Spread of (Mis)Information over Social Networks

Daron Acemoglu, Asu Ozdaglar, Ali ParandehGheibi

January 2009

Department of Economics
Department of Electrical Engineering & Computer Science

Massachusetts Institute of Technology, USA

Motivation

- Misinformation is everywhere and often spread by certain individuals, groups, or news outlets.
- **Examples:**
  - During the 2004 presidential elections a large fraction of independent voters came to question Senator Kerry's Vietnam war record as a result of Swift Boat Ads.
  - A large fraction of the population in the Middle East believes that 9/11 was a US conspiracy (including 28% of US Muslims)
  - As of 2007, 41% of the US population still believes that Saddam Hussein was directly involved in 9/11

Question

- How does information spread in a society consisting of individuals communicating and sharing information?
- How does misinformation spread and affect beliefs?
- What types of societies and communication structures are "robust" to misinformation?

Approach and Model: Bayesian versus Non-Bayesian

- Model society as a social network of communicating agents.
- "Learning": forming correct beliefs about the underlying state
- Since we focus on spread of misinformation, we adopt a non-Bayesian framework, where some agents are able to "influence" the views of others.
- Why not Bayesian learning?
- With Bayesian learning, the influence on the views of an agent will depend on his belief about whether the person communicating with him is trying to influence him or not
  - Both complicated and limiting the extent of misinformation
  - Non-Bayesian learning similar to "worst-case"

Approach and Model: Bayesian versus Non-Bayesian

- With Bayesian learning, a finite number of influential agents would have no impact on asymptotic beliefs
  - **Acemoglu, Dahleh, Lobel and Ozdaglar (08):**
    - Model of Bayesian learning over an arbitrary social network
    - **Main result**: with unbounded beliefs (i.e., unbounded likelihood ratios) and weak regularity conditions on the structure of the network, asymptotic learning
    - **Additional result**: asymptotic learning will not be disrupted by a finite number of agents with misinformation, even if the purposefully try to manipulate learning

Approach and Model: Interaction Structure

- Focus on non-Bayesian or rule-of-thumb learning
- Distinguish between two kinds of agents:
  - **Regular**
  - **Forceful**: opinionated individuals, news sources, community leaders, political parties...
- Random matching according to an arbitrary communication matrix P, capturing social connections and informational links.
- p_{ij}: probability that agent i observes j.
  - If j is a regular agent, then meeting $\approx$ exchange of information—i and j exchange information and agree with some probability, in which case they take an average of their beliefs.
  - If j is a forceful agent, then meeting $\approx$ i being influenced by j (e.g., listening to the news)—with some probability i adopts j's belief (with ϵ weight on his own belief.

Approach and Model: Method of Analysis

- Transform the evolution of beliefs into transitions of a non-homogeneous Markov chain.
- Convergence analysis using a Lyapunov function argument
- Decompose the mean transition matrix of the Markov chain into the sum of a doubly stochastic matrix and an influence matrix (reflecting the influence of forceful agents).
- Develop bounds on the stochastic behavior of left eigenvectors as a function of the doubly stochastic and the influence matrices
  - Using perturbation theory for Markov chains **(Schweitzer 68)**, spectral graph theory **(Cheeger 70)**, and min cut-max flow theorem **(Ford-Fulkerson 56)**

Approach and Model: Results I

- **Notion of Learning**: (almost sure) convergence to consensus with $1/n$ weight on the initial beliefs of each agent
  - Capturing aggregation of decentralized information across agents
- In the absence of forceful agents, in a society with n agents, beliefs converge to $1/n$-weighted average almost surely.
- **Result 1**: With forceful agents, beliefs still converge to consensus almost surely, but this consensus value is a random variable.
- **Question**: How far is this random consensus from the $1/n$-weighted average?
- **Assumption**: nonzero (small) probability that even forceful agents obtain information from (or be influenced by) others.

