Naive Learning in Social Networks and the Wisdom of Crowds

by Benjamin Golub and Matthew O. Jackson

Jakub Pawelczak
University of Minnesota

October 6, 2020

Table of contents

Introduction

De Groot model

Wisdom characterization

Prominent groups

Speed vs wisdom

Final comments

Introduction

Questions
Can large societies whose agents are naive individually be smart in the aggregate?
What
Social networks which capture transfer of information, influence . Societies which converge to true value are called *wise*.

Introduction

Questions
Can large societies whose agents are naive individually be smart in the aggregate?
What
Social networks which capture transfer of information, influence . Societies which converge to true value are called *wise*.
How
Nonstrategic agents with bounded rationally in essential the same way. Opinions are formed based on initial beliefs updated by imposed rule. They start with agents which receive independent noisy signal about true value and then communicate in network by taking average of neighbours.

Main results

- They characterize wise networks.
- If there is enough dispersion in the people to whom they listen, and if they avoid concentrating too much on any small group of agents.
- On the other hand, their sufficient conditions can fail if there is just one group that receives too much weight or is too insular.
- More generaly they proved that having bounded number of agents who are *prominent* (with nonvanishing attention from rest of agents in network) causes learning to fail.
- **Main results** Sufficient condition for wisdome is satisfying balance and minimal out-dispersion

Definition of model

- A finite set $N = (1, \dots, n)$ of agents interact according to social network
- Interactions capture by stochastic matrix $T = [T_{ij}]$
- $T_{i,j}$ is weight or trust that agent i places on the current belief of agent j
- $p_i^{(t)} \in [0, 1]$ is belief of agent i in time t
- updating rule:

$$p^{(t)} = Tp^{(t-1)} \tag{1}$$

Motivation for the evolution of beliefs

- At $t = 0$ each agent receives noisy signal ($\mathbb{E}e_i = 0$, $\sigma^2 \leq \text{VAR}e_i < 1$, IID)

$$p_i^{(0)} = \mu + e_i$$

- i has subjective prob π_{ij} which assigns to agents which opinions he/she hears

Motivation for the evolution of beliefs

- At $t = 0$ each agent receives noisy signal ($\mathbb{E}e_i = 0$, $\sigma^2 \leq$VAR$e_i < 1$, IID)

$$p_i^{(0)} = \mu + e_i$$

- i has subjective prob π_{ij} which assigns to agents which opinions he/she hears
- Bayesian updating from independent signal at $t = 1$ gives

$$T_{i,j} = \frac{\pi_{ij}}{\sum_k \pi_{ik}}$$

- focus on direct communication due to cost or constraints
- updating rule stable over time but beliefs are not
- possible repetition of information

Convergence

- Walk, path, cycle, simple cycle.
- Matrix T is strongly connected (irreducible) if there is a path in T from any node to any other
- Matrix T is aperiodic if GCD of lengths of simple cycles is 1
- A matrix T is **convergent** if $\lim_{t\to\infty} T^t p$ exists for all $p \in [0,1]$

Convergence

- Walk, path, cycle, simple cycle.
- Matrix T is strongly connected (irreducible) if there is a path in T from any node to any other
- Matrix T is aperiodic if GCD of lengths of simple cycles is 1
- A matrix T is **convergent** if $\lim_{t\to\infty} T^t p$ exists for all $p \in [0, 1]$

Examples

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ aperiodic $\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $s = (2/5, 2/5, 1/5)$

Theorem

A strongly connected, stochastic matrix T is convergent $\iff$ aperiodic $\iff$ there is unique left eigenvector s corresponding to eigenvalue 1, so

$$(\lim_{t \to \infty} T^t p)_i = sp$$

where s_i is called *influence weight* of agent i. Additionally we have

$$s = sT \quad s_i = \sum_j T_{ji} s_j$$

So influence of i is a weighted sum of j's who pay attention to i weighted by T_{ji} trust that j places on i

Undirected networks with equal weights- Example 1

- Consider a network with G is adjancy matrix
- degree if nr of neighbors $d_i(g) =_j G_{ij}$
- Consider a model where each agent equally split attention among neighbors ($T_{ij} = \frac{g_{ij}}{\sum_k g_{ik}}$)
- then

$$s_i = \frac{d_i(g)}{\sum_j d_j(g)}$$

- so the influence is directly proportional to degree.

