Social Choice and Social Networks

Consensus, Bribe and Marketing

Draft - All right reserved

Elchanan Mossel
UC Berkeley
Admin tasks of the day

- Scribe notes
- To get a grade you must do one.
- Today?
- Last week?
- Previous notes?
Today topics

• Topic 1:
  • Randomized consensus protocols
  • The “voter model” and generalizations.

• Topic 2:
  • Which voters to “buy” in the voter model.
  • The viral marketing problem.

• Next week/month: a new topic - unbiased signals and social choice.
Randomized Consensus protocols

• As before we will consider a social network which is a graph.
  • G=(V,E)

• We will work in continuous time and asynchronous fashion.
• At time 0 people hold opinions $X_0(v)$, $v \in V$.

• In vertex models:
  • Update times of each vertex is a Poisson(1) process.
  • Equivalent to: after each update – waiting $\text{Exp}(1)$ time.
  • At update time: update $X_t(v)$ according to neighbors, current opinion and randomness

• Edge models similar:
  • Edges (u,v) have update times – both end points of edge update $X_u(t)$ and $X_v(t)$
Randomized Consensus protocols

• Vertex models:
  Individual decisions.

• Edge models:
  Decisions are result of interactions with models.

• Natural to consider also clique models.

• “Goal”: find “good models” that converge to consensus. What are good models?
Good consensus models

• “Goal”: find “good models” that converge to consensus.

• Property 1: Fairness with respect to alternatives.

• Property 2: Consensus is a fixed point.

• Property 3: Simple.

• Many of the following slides due to G. Schoenebeck
Big Question/General Goals

Courtesy of G. Schoenebeck
Series of Experimental Work

• Latane L’Herrou[96]
  – Try to play majority
• Kearns, Judd, Tan, Wortmann [09]
  – Consensus with different payoffs
• Kearns, Suri, Montefort [06]
  – coloring
  – Enemark, McCubbins, Paturi [09]
Our Model

\[ T = T'(S_{Casey}, S_{Grant}, R, \text{Advice}) \times T'(S_{Grant}, S_{Casey}, R, \text{Advice}) \]

Courtesy of G. Schoenebeck
Problems Studied

- Coordination: Arrive at consensus
- Majority Coordination:
  - Arrive at consensus which equals majority of original opinions.
  - Protocols have to be symmetric with respect to the two states.
Definitions: Broadcast and Collision time

• Broadcast Time: Time for a message to flood network.
  – More like Expansion than Diameter

• Collision Time: Time until for every pair of people, someone has received both of their messages.
  – Provides trivial lower bound

Courtesy of
G. Schoenebeck
Related Work

• Similar to Distributed Computing
  – Usually a different time metric
    • Synchronous
    • Worst case
  – Usually different symmetry condition
• Coordination Games in Economics
• Similar to Simulations in Social Networking literature.
• More to come in context

Courtesy of G. Schoenebeck
Coordination using the Voter Model

- Voter Model is an edge model where:
- If \( X_v(t-) = X_u(t-) \) then: \( X_v(t) = X_u(t) = X_v(t-) = X_u(t-) \)
- If \( X_v(t-) \neq X_u(t-) \) then:
  - \( \frac{1}{2} \) Prob: \( X_v(t) = X_v(t-) \neq X_u(t) = X_u(t-) \)
  - \( \frac{1}{4} \) Prob: \( X_v(t) = X_u(t) = X_u(t-) \)
  - \( \frac{1}{4} \) Prob: \( X_u(t) = X_v(t) = X_v(t-) \).
- If \( X_v(t-) = X_u(t-) \) then: \( X_v(t) = X_u(t) = X_v(t-) = X_u(t-) \)
- P. Cliford and A. Sudury. A model for spatial conflict (1973) + Liggett
- A. Holley and T. M. Ligget. Ergodic theorems for weakly interacting infinite systems and the voter
- Model (1975)
Voter Model

Courtesy of G. Schoenebeck
Convergence of the voter model

Claim: The voter model converges to consensus. Moreover, the convergence time is $O(|V|^2)$.

Pf: ??

Courtesy of
G. Schoenebeck
Convergence of the voter model

Claim: The voter model converges to consensus. Moreover, the convergence time is $O(|V|^2)$.