Approach and Model: Results II

- **Result 2**: General bounds on the gap between the mean consensus and the $1/n$-weighted average as a function of:
  - Size of society (as n gets large holding number of forceful agents constant, consensus arbitrarily close to $1/n$-weighted average in some topologies)
  - Number and connections of forceful agents (as the probability that others observe forceful agents diminishes, consensus closer to $1/n$-weighted average).
  - Network topology (in particular, whether the induced Markov chain is slow-mixing or fast-mixing).
- **Result 3**: Exact characterization of the difference between the mean consensus and the $1/n$-weighted average as a function of:
  - Mean passage times of the induced Markov chain
  - "Relative minimum cuts" between regular and forceful agents: minimum number of edges between subsets of nodes that include the regular and forceful agents in the network

Related Literature

- Most closely related non-Bayesian learning models:
  - **DeGroot (74)**, **DeMarzo, Vayanos, Zwiebel (03)**, **Golub and Jackson (07), (08)**
  - Beliefs updated using simple averaging rules
  - Conditions on the network structure that lead to asymptotic learning
- We have an alternative model of spread of misinformation, which turns out to be more tractable to characterize and quantify the impact of influential agents on the asymptotic belief distributions

Model: Agents and Beliefs

- Finite society consisting of a set $\mathcal{N} = \{1, \ldots, n\}$ of agents.
- Each agent i endowed with initial belief $x_i(0)$.
- With a law of large numbers reasoning, we are interested in whether "social beliefs" or "consensus" across agents will reflect

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i(0).$$

- **Interpretation**: an extreme $x_i(0)$ will have little "influence" in μ.
- But it may have a large influence when beliefs converge to some other random variable (or fail to converge).
- **Forceful agents**: those with extreme $x_i(0)$'s having potential influence on others' beliefs.

Model: Communication and Information Exchange

- Time is continuous. Each agent is recognized according to iid Poisson processes.
- Let $k = 1, 2, 3, ...$ index dates of communication.
- $x_i(k)$: belief of agent i after k^{th} communication.
- Conditional on being recognized, agent i observes agent j with probability p_{ij}.
- Conditional on i observing agent j:
  - With probability β_{ij}, the two agents agree and exchange information

    $$x_i(k+1) = x_j(k+1) = (x_i(k) + x_j(k))/2.$$

  - With probability γ_{ij}, disagreement and no exchange of information.
  - With probability α_{ij}, i is influenced by j

    $$x_i(k+1) = \epsilon x_i(k) + (1-\epsilon)x_j(k)$$

    for some $\epsilon > 0$ small. Agent j's belief remains unchanged.
- We say that j **is a forceful agent if** $\alpha_{ij} > 0$ **for some** i.

Model: Notation

- Let $x(k) = (x_1(k), \ldots, x_n(k))$ denote the vector of agent beliefs at time k.
- The agent beliefs updated according to

$$x(k+1) = W(k)x(k),$$

where $W(k)$ is a random matrix given by

$$W(k) = \begin{cases} A_{ij} \equiv I - \frac{(e_i - e_j)(e_i - e_j)'}{2} & \text{with probability } p_{ij}\beta_{ij}/n, \\ J_{ij} \equiv I - (1-\epsilon)\, e_i(e_i - e_j)' & \text{with probability } p_{ij}\alpha_{ij}/n, \\ I & \text{with probability } p_{ij}\gamma_{ij}/n. \end{cases}$$

- The matrix $W(k)$ is a **(row) stochastic matrix** for all k (i.e., $\sum_{j=1}^{n}[W(k)]_{ij} = 1$ for all i, k), and is independent and identically distributed over all k.
- We introduce the **transition matrices**

$$\Phi(k,s) = W(k)W(k-1)\cdots W(s+1)W(s) \qquad \text{for all } k \text{ and } s \text{ with } \; k \geq s,$$

- The belief update rule can be written as

$$x_i(k+1) = \sum_{j=1}^{n} [\Phi(k,s)]_{ij}\, x_j(s) \quad \text{for all } k \geq s, \text{ and all } i.$$

Model: Assumptions

Assumption *(Communication Probabilities)*

(a) For all i, the probabilities p_{ii} are equal to 0.

