Economics of Networks
Social Learning

Evan Sadler

Massachusetts Institute of Technology

Agenda

- Recap of rational herding
- Observational learning in a network
- DeGroot learning

Reading: Golub and Sadler (2016), "Learning in Social Networks"

Supplement: Acemoglu et al., (2011), "Bayesian Learning in Social Networks;" Golub and Jackson (2010), "Naïve Learning in Social Networks and the Wisdom of Crowds"

The Classic Herding Model

Two equally likely states of the world $\theta \in \{0, 1\}$

Agents $n = 1, 2, ...$ sequentially make binary decisions $x_n \in \{0, 1\}$

Earn payoff 1 for matching the state, payoff 0 otherwise

Each agent receives a binary signal $s_n \in \{0, 1\}$, observes history of actions

Signals i.i.d. conditional on the state:

$$\mathbb{P}(s_n = 0 \,|\, \theta = 0) = \mathbb{P}(s_n = 1 \,|\, \theta = 1) = g > \frac{1}{2}$$

Rational Herding

Last time we showed in any PBE of the social learning game, we get herd behavior

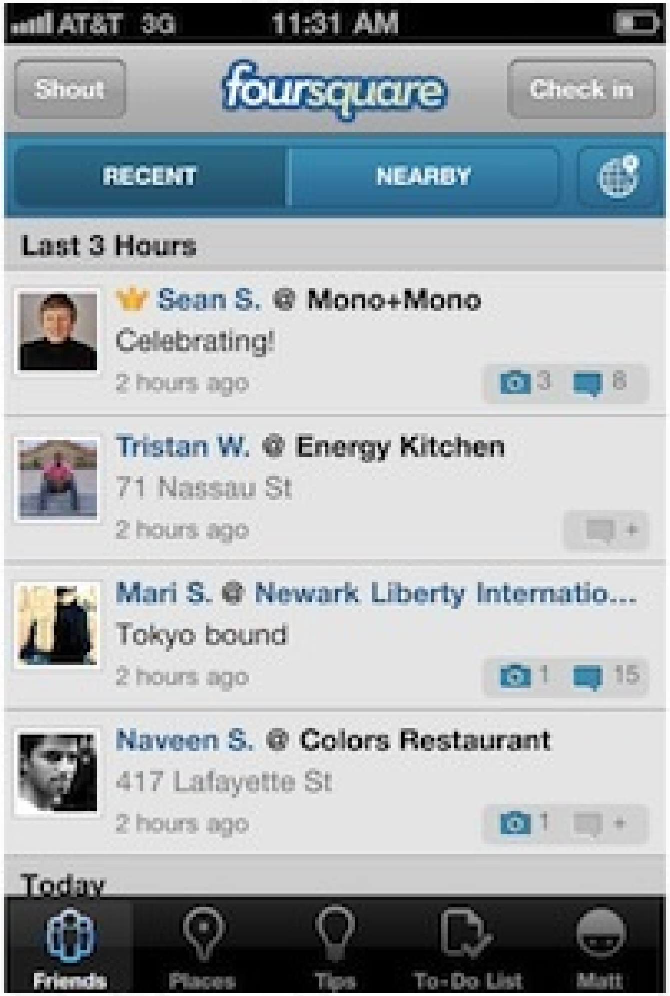
- All agents after some time t choose the same action

With positive probability, agents herd on the wrong action

Inefficiency reflects an informational externality

- Agents fail to internalize the value of their information to others

Observational Learning: A Modern Perspective



Observational Learning: A Modern Perspective

We observe more of what other people do...

...but observing entire history is less reasonable

How does the observation structure affect learning?

