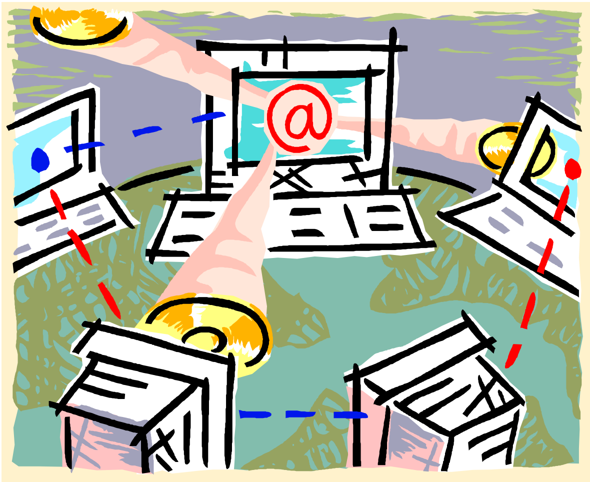
Probability Models of Information Exchange on Networks
Lecture 2

Elchanan Mossel
UC Berkeley



Lecture Plan

• We will see the first simple models of information exchange on networks.

• Assumptions:

• $P[S = +] = P[S = -] = 0.5$.

• N independent signals S_i with $P[S_i = S \mid S] = p > 0.5$.

. Agents update their opinions in various ways …

Lecture Plan

• Basic questions:

• <u>Agreement</u>: Do all agents converge to the same belief/action?

• <u>Learning</u>: Do the agents converge to the correct state of the world (with high probability)?

• Compare to the Jury Theorem where we have agreement (the majority value) and learning (by the Jury Theorem)

The DeGroot Model

- DeGroot model: Repeated averaging of probabilities of self and neighbors.

- Formal definition: n individuals denoted 1,…,n.
- At time 0 some initial beliefs: $s(i) = s(i,0)$ for $i \in [n]$.
- Averaging weights: $w(i,j) \geq 0$ satisfy $\sum_j w(i,j) = 1$
- Update rule: $s(i,t+1) = \sum_j w(i,j)\, s(j,t)$
- In matrix notation: $s(t+1) = W\, s(t)$

- Linear model, linear updates …
- Introduced by De Groot (Penn Statistician, 1931-1989) in 1974.

The DeGroot Model and Markov Chains

- In matrix notation: $s(t+1) = W\, s(t)$

- Let $X(i,t)$ be the Markov chain defined by W started at i and run for t steps.

- Claim: $s(i,t) = E[\, s(X(i,t))\,]$

- Recall that W is called ergodic if there is a power t such that all entries of W^t are positive.

Markov (wikipedia)

Kolmogorv (wikipedia)

Agreement in the DeGroot Model

- Corollary: If W is ergodic with stationary distribution π then $s(i,t) \rightarrow_{t \rightarrow \infty} \sum_j \pi(j)\, s(j)$. (all agents converge to same s)

- Interpretation: $\pi(j)$ is the centrality of node j (it's PageRank)

Agreement in the DeGroot Model

- Corollary: If W is ergodic with stationary distribution π then $s(i,t) \to_{t \to \infty} \sum_j \pi(j)\, s(j)$. (all agents converge to same p)

- Network interpretation: Suppose w(i,i) > 0 for all i. Let G = ([n],E) be the directed graph where $(i,j) \in E$ iff w(i,j) > 0. Then if G is strongly connected then W has a unique stationary distribution so agreement is reached.

Agreement in the DeGroot Model

- Corollary: If W is ergodic with stationary distribution π then $s(i,t) \to_{t \to \infty} \sum_j \pi(j)\, s(j)$. (all agents converge to same p)

- Network interpretation: Suppose w(i,i) > 0 for all i. Let G = ([n],E) be the directed graph where $(i,j) \in E$ iff w(i,j) > 0. Then if G is strongly connected then P has a unique stationary distribution so agreement is reached.

- Exercise: infinite graphs ??

- Next question: Do the agents also learn?

Learning in the DeGroot Model

- Def: Consider the DeGroot model W corresponding to a chain with a unique stationary distribution π and signal of bias p. Assume the initial signals S(1),..,S(n) are conditionally i.i.d. with P[S(i) = S | S] = p > ½. Then:

- The probability of learning in the model is:
- P_p[Learning in W] := $P_p[S = \mathrm{sgn}(\sum \pi_j S(j))]$

- Note – this definition is somewhat arbitrary. Ideally we would like the definition to also measure how convinced the agents are.