Wisdom of the crowds

- How decentrailized DeGroot process of communication correctly aggregate the diverse informmation initially held by the different agents?
- In particular consider large societies. Size of network $n \to \infty$
- limit of this sequance is given by $p_i^{\infty}(n)$

Wisdom of the crowds

- How decentrailized DeGroot process of communication correctly aggregate the diverse informmation initially held by the different agents?
- In particular consider large societies. Size of network $n \to \infty$
- limit of this sequance is given by $p_i^\infty(n)$
- The sequence of societies with $T(n)$ is **wise** if

$$\text{plim}_{n\to\infty} \max_{i\leq n} |p_i^\infty(n) - \mu| = 0$$

WISDOM OF THE CROWDS

- How decentrailized DeGroot process of communication correctly aggregate the diverse informmation initially held by the different agents?
- In particular consider large societies. Size of network $n \to \infty$
- limit of this sequance is given by $p_i^\infty(n)$
- The sequence of societies with $T(n)$ is **wise** if

$$\text{plim}_{n\to\infty} \max_{i\leq n} |p_i^\infty(n) - \mu| = 0$$

Lemma- Wisdom and influence

If $T(n)$ a sequence of convergent stochastic matrices then it is wise $\iff$ (s(n)) (sequence of influence vectors (stationary)) are such that $s_1(n) \to^n 0$.

Proof

$\Rightarrow$variance of $p_i^{(0)}(n) \in (\sigma^2, 1)$

Let $X(n) = \sum_i s_i(n) p_i^{(0)}(n)$. Then $\text{var}(X(n)) \leq \bar{\sigma}^2 \sum_i s_i(n)^2$

First, suppose $s_1(n) \to 0$. since $s_i(n) \geq s_{i+1}(n) \geq 0$ for all i and n, it follows that

$$\text{var}(X(n)) \leq \bar{\sigma}^2 \sum_i s_i(n)^2 \leq \bar{\sigma}^2 s_1(n) \sum_i s_i(n) = \bar{\sigma}^2 s_1(n) \to 0$$

By Chebychev's inequality, fixing any $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\sum_i s_i(n) p_i^{(0)}(n) - \mu\right| > \varepsilon\right] \leq \frac{\text{var}(X(n))}{\varepsilon^2} \to 0$$

Proof

$\Rightarrow$variance of $p_i^{(0)}(n) \in (\sigma^2, 1)$

Let $X(n) = \sum_i s_i(n) p_i^{(0)}(n)$. Then $\mathrm{var}(X(n)) \leq \bar{\sigma}^2 \sum_i s_i(n)^2$

First, suppose $s_1(n) \to 0$. since $s_i(n) \geq s_{i+1}(n) \geq 0$ for all i and n, it follows that

$$\mathrm{var}(X(n)) \leq \bar{\sigma}^2 \sum_i s_i(n)^2 \leq \bar{\sigma}^2 s_1(n) \sum_i s_i(n) = \bar{\sigma}^2 s_1(n) \to 0$$

By Chebychev's inequality, fixing any $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\sum_i s_i(n) p_i^{(0)}(n) - \mu\right| > \varepsilon\right] \leq \frac{\mathrm{var}(X(n))}{\varepsilon^2} \to 0$$

$\Leftarrow$ suppose (taking a subsequence if necessary) $s_1(n) \to s > 0$. Since each $p_i^0(n)$ has a variance bounded below, it then follows that there exists $\delta > 0$ such that $\mathrm{var}(X(n)) > \delta$ for all n. For uniformly bounded random variables, convergence in probability to 0 implies that the same holds in L^2, which means that the $X(n)$ cannot converge to 0 in probability

Wisdom - Example 1

So the society is wise iff most important agent's influence diminish with increase of society.
This is very important notion, society will fail to converge to truth if leader has too much power on forming opinions of others.
In case of our example G(n) is wise $\iff$

$$\max_{1\leq i\leq n} \frac{d_i(G)}{\sum_j d_j(G)} \to^n 0 \tag{2}$$

In other words: disproportional popularity of some agent is the only obstacle to wisdom.

