From Consensus to Social Learning in Complex Networks

Ali Jadbabaie

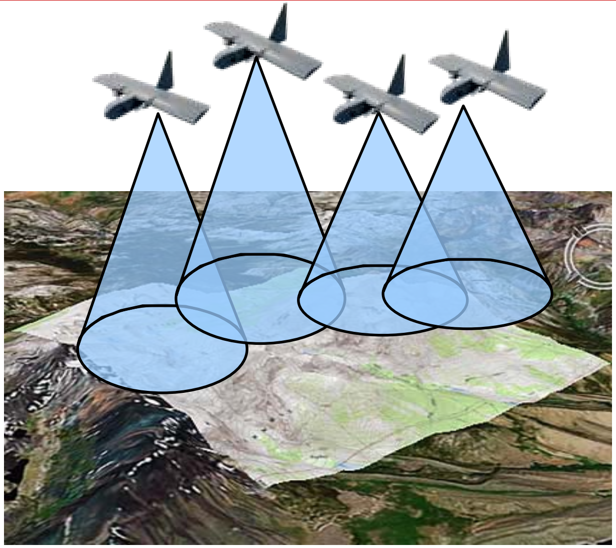
Department of Electrical and Systems Engineering and GRASP Laboratory

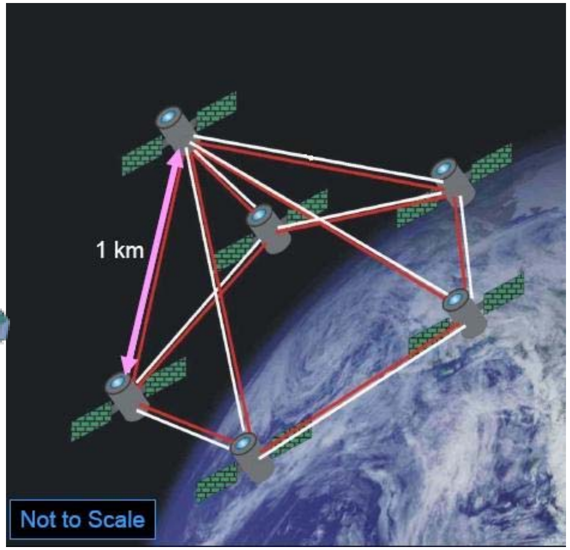
University of Pennsylvania

with Alireza Tahbaz-Salehi (MIT)

Alvaro Sandroni (Kellogg School, Northwestern University)

Consensus, flocking, synchronization





Emergence of collective decisions/actions/behaviors

Social and Economic Networks

- Epidemics and Pandemics
- Bubbles
- Bank Runs

An intuitive model (Vicsek' 1995)

The heading value updated (in discrete time) as a **weighted average** of the value of its neighbors: move one step along updated direction

$$\theta_i(k+1) = <\theta_i(k)>_r := \operatorname{atan} \frac{(\sum_{j \in \mathcal{N}_i(k)} \sin\theta_j(k)) + \sin\theta_i(k)}{(\sum_{j \in \mathcal{N}_i(k)} \cos\theta_j(k)) + \cos\theta_i(k)}$$

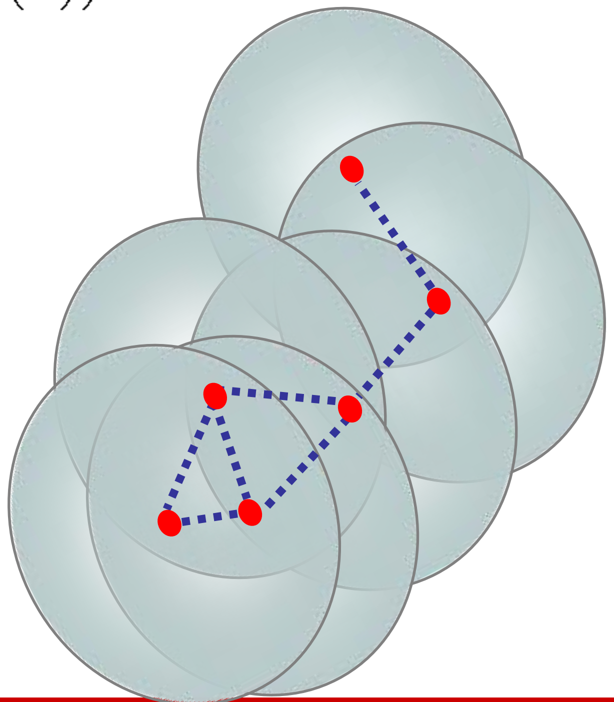
Locally:

