Modeling and learning social influence from opinion dynamics under attack

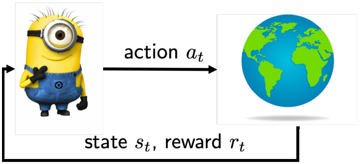
Anna Scaglione

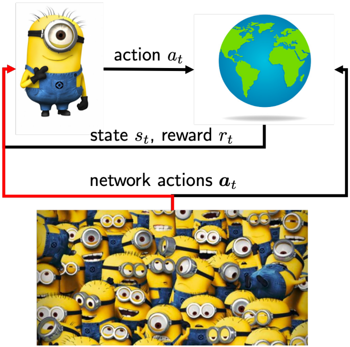
Arizona State University

One World Signal Processing Seminar

Talk for the SPS Distinguished Lecturer series 2019-2020

Cognitive agents: Autonomous and multi-agent systems





- Conceptually the components within each the agents are the same:
  - A learning module, processing sensor information
  - A decision module, optimizing objectives leveraging knowledge from the environment to select an action
- The difference is that in multi-agent systems agents exchange messages to improve learning and decisions performance

Talk Objective

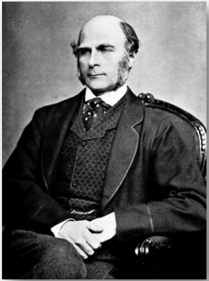
Basic questions:
How can we model swarm behavior in social networks?
Can we model disfunction in the social discourse and polarization?
Can we uncover who influences whom?

What comes next:

- **Part 1: Analysis** — Modeling consensus & polarization of opinions.
- **Part 2: Identification** — Can we infer the social system using these models solving a *system identification* problem?

Understanding social network systems

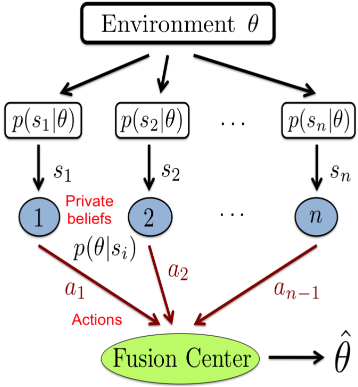
- *Opinion dynamics* model how individuals influence each others



- Early interest in them: Nicolas Marquis de Condorcet, Francis Galton famous paper "Vox populi" [Galton, 1907] sought to prove mathematically that *crowds were wiser* than each individual

Learning under social pressure

- Galton experiment reported in [Galton, 1907] averaged many answers about the weight of an ox, noting the average was very close to the true value
- **Modern opinion dynamics** recognize the critical difference between voting by individuals who just express their preference based on their individual information and individuals who observe the votes of others
- Agents are influenced by their observation of other votes before choosing their action



Models for Opinion Diffusion

Opinion Dynamics

The social systems equations:
Describe through a deterministic or a probabilistic dynamical model how a set of agents combine their private beliefs with observations (direct or indirect) of the peers' beliefs. The goal is understanding the *emergent behavior* of society.

The models are relatively simple:

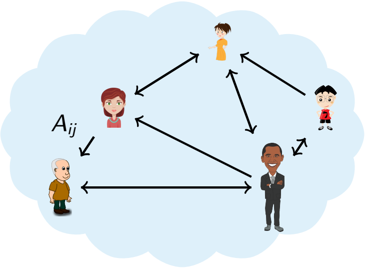
- Agent $i = 1, \dots, n$ holds an initial opinion $x_i(0)$ on a topic θ. For example, for agent i it can be $x_i(0) = p(\theta|s_i)$ or $x_i(0) = \mathbb{E}[\theta|s_i]$
- There is a network process (social activities) through which these opinions are shared and updated

$$\text{agents beliefs at time } t \quad \boldsymbol{x}(t) = (x_1(t), \dots, x_n(t))$$

$$\text{continuous time} \mapsto \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \qquad \text{discrete time} \mapsto \boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t))$$

Social Network Model

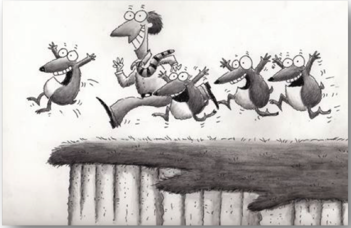
- Associated to the social network is a directed weighted graph $G = (V, E, \boldsymbol{A})$ that is useful to model the interactions.
- **The opinion** of agent i on the kth topic θ_k after the t interactions is $x_i(t; k)$

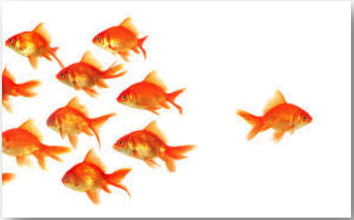


- $V = \{1, \ldots, n\}$ — set of agents (or individuals).
- $E \subseteq V \times V$ — set of edges (or friendship), with $|E| \ll n^2$.
- Each edge has a weight $A_i j$. $\boldsymbol{A} \in \mathbb{R}_+^{n \times n}$ — weighted adjacency matrix, with $\boldsymbol{A}\mathbf{1} = \mathbf{1}$ (each row adds up to one).
- We study **opinion dynamics** where agents interact with each other, with focus on analyzing **steady-state opinions** across different topics.

Classes of Opinion Dynamics models

- Alignment of social behavior:
  - Consensus models
    - ★ Herding behavior (Economics)
    - ★ Fad or trend behavior (Social Psychology)
    - ★ Bandwagon effect (political science)
- Disagreement, polarization
  - Social influence models for "mavens" or "stubborn" agents
  - Bounded confidence models





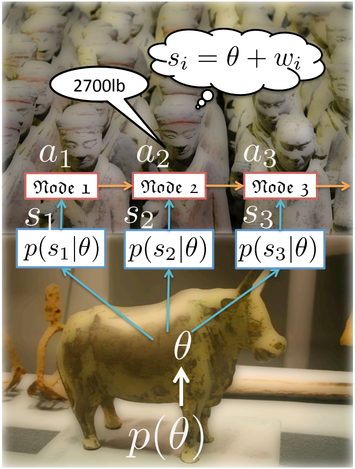
"Vox populi" [Galton, 1907]

- The kth topic θ_k is Gaussian $\theta_k \sim \mathcal{N}(0,\sigma_o^2)$
- The private signals signal are $s_i = \theta_k + w_i$ where $w_i \sim \mathcal{N}(0,\sigma_i^2)$. For simplicity $\sigma_i = 1$
- Each agent original opinion is $x_i(0) = s_i$
- The communication occurs in a sequence
- The optimal weights to minimize the mean squared error are: $A_{i,j} = \frac{t-1}{t}$ for $(i,j) = (t,t-1)$, else $A_{i,j} = 0$
- Update:

$$x(t;k) = (1 - A_{t,t-1})s_t + A_{t,t-1}x(t-1;k).$$

- The opinion converges to the true value in the mean square sense

$$x(t;k) \sim \mathcal{N}\left(\theta_k, \frac{\sigma_o^2}{t}\right)$$

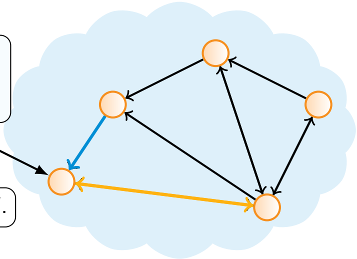


DeGroot Opinion Dynamics, [DeGroot, 1974]

- Topics of discussions indexed by $k = 1, ..., K$. Let $t \geq 0$ be the time index:

$$x_i(t+1;k) = \sum_{j=1}^{n} A_{ij} x_j(t;k)$$

$x_i(t;k)$ — opinion of agent i.



