From Consensus to Social Learning

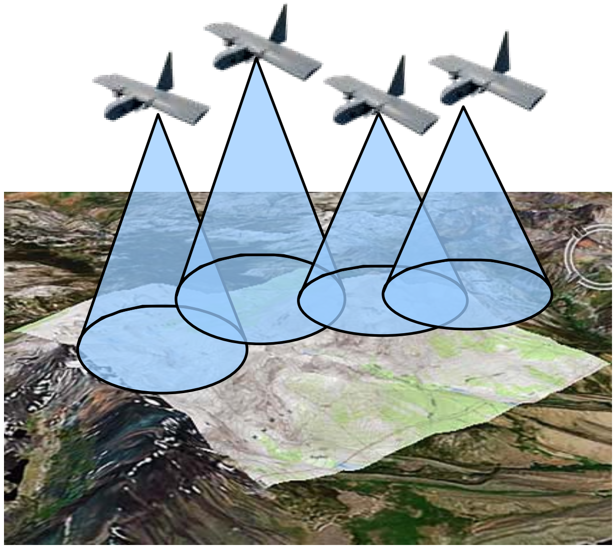
Ali Jadbabaie

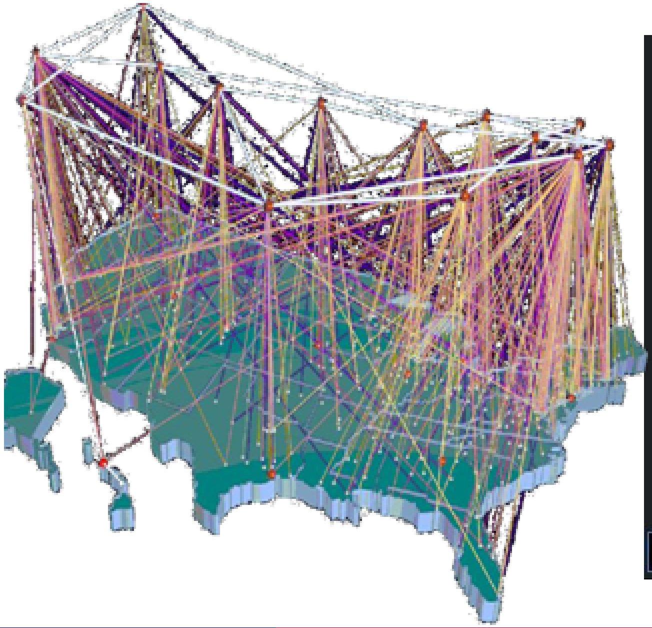
Alfred Fitler Moore Professor of Network Science
Electrical and Systems Engineering, Computer & Information Sciences
Operations & Information Management
University of Pennsylvania

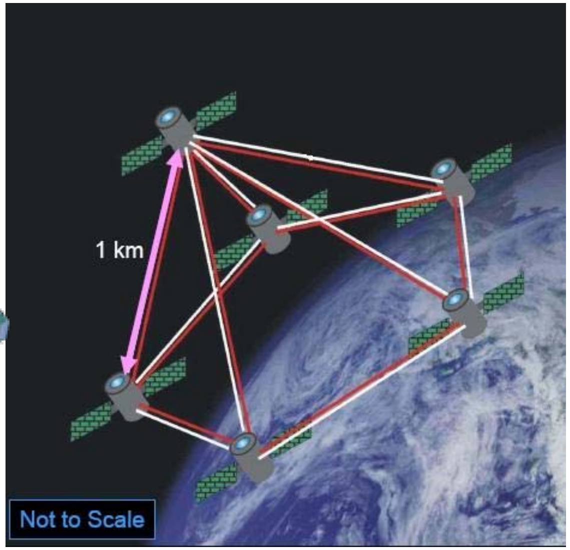
Research supported by NSF, ARO, ONR, and AFOSR
with Pooya Molavi (MIT Economics), Alireza Tahbaz Salehi(Columbia GSB)