Learning in the DeGroot Model

- Def: Consider the DeGroot model W corresponding to a chain with a unique stationary distribution π and signal of bias p. Assume the initial signals S(1),..,S(n) are conditionally i.i.d. with P[S(i) = S | S] = p > ½. Then:
- The probability of learning in the model is:
- P_p[Learning in W] := $P_p[S = \mathrm{sgn}(\sum \pi_j S(j))]$
-
- Def 2: Given a sequence of models W^n and a bias p, we say that learning holds if P_p[Learning in W^n] $\to 1$ as $n \to \infty$

- Note: Learning depends only on stationary distribution(s).

- Claim: Learning holds for all p>1/2 for the models W^n if and only if $\limsup_n \max_j \pi^n(j) = 0$.

Learning in the DeGroot Model

- Claim: Learning holds for all $p>1/2$ for the models W^n if and only if $\limsup_n \max_j \pi^n(j) = 0$ (“Egalitirism leads to learning”)

Learning in the DeGroot Model

- Claim: Learning holds for all $p>1/2$ for the models W^n if and only if $\limsup_n \max_j \pi^n(j) = 0$ ("Egalitirism leads to learning")

- Pf: If $\limsup_n \max_j \pi^n(j) = 0$ apply the weak law of large numbers.

- Other direction: exercise!

- Q: Examples of sequences of graphs where:
- 1. There is learning
- 2. There is no learning.
- 3. There is learning for some values of p but not for others.

Learning in the DeGroot Model

- Claim: Learning holds for all $p>1/2$ for the models W^n if and only if $\limsup_n \max_j \pi^n(j) = 0$ ("Egalitirism leads to learning")

- Pf: If $\limsup_n \max_j \pi^n(j) = 0$ apply the weak law of large numbers.

- Other direction: exercise!

- Example 1: Consider $w = 1$ for all edges (including loops). Then learning for the DeGroot model holds for any sequence of connected bounded degree graphs G_n with $|V_n| \to \infty$.
- This is true since $\max \pi^n(i) \leq \max \deg / (|V_n|-1) \to 0$

- Example 2: No learning for (G_n) = (star on n vertices)

Rate of Convergence in the DeGroot Model

- Next we briefly discuss the rate of convergence in the Degroot model.

- Assume that the chain corresponding to W has a unique stationary distribution.

- Def: The total variation distance between two distributions P,Q is:

- $|P-Q|_1 = 0.5 \sum_x |P(x)-Q(x)|$

Rate of Convergence

- In the theory of M.C. there are many ways to measure the rate of convergence of a chain.
- We can directly use any of them to obtain bounds on the rate of convergence. For example:

- <u>Claim:</u> For all $s \in [1/2-\delta, 1/2+\delta]^n$ it holds that
- $| s(i,t) - E_\pi s| = | E s (X(i,t)) - E_\pi s| \leq 2 \delta |P(i,t) - \pi|_1$

- $P(i,t)$ is the distribution of $X(i,t)$.
- Long after mixing time, agents have almost converged.

- Some of the techniques in the theory of Markov chain include: Spectral techniques and conductance, coupling, log sobolev inequalities etc.

Rate of Convergence

- <u>Condoctance bounds:</u>
- $\pi(\partial A) := \sum \{w(x,y)\, \pi(x) + w(y,x)\, \pi(y) : (x,y) \in (A,A^c)\}$
- $I = \min \{Q(\partial A)/ \pi(A) : \pi(A) \leq ½\}$.
- Then gap $\geq I^2/8$ (Cheeger's inequality) and therefore
- By time $t = 8\, s\, (1+\max_i \ln(1/\pi(i)))/I^2$ we have:
- $|\, P(i,t) - \pi|_1 \leq \exp(-s)$ for all i.

• This means if there are no "isolated communities" then convergence is quick.

- In the minimum suffices to look at connected sets.

Cheating in the DeGroot Model

- A basic question:
- What happens if somebody cheats? Can they convince the rest of the group to reach whatever value they want?

Cheating in the DeGroot Model

- A basic question:
- What happens if somebody cheats? Can they convince the rest of the group to reach whatever value they want?

- <u>Claim:</u> If the chain is ergodic and there is one cheater by repeatedly stating a fixed value all opinions will converge to that value.