- Agents takes a weighted average of neighbors' opinions (recall $\boldsymbol{A}\mathbf{1} = \mathbf{1}$).

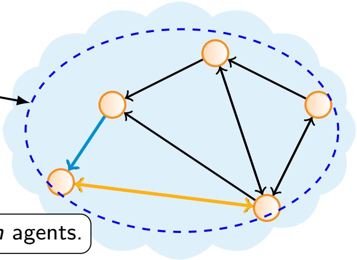
Fact: *if G is* **strongly connected** *and $\boldsymbol{A}\mathbf{1} = \mathbf{1}$, $\lambda_1(\boldsymbol{A}) = 1$, $\lambda_2(\boldsymbol{A}) < 1$*

$$\lim_{t \to \infty} \boldsymbol{x}(t;k) = c(k)^{\star}\mathbf{1} \Longrightarrow \textit{consensus is reached.} \quad (1)$$

DeGroot Opinion Dynamics, [DeGroot, 1974]

- Topics of discussions indexed by $k = 1, ..., K$. Let $t \geq 0$ be the time index:

In matrix form:
$\boldsymbol{x}(t+1;k) = \boldsymbol{A}\boldsymbol{x}(t;k)$



$\boldsymbol{x}(t;k) \in \mathbb{R}^n$ — opinions of n agents.

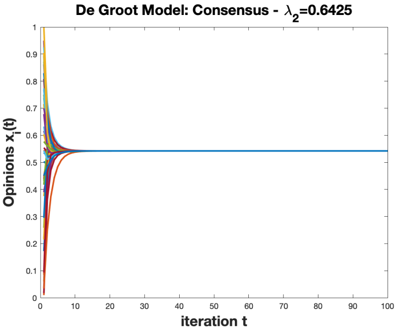
- Agents takes a weighted average of neighbors' opinions (recall $\boldsymbol{A}\mathbf{1} = \mathbf{1}$).

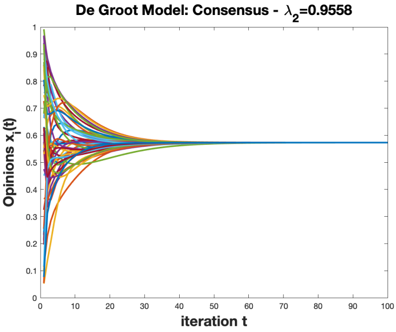
Fact: *if G is* **strongly connected** *and* $\boldsymbol{A}\mathbf{1} = \mathbf{1}$, $\lambda_1(\boldsymbol{A}) = 1$, $\lambda_2(\boldsymbol{A}) < 1$

$$\lim_{t \to \infty} \boldsymbol{x}(t;k) = c(k)^{\star}\mathbf{1} \Longrightarrow \textit{consensus is reached}. \tag{1}$$

Evolution of the opinions

- The DeGroot model is identical to Average Consensus Gossiping (ACG) algorithm studied extensively in computer science, communications, control, signal processing.
- Its convergence is geometric (these are linear dynamics, $\lambda_1(\boldsymbol{A}) = 1$) and the rate is the second largest eigenvalue $\lambda_2(\boldsymbol{A}) < 1$ - the more meshed the network is, the lower is $0 \leq \lambda_2 \leq 1$. The algebraic connectivity of the graph is $1 - \lambda_2(\boldsymbol{A})$.





- A popular variant is randomized...

Randomizing the opinion exchange

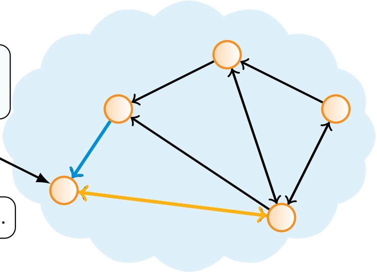
- Pairs of neighbors wake up at random, exchange and then update their opinions to the average of their current ones, i.e.:

$$x_i(t+1;k) = x_j(t+1;k) = 0.5x_i(t;k) + 0.5x_j(t;k)$$

Note the trust matrix is switching randomly (I.e. $\boldsymbol{A}(t)\mathbf{1} = \mathbf{1}$).

$x_i(t+1;k) = 0.5x_i(t;k) + 0.5x_j(t;k)$

$x_i(t;k)$ — opinion of agent i.



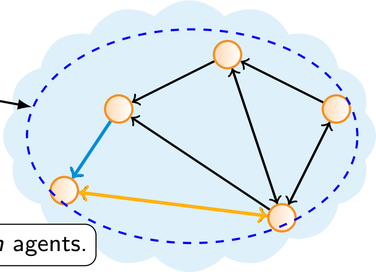
Randomizing the opinion exchange

- Pairs of neighbors wake up at random, exchange and then update their opinions to the average of their current ones, i.e.:

$$x_i(t+1;k) = x_j(t+1;k) = 0.5x_i(t;k) + 0.5x_j(t;k)$$

Note the trust matrix is switching randomly (I.e. $\boldsymbol{A}(t)\mathbf{1} = \mathbf{1}$).

In matrix form:
$\boldsymbol{x}(t+1;k) = \boldsymbol{A}(t)\boldsymbol{x}(t;k)$



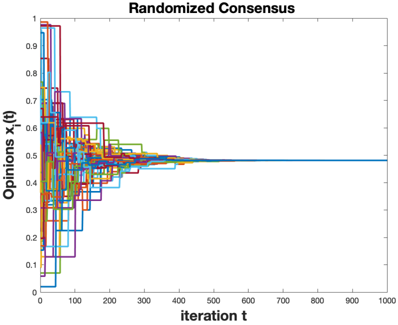
$\boldsymbol{x}(t;k) \in \mathbb{R}^n$ — opinions of n agents.