Penn
UNIVERSITY of PENNSYLVANIA

Collective Phenomena:
Consensus motion coordination, and coverage

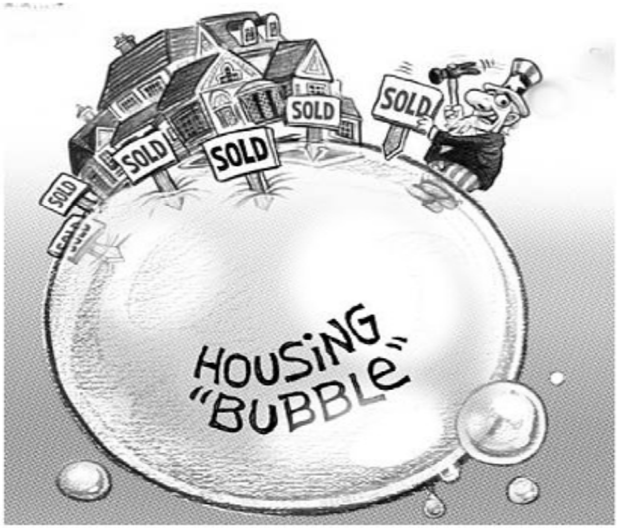






Collective Phenomena:

Social and Economic Networks



learning, spreading, and cascades

Consensus and aggregation of subjective probabilities

Reaching a Consensus

MORRIS H. DeGROOT*

Consider a group of individuals who must act together as a team or committee, and suppose that each individual in the group has his own subjective probability distribution for the unknown value of some parameter. A model is presented which describes how the group might reach agreement on a common subjective probability distribution for the parameter by pooling their individual opinions. The process leading to the consensus is explicitly described and the common distribution that is reached is explicitly determined. The model can also be applied to problems of reaching a consensus when the opinion of each member of the group is represented simply as a point estimate of the parameter rather than as a probability distribution.

distribution over Ω for which the probability of any measurable set A is $\sum_{i=1}^{k} p_i F_i(A)$. Some of the writers previously mentioned have suggested representing the overall opinion of the group by a probability distribution of the form $\sum_{i=1}^{k} p_i F_i$. Stone [13] has called such a linear combination an "opinion pool." The difficulty in using an opinion pool to represent the consensus of the group lies, of course, in choosing suitable weights $p_1, \cdots, p_k$. In the model that will be presented in this article, the consensus that is reached by the group will

Journal of the American Statistical Association, March 1974

DeGroot Learning Models

DeGroot based models are tractable:

- DeGroot (1974)
- Tsitsiklis (1984)
- DeMarzo, Vayanos, Zwiebel (2003)
- Jadbabaie, Lin, Morse (2003)
- Acemoglu, Ozdaglar, ParandehGheibi (2010)
- Golub and Jackson (2010)
- Mossel and Tamuz (2013)

DeGroot Learning Models

DeGroot based models are tractable:

- DeGroot (1974)
- Tsitsiklis (1984)
- DeMarzo, Vayanos, Zwiebel (2003)
- Jadbabaie, Lin, Morse (2003)
- Acemoglu, Ozdaglar, ParandehGheibi (2010)
- Golub and Jackson (2010)
- Mossel and Tamuz (2013)

There is empirical evidence in favor of DeGroot models:

- Chandrasekhar, Larreguy, Xandri (2012)

The Problem

Research question

How do the network structure and agents' information structure determine the extent of information aggregation?

Our model: an extension of DeGroot's learning model with

- continuous flow of new information
- heterogenous observations
- asymptotic agreement with the Bayesian benchmark

Information aggregation and social learning

Challenges

- Information is often dispersed through out the network
- No central mechanisms for aggregation
- Interactions are local
- Related: Diffusion, gossip in
  - face to face communications
  - online social media
- examples:
  - diffusion of micro finance programs (Banerjee et al. 2013)
  - Coordination during popular uprisings (Hassanpour (2012))
  - decision making in organizations (Calvó-Armengol, Beltran('09))
  - Making consumption decisions (Kotler ('86))
  - Learning new agricultural techniques (Hagerstrand ('69), Rogers ('83))

Model (agents and observations)

- $\{1, \ldots, n\}$: finite set of agents
- Agents want to learn an underlying state $\theta \in \Theta$.
- $t \in \mathbb{N}$: discrete time. 'State" drawn at $t = 0$ according to agents' common prior.
- $\omega_{it} \in S$: private observations of agent i at time t
- Conditional on θ being realized, $\omega_{it} \sim \ell_i^\theta \in \Delta S$.
- $\ell_i = \{\ell_i^\theta\}_{\theta \in \Theta}$: agent i's signal structure: what is the likelihood of $\omega_{it} \in S$, if θ is the truth?