Definition The weight that group B places on group C

$$T_{B,C} = \sum_{i \in B, j \in C} T_{ij}$$

The group B is prominent in t steps (related to T) if

$$\forall i \notin B \qquad (T^t)_{i,B} > 0 \tag{3}$$

Definition The weight that group B places on group C

$$T_{B,C} = \sum_{i \in B, j \in C} T_{ij}$$

The group B is prominent in t steps (related to T) if

$$\forall i \notin B \qquad (T^t)_{i,B} > 0 \tag{3}$$

- with $\pi(T,t) = \min_{i \notin B}(T^t)_{i,B}$ t-step prominence (of B relative to T)
- So a group that is prominent in t steps is one such that each agent outside of it is influenced by at least someone in B group in t steps.

Definition The uniformly prominent family $\{B_n\}$ is such that there exists threshold > 0 that $\forall n \quad \pi_{B_n}(T(n), t) \geq \alpha$

Prominent groups as an obstacle to wisdom

Theorem

If there is a finite, uniformly prominent family with respect to $T(n)$ then the sequence of networks is not wise

It's not full characterization, they gave example when wisdom fails and we do not have prominent group.

PROOF

Observation $k(T) = \sum_i s_i$ where $k(T)$ is nr of closed and strongly connected groups relative to T.
Lemma. For any $B \subseteq N$ and natural number t,

$$\max_{i \in N} s_i \geq \frac{\kappa(\mathbf{T})\pi_B(\mathbf{T};t)}{|B|\left(1+\pi_B(\mathbf{T};t)\right)}$$

Proof of lemma: :since s is a row unit eigenvector of $\mathbf{T}^t$, it follows that

$$\sum_{i \in B} s_i \geq \sum_{i \in B} \sum_{j \notin B} T_{ji}^{(t)} s_j$$

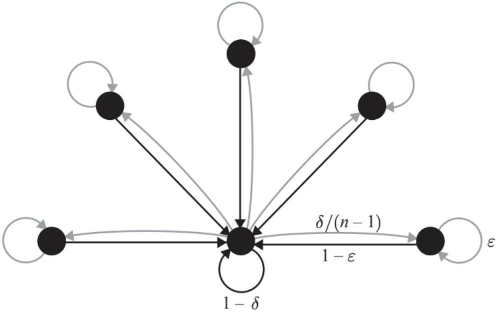
$$= \sum_{j \notin B} s_j \sum_{i \in B} T_{ji}^{(t)} \geq \pi_B(\mathbf{T};t) \sum_{j \notin B} s_j$$

Then, by observation we know that

$$\sum_{j \notin B} s_j = \kappa(\mathbf{T}) - s_B \quad s_B \geq \pi_B(\mathbf{T}; t)\left(\kappa(\mathbf{T}) - s_B\right)$$

which yields the first claim of the lemma.
Proof of theorem:
The fact that $s_1(n)$ does not converge to 0 as $n \to \infty$ follows immediately upon applying Lemma to each matrix in the sequence.
We use the finiteness of (B_n) to prevent the denominator in the lemma from exploding,
and the uniform lower bound on the prominence of each B_n relative to $\mathbf{T}(n)$ to keep the numerator from going to 0.

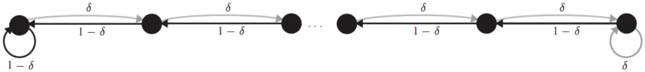
Example 2



$s_1(n) = \frac{1-\epsilon}{1-\epsilon+\delta}$
$s_i(n) = \frac{\delta}{(n-1)(1-\epsilon+\delta)}$

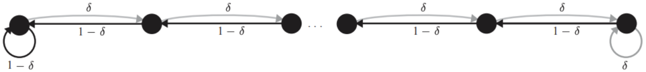
- This won't converge to truth since $s_1(n)$ is constant . Leaders info s weighted heavily enough that it biases the final belief (nonvanishing influence of central agent)
- However, even if $1\epsilon \to 0$ so that people are not weighting the center much, it has nonvanishing influence as long as $1\epsilon \sim \delta$
- Not simply the total weight on a given individual matters, but **the relative weights coming in and out** groups of nodes.
- In particular, if the weight on the center decays (so that nobody is prominent in one step), wisdom may still fail - check out next example.