$$<\theta_i(k)>_r = \frac{1}{d_i(k)+1}\Big(\sum_{j \in \mathcal{N}_i(k)} w_{ij}\theta_j(k) + w_{ii}\theta_i(k)\Big)$$

Neighborhood relation depends on heading value, resulting in change in topology

MAIN QUESTION : ***When do all headings converge to the same value?***

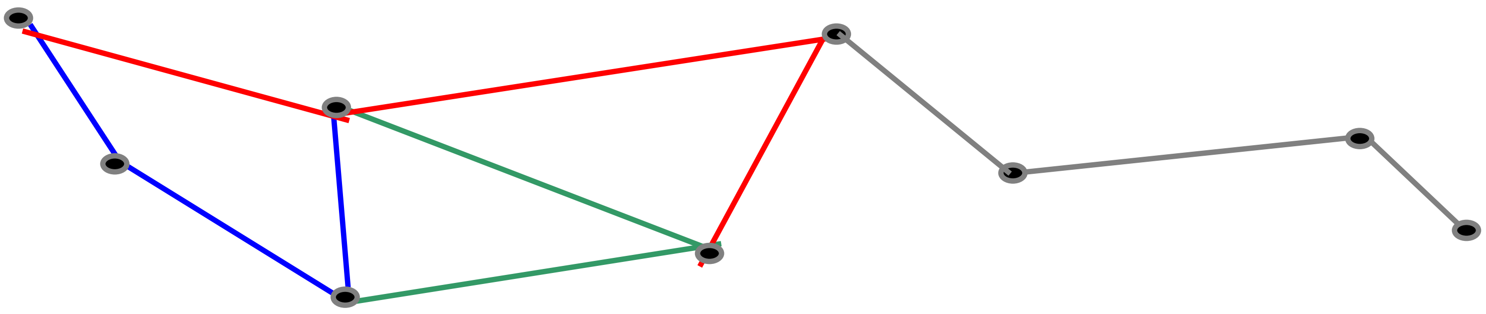
A network which changes as a result of node dynamics



Consensus in changing networks

Theorem (Jadbabaie et al. 2003, Tsitsiklis'84): *If there is a sequence of bounded, non-overlapping time intervals T_k, such that over any interval of length T_k, the network of agents is "jointly connected", then all agents will asymptotically reach consensus.*

- **Special case: network is connected "once in a while"**
  - Similar result for continuous time, leader follower,
  - Time-delays, dynamic agents, nonlinear averaging….



Consensus and Information aggregation

Do consensus algorithms aggregate information correctly?

Sometimes.

- Computing the maximum likelihood estimator [Boyd, Xiao, and Lall 2006]
- Learning in large networks [Golub and Jackson 2008]

In many scenarios agreement is not sufficient.
Agents need to agree on the "right" value: learning.