A Souped-up Model

Two states of the world $\theta \in \{0, 1\}$, common prior $q_0 = \mathbb{P}(\theta = 1)$

Agents $n = 1, 2, ...$ sequentially make binary decisions $x_n \in \{0, 1\}$

Earn payoff $u(x_n, \theta)$, arbitrary function satisfying

$$u(1,1) > u(0,1), \quad u(0,0) > u(1,0)$$

Each agent receives a signal $s_n \in \mathcal{S}$ in an arbitrary metric space

Signals are conditionally i.i.d. with distributions $\mathbb{F}_\theta$

The Observation Structure

Agent n has a neighborhood $B(n) \subseteq \{1, 2, ..., n-1\}$, observes x_k for $k \in B(n)$

Information set $\mathcal{I}_n = \{s_n, B(n), x_k \,\forall\, k \in B(n)\}$

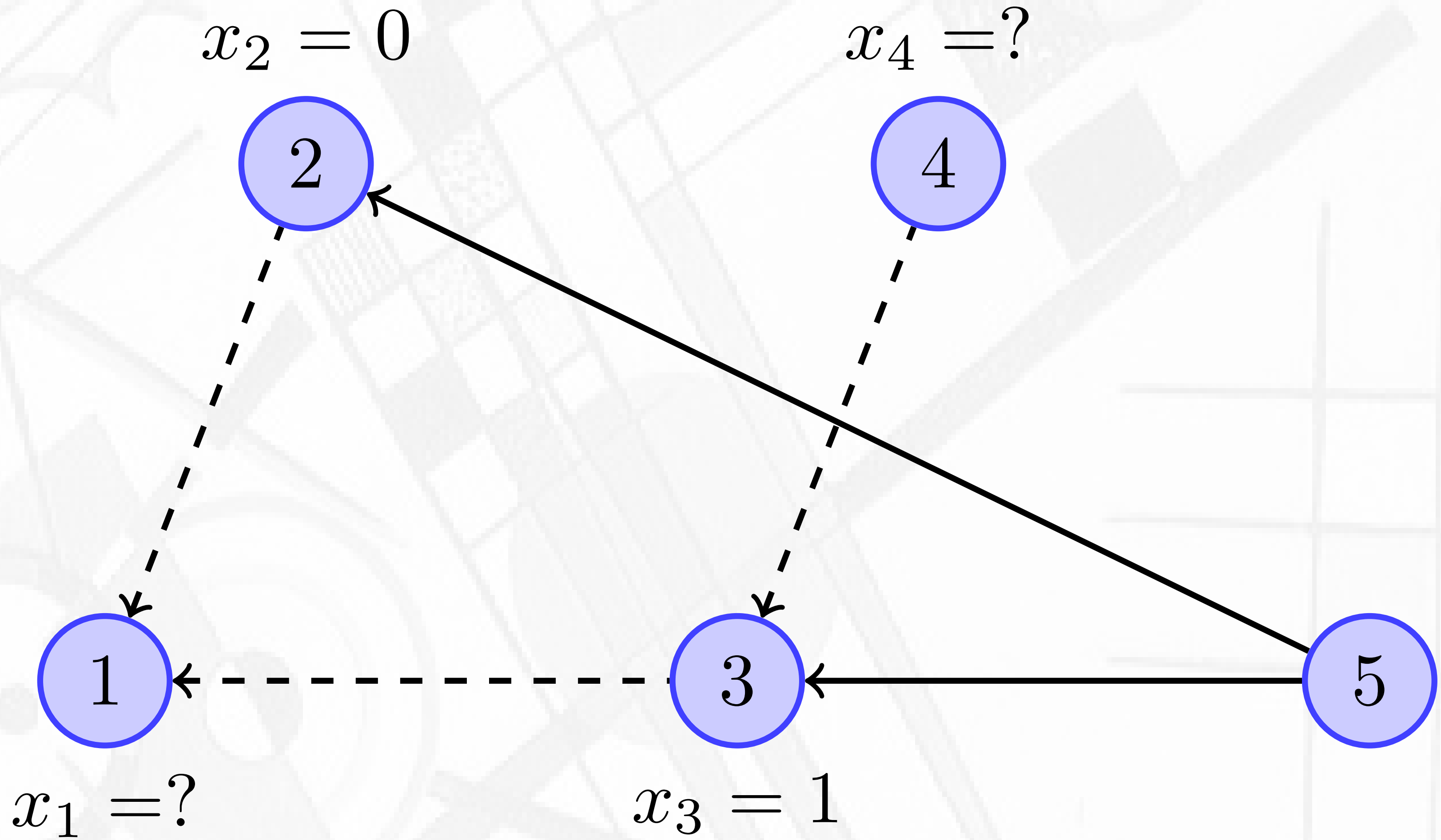
Neighborhoods drawn from a joint distribution $\mathbb{Q}$ that we call the network topology