EXAMPLE 3



$$s_i(n) = \left(\frac{\delta}{1-\delta}\right)^i \frac{1-\frac{\delta}{1-\delta}}{1-\left(\frac{\delta}{1-\delta}\right)^n}$$

EXAMPLE 3



$$s_i(n) = (\frac{\delta}{1-\delta})^i \frac{1 - \frac{\delta}{1-\delta}}{1 - (\frac{\delta}{1-\delta})^n}$$

So $s_1(n)$ can be as close to 1 as small δ is and wisdom is not obtained. It is not direct weigh on 1 but indirect Thus while agent 1 is not prominent in any steps (less than $n-1$) the agent influence can exceeed the sum of other influences . Not only direct but inflow and outflow influence can be not important. Thus entire structure is relevant.

WE ARE ALMOST DONE- SUFFICIENT CONDITIONS FOR WISE CROWD

Definition We say family $\{B_n\}$ is **balanced** if $\exists j(n) \to \infty$ s.t. $|B_n| \leq j(n)$ then

$$\sup_n \frac{T_{B_n^c, B_n}}{T_{B_n, B_n^c}} < \infty$$

The balance condition says that no family below a certain size can be getting infinitely more weight from the remaining agents than it gives to the remaining agents. It rules out example 2.

We are almost done- sufficient conditions for wise crowd

Definition We say family $\{B_n\}$ is **balanced** if $\exists j(n) \to \infty$ s.t. $|B_n| \leq j(n)$ then

$$\sup_n \frac{T_{B_n^c,B_n}}{T_{B_n,B_n^c}} < \infty$$

The balance condition says that no family below a certain size can be getting infinitely more weight from the remaining agents than it gives to the remaining agents. It rules out example 2.

From Example 3, it is not enough to rule out situations where there is infinitely more direct weight into some family of agents than out. Take care of large-scale asymmetries (f.e. small groups focusing their attention too narrowly). The next condition deals with this

Definition We say **minimal out-dispertion** holds if there is $q \in \mathbb{N}$ and $r > 0$ s.t. if B_n finite , $|B_n| \geq q$ and $C_n/n \to 1$ then $T_{B_n.C_n} > r$

Definition We say **minimal out-dispertion** holds if there is $q \in \mathbb{N}$ and $r > 0$ s.t. if B_n finite , $|B_n| \geq q$ and $C_n/n \to 1$ then $T_{B_n.C_n} > r$
The minimal out-dispersion condition requires that any large enough finite family must give at least some minimal weight to any family which makes up almost all of society.

Sufficiency for wisdom

Theorem

If $T(n)$ is sequence of convergent stochastic matricies satisfying balance and minimal out-dispersion then it's wise

Note that in example 2 we have balance but not minimal out-dispersion. In example 3 we have minimal out-dispersion but not balanced. In both society fails to be wise

SPEED OF CONVERGENCE AND WISDOM

They find lack of any necessary relationship between convergence and wisdom. Observe that:

- Example 1 -This society is wise and has immediate convergence.
- consider a society where all agents weight just one agent. Here, we have immediate convergence but no wisdom
- Society in which all agent have $1 - \epsilon$ on themselves and distribute the rest equally. This is wise but can be very slowly converging to wisdom.
- Society in which all agent have $1 - \epsilon$ on themselves and gives the rest to one guy (different for different people). There is neither wisdom nor fast convergence.

However we can think about bounds on convergence. But first how to even measure it?