Naïve Social Learning (Golub & Jackson)

- There are n agents in the society.
- Each agent receives **one** noisy signal about the state.
- Agent's initial belief is equal to the signal observed.
- Update the belief as the average belief of the neighbo

Special case of
[Boyd, Xiao, Lall 2006]

$$s_i = \theta^* + \epsilon_i$$

$$\mu_{i,0} = s_i$$
$$\mu_{t+1} = A\mu_t$$

- Law of large numbers guarantees that this average asymptotic belief converges to the true state as $n \to \infty$, if no finite group of agents are overly influential.

$$\lim_{n\to\infty} \lim_{t\to\infty} \mu_{i,t} = \theta^* \qquad \forall i$$

Bayesian (rational) learning

Θ	A finite set of possible states of the world
$\theta^* \in \Theta$	The unobservable true state of the world
$s_t \in S$	The noisy signal observed by the rational agent
$\ell(s\|\theta)$	The likelihood function, known to the individual
$\mu_t(\theta)$	Time t beliefs of the agent
$\mu_0(\theta)$	The prior beliefs of the agent

Time t forecasts of the next observation:

$$m_t(s_{t+1}) = \int_\Theta \ell(s_{t+1}|\theta) d\mu_t(\theta)$$

Bayesian updating of beliefs

Rational update of the beliefs:

$$\mu_{t+1}(\theta) = \mu_t(\theta)\frac{\ell(s_{t+1}|\theta)}{m_t(s_{t+1})}$$

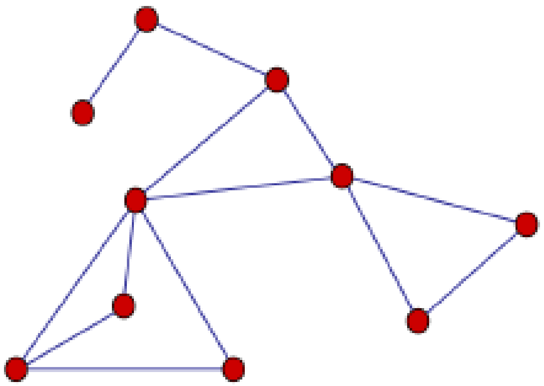
Define: $\mathbb{P}^* = \otimes_{t=1}^{\infty}\ell(\cdot|\theta^*)$.

Theorem

If Θ is finite and $\mu_0(\theta^*) > 0$, then the forecasts of the Bayesian agent are eventually correct with $\mathbb{P}^*-$probability one.

D. Blackwell and L. Dubins
Merging of Opinions with Increasing Information
Annals of Mathematical Statistics, 1962.

Bayesian learning on Networks



$$\mu_{i,t}(\theta) = \mathbb{P}\left[\theta = \theta^* | \mathcal{F}_{i,t}\right]$$

where

$$\mathcal{F}_{i,t} = \sigma\left(s_1^i, \ldots, s_t^i, \{\mu_{j,k} : j \in \mathcal{N}_i, k \leq t\}\right)$$