Pf of Convergence:
Each of the consensus configurations is a fixed point. It is also clearly reachable from any other configurations.

Pf idea of Convergence:
Let opinions be +,-. Then $X(t) = \sum X_v(t)$ is a martingale.
Convergence of the voter model

- Convergence in terms of expansion:
  - Let \( r = \) smallest cut in the graph. Then for any starting configuration will converge to consensus by time \( n^2/r \) with probability at least \( \frac{1}{2} \).
  - Pf: \( X(t) \) is a martingale. Let \( P(t) = P(\text{Cons. by } t) \).
  - \( f(t) = \text{Var}[X(t)] \) satisfies:
  - \( f'(t) = \lim E[(X(t+h)-X(t))^2]/h \)
  - \( f'(t) \geq r P(\text{No consensus at time } t) \)
  - If \( P(4 \frac{n^2}{r}) \leq \frac{1}{2} \) then
  - \( n^2 \geq f(4 \frac{n^2}{r}) \geq \frac{1}{2} r 4 \frac{n^2}{r} - \) contradiction.
Coordination by choosing a leader

• Each player chooses at random how strong $S(v)$ her opinion is in the range $[n^{10}]$.

• $X_v(0+) = X_v(0) \times S(v)$.

• In edge update: weak copies strong.

• In case of tie – voter like dynamics.
Greatest Element Dynamics

Courtesy of G. Schoenebeck
Coordination by choosing a leader

• Each player chooses at random how strong $S(v)$ her opinion is in the range $[n^{10}]$.

• $X_v(0+) = X_v(0) \times S(v)$.

• In edge update: weak copies strong.

• In case of tie – voter like dynamics.

• Analysis: in case of single leader – broadcast time.

• In the case of more than one – like voter model.

• Expected time := broadcast time.
# Coordination Summary

<table>
<thead>
<tr>
<th>Problem</th>
<th>Memory</th>
<th>Time</th>
<th>Required Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voter Model</td>
<td>1</td>
<td>$n^2$</td>
<td>none</td>
</tr>
<tr>
<td>Greatest Element</td>
<td>$O(\log(n))$</td>
<td>broadcast</td>
<td>$\Theta(\log(</td>
</tr>
<tr>
<td>Wait-and-See</td>
<td>expected $O(1)$</td>
<td>$O(\text{broadcast})$</td>
<td>$\Theta(\text{broadcast} \cdot</td>
</tr>
</tbody>
</table>

Courtesy of G. Schoenebeck
The Majority Coordination problem

- Claim: Cannot be done with no memory.
- Pf: ??
The Majority Coordination problem

• Claim: Cannot be done with no memory.
• Pf: Look at a configuration resulting in a change.
• Make it change majority.
The Majority Coordination problem

• Claim: Can be done with 2 bits of memory.
• Pf: ??
Strong Weak Voter

• 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
    • STRONG/weak swap strengths

Courtesy of
G. Schoenebeck
<table>
<thead>
<tr>
<th></th>
<th>Memory</th>
<th>Time</th>
<th>Required Advice</th>
</tr>
</thead>
<tbody>
<tr>
<td>[LB95]</td>
<td>1</td>
<td>impossible</td>
<td></td>
</tr>
<tr>
<td>[BTV09]</td>
<td>2</td>
<td>$&lt; \infty$</td>
<td>none</td>
</tr>
<tr>
<td>Strong-Weak</td>
<td>2</td>
<td>$O(n^3)$</td>
<td>none</td>
</tr>
<tr>
<td>[KT08]</td>
<td>$O(\log(n))$</td>
<td>$O(n^7)$</td>
<td>$</td>
</tr>
<tr>
<td>Wait-and-See</td>
<td>expected $O(\log(\Delta))$</td>
<td>$O(d + \log(n)) \cdot \log(n)$</td>
<td>$\Theta(\text{broadcast} \cdot</td>
</tr>
</tbody>
</table>

Courtesy of G. Schoenebeck
Next topic: which voters to buy

Q: Consider the voter model on a graph.
Suppose you can change the opinions of $r$ people from $-$ to $+$.
Which opinions should you change?
Want to maximize the probability that convergence to all $+$. 
Which voters to buy - a different model

• Q: Consider the voter model on a graph.
• Suppose you can change the opinions of r people from - to +.
• Which opinions should you change?
• Want to maximize the probability that convergence to all +.