Speed of convergence

In this section they do not provide any bound for speed of convergence which is one of main drawbacks of paper. Especially lower bound would be interesting in this case. So I did my research and prove on my own:

Definition Distance from stationty distribution in time t is defined as:

$$d(t) := \max_{x \in \Omega} ||T^t(x, \cdot) - \pi(\cdot)||_{tv} = \max_{x \in \Omega} \frac{1}{2} \sum_{x \in \Omega} |T^t(x, y) - \pi(y)| \tag{4}$$

Definition Mixing time with respect to ϵ is minimal time after which we will be at least ϵ close to stationary:

$$t_{mix}(\epsilon) = \min\{t \geq 0 : d(t) \leq \epsilon\} \tag{5}$$

π-scalar product on $\mathbb{R}^{\Omega}$ we call:

$$< f, g >_{\pi} = \sum_{x \in \Omega} f(x)\pi(x)g(x) \tag{6}$$

$$\Pi = \begin{bmatrix} \pi(x_1) & 0 & \dots & 0 \\ 0 & \pi(x_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi(x_k) \end{bmatrix} = diag(\pi(x))_{x \in \Omega}$$

In matrix notation

$$< f, g >_{\pi} = f^T \Pi g \tag{7}$$

Spectral decomposition

For matrix T there is decomposition :

$$T = F\Lambda F^T \Pi \tag{8}$$

where

1. $F\Pi F^T = I$, i.e. matrix F is π-orthogonal F
2. $\Lambda = diag(1, \lambda_2, ..., \lambda_n)$

Weak Ergodic Theorem

Strongly connected,aperiodic chain converges to stationary distribtuion π and :

$$\lim_{t \to \infty} T^t = 1\pi^T$$

My remarks

Theorem

For $\pi_{min} = \min_{x\in\Omega} \pi(x)$ and $\gamma^* = 1 - \lambda_*$ where $\lambda^* = \max\{|\lambda| : Tf = \lambda f, \lambda \neq 1\}$. Then mixing time is bounded by:

$$\left(\frac{1}{\gamma_*} - 1\right) log\left(\frac{1}{2\epsilon}\right) \overset{L}{\leq} t_{mix} \overset{R}{\leq} log\left(\frac{1}{\epsilon\pi_{min}}\right)\frac{1}{\gamma_*} \tag{9}$$

Lemma If $< f, 1 >_\pi = 0$ then

$$||Tf||_\pi \leq |\lambda| ||f||_\pi$$

Let f be eigenvector for $\lambda \neq 1$.Then $Tf = \lambda f$. eigen vectors are π-orthogonal so we use lemma:

$$0 = < f, 1 >_\pi = \sum_{y \in \Omega} \pi(y) f(y)$$

.

$$T^k = F \Lambda^k F^T \Pi = 1\pi^T + \sum_{i=2}^{n} \lambda_i^k f_i f_i^T \Pi$$

$$\begin{aligned} < Tf, Tf >_\pi = < f, T^2 f >_\pi \\ = & f^T \Pi (1\pi^T + \sum_{i=2}^{n} \lambda_i^2 f_i f_i^T \Pi) f \\ = & f^T \Pi 1 \pi^t f + f^T \Pi \sum_{i=2}^{n} \lambda_i^2 f_i f_i^T \end{aligned} \tag{10}$$

$$
\begin{aligned}
|\lambda^t f(x)| =& |T^t f(x)| = |\sum_{y\in\Omega} \left(T^t(x,y)f(y) - \pi(y)f(y)\right)| \\
\leq& \sum_{y\in\Omega} |T^t(x,y) - \pi(y)||f(y)| \leq ||f||_\infty \sum_{y\in\Omega} |T^t(x,y) - \pi(y)| \\
=& ||f||_\infty \cdot 2 \cdot ||T^t(x,\cdot) - \pi(\cdot)||_{tv} \\
\leq& 2 \cdot ||f||_\infty \max_{x\in\Omega} ||T^t(x,\cdot) - \pi(\cdot)||_{tv} = 2 \cdot ||f||_\infty d(t)
\end{aligned}
\tag{11}
$$

by taking sup with respect to x we obtain:

$$|\lambda^t| \leq 2d(t)$$

$$|\lambda^t| \leq 2d(t)$$

let's take $t = t_{mix}(\epsilon)$ so $d(t) \leq \epsilon$ hence:

$$|\lambda|^{t_{mix}} \leq 2\epsilon$$

$$\log(\frac{1}{2\epsilon}) \leq \log(\frac{1}{|\lambda|}) \cdot t_{mix}(\epsilon) \leq t_{mix}(\epsilon)(\frac{1}{|\lambda|} - 1)$$

To sum up we end up with:

$$t_{mix}(\epsilon) \geq \log(\frac{1}{2\epsilon})\frac{|\lambda|}{1-|\lambda|} \geq (\frac{1}{\gamma_*} - 1)\log(\frac{1}{2\epsilon})$$

My remarks

To prove righthandside inequality we need following lemmas.