Cheating in the DeGroot Model

- A basic question:
- What happens if somebody cheats? Can they convince the rest of the group to reach whatever value they want?

- Claim: If the chain is ergodic and there is one cheater by repeatedly stating a fixed value all opinions will converge to that value.

- Pf: Follows from MC interpretation:
- Let the cheater play v. Let A(i,t) be the event that X(i,0),..,X(i,t) hit the cheater by time t then:
- $p(t,i) = E[1(A(i,t)) \times v + (1-1(A(i,t)) \times X(i,t)]$

- Exercise / Research Question: Multiple cheaters?

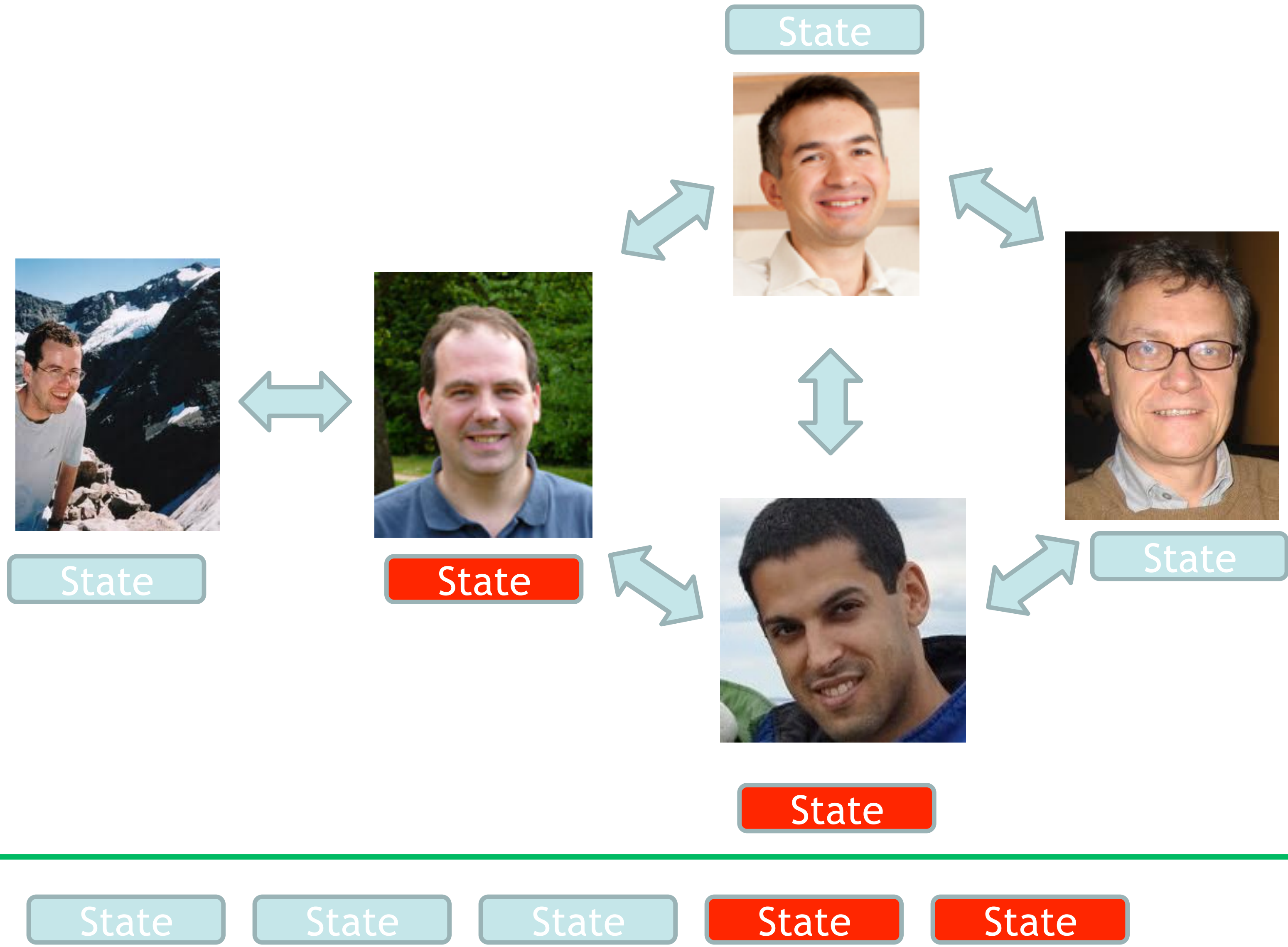
The Voter Model

- The Voter model: Repeated sampling of opinions of neighbors.