(b) For all i, the probabilities p_{ij} are nonnegative for all j and they sum to 1 over j,

$$p_{ij} \geq 0 \quad \text{for all } i, j, \qquad \sum_{j=1}^{n} p_{ij} = 1 \quad \text{for all } i.$$

- Natural assumption
- The communication matrix $P = [p_{ij}]_{i,j \in \mathcal{N}}$

Mean connectivity graph: $(\mathcal{N}, \mathcal{E})$, where $\mathcal{E}$ is the set of edges induced by the positive communication probabilities p_{ij}, i.e.,

$$\mathcal{E} = \{(i, j) \mid p_{ij} > 0\}.$$

Assumption *(Connectivity)* The graph $(\mathcal{N}, \mathcal{E})$ is connected, i.e., for all $i, j \in \mathcal{N}$, there exists a directed path connecting i to j with edges in the set $\mathcal{E}$.

- This assumption ensures that information (or misinformation) does not get trapped in a subnetwork.

Model: Assumptions (continued)

Assumption *(Interaction Probabilities)* For all $(i, j) \in \mathcal{E}$, the sum of the *averaging probability* β_{ij} and the *influence probability* α_{ij} is positive, i.e.,

$$\beta_{ij} + \alpha_{ij} > 0 \qquad \text{for all } (i, j) \in \mathcal{E}.$$

- Positive probability that even forceful agents eventually exchange information.
- For example, they obtain their own information from the other agents in the society.
  - Otherwise, not a "connected network"

Preliminary Result

Theorem: Let Communication Probabilities, Connectivity, and Interaction Probabilities assumptions hold and suppose that **there are no forceful agents**, i.e., $\alpha_{ij} = 0$ for all $(i, j) \in \mathcal{E}$. Then, the beliefs $\{x_i(k)\}$, $i \in \mathcal{N}$ converge to a consensus belief of $\frac{1}{n}\sum_{i=1}^{n} x_i(0)$, i.e.,

$$\lim_{k\to\infty} x_i(k) = \frac{1}{n}\sum_{i=1}^{n} x_i(0) \qquad \text{for all } i \text{ with probability one.}$$

- Well-known result as a benchmark for comparison

Main Theorems (I): Convergence to Consensus

Theorem: Let Communication Probabilities, Connectivity, and Interaction Probabilities assumptions hold. Then, the beliefs $\{x_i(k)\}$, $i \in \mathcal{N}$ converge to a **consensus belief**, i.e., there exists a scalar random variable $\bar{x}$ such that

$$\lim_{k\to\infty} x_i(k) = \bar{x} \qquad \text{for all } i \text{ with probability one.}$$

- Convergence to consensus guaranteed.
- But with forceful agents, consensus belief is a random variable.
- **Rate of convergence** can be written as a function of the number of agents and the spectral properties of the underlying mean connectivity graph.
  - Due to lack of *doubly stochasticity* (i.e., both row and column stochasticity) of the evolution matrix $W(k)$, convergence rate can be slow.

Proof Sketch

- With positive probability (uniformly bounded away from 0), there exists a scalar $\eta > 0$ such that

$$[\Phi(s + n^2 d - 1, s)]_{ij} \geq \eta^{n^2 d}, \text{ for all } i, j, \text{ and } s \geq 0,$$

where d is the maximum shortest path length over any (i, j) in the mean connectivity graph.

- Let $\{x(k)\}$ denote the belief sequence.
- Define the Lyapunov function $V(k) = M(k) - m(k)$, where

$$M(k) = \max_{i \in \mathcal{N}} x_i(k), \qquad m(k) = \min_{i \in \mathcal{N}} x_i(k).$$

- Show $V(k)$ strictly decreases with positive probability for all k.

Characterization of Social Influence

- We are interested in providing an upper bound on

$$E\left[\bar{x} - \frac{1}{n}\sum_{i=1}^{n} x_i(0)\right],$$

where $\bar{x}$ is the stochastic consensus belief.