Assumption (identifiability)

For all $\theta, \hat{\theta} \in \Theta$, there exists i such that $\ell_i^\theta \neq \ell_i^{\hat{\theta}}$. Globally, there is enough to discover the truth

Question:

Role of network and information structure?

Classical setting, no networks, What to expect?

Doob (1949), Blackwell and Dubins (1962)

Merging of opinions with increasing information: The belief of a Bayesian agent i with *absolutely continuous* prior observing a stream of signals will *merge to the truth*; i.e., she will learn the likelihood function ℓ_i.

Classical setting, no networks, What to expect?

Doob (1949), Blackwell and Dubins (1962)

Merging of opinions with increasing information: The belief of a Bayesian agent i with *absolutely continuous* prior observing a stream of signals will *merge to the truth*; i.e., she will learn the likelihood function ℓ_i.

Geanakoplos and Polemarchakis (1982)

We can't disagree forever: Two agents with a common prior exchanging beliefs repeatedly will reach agreement; moreover, their consensus belief will generically be as if they *commonly* knew each others' private information.

What happens in the networked case?

The Bayesian Benchmark: Multi-agent setting

- Let $\mathcal{X} = \overbrace{\Theta}^{\text{state}} \times \overbrace{\Omega}^{\text{signals}} \times \overbrace{\Gamma}^{\text{network}}$ be the measurable space that captures *all* uncertainty.
- Assume agents have a *common* prior over the $\mathcal{X}$.

The Bayesian Benchmark: Multi-agent setting

- Let $\mathcal{X} = \overbrace{\Theta}^{\text{state}} \times \overbrace{\Omega}^{\text{signals}} \times \overbrace{\Gamma}^{\text{network}}$ be the measurable space that captures *all* uncertainty.
- Assume agents have a *common* prior over the $\mathcal{X}$.

Theorem

Assume

(a) *agents' common prior has full support over $\mathcal{X}$;*

(b) *the realized network is strongly connected;*

(c) *the realized state is identifiable.*

Then all agents learn the true state asymptotically almost surely; i.e., $\mu_{it} \longrightarrow \mathbf{1}_{\theta^*} \qquad \text{for all } i \in \mathcal{N}$.

Agents need to reason about too many things. Is there a simpler behavioral model?

Alternative?

What is a "reasonable" Non-Bayesian alternative?
Extension/modification of DeGroot learning model (Golub and Jackson 2010) with these features:

- continuous flow of new information
- heterogenous stream of private observations
- asymptotic agreement with the Bayesian benchmark

  - Implications of the rate analysis
  - Axiomatic construction of non-Bayesian models

Model (learning rule)

- At $t \in \mathbb{N}$ agents also observe beliefs of their neighbors.
- $\mu_{it} \in \Delta\Theta$: belief of agent i at t

Model (learning rule)

- At $t \in \mathbb{N}$ agents also observe beliefs of their neighbors.
- $\mu_{it} \in \Delta\Theta$: belief of agent i at t
- The update rule:

$$\mu_{it+1} = a_{ii}\underbrace{\mathrm{BU}(\mu_{it};\omega_{it+1})}_{i} + \sum_{j\neq i} a_{ij} \underbrace{\mu_{jt}}_{ii}.$$

- i: Bayesian posterior belief conditioned on private signal
- ii: beliefs of the neighbors
- Weights sum to one representing network connections.
- Is there a behavioral foundation for this model?

Research question

How do the network structure and agents' information endowments determine the extent of information aggregation?