- $\mathbb{Q}$ is common knowledge
- For this class, assume $\{B(n)\}_{n \in \mathbb{N}}$ are mutually independent

Study perfect Bayesian equilibria σ of the learning game:

$$\sigma_n = \arg\max \mathbb{E}_\sigma \left[u(x, \theta) \,|\, \mathcal{I}_n \right]$$

A Complex Inference Problem



Learning Principles

Cannot fully characterize decisions, focus on asymptotic outcomes

Two learning principles:

- The improvement principle
- The large-sample principle

Corresponding learning metrics: diffusion vs. aggregation

Private and Social Beliefs

Define the private belief $p_n = \mathbb{P}(\theta = 1 \,|\, s_n)$, distribution $\mathbb{G}_\theta$

Social belief $q_n = \mathbb{P}\left(\theta = 1 \,|\, B(n),\, x_k,\, k \in B(n)\right)$

Support of private beliefs $[\underline{\beta}, \overline{\beta}]$

$$\underline{\beta} = \inf\{r \in [0,1] \,:\, \mathbb{P}(p_1 \leq r) > 0\}$$

$$\overline{\beta} = \sup\{r \in [0,1] \,:\, \mathbb{P}(p_1 \leq r) < 1\}$$

The expert signal $\tilde{s}$, binary with

$$\mathbb{P}(\theta = 1 \,|\, \tilde{s} = 0) = \underline{\beta}, \quad \mathbb{P}(\theta = 1 \,|\, \tilde{s} = 1) = \overline{\beta}$$

Learning Metrics

Information diffuses if

$$\liminf_{n\to\infty} \mathbb{E}_\sigma \left[u(x_n, \theta)\right] \geq \mathbb{E}[u(\tilde{s}, \theta)] \equiv u^*$$

Information aggregates if

$$\lim_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) = 1$$

A network topology $\mathbb{Q}$ diffuses (aggregates) information if diffusion (aggregation) occurs for *every* signal structure and *every* equilibrium strategy profile

Diffusion vs. Aggregation

If $1 - \underline{\beta} = \overline{\beta} = 1$, the two metrics coincide

- We say private beliefs are unbounded

If $\underline{\beta} > 0$ and $\overline{\beta} < 1$, private beliefs are bounded

- Diffusion is weaker condition than aggregation

In complete network, aggregation iff unbounded private beliefs (Smith and Sorensen, 2000)

Our definition emphasizes role of network

- Complete network diffuses, does not aggregate, information

Necessary Conditions for Learning

Basic requirement: sufficient connectivity

An agent's personal subnetwork $\hat{B}(n)$ includes all $m < n$ with a directed path to n

Theorem

If $\mathbb{Q}$ diffuses information, we must have expanding subnetworks:

$$\lim_{n \to \infty} \mathbb{P}(|\hat{B}(n)| < K) = 0$$

for all $K \in \mathbb{N}$

The Improvement Principle

Intuition: I can always pick a neighbor to copy

- Whom do I imitate?
- Can I improve?