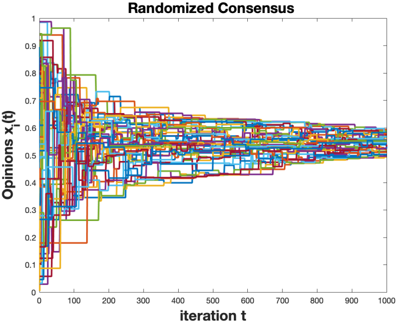
Convergence of Randomized Interaction

- If the process $\boldsymbol{A}(t)$ is stationary (e.g. the pair choice is i.i.d. vs t) in expectation the dynamics look exactly the same. Let $\overline{\boldsymbol{x}}(t;k) = \mathbb{E}[\boldsymbol{x}(t;k)|\theta_k]$ and $\overline{\boldsymbol{A}} = \mathbb{E}[\boldsymbol{A}(t)]$

$$\overline{\boldsymbol{x}}(t+1;k) = \overline{\boldsymbol{A}}\overline{\boldsymbol{x}}(t;k)$$

- In this case if the graph is strongly connected and $\overline{\boldsymbol{A}}\mathbf{1} = \mathbf{1}$ we have almost sure convergence to consensus

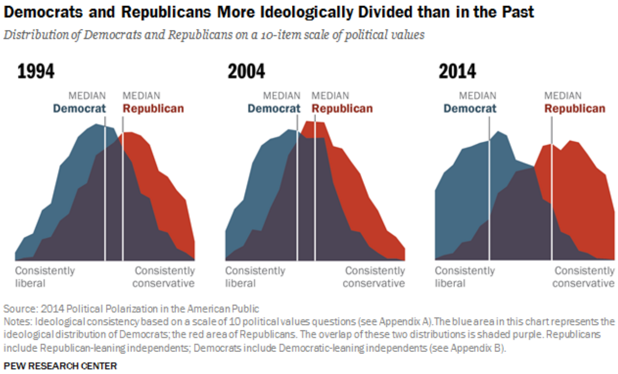




Disagreement and polarization

- Consensus is more the exception than the rule (polarization)

Social network data / opinion:



- **Polarization** is common in social networks' opinions (democrats vs. republicans)

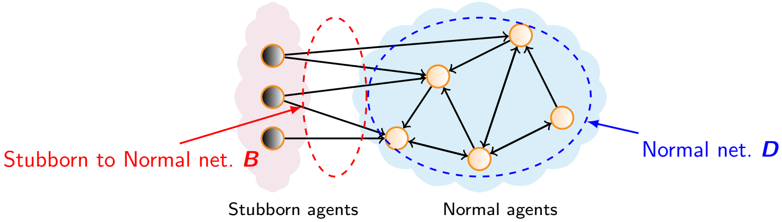
DeGroot Model with Stubborn Agents

- Expand $G = (V, E, \boldsymbol{A})$ to include S stubborn agents with $S \ll n$.
- **Stubborn agents** are agents whose opinions do not change, e.g., the leaders, guiding the rest towards different actions. Mathematically, we have:

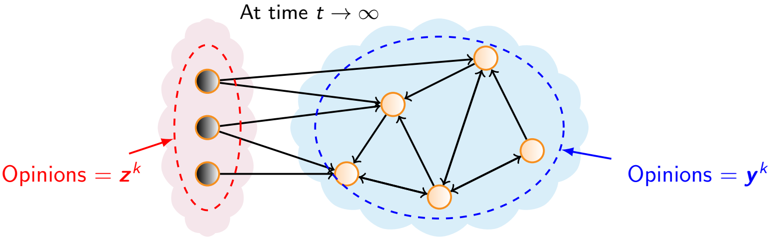
$$x_i(t+1;k) = x_i(t;k), \ \forall \ t, \ \text{if agent } i \text{ is stubborn.} \tag{2}$$

- *DeGroot Opinion Dynamics* — let us partition $\boldsymbol{x}(t;k) = (\boldsymbol{z}(t;k), \boldsymbol{y}(t;k))$

$$\begin{pmatrix} \boldsymbol{z}(t+1;k) \\ \boldsymbol{y}(t+1;k) \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{D} \end{pmatrix}}_{\text{weighted adjacency matrix } \boldsymbol{A}} \begin{pmatrix} \boldsymbol{z}(t;k) \\ \boldsymbol{y}(t;k) \end{pmatrix} \tag{3}$$



Steady-state with Stubborn Agents



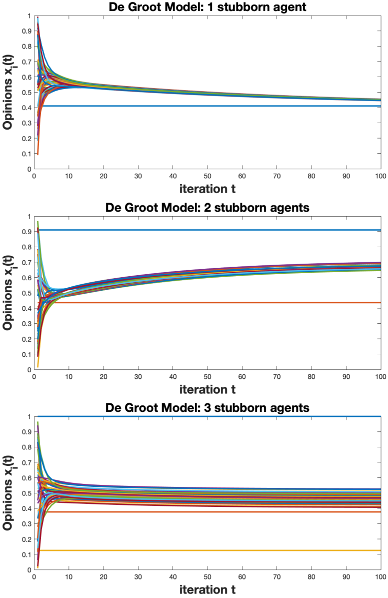
Steady-state: *if sub-graph $G[V \setminus [S]]$ is* **strongly connected** *(+other conditions):*

$$\boldsymbol{y}^k = (\boldsymbol{I} - \boldsymbol{D})^{-1} \boldsymbol{B} \boldsymbol{z}^k . \tag{S2}$$

- Stubborn agents are the **opinion leaders** ($\boldsymbol{y}^k$ depends on $\boldsymbol{z}^k$).

Evolution of the opinion

- The opinions $\boldsymbol{y}^k$ is an output of the linear transformation '$(\boldsymbol{I} - \boldsymbol{D})^{-1}\boldsymbol{B}$'.
- The observations $\{\boldsymbol{y}^k, \boldsymbol{z}^k\}_{k=1}^K$ is at most rank $\min\{S, K\} \Longrightarrow$ rank of data $= S =$ no. of stubborn agents.
- In general there is no consensus. It occurs if there is only one stubborn agent, or all stubborn agents have the same opinion.

