University of Maryland September 30, 2011

# Non-Bayesian Social Learning and Information Dissemination in Complex Networks


**Ali Jadbabaie**

**Department of Electrical and Systems Engineering and GRASP Laboratory**

UNIVERSITY *of* PENNSYLVANIA

Pooya Molavi, Victor Preciado (Penn)

Alireza Tahbaz-Salehi (Columbia GSB), Kamiar R Rad (Columbia Statistics)

Alvaro Sandroni (Northwestern University, Kellogg School)

# Motivation

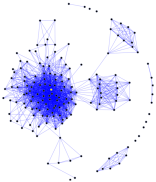
- Individuals form opinions about social, economic, and political issues
- These opinions influence their decisions when faced with choices:
  - Choice of agricultural products.
  - Buy Mac or PC?
  - Smoke or not to smoke?
  - Vote Democrat or Republican?
- What is the role of social networks in forming opinions?

# Motivation

- Individuals form opinions about social, economic, and political issues
- These opinions influence their decisions when faced with choices:
  - Choice of agricultural products.
  - Buy Mac or PC?
  - Smoke or not to smoke?
  - Vote Democrat or Republican?
- What is the role of social networks in forming opinions?

# Motivation

- Individuals form opinions about social, economic, and political issues
- These opinions influence their decisions when faced with choices:
  - Choice of agricultural products.
  - Buy Mac or PC?
  - Smoke or not to smoke?
  - Vote Democrat or Republican?



- What is the role of social networks in forming opinions?

# Naïve Opinion Dynamics Models

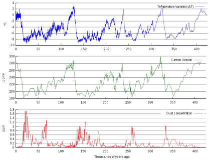
- How do agents form subjective opinions and how these opinions are diffused in social networks? e.g., fashion trends, consumption tastes, ...
- In most cases, there is *no* underlying "true state".
- In some scenarios there *is* a true state that can be identified through observations, e.g., climate change.
- Is it "man-made" or "the wavy arm thing"?

# Problem Description

- Under what assumptions can we be sure that the agents can learn the true state of the world?
- How can this be implemented tractably?
- Rapidly growing interest in the topic in Economics and Game Theory literature: Ellison & Fudenberg '93, '95, Smith & Sorensen '98, Banerjee '98, Acemoglu *et al.* 2008, Bala & Goyal ' 98, 2001, DeMarzo *et al.* 2003, Gale & Kariv '2003, and many others
- Also studied in the context of estimation and detection, Tsitsiklis '85-'95, Borkar & Varaiya '78

# Naïve Opinion Dynamics Models

- How do agents form subjective opinions and how these opinions are diffused in social networks? e.g., fashion trends, consumption tastes, ...
- In most cases, there is *no* underlying "true state".
- In some scenarios there *is* a true state that can be identified through observations, e.g., climate change.
- Is it "man-made" or "the wavy arm thing"?




# Bayesian Learning (Blackwell and Dubins '62)

Given outcome of an i.i.d. coin toss process, what can one learn?

H T H H H T H T T H H T T T T H T ...

- The next toss has roughly 50% chance of being H.
  $\longrightarrow$ Week merging
- Q: Can two coins by distinguished by observing coin toss outcomes?
  A: Only if $\mathbb{P}(H|\text{Coin 1}) \neq \mathbb{P}(H|\text{Coin 2})$.
  $\longrightarrow$ Observational distinguishability
- Q: How many observations we need to distinguish them?
  A: Depends on Kullback-Leibler divergence of conditional distributions.

# Bayesian Learning (Blackwell and Dubins '62)

Given outcome of an i.i.d. coin toss process, what can one learn?

H T H H H T H T T H H T T T T H T ...

- The next toss has roughly 50% chance of being H.
  $\longrightarrow$ Week merging
- Q: Can two coins by distinguished by observing coin toss outcomes?
  A: Only if $\mathbb{P}(H|\text{Coin }1) \neq \mathbb{P}(H|\text{Coin }2)$.
  $\longrightarrow$ Observational distinguishability
- Q: How many observations we need to distinguish them?
  A: Depends on Kullback-Leibler divergence of conditional distributions.

# Bayesian Learning (Blackwell and Dubins '62)

Given outcome of an i.i.d. coin toss process, what can one learn?

H T H H H T H T T H H T T T T H T ...