A heuristic approach: look at neighbor with largest index $\overline{B}(n)$

- If we have expanding subnetworks, then $\mathbb{P}(\overline{B}(n) < K) \to 0$ as $n \to \infty$ for any fixed K
- Key idea: imitate this neighbor if my signal is weak, follow my signal if it is strong

Suboptimal rule, but it gives a lower bound on performance

- Rational agents must do (weakly) better

Two Lemmas

Lemma

Suppose $\mathbb{Q}$ has expanding subnetworks, and there exists a continuous increasing $\mathcal{Z}$ such that $\mathcal{Z}(u) > u$ for all $u < u^*$, and

$$\mathbb{E}_\sigma[u(x_n, \theta)] \geq \mathcal{Z}(\mathbb{E}_\sigma[u(x_{\overline{B}(n)}, \theta)])$$

Then $\mathbb{Q}$ diffuses information.

Lemma

There exists a continuous increasing $\mathcal{Z}$ with $\mathcal{Z}(u) > u$ for all $u < u^*$ such that

$$\mathbb{E}_\sigma[u(x_n, \theta)] \geq \mathcal{Z}(\mathbb{E}_\sigma[u(x_m, \theta)])$$

for any $m \in B(n)$.

A Key Assumption: Independent Neighborhoods

Our two lemmas imply that information diffuses in any sufficiently connected network

- Relies on independence of neighborhoods

If neighborhoods are correlated, the fact that I observe someone is related to how informative their choice is

Failure to Aggregate

Proposition (Acemoglu et al., 2011, Theorem 3)

The topology $\mathbb{Q}$ fails to aggregate information if any of the following conditions hold:

- $B(n) = \{1, 2, ..., n-1\}$
- $|B(n)| \leq 1$ *for all* n
- $|B(n)| \leq M$ *for all* n *and some* $M \in \mathbb{N}$*, and*

$$\lim_{n\to\infty} \max_{m \in B(n)} m = \infty \quad \textit{almost surely}$$

The Large-Sample Principle

Intuition: I can always learn from many independent observations

Limiting connectively can create "sacrificial lambs:" $B(m) = \emptyset$

Proposition

Suppose there exists a subsequence $\{m_i\}$ *such that*

$$\sum_{i \in \mathbb{N}} \mathbb{P}(B(m_i) = \emptyset) = \infty, \text{ and } \lim_{n \to \infty} \mathbb{P}(m_i \in B(n)) = 1$$

for all i. *Then* $\mathbb{Q}$ *aggregates information.*

Follows from a martingale convergence argument

Heterogeneous Preferences

Key limitation so far: everyone has the same preferences

Give each agent n a type $t_n \in (0,1)$

Payoffs

$$u(x,\theta,t) = \begin{cases} 1-\theta+t & \text{if } x=0 \\ \theta+1-t & \text{if } x=1 \end{cases}$$

The type t parameterizes the relative cost of error in each state

Failure of the Improvement Principle

Copying a neighbor no longer guarantees same utility

- Copying works better when neighbor's preferences are close to own

Assume

- $B(n) = \{n-1\}$ for all n
- Odds have type $\frac{1}{5}$, evens have type $\frac{4}{5}$
- $\mathbb{G}_0(r) = 2r - r^2$ and $\mathbb{G}_1(r) = r^2$

Can show inductively that all odds (evens) err in state 0 (state 1) with probability at least $\frac{1}{4}$ (homework problem)

Robust Large-Sample Principle

With full support in preference distribution, preferences can counterbalance social information

- Some agents will act on signals
- No need for sacrificial lambs

Proposition

Suppose preference types are i.i.d. with full support on $(0,1)$*, and there exists an infinite sequence* $\{m_i\}$ *such that*

$$\lim_{n\to\infty} \mathbb{P}(m_i \in B(n)) = 1$$

for all i*. Then information aggregates.*

Remarks on the SSLM

Clear understanding of learning mechanisms

- Improvement vs. Large samples
- Different effects of preference heterogeneity

Rationality is a very strong assumption...