• A: It doesn’t matter since it is the same Random walk.

• Consider a synchronous model where
• $X_v(t+1) = +/-$ with probabilities $\# \{w \in N(v) : X_w(t) = +/-\}$

• Who should we choose here?
• Difference between vertex model and edge model.
• Question asked by: Even Dar and Shapira
Which voters to buy - a different model

• Consider a synchronous model where
• $X_v(t+1) = +/-$ with probabilities $\# \{w \in N(v) : X_w(t) = +/-\}$
• Want to change $X_v(0)$ for s nodes to maximize probability of final vote to be +. Who should we choose?

• Who should we choose here?
• Difference between vertex model and edge model.
Which voters to buy - a different model

• Consider a synchronous model where
  \( X_v(t+1) = +/- \) with probabilities \# \{\( w \in N(v) : X_w(t) = +/- \}\}
• Want to change \( X_v(0) \) for \( s \) nodes to maximize probability of final vote to be +. Who should we choose?

• Who should we choose here?
• Claim: we should choose the highest degree nodes that are-.
• Claim: \( \sum_v d_v X_v(t) \) is a martingale.

• This problem leads us naturally to other problems involving changing people opinions. Next we discuss the “Viral Marketing Problem”
Which voters to buy - a different model

• Consider a synchronous model where
  \[ X_v(t+1) = +/- \text{ with probabilities } \# \{w \in N(v) : X_w(t) = +/-\} \]
• Want to change \( X_v(0) \) for \( s \) nodes to maximize probability of final vote to be +. Who should we choose?

• Who should we choose here?
• Claim: we should choose the highest degree nodes that are -.
• Claim: \( \sum_v d_v X_v(t) \) is a martingale.

• This problem leads us naturally to other problems involving changing people opinions. Next we discuss the “Viral Marketing Problem”
• Most of the next slides are due to Sebastien Roch from a joint paper on viral marketing.
Some Network Optimization Problems

• **Problem:**
  - Optimization over stochastic models defined on networks.

• **Examples:**
  - Which Genes to knock out in order to kill a cancer cell?
  - Which computers to immune in order make a networks robust?
  - Which computers to attack in order to fail the network?
  - Which individuals to immune to stop a disease from spreading.

  - **Viral Marketing:** Which individuals to expose to a product so as to maximize its distribution?
models of collective behavior

- **examples:**
  - joining a riot
  - adopting a product
  - going to a movie

- **model features:**
  - binary decision
  - cascade effect
  - network structure

Courtesy of S. Roch
viral marketing

• referrals, word-of-mouth can be very effective
  - ex.: Hotmail
• viral marketing
  - goal: mining the network value of potential customers
  - how: target a small set of trendsetters, seeds
• example [Domingos-Richardson’02]
  - collaborative filtering system
  - use MRF to compute “influence” of each customer

Courtesy of S. Roch
independent cascade model

- when a node is activated
  - it gets one chance to activate each neighbour
  - probability of success from $u$ to $v$ is $p_{u,v}$

Courtesy of S. Roch
Independent Cascades Model

- graph $G=(V,E)$; initial activated set $S_0, S_1$ or $S_0 \cup S_1, S_0 \cap S_1$
- $a,b$ and $c,d$ the expected size of the marketed set given starting at the 4 sets then:
- **Claim:** $c+d \leq a+b$
- **Pf:** ???
Independent Cascades Model

- graph $G=(V,E)$; initial activated set $S_0$, $S_1$ or $S_0 \cup S_1$, $S_0 \cap S_1$
- $a$, $b$ and $c$ the expected size of the marketed set given starting at the three sets then:
  - **Claim:** $c + d \leq a + b$
  - **Pf:** Use the same randomness to decide which edges copy and which not.
    - Let $G'$ be the (random) induced graph.
    - Then $a = E[\text{vertices connected to } S_0 \text{ in } G]$
    - And $b = E[\text{vertices connected to } S_1 \text{ in } G]$
    - And $c = E[\text{vertices connected to } S_0 \cup S_1 \text{ in } G]$
    - And $d = E[\text{vertices connected to } S_0 \cup S_1 \text{ in } G]$