- The next toss has roughly 50% chance of being H.
  $\longrightarrow$ Week merging
- Q: Can two coins by distinguished by observing coin toss outcomes?
  A: Only if $\mathbb{P}(H|\text{Coin }1) \neq \mathbb{P}(H|\text{Coin }2)$.
  $\longrightarrow$ Observational distinguishability
- Q: How many observations we need to distinguish them?
  A: Depends on Kullback-Leibler divergence of conditional distributions.

# Bayesian Learning (Blackwell and Dubins '62)

Given outcome of an i.i.d. coin toss process, what can one learn?

H T H H H T H T T H H T T T T H T ...

- The next toss has roughly 50% chance of being H.
  $\longrightarrow$ Week merging
- Q: Can two coins by distinguished by observing coin toss outcomes?
  A: Only if $\mathbb{P}(H|\text{Coin } 1) \neq \mathbb{P}(H|\text{Coin } 2)$.
  $\longrightarrow$ Observational distinguishability
- Q: How many observations we need to distinguish them?
  A: Depends on Kullback-Leibler divergence of conditional distributions.

# Rate of Bayesian Learning

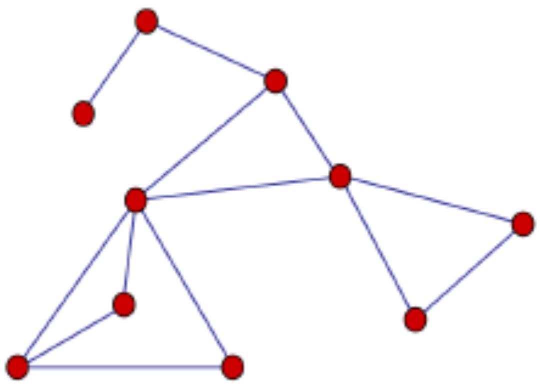
$$D_{\mathsf{KL}}(\mathbb{P}_1 \| \mathbb{P}_2) \stackrel{\text{def}}{=} \int \log \frac{\mathbb{P}_1(x)}{\mathbb{P}_2(x)} d\mathbb{P}_1(x) \geq 0 \quad \& \quad = 0 \qquad \text{iff} \qquad \mathbb{P}_1 = \mathbb{P}_2 \text{ a.s.}$$

**Lemma (Chernoff-Stein)**

*The probability of error goes to zero exponentially fast.*

$$\limsup_{t\to\infty} \frac{1}{t} \log|e| \leq -D_{KL}(\mathbb{P}(\cdot|\mathit{Coin\ 1}) \| \mathbb{P}(\cdot|\mathit{Coin\ 2})).$$

# 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 Description

- Agents repeatedly communicate with their neighbors in a social network, e.g., colleagues, friends,... In a contact agents average their beliefs.
- Sometimes, some agents make private observations and incorporate the observations in their beliefs in a Bayesian way.
- What is the outcome of this process?

# Problem with Bayesian Social learning

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

**Borkar and Varaiya'78**

# The Model

Assumptions:

- At each time period some agents receive signals and incorporate them in their beliefs.
- Then any agent averages her belief with those of her neighbors.
- Observations are i.i.d.

# Notation

- $\Theta$: the finite set of possible states of the world
- $\theta^*$: true state of the world
- $\mu_{i,t}^{\theta}$: beliefs of the agents
- $m_{i,t}(s_i)$: agent $i$'s forecast at time $t$ that signal $s_i$ will be observed next
- $S_i$: agent $i$'s signal space
- Signals are generated according $\ell(\cdot|\,\theta^*)$.
- $\ell_i(\cdot|\,\theta^*)$: the $i$th marginal of $\ell(\cdot|\,\theta^*)$
- $\bar{\Theta}_i = \{\theta : \ell_i(\cdot|\theta) = \ell_i(\cdot|\theta^*)\}$: the set of signals that are observationally equivalent to $\theta^*$ from the point of view of $i$

# Notation, cont'd

- $\mathbb{P} = \ell(\cdot\,|\,\theta^*)^{\mathbb{N}}$: the product measure
- $(\Omega, \mathcal{F}, \mathbb{P})$: the probability triple
- $\mathcal{F}_t$: the filtration generated by observations to time $t$
- $\omega \in \Omega$: the infinite sequence of signals