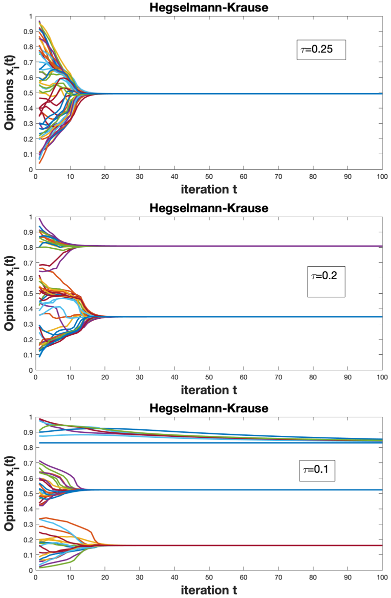


Bounded Confidence and Polarization

- Are stubborn agents necessary to create social division?
- R. Hegselmann and U. Krause in [Hegselmann and Krause, 2002] introduced a model that shows polarization can emerge on its own
- $d_{ij}(t;k)$ is the opinion distance of agents i and j
- Agents "talk" iff $d_{ij}(t;k) < \tau \rightarrow$ update is non linear
- With $u(x)$ the unit step:

$$x_i(t+1;k) = \sum_{j=1}^{n} A_{ij}\, u(\tau - d_{ij}(t;k))\, x_j(t;k)$$

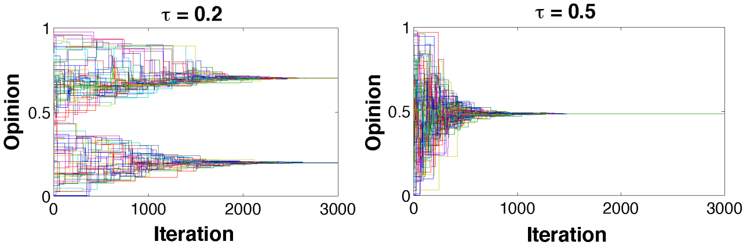
$$\boldsymbol{x}(t+1;k) = \boldsymbol{A}(\boldsymbol{x}(t;k))\boldsymbol{x}(t;k)$$



Alternative: Bounded Confidence Randomized

- The randomized variant is due to G. Deffuant [Deffuant et al., 2000]

$$x_i(t+1;k) = x_i(t;k) + 0.5u(\tau - d_{ij}(t;k))(x_j(t;k) - x_i(t;k))$$
$$x_j(t+1;k) = x_j(t;k) + 0.5u(\tau - d_{ij}(t;k))(x_i(t;k) - x_j(t;k))$$



Theorem [Li et al., 2013]: phase transition to consensus

Let $\overline{A}_{ij}$ probability of the interaction and $d_{ij}(t;k) = |x_i(t;k) - x_j(t;k)|$. A **necessary** condition for the system to converge almost surely to consensus:

$$\tau > \overline{d}_k := \sum_{ij} \overline{A}_{ij} d_{ij}(0;k) \quad (4)$$

Bounded opinions: Voter model

- Agents have a prior belief $x_i(0; k) = p(\theta_k | s_i)$ on a topic θ_k
- Agents can only observe actions $a_i(t; k)$ (typically binary, i.e. $\in \{1, 0\}$) and update their beliefs
- The fact that the observations are quantize makes a significant difference

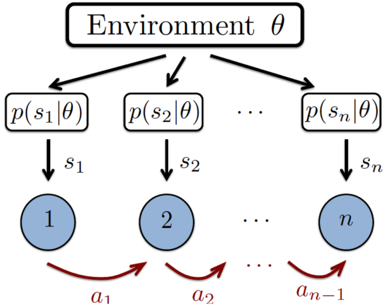
Variants of the voter model

- Rational agents with bounded opinions - Bikhchandani, Hirshleifer and Welch (BHW) [Bikhchandani et al., 1992]
- Binary opinion dynamics - Peter Clifford and Aidan Sudbury [Clifford and Sudbury, 1973] is inspired by Ising model for spin alignment

Both of them lead to consensus or herding

Herding of rational agents

- Bikhchandani, Hirshleifer and Welch (BHW) [Bikhchandani et al., 1992]: θ is randomly determined between the values of 1 ("good state") and 0 ("bad state")
- The agents gain c if $a_i = \theta = 1$ they loose c if $a_i = 1$ and $\theta = 0$ and $a_i = 0$ yields no gain or cost



- The updates are in sequence. The tth agent to update observed s_t and the actions of the previous agents to decides

$$a(t;k) = \text{argmax}_{\theta \in \{0,1\}} P(\theta | s_t, a(t-1;k), \ldots, a(1;k)) \;\; \rightarrow \text{rational choice}$$

Information cascade

The update rule (for cond. independent s_i) for the likelihood ratio is a martingale. In finite iterations the agents *herd* and make the same decision irrespective of their private belief, which has finite probability of being wrong!

Voter model

- Introduced by Peter Clifford and Aidan Sudbury [Clifford and Sudbury, 1973] analyzed first by Richard A Holley and Thomas M Liggett [Holley and Liggett, 1975]
- In interacting social agents copy at random the action of a neighbor
- Equivalent to having the next action $a_i(t;k)$ a Bernoulli random variable:

$$a_i(t;k) \sim \mathcal{B}(x_i(t;k)), \qquad x_i(t+1;k) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i}^{n} a_j(t;k)$$

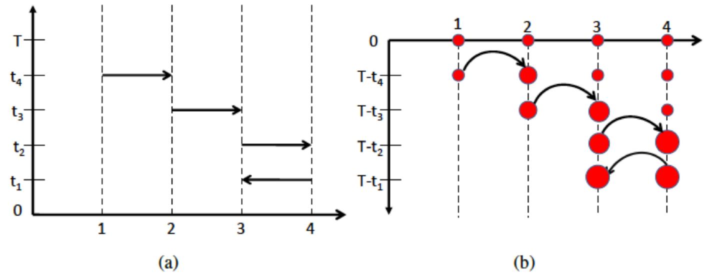
Voter model at $t \to \infty$

Consensus: Reached over connected finite graphs and $1-2$D lattices. All actions are almost surely the same, i.e. $t \to \infty$ $\boldsymbol{a}(t;k) = a^\infty$ a.s..
The final collective action can be $a^\infty = 0$ or $a^\infty = 1$. Given $x_i(0;k)$:

$$P(a^\infty = 1) = \frac{1}{n} \mathbf{1}\mathbf{1}^T \, \boldsymbol{x}(0;k).$$

Analysis of the voter model

- The key analysis methods is the dual approach: it tracks the origin of an opinion at node i at time T back in time, by viewing it as a random *walk*
- The state of each agent i at time T can be traced back to an opinion jumping from a neighbor to another
- If the origins of opinions of all agents at some in time point coalesce, the walks will continue together
- The number of steps backward when each of the random walks coalesces is the consensus convergence time



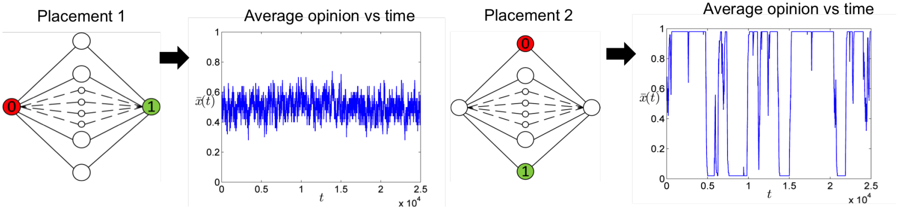
Voter model with stubborn agents

- No consensus. The network continues to randomly change votes
- Mobilia in [Mobilia, 2003] mean field study came first.