- Formal definition: n individuals denoted 1,…,n.
- At time 0 some initial -/+ beliefs: s(i) = for i $\in$ [n].
- Averaging weights: w(i,j) $\geq$ 0 satisfy $\sum_j$ w(i,j) =1
- Update rule: P[s(i,t+1) = s(j,t)] = w(i,j)

- Introduced in the 1970s by P. Cliford and A. Sudury and studied by Holley and Liggett

Voter Model


State
State
State
State
State
State
State
State
State
State
State

Agreement in the Voter Model

- <u>Claim:</u> Let G be the directed graph defined by letting i -> j $\in$ E iff w(i,j) > 0. Assume that G is strongly connected and w(i,i)>0 for all i. Then the voter model on G converges a.s. to all + or all -.

Agreement in the Voter Model

- Claim: Let G be the directed graph defined by letting $i \to j \in E$ iff $w(i,j) > 0$. Assume that G is strongly connected and $w(i,i)>0$ for all i. Then the voter model on G converges a.s. to all + or all -.

- Pf Sketch:
- The voter model is a Markov chain.
- All + and all – are fixed point of the chain.
- Under the conditions from any other state it is possible to reach the all + or all – states.

(Non) Learning in the Voter Model

• Setup: Assume self loops and that G is strongly connected.

• Claim: Suppose $p < P[S(i) = S] < q$. Let r be the probability that the voter model converges to the correct state then $p < r < q$ (no learning)

(Non) Learning in the Voter Model

- Setup: Assume self loops and that G is strongly connected.

- Claim: Suppose $p < P[S(i) = S] < q$. Let r be the probability that the voter model converges to the correct action then $p < r < q$ (no learning)

- Claim: For all S, $P[S(i,t) = 1] = DG(i,t)$ where $DG(i,t)$ is the value of node i at iteration t in the DeGroot model with the same starting conditions S.

Convergence Rate of the Voter Model

- There are numerous ways of analyzing the convergence rate of the voter model.

- Coalescence of Random walks:
- S(i,t) = S(X(i,t))
- If we can couple X(i,t) so that by time T we have X(i,T) = X(j,T) for all i and j, then agreement must have occurred by time T!

- See Liggett, Durrett, …

Convergence Rate of the Voter Model

- Martingale arguments:

- Example: suppose G is undirected and connected.

- Claim: $M(t) = \sum_{\iota} d(i)\, S(i,t)$ is a martingale.

Convergence Rate of the Voter Model

- Martingale arguments:

- Example: suppose G is undirected and connected.

- Claim: $M(t) = \sum_{\iota} d(i)\, S(i,t)$ is a martingale.

- Pf Sketch: For each edge e, the contribution of that edge is a martingale.

- Then: use Wald's 2^{nd} equation to bound the convergence time.

- Exercise: Do it: show that converges in time $O(|V|^2)$
- Get better bounds when the graph expands.

Bribery in the Voter Model

- <u>Martingale arguments:</u>

- Example: suppose G is undirected and connected.

- $M(t) = \sum_{\iota} d(i)\, S(i,t)$ is a martingale.

- <u>Exercise:</u> Suppose you can bribe k nodes who are – and change their value to +. Which nodes will you bribe?

(Non) Learning in the Voter Model

- Recall that we saw that if the original signals satisfy
- Suppose $p < P[S(i) = S] < q$ then the probability r of convergence to the correct action also satisfies $p < r < q$ (no learning)

- Question: is there a dynamic "like" the voter model where learning holds?

- Stronger question: is there a dynamic "like" the voter model which converges to the majority of the original signals?

- <u>Claim:</u> No local dynamics without additional memory.

Strong Weak Voter Model

- It’s possible with one extra bit of memory in a non-synchronous model:

Strong Weak Voter Model

- It's possible with one extra bit of memory in a non-synchronous model:
- Edges are chosen according to Poisson process.
- All Voters have **opinion** (red/blue) and **strength** of opinion (STRONG/weak). Originally all strong.
- When they meet,
  - Update color:
    - STRONG influence weak
    - Otherwise voter model
  - Update Strengths:
    - Two STRONGS of different colors cancel to weak
    - Otherwise stay the same
    - Two sides swap location with probability 0.5.