- Network is represented by a weighted directed graph.
- $a_{ij}$: the weight $i$ assigns to the belief of $j$
- $A = [a_{ij}]$: the weighted graph matrix
- $\mathcal{N}_i$: neighbors of agent $i$
- $a_{ii}$: self reliance of agent $i$

# Model, cont'd

Agent updates her belief to the convex combination of her Bayesian posterior and her neighbors' beliefs:

$$\mu_{i,t+1}^{\theta} = a_{ii}\mu_{i,t}^{\theta}\frac{\ell_i(\omega_{i,t+1}|\,\theta)}{m_{i,t}(\omega_{i,t+1})} + \sum_{j\in\mathcal{N}_i} a_{ij}\mu_{j,t}^{\theta},$$

where $\omega_{i,t}$ is observation of agent $i$ at time $t$.

$m_{i,t}(\cdot)$ is the *one step forecast* of agent $i$ defined as:

$$m_{i,t}(s_i) = \sum_{\theta\in\Theta} \ell_i(s_i|\,\theta)\mu_{i,t}^{\theta}.$$

# Weak Social Learning

**Theorem (Jadbabaie, Sandroni, Tahbaz Salehi 2010)**

*Assume:*

(a) *The social network is strongly connected.*

(b) *There exists an agent with positive prior belief on the true parameter $\theta^*$.*

*Then agents with positive self-reliance will eventually forecast immediate future correctly.*

**Sketch of proof**

- $v^T \mu_t(\theta^*)$ is a bounded submartingale that converges.
- $v^T \log \mu_t(\theta^*)$ is a bounded submartingale that converges.
- The submartingale increments go to zero $\mathbb{P}$-almost surely.

# Previous Result

**Theorem**

*Assume:*

(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 true parameter $\theta^*$.*

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

*Then all the agents learn the true state of the world with $\mathbb{P}$-probability one.*

# Sketch of Proof (1)

- Look at the $k$ step forecast of the agent:

$$m_{i,t}(s_{i,1}, s_{i,2}, \ldots, s_{i,k}) = \sum_{\theta \in \Theta} \mu_{i,t}^{\theta} \ell_i(s_{i,1}, s_{i,2}, \ldots, s_{i,k} | \theta),$$

where
$\ell_i(s_{i,1}, s_{i,2}, \ldots, s_{i,k} | \theta) = \ell_i(s_{i,1} | \theta)\ell_i(s_{i,2} | \theta) \ldots \ell_i(s_{i,k} | \theta)$.

- Asymptotically $\mathbb{P}$-almost surely the $k$ step forecast decomposes into products of $k$ one step forecasts, i.e.

$$\begin{aligned} m_{i,t}(s_{i,1}, s_{i,2}, \ldots, s_{i,k}) &\overset{a.a.s.}{=} m_{i,t}(s_{i,1}) m_{i,t+1}(s_{i,2}) \ldots m_{i,t+k}(s_{i,k}) \\ &\overset{a.a.s.}{=} \ell_i(s_{i,1}, s_{i,2}, \ldots, s_{i,k} | \theta^*). \end{aligned}$$

# Sketch of Proof (2)

**Lemma**

*Asymptotically $\mathbb{P}$-almost surely, the dynamic of opinions follow consensus update in expectation, i.e.*

$$\mathbb{E}(\mu_{t+1}(\theta)|\mathcal{F}_t) \overset{a.a.s.}{=} A\mu_t(\theta).$$

**Lemma**

*Asymptotically $\mathbb{P}$-almost surely, the $k$ step forecast decomposes as*

$$m_{i,t}(\omega_{i,t+1}, s_{i,2}, \ldots, s_{i,k}) \overset{a.a.s.}{=} m_{i,t}(\omega_{i,t+1}) m_{i,t}(s_{i,2}, \ldots, s_{i,k}).$$

Proof: Induction on $k$.

**Claim**

The result is also true for arbitrary $s_{i,1} \in S_i$. This is intuitive because of independence.

# Sketch of Proof (3)

**Lemma**

*If the true state is distinguishable, there exists a finite number $\hat{k}_i$ and signals $\hat{s}_{i,1}, \hat{s}_{i,2}, \ldots, \hat{s}_{i,\hat{k}_i}$ such that*

$$\frac{\ell_i(\hat{s}_{i,1}, \hat{s}_{i,2}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta)}{\ell_i(\hat{s}_{i,1}, \hat{s}_{i,2}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta^*)} \leq \delta_i < 1 \qquad \forall \theta \notin \bar{\Theta}_i,$$

*for some $\delta_i \geq 0$.*

**Claim**

The signal sequence in which $s_i$ appears with frequency $\ell_i(s_i | \theta^*)$ has this property.