- but proofs are based on heuristic benchmarks

Can't say much about rate of learning, influence

A Different Approach

Look at a model of heuristic learning based on DeGroot (1974)

Finite set N of agents, time is discrete

At time t, agent i has a belief or opinion $x_i(t) \in [0, 1]$

- How likely is it the state is 1?
- How good is politician X?

A simple update rule:

$$x_i(t) = \sum_{j \in N} W_{ij} x_j(t-1)$$

Think of W as a weighted graph

DeGroot Updating

Assumptions:

- The $x_i(0)$ are given exogenously
- The matrix W is an $n \times n$ matrix with non-negative entries
- For each i we have $\sum_{j \in N} W_{ij} = 1$

Take a weighted average of friends' opinions

Simple example:

- Consider an unweighted graph G, agent i has degree d_i
- $W_{ij} = \frac{1}{d_i}$ for each neighbor j of i, and $W_{ij} = 0$ for each non-neighbor

Matrix Powers and Markov Chains

Can rewrite the update rule as

$$\mathbf{x}(t) = W\mathbf{x}(t-1) \quad \Longrightarrow \quad \mathbf{x}(t) = W^t\mathbf{x}(0)$$

Reduction to dynamics of matrix powers

Entries in each row sum to 1, so this is a row-stochastic matrix

- Correspond to transition probabilities for an n-state Markov chain

How to think about W_{ij}^t

- $\frac{\partial x_i(t)}{\partial x_j(0)} = W_{ij}^t$: influence of j on i's time t opinion
- W_{ij}^t sums over all paths of indirect influence

The Long-Run Limit

Does each individual's estimate settle down to a long-run limit?

- Does $\lim_{t\to\infty} x_i(t)$ exist?

Do agents reach a consensus? If so, what does it look like?

- How do long-run beliefs depend on W and the initial estimates $\mathbf{x}(0)$?

Start with strongly connected networks

- The network W is strongly connected if there is a directed path from i to j for every $i, j \in N$

Call W *primitive* if there exists q such that every entry of W^q is strictly positive

- Equivalent to aperiodicity in the network

The Long-Run Limit

Theorem

Suppose W is strongly connected and aperiodic. The limit $\lim_{t\to\infty} x_i(t)$ exists and is the same for each i.

Proof:

- The sequence $\max_i x_i(t)$ is monotonically decreasing
- The sequence $\min_i x_i(t)$ is monotonically increasing
- Primitivity ensures the two extreme agents put at least weight $w > 0$ on each other after q steps
- Distance between max and min decreases by factor at least $1 - w$ after every q steps

Influence on the Consensus

$$\lim_{t\to\infty} \mathbf{x}(t) = \lim_{t\to\infty} W^t \mathbf{x}(0)$$

The matrix powers must converge

Moreover, since agents reach consensus, it must be that all rows of W^t converge to the same vector $\boldsymbol{\pi}$

$$\mathbf{x}(\infty) = \boldsymbol{\pi}^T \mathbf{x}(0) = \sum_{i\in N} \pi_i x_i(0)$$

The coefficient π_i gives the influence of agent i on the consensus

- Depends only on the network W, not on initial estimates $\mathbf{x}(0)$

Vector $\boldsymbol{\pi}$ must satisfy

$$\boldsymbol{\pi}^T W = \boldsymbol{\pi}$$

Left eigenvector with eigenvalue 1

Influence on the Consensus

Theorem

If W is strongly connected and primitive, then for all i

$$\lim_{t\to\infty} x_i(t) = \sum_{i\in N} \pi_i x_i(0)$$

where π_i is the left eigenvector centrality of i in W

Note vector $\boldsymbol{\pi}$ is also the unique stationary distribution of the Markov chain with transition probabilities given by W

Can also be seen as a consequence of the Perron-Frobenius Theorem from linear algebra

Beyond Strong Connectedness

If network not strongly connected, can decompose into strongly connected subgraphs

- Equivalent to reduction of a Markov chain to closed communicating classes
- Analyze each subgraph separately using earlier result

Agents i and j are is same communicating class if there is a directed path from i to j and vice versa

No longer guarantee consensus

- Consensus within communicating classes, not necessarily across

Small amount of communcation across classes makes large (discontinuous) difference in asymptotic outcomes

When is Consensus Correct?

