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)



Collective Phenomena: Consensus motion coordination, and coverage















Collective Phenomena:

Social and Economic Networks



learning, spreading, and cascades

Preciado Zaroham Enivoha Jadhahaje 2013





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, \dots, 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 ∈ 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, $\omega_{it} \sim \ell_i^{\theta} \in \Delta S$.

ℓ_i = {ℓ^θ_i}_{θ∈Θ}: agent i's signal structure: what is the likelihood of ω_{it} ∈ 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 $\mathcal{X} = \Theta \times \Omega \times \Gamma$ 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., $\mu_{it} \longrightarrow \mathbf{1}_{\theta^*}$ 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:

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

- ▶ 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)

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

aii

Model (social network)

- ► $a_{ij} > 0 \quad \Leftrightarrow \quad \text{Agent } j \text{ is a neighbor of agent } j.$
- $A = [a_{ij}]$ row-stochastic social interaction matrix \downarrow_i

aii

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

- 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(θ, θ̂) 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 ⇒ 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:

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

Definition (rate of learning)

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

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)

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

• λ depends on

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

Relative Entropy and Eigenvector Centrality

Definition (relative entropy)

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

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

- ► $h_i(\theta, \hat{\theta})$: information in favor of θ against $\hat{\theta}$ when θ is realized
- ► $h_i(heta, \hat{ heta}) = 0$ \Rightarrow agent *i* cannot distinguish heta and $\hat{ heta}$

► larger $h_i(\theta, \hat{\theta}) \Rightarrow$ 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



- ▶ 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?



- 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 $(\theta, \hat{\theta})$:

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

specialty of agent i:

$$E_i = \{(heta, \hat{ heta}) : heta
eq \hat{ heta} \text{ and } \gamma_i(heta, \hat{ heta}) \ge 1\}$$

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?





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?

PropositionSuppose agents' signals are comparable with respect to
$$\succeq_{UI}$$
. Then, $A \succeq_{reg} A' \Rightarrow r^* \leq r'^*$

Network Regularity and Learning

Consider the rate under the best allocation: r*.

Is r* higher for regular or irregular networks?



The gap can grow unboundedly in large networks



Ordering of networks is reversed with expert agents!

The gap does not grow unboundedly.

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

NETWORK SCIENCE AND ENGINEERING

The *IEEE Transactions on Network Science and Engineering (TNSE)*, is committed to timely publishing of peer-reviewed technical articles that deal with the theory and applications of network science and the interconnections among the elements in a system that form a network. In particular, the *IEEE Transactions on Network Science and Engineering* publishes articles on understanding, prediction, and control of structures and behaviors of networks at the fundamental level. The types of networks covered include physical or engineered networks, information networks, biological networks, semantic networks, economic networks, social networks, and ecological networks. Aimed at discovering common principles that govern network structures, network functionalities and behaviors of networks, the journal seeks articles on understanding, prediction, and control of structures and behaviors of networks for transactions on *Network Science and Engineering* prediction, and control of structures and behaviors of networks. Another trans-disciplinary focus of the *IEEE Transactions on Network Science and Engineering* is the interactions between and co-evolution of different genres of networks.



TNSE EDITOR-IN-CHIEF

Ali Jadbabaie

Alfred Fitler Moore Professor of Network Science, Dept. of Electrical and Systems Engineering, University of Pennsylvania

TNSE Steering Committee Members

LAUNCHING

IEEE COMPUTER SOCIETY Mung Chiang (SC Chair)

Kathleen Carley John C.S. Lui Ananthram Swami Bolek Szymanski

IEEE COMMUNICATIONS SOCIETY Steven Low Muriel Médard Michele Zorzi

IEEE CIRCUITS AND SYSTEMS SOCIETY Giuseppe Grassi Ching-Yung Lin

For more information please visit: http://www.computer.org/tnse









TNSE EDITORIAL BOARD

EDITOR-IN-CHIEF

ALI JADBABAIE

Alfred Fitler Moore Professor of Network Science, Dept. of Electrical and Systems Engineering, University of Pennsylvania jadbabai@seas.upenn.edu

ASSOCIATE EDITORS-IN-CHIEF

MAURICIO BARAHONA Imperial College London m.barahona@imperial. lb19@cornell.edu ac.uk

LARRY BLUME Cornell University and Santa Fe Institute

BABAK HASSIBI California Institute of Technology hassibi@caltech.edu

STEVEN LOW California Institute of Technology slow@caltech.edu

BALAJI PRABHAKAR **Stanford University** balaji@stanford.edu

ASSOCIATE EDITORS

DAVE ALDERSON Naval Postgraduate School dlalders@nps.edu

VIVEK BORKAR Indian Institute of Technology, Mumbai borkar.vs@gmail.com

CONSTANTINE CARAMANIS University of Texas at Austin constantine@utexas.edu

JORGE CORTÉS University of California, San Diego cortes@ucsd.edu

CHENG-SHANG CHANG National Tsing Hua University cschang@ee.nthu.edu.tw

FAN CHUNG GRAHAM University of California, San Diego fan@math.ucsd.edu

DOMITILLA DEL VECCIO Massachusetts Institute of Technology ddv@mit.edu

JOHN C. DOYLE California Institute of Technology doyle@caltech.edu

RAISSA D'SOUZA University of California, Davis raissa@cse.ucdavis.edu

FABIO FAGNANI Politecnico of Torino fabio.fagnani@polito.it

MASSIMO FRANCESCHETTI University of California, San Diego massimo@ece.ucsd.edu

ZACHARY IVES Pennsylvania State University zives@cis.upenn.edu

RAHUL JAIN University of Southern California rahul.jain@usc.edu

TARA JAVIDI University of California, San Diego tjavidi@ucsd.edu

ERIC KOLACZYK **Boston University** kolaczyk@bu.edu

MARC LELARGE INRIA marc.lelarge@ens.fr

JURE LESKOVEC Stanford University jure@cs.stanford.edu

HUAN LIU Arizona State University huan.liu@asu.edu

MEHRAN MESBAHI University of Washington mesbahi@aa.washington.edu

Andrea Montanari Stanford University montanari@stanford.edu

ELCHANANN MOSSEL University of California, Berkeley mossel@stat.berkeley.edu

ERIC MOULINES ParisTech moulines@telecom-paristech.fr FABIO PAMMOLLI IMT Lucca pammolli@gmail.com

MASON A. PORTER **Oxford University** porterm@maths.ox.ac.uk

SANDIP ROY Washington State University sroy@eecs.wsu.edu

AMIN SABERI Stanford University saberi@stanford.edu

RODOLPHE **S**EPULCHRE Cambridge University r.sepulchre@eng.cam.ac.uk

DEVAVRAT SHAH Massachusetts Institute of Technology devavrat@mit.edu

AMIT SINGER Princeton University amits@math.princeton.edu

ADAM WIERMAN California Institute of Technology adamw@caltech.edu

WALTER WILLINGER Niksun wwillinger@niksun.com

Edmund Yeh Northeastern University eyeh@ece.neu.edu