Theorem – E. M. Yildiz in [Yildiz et al., 2011]

If every non-stubborn node in the network has at least one directed path to a stubborn agent votes $\boldsymbol{a}(t; k)$ converges in distribution, i.e. for $t \to +\infty$ $\boldsymbol{a}(t; k) \sim \mathcal{B}(\boldsymbol{x}(+\infty))$.

- Placement of stubborn nodes makes a huge difference on the variance



Optimal placement

The problem of maximizing social bias is a sub-modular problem [Yildiz et al., 2011] $\to$ near optimal solution is greedy

The Identification of Social Network Systems

Graphical Model Identification

- The effect of the dynamics is to create a statistical dependency among the different components of $\boldsymbol{x}^k$
- A popular non-parametric approach is to model the data $\{\boldsymbol{x}^k\}_{k=1}^K$ with a **graphical model**.

Observation:
The r.v.s $[\boldsymbol{x}^k]_i$ and $[\boldsymbol{x}^k]_j$ are **conditionally independent** if $(i,j) \notin E$. Equivalently, we have $[\boldsymbol{C}_X^{-1}]_{ij} = 0$ where $\boldsymbol{C}_X = \mathbb{E}[\boldsymbol{x}^k(\boldsymbol{x}^k)^\top]$.

- **Graphical LASSO heuristic** [Friedman et al., 2008] — first estimate the covariance matrix:

$$\hat{\boldsymbol{C}}_X = (1/K) \sum_{k=1}^K \boldsymbol{x}^k(\boldsymbol{x}^k)^\top \ . \tag{5}$$

- then we solve the covariance selection problem:

$$\min_{\boldsymbol{A} \in \mathbb{R}^{n \times n}} \underbrace{-\log\det \boldsymbol{A} + \mathrm{Tr}(\hat{\boldsymbol{C}}_X \boldsymbol{A})}_{\text{yields } \boldsymbol{A} \approx \hat{\boldsymbol{C}}_X^{-1}.} + \rho \cdot \underbrace{\|\mathrm{vec}(\boldsymbol{A})\|_1}_{\text{enforce sparsity.}} \ . \tag{6}$$

Related work

Graphical Model Identification: the main trend among statisticians

- [Hsieh et al., 2014, d'Aspremont et al., 2008, Scheinberg et al., 2010] develop fast algorithm for large networks.
- [Banerjee et al., 2008] analyzes the sample complexity for G-LASSO.
- [Segarra et al., 2016] infer sparse graphs from spectral templates.
- [Bresler, 2015] considers *binary* observations (a.k.a. Ising model).

Other statistical methods: popular among bio-informaticists

- ANOVA [Küffner et al., 2012]; Feature selection with TIGRESS [Haury et al., 2012] and GENIE3 [Huynh-Thu et al., 2010], mostly empirical studies.

Big-Data Challenge?

- Graphical model (+other stat. models) requires **BIG** + **High-Rank** data, e.g.,

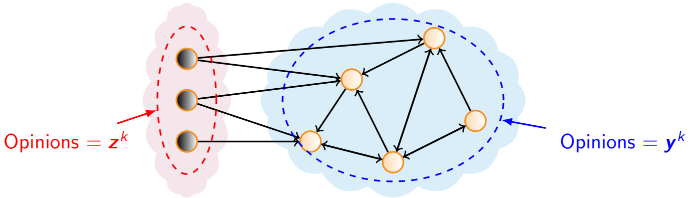
$$\boldsymbol{x}^k = \boldsymbol{C}_X^{1/2} \boldsymbol{z}^k, \;\; \mathbb{E}[\boldsymbol{z}^k (\boldsymbol{z}^k)^\top] = \boldsymbol{I} \in \mathbb{R}^{n \times n} \,. \tag{7}$$

Social System Identification of Opinion Dynamics

- Alternative? [Timme, 2007, Wang et al., 2011] suggest using the DeGroot model for social system id. $\mapsto$ linear equations $\boldsymbol{x}(t; k) = \boldsymbol{A}\boldsymbol{x}(t-1; k)$, not hard...

$$\textbf{Data required}: \ \text{(est. of) } \boldsymbol{x}(t_0; k),\ \boldsymbol{x}(t_0+1; k),\ \boldsymbol{x}(t_0+2; k),\ \cdots$$

- HARD to collect the data in the transient
  1. **Observability** - 'opinion updates' are random and happen in humans' brains
  2. **Stationarity** - semantic analysis requires the latent opinions to be stationary



Our idea: Social RADAR [Wai et al., 2016]

Steady-state: if sub-graph $G[V \setminus [S]]$ is **strongly connected** at steady state:

$$\boldsymbol{y}^k = (\boldsymbol{I} - \boldsymbol{D})^{-1} \boldsymbol{B} \boldsymbol{z}^k \ .$$

The signal from stubborn agents helps *reveal* the social system $\boldsymbol{A}$.

Network Identification Formulation

- **Main idea**: fitting the model $(\boldsymbol{I} - \boldsymbol{D})\boldsymbol{y}^k = \boldsymbol{B}\boldsymbol{z}^k, \forall\ k$. We define:

$$J_i(\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i) := \sum_{k=1}^{K} |(\boldsymbol{e}_i - \hat{\boldsymbol{d}}_i)^\top \boldsymbol{y}^k - \hat{\boldsymbol{b}}_i^\top \boldsymbol{z}^k|^2 \,, \tag{8}$$

  where $\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i$ are estimate of the ith row of $\boldsymbol{B}, \boldsymbol{D}$ and $\boldsymbol{e}_i$ is the ith coordinate vector.
- We solve the following for each $i \in [n]$ in *parallel*: $(\rho > 0)$

$$\min_{\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i} \underbrace{J_i(\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i)}_{\text{LS fitting}} + \rho \underbrace{\|\hat{\boldsymbol{b}}_i\|_1}_{\text{regularizer}} \text{ s.t. } \underbrace{\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i \geq \boldsymbol{0},\ \hat{\boldsymbol{b}}_i^\top \boldsymbol{1} + \hat{\boldsymbol{d}}_i^\top \boldsymbol{1} = 1,\ [\hat{\boldsymbol{d}}_i]_i = 0}_{[\boldsymbol{B}, \boldsymbol{D}] \text{ is stochastic matrix.}} . \tag{9}$$

- Identifiability? Eq. (9) solves an **underdetermined system** as $S \ll n$.
- Intuitively, we expect **identifiability** if the data's rank S grows proportionally to the **sparseness** of the network.

† 'Identifiability' refers to having a unique minimum at $\boldsymbol{b}_i = \hat{\boldsymbol{b}}_i$ and $\boldsymbol{d}_i = \hat{\boldsymbol{d}}_i$ for all i.

Identifiability Condition — 'Active Sensing' Setup

- To gain insight we studied a simplified problem: the **'Active Sensing'** setting where stubborn agents are implanted ($\sim$ social experiment):

Placement of stubborn agents

Each normal agent is connected to **exactly** ℓ stubborn agents, chosen at random the non-zero support Ω_B of matrix $\boldsymbol{B}$ is known. These connections are **known** to the network identification problem.