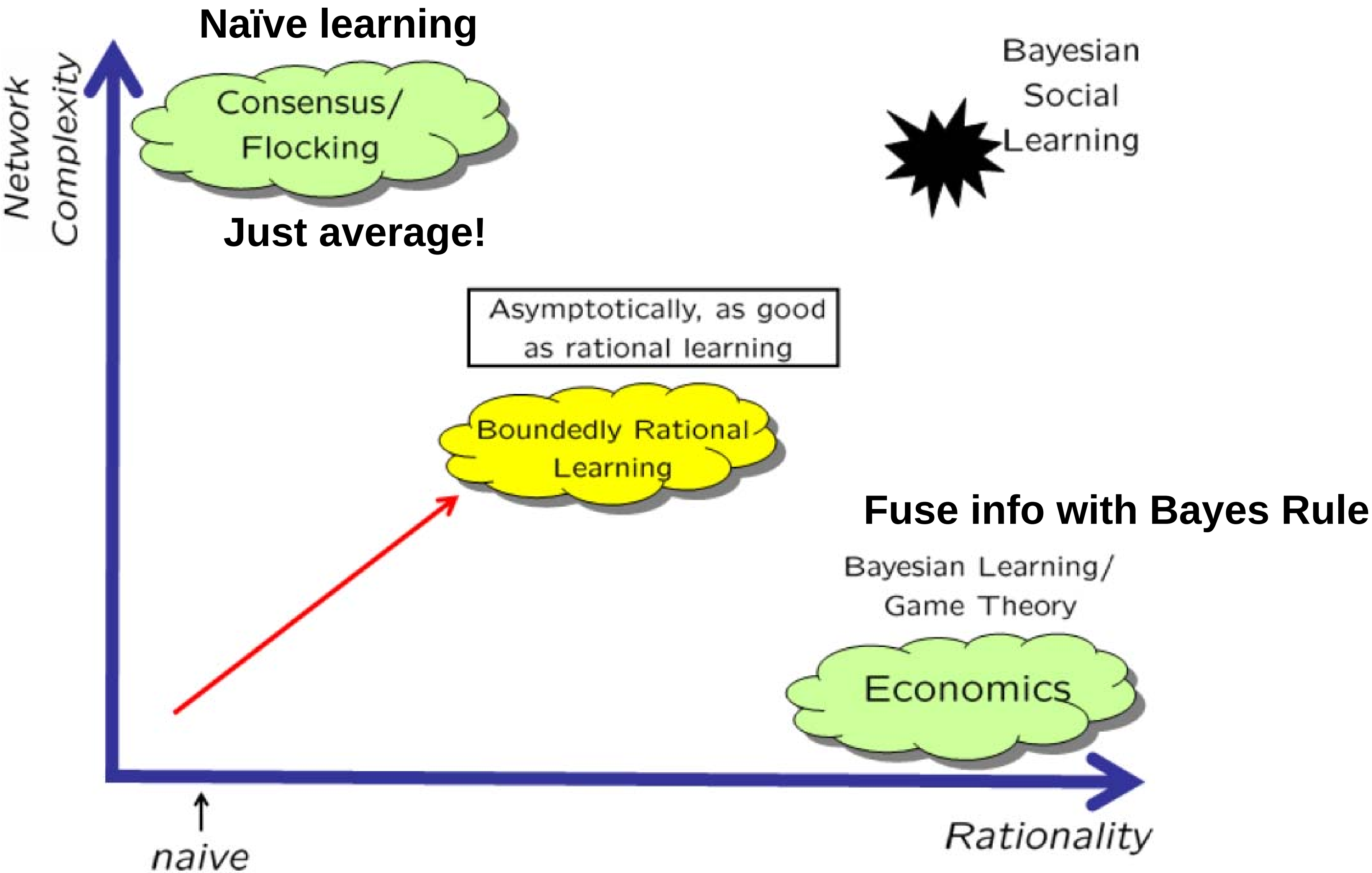
is the information available to agent i up to time t.

Agents need to make rational deductions about everybody's beliefs based on only observing neighbors' beliefs:

Problem with Bayesian Social learning

1. Incomplete network information
2. Incomplete information about other agents' signal structures
3. Higher order beliefs matter ▸ Example **Borkar and Varaiya'78**
4. The source of each piece of information is not immediately clear

Naïve vs. Rational learning



Locally rational, globally naïve: Bayesian learning under peer pressure

Need a local and computationally tractable update, which hopefully delivers asymptotic social learning.

Agent i is

- Bayesian when it comes to her observation
- non-Bayesian when incorporating others information

[Tahbaz-Salehi, Sandroni, and Jadbabaie 2009]

Model Description

$\mathcal{N} = \{1, 2, \ldots, n\}$	individuals in the society
$G = (\mathcal{N}, \mathcal{E})$	social network
Θ	finite parameter space
$\theta^* \in \Theta$	the unobservable true state of the world
$s_t = (s_t^1, \ldots, s_t^n)$	s_t^i is the signal observed by agent i at time t
$S = S_1 \times S_2 \times \cdots \times S_n$	signal space
$\ell(s\|\theta)$	the likelihood function (prob. of observing s if the true state is θ)
$\ell_i(s^i\|\theta)$	the marginal likelihood function

Model Description

$\mu_{i,t}(\theta)$	time t beliefs of agent i (a probability measure on Θ)
$\mu_{i,0}(\theta)$	agent i's prior belief
$\mathbb{P}^* = \otimes_{t=1}^{\infty} \ell(\cdot\|\theta^*)$	the true probability measure

Agent i's time t forecasts of the next observation profile:

$$m_{i,t}(s_{t+1}) = \int_{\Theta} \ell(s_{t+1}|\theta) d\mu_{i,t}(\theta)$$

What do we mean by learning?