- This means that the expected size of the infected set is a submodular functions of the set.
generalized models

- graph $G=(V,E)$; initial activated set $S_0$

- **generalized threshold model** [Kempe-Kleinberg-Tardos’03,’05]
  - activation functions: $f_u(S)$ where $S$ is set of activated nodes
  - threshold value: $\theta_u$ uniform in $[0,1]$
  - dynamics: at time $t$, set $S_t$ to $S_{t-1}$ and add all nodes with $f_u(S_{t-1}) \geq \theta_u$
    (note the process stops after (at most) $n-1$ steps)

- **generalized cascade model** [KKT’03,’05]
  - when node $u$ is activated:
    - gets one chance to activate each of the neighbours
    - probability of success from $u$ to $v$: $p_u(v,S)$ where $S$ is set of nodes who have already tried (and failed) to activate $u$
  - assumption: the $p_u(v,\cdot)$’s are “order-independent”

- **theorem** [KKT’03] - the two models are equivalent

Courtesy of
S. Roch
influence maximization

- **definition** - the influence $\sigma(S)$ given the initial seed $S$ is the expected size of the infected set at termination

$$\sigma(S) = E_S [S_{n-1}]$$

- **definition** - in the influence maximization problem (IMP), we want to find the seed $S$ of fixed size $k$ that maximizes the influence

$$S^* = \operatorname{arg\ max} \{\sigma(S) : S \subseteq V, |S| = k\}$$

- **theorem** [KKT’03] - the IMP is **NP-hard**
  - reduction from *Set Cover*: ground set $U = \{u_1, \ldots, u_n\}$ and collection of cover subsets $S_1, \ldots, S_m$

$$(u_i, S_j) \in E \iff u_i \in S_j \quad \exists S, |S| = k, \sigma(S) \geq n + k?$$

*Courtesy of S. Roch*
submodularity

• **definition** - a set function \( f : V \to \mathbb{R} \) is **submodular** if for all \( A, B \) in \( V \)

\[
f(A) + f(B) \geq f(A \cap B) + f(A \cup B)
\]

• example: \( f(S) = g(|S|) \) where \( g \) is concave

• interpretation: “discrete concavity” or “diminishing returns”, indeed submodularity equivalent to

\[
\forall S \subseteq T, \forall v \in V, \quad f(T \cup \{v\}) - f(T) \leq f(S \cup \{v\}) - f(S)
\]

• **threshold models:**
  - it is natural to assume that the **activation functions have diminishing returns**
  - supported by observations of [Leskovec-Adamic-Huberman’06] in the context of viral marketing

Courtesy of S. Roch
main result

• **theorem** [M-Roch’06; first conjectured in KKT’03] - in the generalized threshold model, if all activation functions are monotone and submodular, then the influence is also submodular

• **corollary** [M-Roch’06] - IMP admits a \((1 - e^{-1} - \varepsilon)\)-approximation algorithm (for all \(\varepsilon > 0\))
  - this follows from a general result on the approximation of submodular functions [Nemhauser-Wolsey-Fisher’78]

• known special cases [KKT’03,’05]:
  - linear threshold model, independent cascade model
  - decreasing cascade model, “normalized” submodular threshold model

\[
\forall S \subseteq T, p_u(v,S) \geq p_u(v,T) \text{ or equiv. } \frac{f_u(S \cup \{v\}) - f_u(S)}{1 - f_u(S)} \geq \frac{f_u(T \cup \{v\}) - f_u(T)}{1 - f_u(T)}
\]

Courtesy of
S. Roch
Easy approximation of sub-modular functions

- **Thm:** Let \( f : 2^{[n]} \rightarrow [0, 1] \) be monotone and submodular.
- Consider finding the set \( S \) maximizing \( f(S) \) under the constraints \(|S| = k\).
- Then the greedy algorithm provides a \((1 - 1/e)\) approximate solution to this problem.