- **Remark**: This is an ideal condition to give *guidelines* in designing network identification experiments.
- We were able to show that a class of random graphs has guaranteed **identifiability** at $S = \Omega(\|\boldsymbol{d}_i\|_0) \Longrightarrow$ network identification is possible with **low-rank** data.

Sufficient condition for identifiability

Theorem 1 – [Wai et al., 2016]

Let $\beta := S/n$ and $\alpha := 2d_{\max}/n$ be the **density** of the number of stubborn agents and maximum *in-degree* of the normal network. Suppose (β, α, ℓ) satisfies

$$\ell - 1 \geq \max\left\{4, \frac{H(\alpha) + \beta' H(\alpha/\beta')}{\alpha \log(\beta'/\alpha)}\right\}, \quad (\min_{ij\in\Omega_B} B_{ij})(2\ell - 3) > 1 + 2(\max_{ij\in\Omega_B} B_{ij}) \;, \tag{10}$$

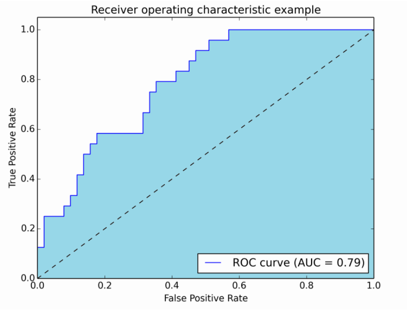
where $\beta' := \beta - \ell/n$ and $H(x) = -x\log x - (1-x)\log(1-x)$, probability of exact recovery:

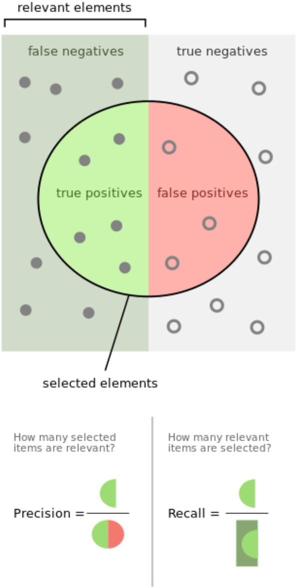
$$\Pr\Big((\hat{\boldsymbol{b}}_i, \hat{\boldsymbol{d}}_i) \neq (\boldsymbol{b}_i, \boldsymbol{d}_i)^1, \ \forall\ i \in [n]\Big) \leq \max_{i\in[n]} \left(\frac{\ell}{\beta}\right)^4 \frac{\ell - 1}{n^2} + \mathcal{O}(n^{2-(\ell-1)(\ell-3)}) \;. \tag{11}$$

The network is **identifiable** if the data's rank satisfies $S = \Omega(d_{\max})$, where $d_{\max}$ is the max. in-degree $\Longrightarrow$ favors **sparse & regular graphs**.

[1] subject to a diagonal scaling ambiguity.

Performance Benchmark





Topology Recovery Performance:

- *Area under ROC curve (AUROC)*: area under the curve of $P_{\text{detection}}$ vs. $P_{\text{false alarm}}$.
- *Area under Precision-Recall (AUPR)*: area under the curve of P_{recall} vs. $P_{\text{precision}}$.

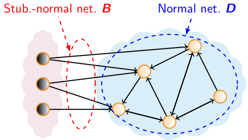
Identification Performance:

- $\text{NMSE} = \|[\boldsymbol{B}, \boldsymbol{D}] - [\hat{\boldsymbol{B}}, \hat{\boldsymbol{D}}]\|_F^2 / \|[\boldsymbol{B}, \boldsymbol{D}]\|_F^2$

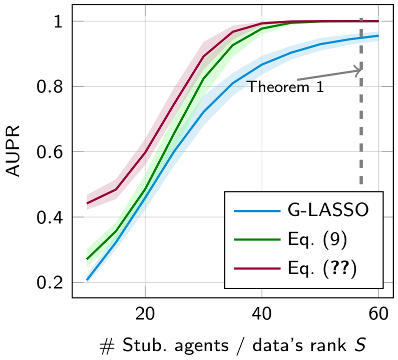
AUROC and AUPR $=1 \rightarrow$ ideal

Synthetic Networks + Data – Topology Recovery

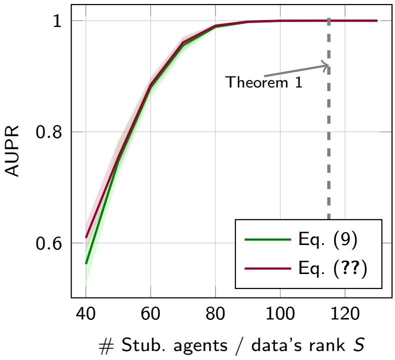
- **Focus**: area under Precision-Recall (= 1 for perfect recover.).
- **Setting**: $\boldsymbol{B} \sim$ each row of $\boldsymbol{B}$ has $\ell = 5$ *non-zeros* at rand. pos..
- **Green — don't know the support of $\boldsymbol{B}$, Ω_B.**
  Red — know the support of $\boldsymbol{B}$, Ω_B. (w/ 'active sensing')



$\boldsymbol{D} \sim$ ER, size $n = 100$, connect. $p = 0.08$



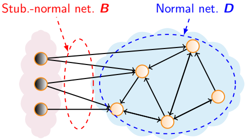
$\boldsymbol{D} \sim$ ER, size $n = 1000$, connect. $p = 0.007$



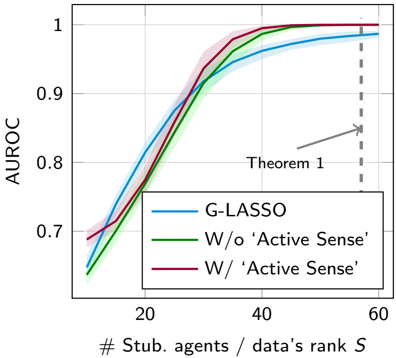
† (Right plot) G-LASSO is numerically unstable as the covariance $\boldsymbol{C}_X$ is extremely low-rank.