Model (social network)

- $a_{ij} > 0 \quad \Leftrightarrow$ Agent j is a neighbor of agent j.
- $A = [a_{ij}]$ row-stochastic social interaction matrix
- weights can be time-varying and belief-dependent.

Model (social network)

- $a_{ij} > 0 \quad \Leftrightarrow \quad$ Agent j is a neighbor of agent j.
- $A = [a_{ij}]$ row-stochastic social interaction matrix
- weights can be time-varying and belief-dependent.

Assumption (strong connectivity)

There is a directed path from any agent to any other one (can be generalized to switching graphs).

- Guarantees that information can flow from any agent to any other.

Asymptotic Learning

Proposition

If identifiability and strong connectivity assumptions are satisfied,

$$\mu_{it}(\cdot) \longrightarrow \mathbf{1}_{\theta}(\cdot)$$

the rate is (up to first order)

$$r \approx \min_{\theta} \min_{\hat{\theta} \neq \theta} \sum_{i=1}^{n} v_i h_i(\theta, \hat{\theta}) + h.o.t$$

- The learning process asymptotically coincides with Bayesian learning
- Unlike Bayesian models, the model is tractable
- Rate is a convex combination of relative entropies $h_i(\theta, \hat{\theta})$ with weights as eigenvector centrality v_i.
- Consistent with empirical and theoretical observations in *Jackson (2013,2014)*

Towards an axiomatic view

What should a reasonable model look like?

- If private signals are uninformative $\Rightarrow$ beliefs updated as in DeGroot '74 (consensus)
- If a signal is evidence in favor of a state, the posterior belief on that state should increase (increasing function of likelihood ratio)
- Update should be separable in terms of private signal and an aggregate of belief of neighbors

- All such updates converge, and have the same asymptotic rate (up to first order)
- One such example: Average log beliefs of neighbors with log private posterior

When signals are uninformative, model reverts to DeGroot.

Rate of Learning

Definition (total uncertainty)

TV distance between agents' beliefs and the true distribution:

$$e_t = \frac{1}{2}\sum_{i=1}^{n} \|\mu_{it}(\cdot) - \mathbf{1}_\theta(\cdot)\|_1$$

Definition (rate of learning)

$$\lambda = \liminf_{t\to\infty} \tfrac{1}{t}|\log e_t|$$

Rate of Learning

Definition (total uncertainty)

TV distance between agents' beliefs and the true distribution:

$$e_t = \frac{1}{2}\sum_{i=1}^{n} \|\mu_{it}(\cdot) - \mathbf{1}_\theta(\cdot)\|_1$$

Definition (rate of learning)

$$\lambda = \liminf_{t\to\infty} \tfrac{1}{t}|\log e_t|$$

- λ depends on
  - agents' information endowments: relative entropy
  - agents' network position: eigenvector centrality

Relative Entropy and Eigenvector Centrality

Definition (relative entropy)

Given $\hat{\theta} \neq \theta$,

$$h_i(\theta, \hat{\theta}) = \sum_{s \in S} \ell_i^{\theta}(s) \log \frac{\ell_i^{\theta}(s)}{\ell_i^{\hat{\theta}}(s)}$$

- $h_i(\theta, \hat{\theta})$: information in favor of θ against $\hat{\theta}$ when θ is realized
- $h_i(\theta, \hat{\theta}) = 0 \quad \Rightarrow \quad$ agent i cannot distinguish θ and $\hat{\theta}$
- larger $h_i(\theta, \hat{\theta}) \quad \Rightarrow \quad$ easier to rule out $\hat{\theta}$ when θ is realized

Eigenvector Centrality

Definition (eigenvector centrality)

Given A, the eigenvector centrality of agent i is

$$v_i = \sum_{j=1}^{n} v_j a_{ji}$$

Uniform Informativeness Order

Definition (uniform informativeness)

$$\ell_i \succeq_{\text{UI}} \ell_i'$$

if

$$h_i(\theta, \hat{\theta}) \geq h_i'(\theta, \hat{\theta}) \text{ for all } \theta, \hat{\theta}$$

- ℓ_i is more informative than ℓ_i' regardless of the realized state.
- a partial order on the set of signal structures
- weaker (more complete) than Blackwell's informativeness

Under which allocation of signals is learning the fastest?