Proof: Maximize over all the probability measures over $S_i$.

# Sketch of Proof (4)

- $m_{i,t}(s_{i,1}, \cdots, s_{i,k}) \rightarrow \ell_i(s_{i,1}, \cdots, s_{i,k} | \theta^*)$ with $\mathbb{P}$-probability one for any sequence of finite length.
- Use the seqeunce in the previous Lemma.
- Therefore,

$$\sum_{\theta} \mu_{i,t}^{\theta} \frac{\ell_i(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta)}{\ell_i(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta^*)} - 1 \longrightarrow 0$$

$$\sum_{\theta \notin \bar{\Theta}_i} \mu_{i,t}^{\theta} \frac{\ell_i(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta)}{\ell_i(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i} | \theta^*)} + \sum_{\theta \in \bar{\Theta}_i} \mu_{i,t}^{\theta} - 1 \longrightarrow 0.$$

And,

$$(1 - \delta_i) \sum_{\theta \notin \bar{\Theta}_i} \mu_{i,t}^{\theta} \longrightarrow 0.$$

# New Result

**Theorem**

*Assume:*

(a) *The social network is strongly connected.*

(b) *There exists an agent with positive prior belief on the true parameter $\theta^*$.*

(c) *For any $\theta \neq \theta^*$, there exist an agent with positive self-reliance who can distinguish $\theta$ from $\theta^*$.*

*Then all the agents learn the true state of the world with $\mathbb{P}$-probability one.*

# Rate of Learning

**Theorem**

*With the same assumptions convergence of $\bar{\mu}_t(\theta)$ to zero is exponential, i.e., for all $\epsilon > 0$ and in a set of $\mathbb{P}$-probability one,*

$$\lambda_1' + \epsilon \leq \limsup_{t \to \infty} \frac{1}{t} \log \|\bar{\mu}_t\| \leq \lambda_1 + \epsilon,$$

*where $\lambda_1 < 0$ is the top Lyapunov exponent of the linearized system and $\lambda_1' < 0$*

- $\bar{\mu}_{i,t}(\theta)$ is the restriction of $\mu_{i,t}^{\theta}$ to $\Theta \setminus \bar{\Theta}$, where $\bar{\Theta} = \bar{\Theta}_1 \cap \cdots \cap \bar{\Theta}_n$.

# Proof Outline

- Look at $\bar{\mu}_t$ as the trajectory of a *Random Dynamical System* (RDS):

$$\bar{\mu}_{t+1} = \varphi_t(\omega; \bar{\mu}_t).$$

- Linearize the dynamics of $\bar{\mu}_{i,t}(\theta)$ at the origin to get $z_{i,t}(\theta)$:

$$\varphi_t(\omega; x) = M_t(\omega)x + F_t(\omega; x),$$

$$z_{t+1} = M_t(\omega) z_t.$$

- A martingale argument shows that $z_t \to 0$ for *all* initial conditions. Thus, $\lambda_1 < 0$.
- Therefore, the nonlinear RDS is exponentially stable in a neighborhood of the origin.

# Bounds on the Rate of Learning

**Theorem**

(a)

$$\lambda_1' \geq -\max_{\theta \notin \bar{\Theta}} \sum_{i \in \mathcal{N}} v_i a_{ii} D_{KL}(\ell_i(\cdot \mid \theta^*) \| \ell_i(\cdot \mid \theta)).$$

(b) *For small distinguishability of the true state,*

$$\lambda_1 \leq -\min_{\theta \notin \bar{\Theta}} \sum_{i \in \mathcal{N}} v_i a_{ii} D_{KL}(\ell_i(\cdot \mid \theta^*) \| \ell_i(\cdot \mid \theta)),$$

*where $v_i$ is the eigenvector centrality of agent $i$.*

- Upper bound is found using an upper bound by Gharavi and Anantharam (2005) on the top Lyapunov exponent (TLE).