Synthetic Networks + Data – Topology Recovery

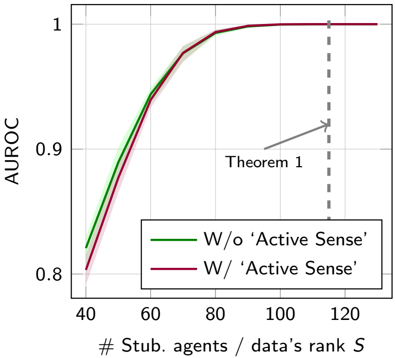
- **Focus**: area under ROC (= 1 for perfect topology recover.).
- **Setting**: $\boldsymbol{B} \sim$ each row of $\boldsymbol{B}$ has $\ell = 5$ *non-zeros* at rand. pos..
- **Green — don't know the support of $\boldsymbol{B}$, Ω_B.**
  Red — know the support of $\boldsymbol{B}$, Ω_B. (w/ 'active sensing')



$\boldsymbol{D} \sim$ ER, size $n = 100$, connect. $p = 0.08$



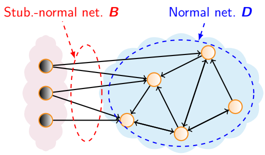
$\boldsymbol{D} \sim$ ER, size $n = 1000$, connect. $p = 0.007$

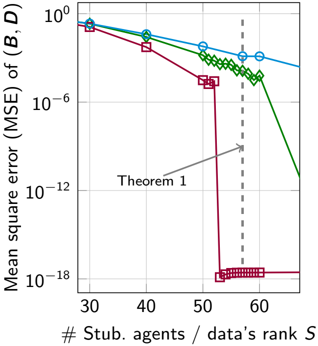


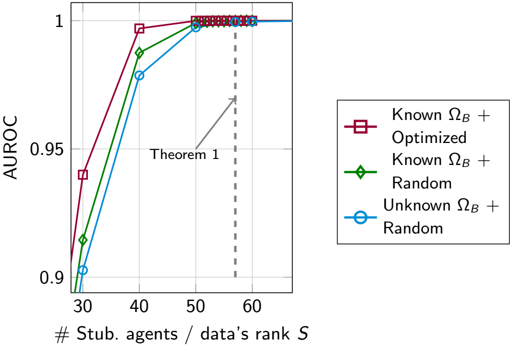
† (Right plot) G-LASSO is numerically unstable as the covariance $\boldsymbol{C}_X$ is extremely low-rank.

Synthetic Networks + Data – Identification performance

- To verify the identification performance from Theorem 1.
- **Settings**: $\boldsymbol{D} \sim$ ER graph with $n = 100$, connectivity $p = 0.08$.
- **Opt.** — each row of $\boldsymbol{B}$ has $\ell = 5$ non-zeros at rand. pos.
  Random — element of $\boldsymbol{B}$ is non-zero with prob. $p = 0.08$.





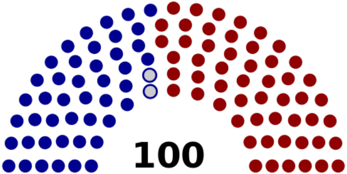


- $S \approx 53$ gives **perfect recovery** under 'active sensing' setting (w/ knowledge of Ω_B).

Real Data from the US Senate

The US Senate

- We consider data from the 114th Congress, including the year 2015 and the portion of 2016 until now.



- Basic facts:
  - There are Republican Senators (red dots), Democratic Senators (blue dots) and Independent Senators (gray dots)
  - The bills are sponsored by a group of congress members, and put forward to the Senate by the Committees
  - The Senators vote for the approval or disapproval of the bill

Data Partition

- Stubborn Senators: whose ideologies are far left or far right, and do not change their own opinions and always try to influence other nodes' opinion.
- We partition the states vector as:

$$\boldsymbol{x}(t;k) = \begin{pmatrix} \boldsymbol{z}(t;k) \\ \boldsymbol{y}(t;k) \end{pmatrix}, \ \boldsymbol{z}(t;k) = \begin{pmatrix} \boldsymbol{s}_R(t;k) \\ \boldsymbol{s}_D(t;k) \end{pmatrix}, \ \boldsymbol{y}(t;k) = \begin{pmatrix} \boldsymbol{r}_R(t;k) \\ \boldsymbol{r}_D(t;k) \\ \boldsymbol{i}(t;k) \end{pmatrix},$$

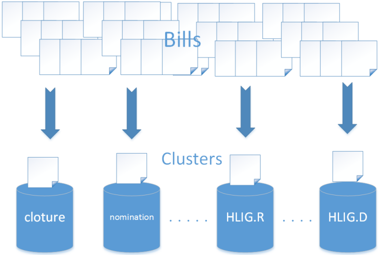
- We partition the transition matrix as:

$$\overline{\boldsymbol{W}} = \begin{array}{c} R_s \\ D_s \\ R_n \\ D_n \\ F \end{array} \overset{\begin{array}{ccccc} R_s & D_s & R_n & D_n & F \end{array}}{\left[\begin{array}{c|c|c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{B}_1 & \mathbf{B}_4 & \mathbf{D}_1 & \mathbf{D}_4 & \mathbf{F}_1 \\ \hline \mathbf{B}_2 & \mathbf{B}_5 & \mathbf{D}_2 & \mathbf{D}_5 & \mathbf{F}_2 \\ \hline \mathbf{B}_3 & \mathbf{B}_6 & \mathbf{D}_3 & \mathbf{D}_6 & \mathbf{F}_3 \end{array}\right]} \qquad (12)$$

E.g., $\mathbf{B}_4$ is the normal Republicans' trust on stubborn Democrats.

Step 1: Clustering - Step 2: Bernoulli sampling

- We cluster bills by the committee and the ideology of the first sponsor of the bill, e.g., (Judiciary, Republican).



- If some of the votes are not specific for a bill or the bill has no committee id (e.g., it is a nomination), we cluster the votes by its vote's category, e.g., 'amendment', 'cloture', 'nomination'.
- We assume that (node n's vote on cluster k) $\sim \mathcal{B}(1, p)$. Then, the node n's likelihood on event k can be estimated as

Data analysis results for the US senate

We can determine according to our model who trusts whom more, and help devise political strategies

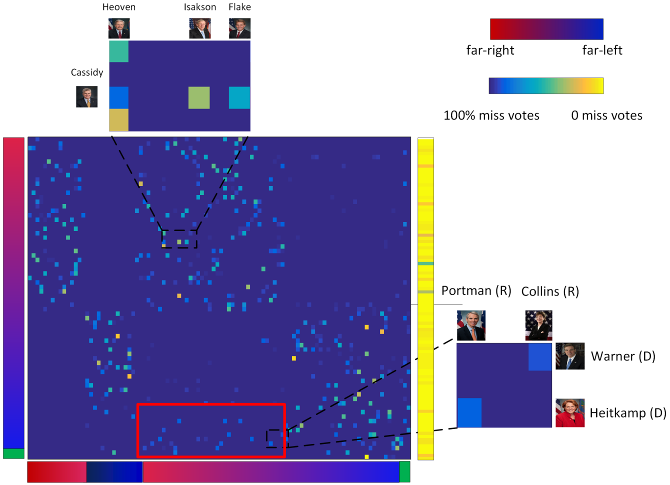


Figure 1: Trust matrix $[\boldsymbol{B}, \boldsymbol{D}]$ with $\mathrm{diag}(\boldsymbol{D}) = \mathbf{0}$.