Lemma $d(t) \leq \max_{y \in \Omega}[1 - \frac{P^t(x,y)}{\pi(y)}]$

Take $B = \{y \in \Omega : T^t(x,y) < \pi(y)\}$ then from definition fo tv:

$$\begin{aligned} ||T^t(x,\cdot) - \pi(\cdot)||_{tv} &= \sum_{y \in \Omega \cap B} [\pi(y) - T^t(x,y)] \\ &= \sum_{y \in \Omega \cap B} \pi(y)[1 - \frac{T^t(x,y)}{\pi(y)}] \\ &\leq \max_{y \in \Omega}[1 - \frac{T^t(x,y)}{\pi(y)}] \\ &= s_x(t) \end{aligned} \tag{12}$$

Lemma2 : $\sum_{i=2}^{k} f_i(x)^2 \leq \frac{1}{\pi(x)}$

$$\begin{aligned}
|\frac{T^t(x,y)}{\pi(y)} - 1| &= \sum_{i=2}^{k} |f_i(x)f_j(y)\lambda_i^t| \leq \sum_{i=2}^{k} |f_i(x)f_j(y)|\lambda_*^t \\
&\leq \lambda_*^t [\sum_{i=2}^{k} fi^2(x) \sum_{i=2}^{k} fi^2(y)]^{1/2} \leq \frac{\lambda^t}{\sqrt{\pi(x)\pi(y)}} \\
&\leq \frac{\lambda^t}{\pi_{min}} = \frac{(1-\gamma_*)^t}{\pi_{min}} \leq \frac{e^{-\gamma_* t}}{\pi_{min}}
\end{aligned} \tag{13}$$

$$d(t) \leq \max_{x\in\Omega} s_x(t) \leq \max_{x\in\Omega} \max_{y\in\Omega} |\frac{P^t(x,y)}{\pi(y)} - 1| = \frac{e^{-\gamma_* t}}{\pi_{min}} \tag{14}$$

$$d(t)\pi_{\min} \leq \frac{e^{-\gamma_* t_{mix}}}{\pi_{min}} \tag{15}$$

so:

$$\log(\frac{1}{\epsilon\pi_{\min}}) \leq \gamma_* t_{mix} \tag{16}$$

and we proved theorem

Theorem

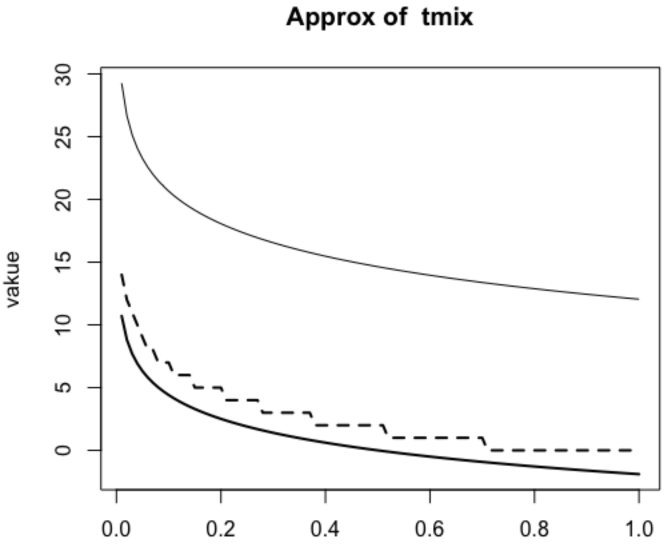
$$\left(\frac{1}{\gamma_*} - 1\right) log\left(\frac{1}{2\epsilon}\right) \overset{L}{\leq} t_{mix} \overset{R}{\leq} log\left(\frac{1}{\epsilon\pi_{min}}\right)\frac{1}{\gamma_*} \tag{17}$$