# Observations About the Rate

- The bounds can be made arbitrarily tight when there are only two states.
- The rate is always smaller than that of an "ideal" observer with access to all observations.
- Learning is faster when central (influential) agents receive better signals.
- While in some large societies rate goes to zero, in others it is bounded below.

# Upper Bound on TLE (Gharavi, Anantharam '05)

- $M^k$: the $k$th possible realization of $M_t(\omega)$
- $p_k$: the probability of $M^k$ being realized
- $H(p)$: the entropy of $p$
- $\mathcal{S} = \{1, \ldots, |S|\}$: an enumeration of possible signal profiles
- $\mathcal{M}$: set of probability distributions over $(\mathcal{N} \times \mathcal{S}) \times (\mathcal{N} \times \mathcal{S})$
- $H(\eta)$: entropy of $\eta \in \mathcal{M}$
- $F(\eta)$: defined for $\eta \in \mathcal{M}$ as

$$F(\eta) = \sum_{\substack{i,j \in \mathcal{N} \\ k,l \in \mathcal{S}}} \eta_{i,j}^{k,l} \log M_{j,i}^{k}.$$

# Upper Bound on TLE (Gharavi, Anantharam '05), cont'd

- An upper bound for the top Lyapunov exponent of a Markovian product of nonnegative matrices using Markovian type counting arguments.
- The bound is expressed as the maximum of a nonlinear concave function over a finite-dimensional convex polytope of probability distributions.

$$
\begin{aligned}
\hat{\lambda}_1 \quad = \quad & \max_{\eta \in \mathcal{M}} \quad H(\eta) + F(\eta) - H(p) \\
\text{subject to} \quad & \eta^{k,l}_{*,*} = p_k p_l \\
& \eta^{k,*}_{i,*} = \eta^{*,k}_{*,i} \\
& \eta^{k,l}_{i,j} = 0 \quad \text{if} \quad M^k_{j,i} = 0.
\end{aligned}
$$

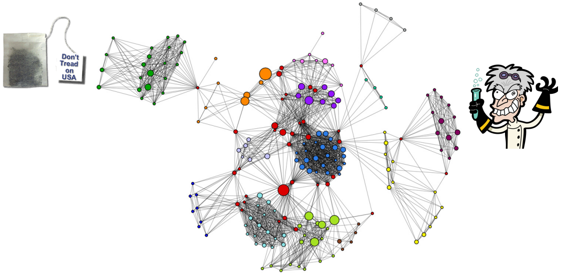
# Clustered Networks

- The social networks are clustered, for instance:
  Climate scientists do not talk to regular people as frequently as they talk to each other.

- The extreme case is when network is not strongly connected.

# Clustered Networks

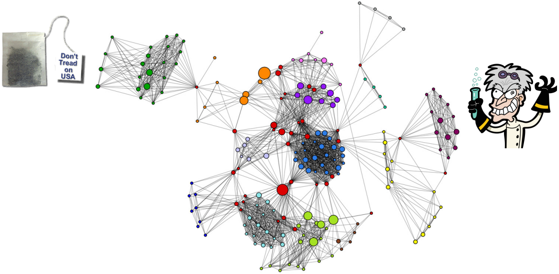
- The social networks are clustered, for instance:
  Climate scientists do not talk to regular people as frequently as they talk to each other.




- The extreme case is when network is not strongly connected.

# Clustered Networks

- The social networks are clustered, for instance:
  Climate scientists do not talk to regular people as frequently as they talk to each other.




- The extreme case is when network is not strongly connected.

# Learning in Non-Strongly-Connected Networks

- Social network can be partitioned into minimal closed groups and agents that belong to no closed minimal group.
- The evolution of beliefs in each "island" is independent of the rest of network.
- Each minimal closed group is strongly connected.
- Beliefs of agents not belonging to groups will be a convex combination of beliefs of agents in minimal closed groups.

# Learning in Non-Strongly-Connected Networks, cont'd

- Assume that agents prior beliefs are stochastic.

**Theorem**

*For almost all prior beliefs and $\mathbb{P}$-almost all observation sequences:*

(a) *In each island and for any $\theta \neq \theta^*$, there exist an agent $i$ with $a_{ii} > 0$ who can distinguish $\theta$ from $\theta^*$.*

$$\Updownarrow$$

(b) *All agents will asymptotically learn the true state.*

- If (b) fails, agents in that island learn with probability zero.
- Agents in different islands will learn with different rates.