Definition Weak merging of opinions

The Forecasts of agent i are eventually correct on a path $\{s_t\}_{t=1}^{\infty}$ if, along that path,

$$m_{i,t}(\cdot) \rightarrow \ell_i(\cdot|\theta^*) \quad \text{as} \quad t \rightarrow \infty.$$

Definition Asymptotic learning

Agent i asymptotically learns the true parameter θ^* on a path $\{s_t\}_{t=1}^{\infty}$ if, along that path,

$$\mu_{i,t}(\theta^*) \rightarrow 1 \quad \text{as} \quad t \rightarrow \infty.$$

- Asymptotic learning, in this setup, is stronger.
- Depends on the information structure.

Our Model: non-Bayesian social learning

$$\mu_{i,t+1}(\theta) = a_{ii} BU(\mu_{i,t}; s^i_{t+1})(\theta) + \sum_{j \in \mathcal{N}_i} a_{ij} \mu_{j,t}(\theta)$$

where

$$BU(\mu_{i,t}; s^i_{t+1})(\theta) = \mu_{i,t}(\theta) \frac{\ell_i(s^i_{t+1}|\theta)}{m_{i,t}(s^i_{t+1})}$$

$$a_{ij} \geq 0 \quad , \quad \sum_{j \in \mathcal{N}_i} a_{ij} = 1$$

- Individuals rationally update the beliefs after observing the signal
- exhibit a bias towards the average belief in the neighborhood

Why this update?

$$\mu_{i,t+1}(\theta) = a_{ii}\mu_{i,t}(\theta)\frac{\ell_i(s^i_{t+1}|\theta)}{m_{i,t}(s^i_{t+1})} + \sum_{i \neq j} a_{ij}\mu_{j,t}(\theta) \qquad \forall \theta \in \Theta$$

- Does not require knowledge about the network.
- Does not require deduction about the beliefs of others.
- Does not require knowledge about other agents' signallings.
- The update is local and tractable.
- If the signals are uninformative, reduces to the consensus update.
- Reduces to the benchmark Bayesian case if agents assign weight zero to the beliefs of their neighbors. [Blackwell and Dubins 1962]

Eventually correct forecasts

$$\mu_{i,t+1}(\theta) = a_{ii}\mu_{i,t}(\theta)\frac{\ell_i(s_{t+1}^i|\theta)}{m_{i,t}(s_{t+1}^i)} + \sum_{i \neq j} a_{ij}\mu_{j,t}(\theta) \qquad \forall \theta \in \Theta$$

Theorem

Suppose that

1. the social network is strongly connected,
2. $a_{ii} > 0$ for all $i \in \mathcal{N}$,
3. there exists an agent i such that $\mu_{i,0}(\theta^*) > 0$.

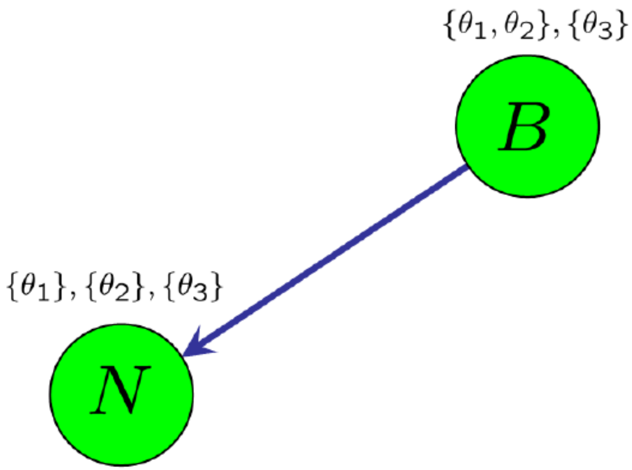
Then the forecasts of all agents are eventually correct $\mathbb{P}^*$-almost surely, that is, $m_{i,t}(\cdot) \rightarrow \ell_i(\cdot|\theta^*)$.