- Consider the mean interaction matrix

$$\tilde{W} = E[W(k)] \qquad \text{for all } k \geq 0.$$

- Under our assumptions, $\tilde{W}$ can be viewed as the transition matrix of an irreducible aperiodic Markov chain (cf. connectivity assumption and positive diagonal assumption, implying self-loops)

- **Implications:**
  - There exists a probability vector π such that $\lim_{k\to\infty} \tilde{W}^k = e\pi'$ (e is the vector of all ones).
  - $E[\bar{x}]$ is given by a convex combination of the initial agent values $x_i(0)$ with weights given by π, i.e.,

$$E[\bar{x}] = \sum_{i=1}^{n} \pi_i x_i(0) = \pi' x(0).$$

Main Theorems (II): Bounds on Limiting Belief Distributions

- Bounds on how far asymptotic beliefs are from $\frac{1}{n}\sum_{i=1}^{n} x_i(0)$.
- Bounds depend on two things: **total influence and the parameter δ related to the mixing time of the graph**

Theorem:

(a) Let π denote the unique stationary distribution of $\tilde{W}$. Then,

$$\left\|\pi - \frac{1}{n}e\right\|_\infty \leq \frac{1}{n}\,\frac{\frac{\sum_{i,j} p_{ij}\alpha_{ij}}{n}}{(1-\delta) - \frac{\sum_{i,j} p_{ij}\alpha_{ij}}{n}},$$

where $\delta > 0$ is a constant given by

$$\delta = (1 - n\xi^d)^{\frac{1}{d}}, \qquad \xi = \min_{(i,j)\in\mathcal{E}} \left\{ \frac{p_{ij}}{n}\,\frac{(1-\gamma_{ij})}{2} \right\},$$

d is the maximum shortest path length in the mean connectivity graph $(\mathcal{N}, \mathcal{E})$.

(b) We have

$$\left| E[\bar{x}] - \frac{1}{n}\sum_{i=1}^{n} x_i(0) \right| \leq \frac{1}{n}\,\frac{\frac{\sum_{i,j} p_{ij}\alpha_{ij}}{n}}{(1-\delta) - \frac{\sum_{i,j} p_{ij}\alpha_{ij}}{n}}\,\|x(0)\|_\infty.$$

Proof Sketch

- Relies on a fundamental result from perturbation theory of Markov Chains
- Consider an irreducible aperiodic Markov Chain (MC) with transition probability matrix T and stationary distribution π
- The **fundamental matrix** of the MC is given by $Z = (I - T - T^{\infty})^{-1}$, where $T^{\infty} = e\pi'$. It is straightforward to show that

$$Z = \sum_{k=0}^{\infty} (T^k - T^{\infty}).$$

- Let $Y = Z - T^{\infty}$ be the **deviation matrix** of the MC.

Theorem [Schweitzer 68] Let D be an $n \times n$ perturbation matrix such that $\sum_{j=1}^{n} [D]_{ij} = 0$ for all i. Assume that the perturbed MC with transition matrix $\hat{T} = T + D$ is irreducible and aperiodic.

Then, the perturbed MC has a unique stationary distribution $\hat{\pi}$, and the matrix $I - DY$ is nonsingular. Moreover, the change in the stationary distributions, $d = \hat{\pi} - \pi$, is given by

$$d = \pi DY(I - DY)^{-1}, \qquad \text{or equivalently} \qquad d = \hat{\pi} DY.$$

Proof Sketch

- In view of the belief update rule, we can write $\tilde{W}$ as

$$\begin{aligned}\tilde{W} &= \frac{1}{n}\sum_{i,j} p_{ij}\Big[\beta_{ij}A_{ij} + \alpha_{ij}J_{ij} + \gamma_{ij}I\Big] \\ &= \frac{1}{n}\sum_{i,j} p_{ij}\Big[(1-\gamma_{ij})A_{ij} + \gamma_{ij}I\Big] + \frac{1}{n}\sum_{i,j} p_{ij}\alpha_{ij}\left[J_{ij} - A_{ij}\right] \equiv T + D\end{aligned}$$

- We obtain a bound on

$$\|DY\|_\infty = \|\sum_{k=0}^{\infty} D(T^k - T^\infty)\|_\infty \leq \sum_{k=0}^{\infty} \|DT^k\|_\infty,$$

where the equality follows since $T^\infty = \frac{1}{n}ee'$, and therefore $DT^\infty = 0$.