- **corollary** [M-Roch’06] - IMP admits a \((1 - e^{-1} - \varepsilon)\)-approximation algorithm (for all \( \varepsilon > 0 \))
  - this follows from a general result on the approximation of submodular functions [Nemhauser-Wolsey-Fisher’78]

- known special cases [KKT’03, ’05]:
  - linear threshold model, independent cascade model
  - decreasing cascade model, “normalized” submodular threshold model

\[
\forall S \subseteq T, p_u(v, S) \geq p_u(v, T) \text{ or equiv. } \frac{f_u(S \cup \{v\}) - f_u(S)}{1 - f_u(S)} \geq \frac{f_u(T \cup \{v\}) - f_u(T)}{1 - f_u(T)}
\]
Easy approximation of sub-modular functions

- **Thm:** Let \( f : 2^n \rightarrow [0,1] \) be monotone and submodular.
- Consider finding the set \( S \) maximizing \( f(S) \) under the constraints \(|S| = k|\).
- Then the greedy algorithm provides a \((1 - 1/e - \varepsilon)\) approximate solution to this problem.

- **Pf:**
  - Let \( S_1, \ldots, S_k \) be the sets chosen by the greedy alg. \( O \) = optimal set.
  - Write \( x_i = f(S_i) - f(S_{i-1}) \)
  - Then \( f(O) \leq f(S_i \cup O) \leq f(S_i) + k \cdot x_{i+1} \)
  - So \( x_{i+1} \geq (f(O) - f(S_i)) / k \).
  - By induction: \( f(S_i) = f(S_{i-1}) + x_i \geq f(O) (1 - (1 - 1/k)^j) \)
  - Taking \( i = k \) we obtain the claim
related work

• **sociology**
  - threshold models: [Granovetter’78], [Morris’00]
  - cascades: [Watts’02]

• **data mining**
  - viral marketing: [KKT’03,’05], [Domingos-Richardson’02]
  - recommendation networks: [Leskovec-Singh-Kleinberg’05], [Leskovec-Adamic-Huberman’06]

• **economics**
  - game-theoretic point of view: [Ellison’93], [Young’02]

• **probability theory**
  - Markov random fields, Glauber dynamics
  - percolation
  - interacting particle systems: voter model, contact process

Courtesy of
S. Roch
proof sketch

Courtesy of S. Roch
coupling

- we use the generalized threshold model
- arbitrary sets $A$, $B$; consider 4 processes:
  - $(A_t)$ started at $A$
  - $(B_t)$ started at $B$
  - $(C_t)$ started at $A \cap B$
  - $(D_t)$ started at $A \cup B$
- it suffices to **couple the 4 processes** in such a way that for all $t$

  $C_t \subseteq A_t \cap B_t$
  $D_t \subseteq A_t \cup B_t$

- indeed, at termination

  $|A_{n-1}| + |B_{n-1}| \geq |A_{n-1} \cap B_{n-1}| + |A_{n-1} \cup B_{n-1}| \geq |C_{n-1}| + |D_{n-1}|$

  (note this works with $|.|$ replaced with any $w$ monotone, submodular)

Courtesy of
S. Roch
proof ideas

- our goal:
  \[ C_t \subseteq A_t \cap B_t \quad (1) \quad D_t \subseteq A_t \cup B_t \quad (2) \]

- antisense coupling
  - obvious way to couple: use same \( \theta_u \)'s for all 4 processes
  - satisfies (1) but not (2)
  - “antisense”: using \( \theta_u \) for \( (A_t) \) and \((1-\theta_u)\) for \( (B_t) \) “maximizes union”
  - we combine both couplings

- piecemeal growth
  - seed sets can be introduced in stages
  - we add \( A \cap B \) then \( A \setminus B \) and finally \( B \setminus A \)

- need-to-know
  - not necessary to pick all \( \theta_u \)'s at beginning
  - can unveil only what we need to know:

\[ \theta_v \in [f_v(S_{t-2}), f_v(S_{t-1})] \]

Courtesy of
S. Roch
piecemeal growth

- process started at $S: (S_t)$
- **partition** of $S$: $S^{(1)},...,S^{(K)}$
- consider the process $(T_t)$:
  - pick $\theta_u$’s
  - run the process with seed $S^{(1)}$ until termination
  - add $S^{(2)}$ and continue until termination
  - add $S^{(3)}$ and so on