Agents will make accurate predictions about the future

Why strong connectivity?

What if the network has a directed spanning tree but is not strongly connected?

- $\mathcal{N} = \{B, N\}$
- $\Theta = \{\theta_1, \theta_2, \theta_3\}$
- $\theta^* = \theta_2$



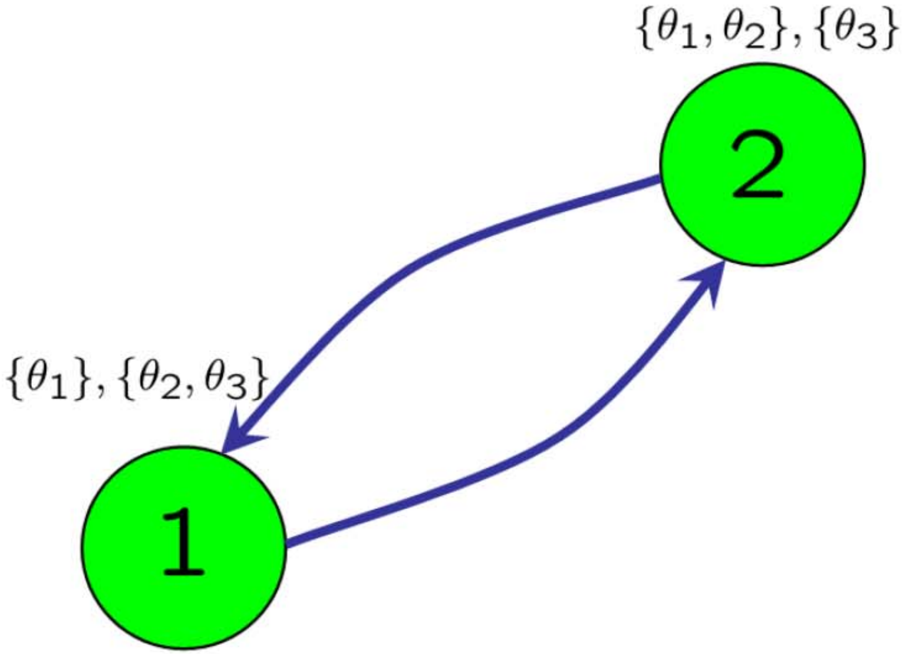
$$\mu_{N,t+1}(\theta) = \lambda \mu_{N,t}(\theta) \frac{\ell_N(s_{t+1}^N|\theta)}{m_{N,t}(s_{t+1}^N)} + (1-\lambda)\mu_{B,t}(\theta) \qquad \forall \theta \in \Theta$$

- No convergence if different people interpret signals differently
- N is misled by listening to the less informed agent B

Example

In any strongly connected social network, forecasts of all agents are correct on almost all sample paths.

- $\mathcal{N} = \{1, 2\}$
- $\Theta = \{\theta_1, \theta_2, \theta_3\}$
- $\theta^* = \theta_2$



One can actually learn from others

Observationally -equivalent states and distinguishability

- A state is *observationally-equivalent with the true state* from the point of view of an agent if the conditional likelihoods are the same, i.e. $\ell_i(s^i|\theta) = \ell_i(s^i|\theta^*)$ for all $s^i \in S_i$
- States that are not equivalent to the true state are distinguishable, i.e., there exists signals and a large enough time such that $$\frac{\ell_i(s^i_{t+1} s^i_t \cdots s^i_1|\theta)}{\ell_i(s^i_{t+1} s^i_t \cdots s^i_1|\theta^*)} \leq \delta < 1$$
- Technical assumption (stronger that what’s needed)