- For any z with $\|z\|_\infty = 1$, $\|DT^k z\|_\infty$ can be upper bounded by

$$\|DT^k z\|_\infty \leq \frac{1}{n}\sum_{i,j} p_{ij}\alpha_{ij}V(k), \quad \text{with} \quad V(k) = \max_l (T^k z)_l - \min_l (T^k z)_l.$$

- $V(k)$ decreases geometrically at a rate δ.

- The result follows by combining this with

$$\|d\|_\infty \leq \|y\|_\infty \frac{\|DY\|_\infty}{1 - \|DY\|_\infty}.$$

Main Theorems (III): Bounds on Limiting Belief Distributions

Theorem: Let π denote the unique stationary distribution of $\tilde{W}$. Then,

$$\left\| \pi - \frac{1}{n} e \right\|_2 \leq \frac{1}{n} \frac{\frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n}}{(1 - \lambda_2) - \frac{\sum_{i,j} p_{ij} \alpha_{ij}}{n}},$$

where λ_2 is the second largest eigenvalue of the matrix T (recall $T = \tilde{W} - D$).

- λ_2 is the second largest eigenvalue of the **doubly stochastic part of the mean interaction matrix**.
- λ_2 related to **mixing time of a Markov Chain** (i.e., an asymptotic measure of the convergence of the state distribution to the uniform stationary distribution)
  - $(1 - \lambda_2)$: **spectral gap**
  - When the spectral gap is large, we say that the Markov Chain induced by the matrix T is **fast-mixing**

Implications and Intuition

Influence in Connected Societies: If $\frac{\sum_i p_{ij}}{n}$ is small for each forceful j,

$$E[\bar{x}] \approx \frac{1}{n} \sum_{i=1}^{n} x_i(0).$$

- If the probability with which forceful agents are observed is small, then this information will not spread fast.

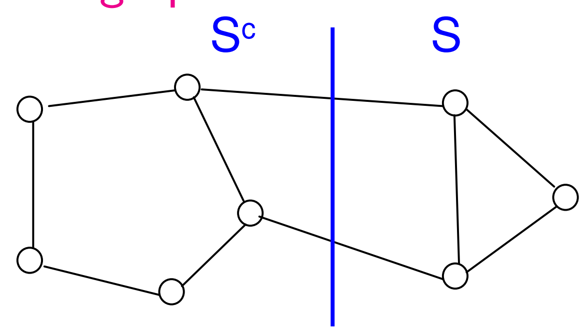
Influence and Social Network Structure: All else equal, the gap between $E[\bar{x}]$ and $\frac{1}{n} \sum_{i=1}^{n} x_i(0)$ is smaller when the Markov chain induced by T is fast-mixing.

- **Intuition**: With a fast-mixing T, forceful agents will themselves be influenced by others (since $\beta_{ij} + \alpha_{ij} > 0$ for all i, j) and misinformation will not spread in the network.

Spectral Gap and Network Properties

Capture network properties in terms of the conductance of a graph

- Given an $n \times n$ symmetric transition probability matrix T, associate an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with $\mathcal{N} = \{1, \ldots, n\}$, and edge weights T_{ij}.