Recent Research Topic

- What can local dynamics compute on networks?
- Motivation from:
- Experimental Sociology Latane L' Herrou (96)
- Sensor networks (Aspens and collaborators 04-…)
- Biological computing (Winfree …, 2005-)
- See also recent preprint by M-Parkash-Valiant

A 3rd type of dynamics

- We've seen two networks dynamics:
- 1st: Average your neighbors (DeGroot)
- 2nd :Sample one of your neighbors (voter)

A 3rd type of dynamics

- We've seen two networks dynamics:
- 1st: Average your neighbors (DeGroot)
- 2nd :Sample on of your neighbors (voter)
- 3rd: Take a majority of your neighbors.

Majority Dynamics

- 3^{rd}: Take a majority of your neighbors.
- Formal definition: n individuals denoted 1,…,n.
- At time 0 some initial -/+ beliefs: s(i) = for $i \in [n]$.
- Weights: $w(i,j) \geq 0$ satisfy $\sum_j w(i,j) = 1$
- Update rule: $s(i,t+1) = \mathrm{sgn}(\sum w(i,j)\, s(j,t))$
- Much harder to analyze
- Not all nodes converge to same value
- Very few tools available – we'll see a few.
- Go to Omer's Talk!

The effect of a voter

Def: $E_p[f_i]$ is called the influence of voter i, where $f_i(x) = f(1,x_{-i})-f(-1,x_{-i})$

Theorem (Talagrand 94):

- Let f be a monotone function.
- If $\delta = \max_r \max_i E_r[f_i]$ and $p < q$ then
- $E_p[f \mid s = +] (1-E_q[f \mid s=+]) \leq \exp(c \ln \delta (q-p))$
- for some fixed constant c>0.

- In particular: if f is fair and monotone, taking p=0.5:

- $E_q[f \text{ is correct}] \geq 1- \exp(c \ln \delta (q-0.5))$

The effect of a voter

. <u>Theorem (Talagrand 94):</u>

- Let f be a monotone function.
- If $\delta = \max_p \max_i E_x[f_i]$ and $p < q$ then
- $E_p[f \mid s = +] (1-E_q[f \mid s = +]) \leq \exp(c \ln \delta (q-p))$
- for some fixed constant $c>0$.

- In particular: if f is fair and monotone, taking $p=0.5$:

- $E_q[f \text{ is correct}] \geq 1- \exp(c \ln \delta (q-0.5))$

- This means that if each voter has a small influence then the function aggregates well!

An important case

Def: A function $f: \{-,+\}^n \rightarrow \{-,+\}$ is transitive if there exists a

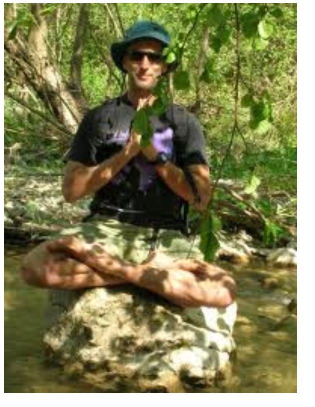
- group G acting transitively on [n] s.t.
- for every $x \in \{-,+\}^n$ and any $\sigma \in G$ it holds that $f(x_\sigma) = f(x)$,

where

- $x_\sigma(i) = x(\sigma(i))$

Thm (Friedgut-Kalai-96) :

- If f is transitive and monotone and
- $E_p[f \mid s= +] > \varepsilon$ then
- $E_q[f \mid s = +] > 1-\varepsilon$ for $q=p+c \log(1/2\varepsilon)/ \log n$



Note: If f is fair transitive and monotone we obtain

$E_q[f \text{ is correct}] > 1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

An important case

Thm (Friedgut-Kalai-96) :

- If f is transitive and monotone and
- $E_p[f] > \varepsilon$ then
- $E_q[f] > 1-\varepsilon$ for $q=p+c \log(1/2\varepsilon)/ \log n$

- Note: If f is fair transitive and monotone we obtain

$E_q[f \text{ is correct}] > 1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

Back to Majority Dynamics

- Note: If f is fair transitive and monotone we obtain E_q[f is correct] > $1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

- Can be applied to: f = Majority of signals at the end of majority dynamics on transitive graphs to show that

- E_q[majority is correct] > $1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

- So information is retained.

- See M-Neeman-Tamuz for generalization to quasi-transitive and pluralities instead majorities.