Proposition

Suppose

- *agents' signals are comparable with respect to* $\succeq_{UI}$*;*
- $\ell_i \succeq_{UI} \ell_j$ *if and only if* $v_i \geq v_j$*.*

Then, no reallocation of signals increases the rate of learning.

- Positive assortative matching of centralities and signal qualities maximizes the rate of learning.
- intuition: Irrespective of the realized state, the most informative signals receive the most attention.

What if information endowments are incomparable?

- relative informativeness of agent i's signals for $(\theta, \hat{\theta})$:

$$\gamma_i(\theta, \hat{\theta}) = \sup\{\beta : h_i(\theta, \hat{\theta}) \geq \beta h_j(\theta, \hat{\theta}) \quad \text{for all } j \neq i\}$$

- specialty of agent i:

$$E_i = \{(\theta, \hat{\theta}) : \theta \neq \hat{\theta} \quad \text{and} \quad \gamma_i(\theta, \hat{\theta}) \geq 1\}$$

Definition (expertise)

- relative expertise: $\gamma_i = \min\{\gamma_i(\theta, \hat{\theta}) : (\theta, \hat{\theta}) \in E_i\}$
- absolute expertise: $\varepsilon_i = \min\{h_i(\theta, \hat{\theta}) : (\theta, \hat{\theta}) \in E_i\}$

Experts

Proposition

Suppose that

- $E_i \neq \emptyset$ *for all* i*;*
- $\varepsilon_i \geq \varepsilon_j$ *if and only if* $v_i \leq v_j$*.*

Then, reallocations of signals do not increase the rate by more than $\alpha(\max_i \varepsilon_i)/(\min_i \gamma_i)$*.*

- 1st condition: Agents are all experts.
- 2nd condition: The least central agents have the highest absolute expertise.

Effect of Network topology?

Definition (regularity)

$$A \quad \succeq_{\text{reg}} \quad A'$$

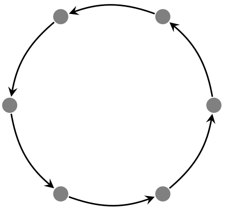
if

$$\textstyle\sum_{i=1}^{k} v_{[i]} \quad \leq \quad \sum_{i=1}^{k} v'_{[i]}$$

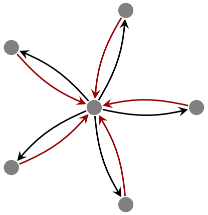
$$\Updownarrow$$

$$v^{\downarrow} \qquad \text{FOSD}^{a} \qquad v'^{\downarrow}$$

[a] first-order stochastically dominates



$$\succeq_{\text{reg}}$$



Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Proposition

Suppose agents' signals are comparable with respect to $\succeq_{UI}$. Then,

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \leq r'^*$$

Network Regularity and Learning

- Consider the rate under the best allocation: r^*.

Is r^* higher for regular or irregular networks?

Proposition

Suppose agents' signals are comparable with respect to $\succeq_{UI}$. Then,

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \leq r'^*$$

- The gap can grow unboundedly in large networks

Network Regularity and Learning: An Example

- $\Theta = \{\theta_0, \theta_1, \ldots, \theta_n\}$
- $S = \{\text{Head}, \text{Tail}\}$
- $\pi > \frac{1}{2}$

$$\ell_i^{\theta}(s) : \quad \begin{array}{c c} & \begin{array}{cc} \text{Head} & \text{Tail} \end{array} \\ \begin{array}{c} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_n \end{array} & \begin{pmatrix} 1-\pi & \pi \\ 1-\pi & \pi \\ \vdots & \vdots \\ \pi & 1-\pi \\ \vdots & \vdots \\ 1-\pi & \pi \end{pmatrix} \end{array}$$

Proposition

$$A \quad \succeq_{reg} \quad A' \qquad \Rightarrow \qquad r^* \geq r'^*$$

- Ordering of networks is reversed with expert agents!
- The gap does not grow unboundedly.

$\Rightarrow$ Rates of learning in all large networks are similar.