Are large populations able to aggregate information?

Suppose there is some true state $\mu \in [0,1]$, and agents begin with noisy estimates of μ

- Suppose the $x_i(0)$ are $i.i.d.$ random variables with mean μ, variance σ^2

Consider an infinite sequence of networks $\{W^{(n)}\}_{n=1}^{\infty}$, population getting larger

If $x^{(n)}(\infty)$ is the consensus estimate in network n, do these estimates converge to μ as $n \to \infty$?

When is Consensus Correct?

Theorem (Golub and Jackson, 2010)

The consensus beliefs $x^{(n)}(\infty)$ converge in probability to μ if and only if

$$\lim_{n\to\infty} \max_i \pi_i^{(n)} = 0.$$

The influence of the most central agent in the network converges to zero

Proof:

- We have $Var\left[x^{(n)}(\infty) - \mu\right] = \sum_{i=1}^{n}(\pi_i^{(n)})^2\sigma^2$
- Converges to zero if and only if $\max_i \pi_i^{(n)} \to 0$
- If not, no convergence in probability
- If it does, Chebyshev's inequality implies convergence in probability

Speed of Convergence

Consensus might be irrelevant if it takes too long to get there

- How long does it take for differences to get “small”?
- What network properties lead to fast or slow convergence?

Note, first question depends both on network and initial estimates

- If we start at consensus, we stay there

Focus on worst-case convergence time, highlight role of network

A Spectral Decomposition

Lemma

For "generic" W, we may write

$$W^t = \sum_{l=1}^{n} \lambda_l^t P_l$$

where

- $1 = \lambda_1, \lambda_2, ..., \lambda_n$ *are* n *distinct eigenvalues of* W
- P_l *is a projection onto the eigenspace of* λ_l
- $P_1 = W^\infty$ *and* $P_1 \mathbf{x}(0) = \mathbf{x}(\infty)$
- $P_l \mathbf{1} = 0$ *for all* $l > 1$*, where* $\mathbf{1}$ *is a vector of all ones*

All other eigenvalues strictly smaller in absolute value than $\lambda_1 = 1$

Speed of Convergence

Theorem

For generic W,

$$\frac{1}{2}|\lambda_2|^t - (n-2)|\lambda_3|^t \leq \sup_{\mathbf{x}(0)\in[0,1]^n} \|\mathbf{x}(t) - \mathbf{x}(\infty)\|_\infty \leq (n-1)|\lambda_2|^t.$$

Note $\|\cdot\|_\infty$ denotes the supremum norm, largest deviation from consensus among all agents

Clear answer to first question: rate of convergence depends on second largest eigenvalue

- Larger λ_2 (i.e. smaller spectral gap) implies slower convergence

Segregation and Slow Convergence

What network features correspond to large $|\lambda_2|$?

On an intuitive level, we get slow convergence in highly "segregated" networks

Define the *bottleneck ratio*

$$(W) = \min_{\substack{M \subseteq N \\ \pi(M) \geq \frac{1}{2}}} \frac{\sum_{i \in M, j \notin M} \pi_i W_{ij}}{\sum_{i \in M} \pi_i}$$

Small when some influential group pays little attention to those outside itself

- Can use to bound size of $|\lambda_2|$

Wrap Up

Limited ability to learn through observation

- Information externality creates inefficiency
- Heterogeneity may help or hurt depending on network properties

Naïve learning model gives measures of influence, learning rate

Next time: moving on to models of diffusion, different influence mechanism