Numerical experiment

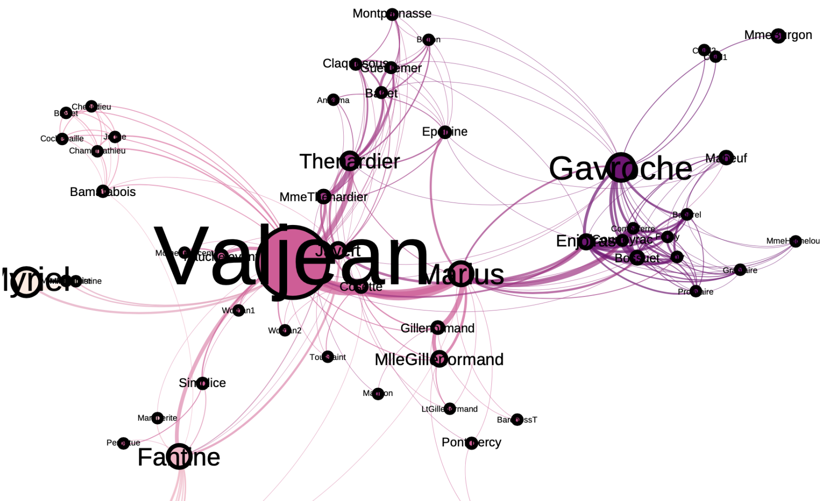
T	1	2	3	4	5	6	7
1	0.51	0.13	0.09	0.18	0.06	0.02	0.01
2	0.24	0.40	0.09	0.18	0.06	0.02	0.01
3	0.21	0.11	0.41	0.18	0.06	0.02	0.01
4	0.12	0.06	0.05	0.68	0.06	0.02	0.01
5	0.06	0.03	0.03	0.09	0.76	0.02	0.01
6	0.06	0.03	0.02	0.09	0.06	0.74	0.01
7	0.05	0.03	0.02	0.08	0.05	0.02	0.74

$$s(7) = [0.2016434, 0.1087785, 0.08323536, 0.3057292, 0.1951048, 0.06566435, 0.0398444]$$

$$\lambda = [1.0000000, 0.7325008, 0.7169051, 0.6957519, 0.4956733, 0.3235656, 0.2715967]$$


Approx of tmix
vakue
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0

Gephi


Montparnasse
Mme Burgon
Claquesous
Gueulemer
Babet
Anzelma
Eponine
Thenardier
Gavroche
Mabeuf
Bamatabois
MmeThenardier
Valjean
Javert
Enjolras
Courfeyrac
Bossuet
Marius
Cosette
Woman1
Woman2
Gillenormand
Toussaint
MlleGillenormand
Simplice
Magnon
LtGillenormand
BaronessT
Marguerite
Perpetue
Fantine
Pontmercy
Chenildieu
Cochepaille
Champmathieu
MmeHucheloup
Grantaire
Prouvaire
Bahorel
Combeferre

Final comments

- The main topic of this paper concerns what types of societies, whose agents get noisy signals at the beginning of the true value of some variable, are able to aggregate information in an approximately efficient way despite their naive and simple and decentralized updating. The existence of prominent groups or lead can mislead agents.
- Sufficient conditions can fail if there is just one group that receives too much weight or is too insular
- The reason is simple opinion of prominent agents are overweighted and will dominate groups opinions over time. Even though authors gave sufficient conditions for wise networks definitions used in theorem are highly impractical and not useful.

Final comments 2

In authors opinion this paper gives more informative conditions on wisdom. But long line of previous papers (f.e. with observation learning , naive learning) suggests that sufficient conditions are hopelessly strong.Well here we pay the price.
In contrary to Acemoglu (2008) paper ((perfect Bayesian) equilibrium of a sequential learning model) it has few similarities and differences driven by differences in agents' rationality.
Above all they forgot to:

- impose technical assumption of non negativity of initial belief vector
- form inequalities for speed of convergence
- try to simplify strong and higly impractical (in my opinion) condition on wise networks

Thank your for your attention! :)