weight assignments are most likely different, the resulting weighted averages that they take during each period are different also. For that reason, we label persistent agents' opinions during time t by the index of those agents, i.e., $p_{s_i}^t$ for $s_i \in S$. This does not affect the cases of type II and double-stubborn agents, where $p_{s_i}^t = p_{s_j}^t = p_s^0$.

Denote the change in non-persistent agents' opinion between 2 consecutive periods by

$$\Delta p_{\bar{s}}^t = |p_{\bar{s}}^t - p_{\bar{s}}^{t-1}|.$$

Also, define the maximum difference between persistent and non-persistent agents' opinions

$$\Delta p^t = \max_{s_i\in S} |p_{s_i}^t - p_{\bar{s}}^t|.$$

Then, by definition of the weighted average updating rule:

$$\begin{aligned}
\Delta p_{\bar{s}}^t &= \left| \sum_{j\in N} \mathbf{T}_{ij}^t p_j^{t-1} - p_{\bar{s}}^{t-1} \right| \\
&= \left| p_{\bar{s}}^{t-1} \sum_{k\notin S} \mathbf{T}_{ik}^t + \sum_{s_j\in S} \mathbf{T}_{is_j}^t p_{s_j}^{t-1} - p_{\bar{s}}^{t-1} \right| \\
&= \left| p_{\bar{s}}^{t-1} \left(\sum_{k\notin S} \mathbf{T}_{ik}^t - 1 \right) + \sum_{s_j\in S} \mathbf{T}_{is_j}^t p_{s_j}^{t-1} \right| \\
&= \left| p_{\bar{s}}^{t-1} \left(-\sum_{s_j\in S} \mathbf{T}_{is_j}^t \right) + \sum_{s_j\in S} \mathbf{T}_{is_j}^t p_{s_j}^{t-1} \right| \\
&= \left| \sum_{s_j\in S} \mathbf{T}_{is_j}^t (p_{s_j}^{t-1} - p_{\bar{s}}^{t-1}) \right| \\
&\leq \sum_{s_j\in S} \mathbf{T}_{is_j}^t |p_{s_j}^{t-1} - p_{\bar{s}}^{t-1}| \quad \text{since } \mathbf{T}_{is_j}^t \geq 0 \\
&\leq \Delta p^{t-1} \sum_{s_j\in S} \mathbf{T}_{is_j}^t.
\end{aligned}$$

Now it comes down to the weight that a non-persistent agent places on persistent agents, which can be written as:

$$\mathbf{T}_{is_j}^{t} = \frac{1/|p_{\bar{s}}^{t-1} - p_{s_j}^{t-1}|}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/|p_{\bar{s}}^{t-1} - p_{s_i}^{t-1}|)},$$

for all $s_j \in S$.

Therefore,

$$\begin{aligned}
&\Delta p^{t-1}\sum_{s_j\in S}\mathbf{T}_{is_j}^{t} \\
&= \Delta p^{t-1}\sum_{s_j\in S}\frac{1/|p_{\bar{s}}^{t-1} - p_{s_j}^{t-1}|}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/|p_{\bar{s}}^{t-1} - p_{s_i}^{t-1}|)} \\
&= \sum_{s_j\in S}\frac{\Delta p^{t-1}(1/|p_{\bar{s}}^{t-1} - p_{s_j}^{t-1}|)}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/|p_{\bar{s}}^{t-1} - p_{s_i}^{t-1}|)} \\
&\le \sum_{s_j\in S}\frac{1}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/|p_{\bar{s}}^{t-1} - p_{s_i}^{t-1}|)} \\
&\le \sum_{s_j\in S}\frac{1}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/\Delta p^{t-1})} \\
&= \frac{\sigma}{\mathbf{T}_{ii}^{t-1} + (n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/\Delta p^{t-1})} \\
&\le \frac{\sigma}{(n-\sigma-1)(1/\underline{d}) + \sum_{s_i\in S}(1/\Delta p^{t-1})} \\
&= \underline{d}\left[\frac{\sigma\Delta p^{t}}{(n-\sigma-1)\Delta p^{t} + \sigma\underline{d}}\right] \\
&< \underline{d}\left[\frac{\sigma}{n-\sigma-1}\right] \quad \text{since } \sigma\underline{d} > 0.
\end{aligned}$$

We know that persistent agents constitute less than 10% of the population, i.e., $\sigma \le 0.1n$. Then, $n-\sigma-1 \ge 0.9n - 1 > 0.1n$ since $n \ge 2$. That is, $n-\sigma-1 > \sigma$. Hence, $\frac{\sigma}{n-\sigma-1} < 1$. We get $\Delta p_{\bar{s}}^{t} < \underline{d}$.