Impact Factor and Influence Factor

Impact Factor

- State Impact Factor: the number of non-zeros entries in each column of the trust matrix $\tilde{\boldsymbol{W}} = [\boldsymbol{B}, \boldsymbol{D}]$:

$$\text{Impact Factor}_i = \|\tilde{\boldsymbol{w}}_i\|_0, \quad \text{where} \quad \tilde{\boldsymbol{w}}_i \quad \text{is the } i\text{th column of } \tilde{\boldsymbol{W}}.$$

- It represents the number of Senators one Senator has an influence on.

Influence Factor

- State Influence Factor: we sum up all the non-zeros entries in each column in the trust matrix $\tilde{\boldsymbol{W}} = [\boldsymbol{B}, \boldsymbol{D}]$:

$$\text{Influence Factor}_i = \|\tilde{\boldsymbol{w}}_i\|_1, \quad \text{where} \quad \tilde{\boldsymbol{w}}_i \quad \text{is the } i\text{th column of } \tilde{\boldsymbol{W}}.$$

Top 10 in Impact Factor and Influence Factor

Table 1: Impact Factor Top 10

Senator	Total Number
David Perdue (GA/R)	13
Dianne Feinstein (CA/D)	10
Bernie Sanders (VT/I)	10
Mike Enzi (WY/R)	9
Rand Paul (KY/R)	9
Ron Johnson (WI/R)	8
John Hoeven (ND/R)	8
Christopher Murphy (CT/D)	8
Patty Murray (WA/D)	7
Richard Shelby (AL/R)	7

Table 2: Influence Factor Top 10

Senator	Total Trust
Timothy Kaine (VA/D)	1.984
Ron Johnson (WI/R)	1.843
Patty Murray (WA/D)	1.603
Bill Cassidy (LA/R)	1.534
Jefferson Sessions (AL/R)	1.469
John Hoeven (ND/R)	1.440
Gary Peters (MI/D)	1.380
Mike Enzi (WY/R)	1.379
Richard Shelby (AL/R)	1.356
Kelly Ayotte (NH/R)	1.350

- Jefferson Sessions (AL/R): Donald Trump's first Attorney General.
- Bernie Sanders (VT/I): former presidential candidate competing with Hillary Clinton
- Tim Kaine (VA/D): Clinton running mate.

Conclusions

- There are a lot of interesting models and results that can be derived looking at social networks as interacting systems governed by a set of system's equations
- Can we move from qualitative to quantitative analysis and confirm these models from data?
- Controlled experiments are often contrived even if useful, for the DeGroot model we showed an approach to attack real data. Can we generalize it?
- Goal: being able to analyze data that are available on the web going beyond latent semantic analysis that is model free and assuming certain social dynamics

References

[Banerjee et al., 2008] Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008).
Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data.
Journal of Machine learning research, 9(Mar):485–516.

[Bikhchandani et al., 1992] Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992).
A theory of fads, fashion, custom, and cultural change as informational cascades.
In *Journal of Political Economy*, volume 100, pages 992–1026.

[Bresler, 2015] Bresler, G. (2015).
Efficiently learning ising models on arbitrary graphs.
In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC '15, pages 771–782, New York, NY, USA. ACM.

[Clifford and Sudbury, 1973] Clifford, P. and Sudbury, A. (1973).
A model for spatial conflict.
Biometrika, 60(3):581–588.

[d'Aspremont et al., 2008] d'Aspremont, A., Banerjee, O., and El Ghaoui, L. (2008).
First-order methods for sparse covariance selection.
SIAM Journal on Matrix Analysis and Applications, 30(1):56–66.

[Deffuant et al., 2000] Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000).
Mixing beliefs among interacting agents.
In *Adv. Compl. Syst.*, volume 3, pages 87–98.

[DeGroot, 1974] DeGroot, M. (1974).
Reaching a consensus.
In *Journal of American Statistcal Association*, volume 69, pages 118–121.

[Friedman et al., 2008] Friedman, J., Hastie, T., and Tibshirani, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics, 9(3):432–441.

[Galton, 1907] Galton, F. (1907).
Vox populi.
Nature, 75:450–451.

[Haury et al., 2012] Haury, A.-C., Mordelet, F., Vera-Licona, P., and Vert, J.-P. (2012).
Tigress: trustful inference of gene regulation using stability selection.
BMC systems biology, 6(1):145.

[Hegselmann and Krause, 2002] Hegselmann, R. and Krause, U. (2002).
Opinion dynamics and bounded confidence models, analysis, and simulation.
Journal of Artificial Societies and Social Simulation, 5(3).

[Holley and Liggett, 1975] Holley, R. A. and Liggett, T. M. (1975).

Ergodic theorems for weakly interacting infinite systems and the voter model.
The annals of probability, pages 643–663.

[Hsieh et al., 2014] Hsieh, C.-J., Sustik, M. A., Dhillon, I. S., and Ravikumar, P. (2014).
Quic: quadratic approximation for sparse inverse covariance estimation.
Journal of Machine Learning Research, 15(1):2911–2947.

[Huynh-Thu et al., 2010] Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010).
Inferring regulatory networks from expression data using tree-based methods.
PLoS ONE, 5(9).

[Küffner et al., 2012] Küffner, R., Petri, T., Tavakkolkhah, P., Windhager, L., and Zimmer, R. (2012).
Inferring gene regulatory networks by anova.
Bioinformatics, 28(10):1376–1382.

[Li et al., 2013] Li, L., Scaglione, A., Swami, A., and Zhao, Q. (2013).
Consensus, polarization and clustering of opinions in social networks.
IEEE J. Sel. Areas Commun., 31(6):1072–1083.

[Mobilia, 2003] Mobilia, M. (2003).
Does a single zealot affect an infinite group of voters?
Physical review letters, 91(2):028701.

[Scheinberg et al., 2010] Scheinberg, K., Ma, S., and Goldfarb, D. (2010).
Sparse inverse covariance selection via alternating linearization methods.
In *Advances in neural information processing systems*, pages 2101–2109.

[Segarra et al., 2016] Segarra, S., Marques, A. G., Mateos, G., and Ribeiro, A. (2016).
Network topology inference from spectral templates.
arXiv preprint arXiv:1608.03008.

[Timme, 2007] Timme, M. (2007).
Revealing network connectivity from response dynamics.
Physical Review Letters, 98(22):1–4.

[Wai et al., 2016] Wai, H.-T., Scaglione, A., and Leshem, A. (2016).
Active sensing of social networks.
2(3):406–419.

[Wang et al., 2011] Wang, W.-X., Lai, Y.-C., Grebogi, C., and Ye, J. (2011).
Network Reconstruction Based on Evolutionary-Game Data via Compressive Sensing.
Physical Review X, 1(2):1–7.

[Yildiz et al., 2011] Yildiz, E., Acemoglu, D., Ozdaglar, A., Saberi, A., and Scaglione, A. (2011).
Discrete opinion dynamics with stubborn agents.
SSRN eLibrary.