- **lemma** - the sets $S_{n-1}$ and $T_{K_{n-1}}$ are the same distribution
antisense coupling

- disjoint sets: $S$, $T$
- partition of $S$: $S^{(1)},...,S^{(K)}$
- piecemeal process with seeds $S^{(1)},...,S^{(K)},T$: $(S_\uparrow)$

- consider the process $(T_\uparrow)$:
  - pick $\theta_u$’s
  - run piecemeal process with seeds $S^{(1)},...,S^{(K)}$ until termination
  - add $T$ and continue with threshold values

$$\theta_v' = 1 - \theta_v + f_v(T_{Kn-1})$$

- lemma - the sets $S_{(K+1)n-1}$ and $T_{(K+1)n-1}$ have the same distribution

Courtesy of
S. Roch
need-to-know

- proof of lemma
  - run the first $K$ stages identically in both processes
  - note that for all $v$ not in $S_{K_{n-1}} = T_{K_{n-1}}$, $\theta_v$ is uniformly distributed in $[f_v(T_{K_{n-1}}), 1]$
  - but $\theta_v' = 1 - \theta_v + f_v(T_{K_{n-1}})$ has the same distribution

simulation 1

simulation 2

Courtesy of S. Roch
Coupling proof I

\[
\begin{align*}
A_0 &= A \cap B \rightarrow A_{n-1} \\
A_n &= A_{n-1} \cup (A \setminus B) \rightarrow A_{2n-1} \\
A_{2n} &= A_{2n-1} \cup \emptyset \rightarrow A_{3n-1} \\
B_0 &= A \cap B \rightarrow B_{n-1} \\
B_n &= B_{n-1} \cup \emptyset \rightarrow B_{2n-1} \\
B_{2n} &= B_{2n-1} \cup (B \setminus A) \rightarrow B_{3n-1}
\end{align*}
\]
**Coupling proof II**

**Phase $A \cap B$**

\[ C_0 = A \cap B \to D_{n-1} \]

**Phase $A \setminus B$**

\[ C_n = C_{n-1} \cup \emptyset \to C_{2n-1} \]

**Phase $B \setminus A$**

\[ C_{2n} = C_{2n-1} \cup \emptyset \to C_{3n-1} \]

\[ A_{3n-1} \cap B_{3n-1} \]

**D**

\[ D_0 = A \cap B \to D_{n-1} \]

\[ D_n = D_{n-1} \cup (A \setminus B) \to D_{2n-1} \]

\[ D_{2n} = D_{2n-1} \cup (B \setminus A) \to D_{3n-1} \]

\[ A_{3n-1} \cup B_{3n-1} \]

**ANTI**
Coupling proof III

• new processes have correct final distribution

• up to time $2n-1$, $B_t = C_t$ and $A_t = D_t$ so that
  
  $$C_t \subseteq A_t \cap B_t \quad D_t \subseteq A_t \cup B_t$$

• for time $2n$, note that
  
  $$B_{2n-1} \subseteq D_{2n-1}$$
  $$B_{2n} = B_{2n-1} \cup (T \setminus S) \quad D_{2n} = D_{2n-1} \cup (T \setminus S)$$

• so by monotonicity and submodularity
  
  $$f_v(B_{2n}) - f_v(B_{2n-1}) \geq f_v(D_{2n}) - f_v(D_{2n-1})$$

• then proceed by induction preserving
  
  $$D_t \setminus D_{2n-1} \subseteq B_t \setminus B_{2n-1} \quad f_v(D_t) - f_v(D_{2n-1}) \leq f_v(B_t) - f_v(B_{2n-1})$$

• At time $t=3n-1$, obtain
  
  $$D_{3n-1} \subseteq D_{2n-1} \cup B_{3n-1} \subseteq A_{3n-1} \cup B_{3n-1}$$
general result

• we have proved:
  **theorem** [Mossel-R’06] - in the generalized threshold model, if all activation functions are submodular, then for any monotone, submodular function $w$, the generalized influence

$$\sigma_w(S) = \mathbb{E}_S[w(S_{n-1})]$$

is submodular

• Note: A closure property for sub-modular functions!
Future Research Directions

- Study optimization problems for other stochastic models defined on networks.