References

Bala, V., Goyal, S., 1998. Learning from neighbors. The Review of Economic Studies 65 (3), 595–621.

Bala, V., Goyal, S., 2001. Conformism and diversity under social learning. Economic Theory 17, 101–120.

Banerjee, A., Fudenberg, D., 2004. Word-of-mouth learning. Games and Economic Behavior 46, 1–22.

Barabási, A.-L., 2003. Linked: How Everything is Connected to Everything Else and What it Means. Plume.

Baron, R., Durieu, J., Haller, H., Solal, P., 2002. Control costs and potential functions for spatial games. International Journal of Game Theory 31, 541–561.

Chan, K.K., Misra, S., 1990. Characteristics of the opinion leader: a new dimension. Journal of Advertising 19, 53–60.

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

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

Durieu, J., Solal, P., 2003. Adaptive play with spatial sampling. Games and Economic Behavior 43, 189–195.

Ellison, G., Fudenberg, D., 1993. Rules of thumb of social learning. The Journal of Political Economy 101 (4), 612–643.

Ellison, G., Fudenberg, D., 1995. Word-of-mouth communication and social learning. The Quarterly Journal of Economics 110 (1), 93–125.

French, J.R.P., 1956. A formal theory of social power. Psychological Review 63 (3), 181–194.

Friedkin, N.E., Johnsen, E.C., 1997. Social positions in influence networks. Social Networks 19, 209–222.

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

Golub, B., Jackson, M.O., 2007. Naive learning in social networks: convergence, influence, and the wisdom of crowds, Revised: May 29th, 2007.

Granovetter, M.S., 1973. The strength of weak ties. American Journal of Sociology 78 (3), 1360–1380.

Hegselmann, R., Krause, U., 2002. Opinion dynamics and bounded confidence, models, analysis and simulation. Journal of Artificial Societies and Social Simulation 5 (3).

Katz, E., Lazarsfeld, P.F., 1955. Personal Influence; The Part Played by People in the Flow of Mass Communications. Free Press.

Krause, U., 2000. A discrete nonlinear and non-autonomous model of consensus formation. In: Communications in Difference Equations: Proceedings of the Fourth International Conference on Difference Equations. Poznan, Poland, August 27–31, 1998. CRC Press, p. 227.

Leider, S., Möbius, M.M., Rosenblat, T., Do, Q.-A., 2007. How much is a friend worth? Directed altruism and enforced reciprocity in social networks, July 23, 2007.

Lorenz, J., 2005. A stabilization theorem for dynamics of continuous opinions. Physica A: Statistical Mechanics and its Applications 355 (1), 217–223.

Pan, Z., 2009. Networks and opinions: How do they affect each other. Working paper.

Pan, Z., Gilles, R.P., 2009. Naive learning and game play in a dual social network framework. Working paper.

Roch, C.H., 2005. The dual roots of opinion leadership. Journal of Politics 67, 110–131.

Seneta, E., 1981. Non-Negative Matrices and Markov Chains. Springer-Verlag New York Inc.

Shiller, R.J., 1995. Conversation, information and herd behavior. The American Economics Review 85 (2), 181–185.

Vernette, E., 2004. Targeting women's clothing fashion opinion leaders in media planning: an application for magazines. Journal of Advertising Research 44, 90–107.

Watts, D.J., Dodds, P.S., 2007. Influentials, networks, and public opinion formation. Journal of Consumer Research 34, 441–458.

Weisbuch, G., Deffuant, G., Amblard, F., Nadal, J.P., 2002. Meet, discuss, and segregate!. Complexity 7 (3), 55–63.