For any agent i, there exists a signal $\hat{s}^i \in S_i$ and a positive number δ_i

$$\frac{\ell_i(\hat{s}^i|\theta)}{\ell_i(\hat{s}^i|\theta^*)} \leq \delta_i < 1 \qquad \forall \theta \notin \bar{\Theta}_i.$$

distinguishable states

Agreement on Beliefs

Proposition

Under the assumptions of the previous proposition:

1. The beliefs of all agents converge with $\mathbb{P}^*$-probability 1.
2. Moreover, all agents have asymptotically equal beliefs $\mathbb{P}^*$-almost surely.

$$\lim_{t \to \infty} \mu_{i,t}(\theta) \text{ exists and is independent of } i.$$

Consensus!

Learning from others

All agents have asymptotically equal forecasts. Therefore,

- Each agent can correctly forecast every other agent's signals.

$$\forall i, j \in \mathcal{N} \quad \int_{\Theta} \ell_j(\cdot|\theta) d\mu_{i,t}(\theta) \longrightarrow \ell_j(\cdot|\theta^*) \qquad \mathbb{P}^* - \text{a.s.}$$

Local information of any agent is revealed to every other agent.

- This does not mean that the agents can forecast the joint distributions. They can only forecast the marginals correctly.
- To be expected: only marginals appear in the belief update scheme.

Theorem

Suppose that:

(a) The social network is strongly connected.

(b) All agents have strictly positive self-reliances.

(c) There exists an agent with positive prior belief on the truth θ^*.

(d) There is no state $\theta \neq \theta^*$ that is observationally equivalent to θ^* from the point of view of all agents in the network.

Then, all agents in the social network learn the true state of the world $\mathbb{P}^*$ almost surely; that is,

$$\mu_{i,t}(\theta^*) \longrightarrow 1 \qquad \mathbb{P}^* - a.s. \quad \forall i.$$

Rate of Convergence

Exponential convergence: $e^{-\lambda t}$

$$F_\theta(s_t) = A + \text{diag}\left(a_{ii}[\frac{\ell_i(s_t^i|\theta)}{\ell_i(s_t^i|\theta^*)} - 1] \right)$$

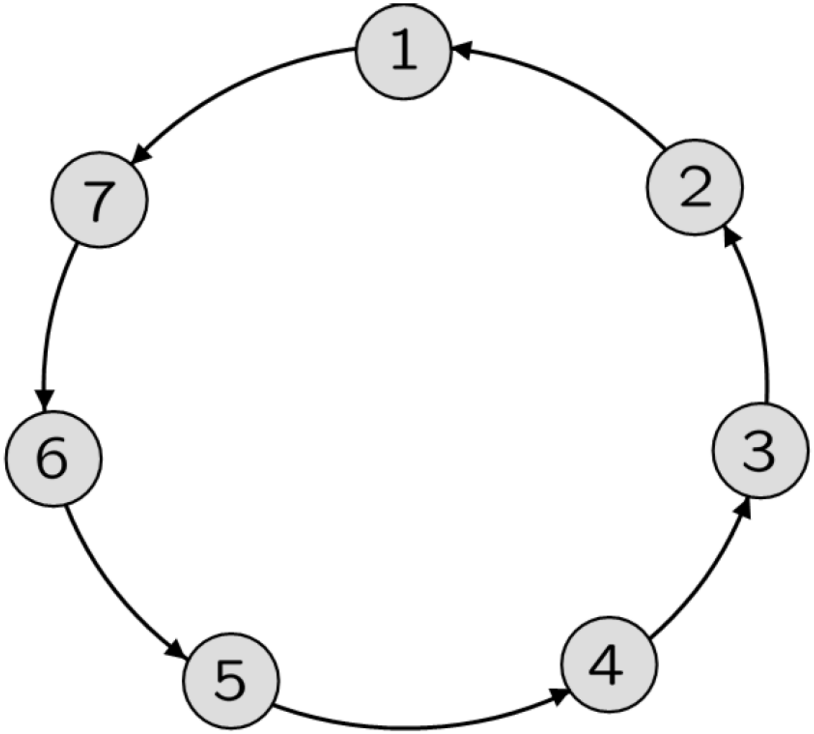
$$\lambda = -\min_{\theta \neq \theta^*} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\log \|F_\theta(s_1) \dots F_\theta(s_T)\|\right]$$

Therefore:

$$-\min_{\theta \neq \theta^*} \log \underline{\rho}\left(\{F_\theta(s) : s \in S\}\right) \geq \lambda \geq -\min_{\theta \neq \theta^*} \log \bar{\rho}\left(\{F_\theta(s) : s \in S\}\right)$$

where $\bar{\rho}(M)$ and $\underline{\rho}(M)$ are the upper and lower spectral radii of the set of matrices M.

Example



$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_7\}$$

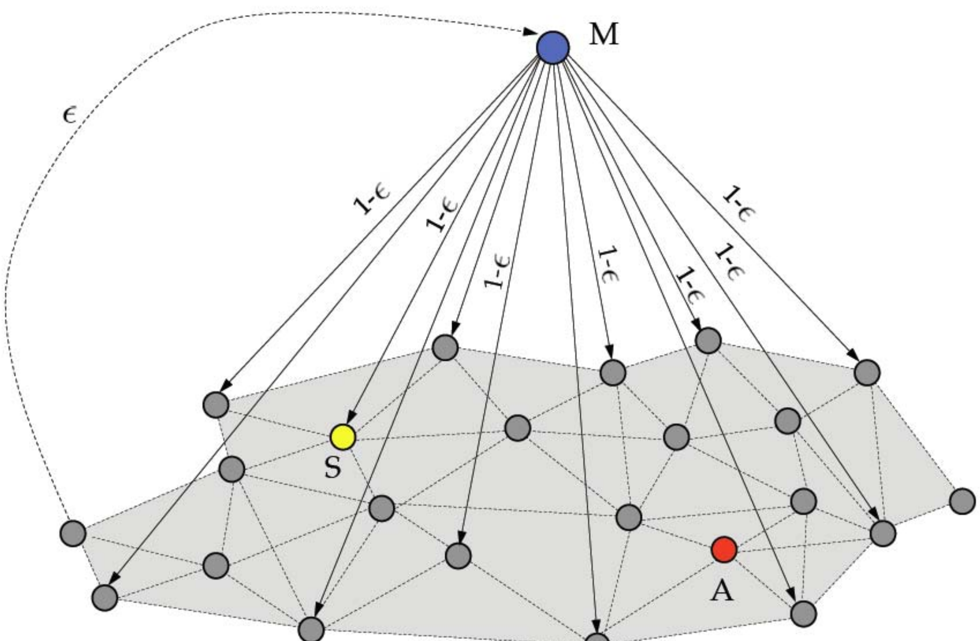
$$\theta^* = \theta_1$$

$$S_i = \{H, T\}$$

$$\ell_i(H|\theta) = \begin{cases} \frac{i}{i+1} & \text{if } \theta = \theta_i \\ \frac{1}{i+1} & \text{otherwise} \end{cases}$$

Local information of every agent is revealed to every other agent.

Can truth prevail despite loud, wrong, highly connected individuals?



- Agent M assigns 0 probability on true state, very opinionated, influential (high out-degree, has no informative private signals and 0 prior on truth
- Agent S is the only agent with informative-enough private signals to resolve identification problems (ie to learn the true state if can get correct forecast), but has zero prior on truth
- Agent A is the only agent with positive prior probability on true state
- Everyone will eventually almost surely get correct forecast and they will all learn the true state!

Summary

How information is aggregated over networks?

- From local information to inference about global uncertainties

Non-Bayesian social learning model

- Learning the true parameter, with little cost
- No information about network topology
- No information on signal structures
- No rational deductions
- Complete learning under mild conditions: Agents learn as if they have access to the observations of all agents at all times.

On-going work:

- Extends to changing graphs under some conditions on weights
- Exponential convergence