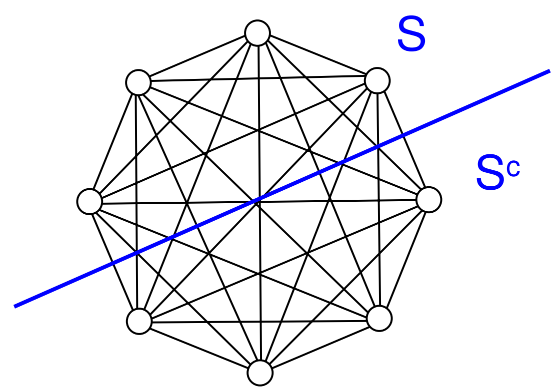
- The conductance $\rho(T)$ of the graph is given by

$$\rho(T) = \min_{\substack{S \subset \mathcal{N} \\ |S| \leq \frac{n}{2}}} \frac{\sum_{i \in S} \sum_{j \in S^c} T_{ij}}{|S|},$$

i.e., "normalized min-cut of the graph"

- For a complete (fully connected) graph,

$$\rho(T) \approx \frac{n^2 \times 1/n}{n} = O(1)$$



Theorem (Cheeger's inequality): The spectral gap, $1 - \lambda_2(T)$, satisfies

$$\frac{\rho(T)^2}{2} \leq 1 - \lambda_2(T) \leq 2\rho(T)$$

Influence in Large Societies

- Consider complete graphs and **expanders**
  - Informally, graphs in which any "small" subset of vertices has a relatively "large" neighborhood.
  - Random graphs under the preferential connectivity model are expanders **Mihail, Papadimitriou, Saberi (03)**.
- Define an agent j to be **locally forceful** if $\sum_{i=1}^{n} p_{ij}\alpha_{ij} = O(1)$
- Assume that there are $M = O(1)$ locally forceful agents.
- For n large,

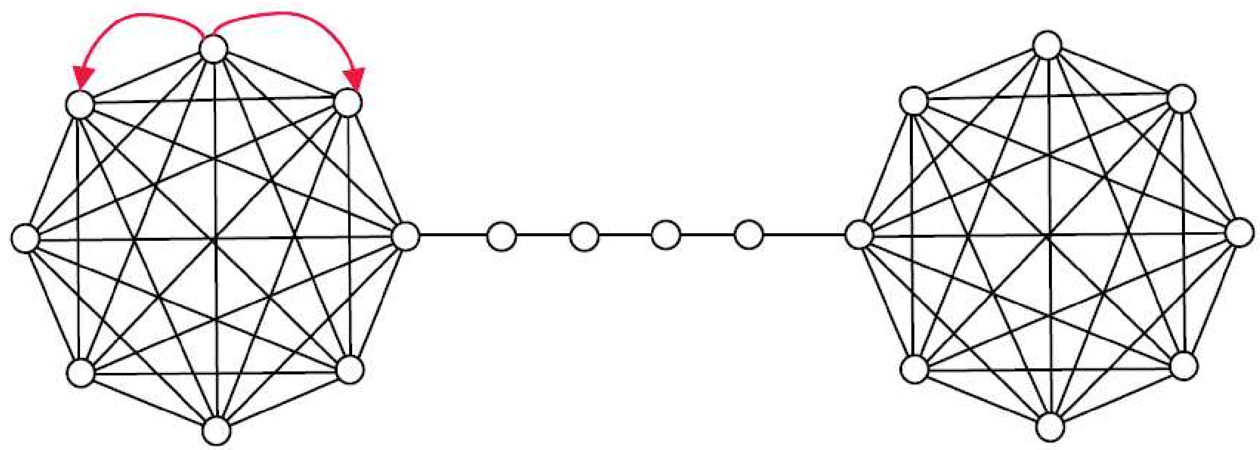
$$E[\bar{x}] \approx \frac{1}{n}\sum_{i=1}^{n} x_i(0).$$

  - In a large connected society, misinformation of locally forceful agents will not spread.
  - **Intuition:** The more connected the network, the less the effect of locally forceful agents

Location of Forceful Agents I

Example: Consider the barbell graph (two complete graphs connected with a line), and one agent influencing two agents in the same cluster

- Related to homophily in societies **Golub and Jackson (08)**



- Intuitively, influence in this graph should be limited (since each cluster is well-connected)
- However, the conductance of the barbell graph is

$$\rho(T) \approx \frac{1}{n} = O\Big(\frac{1}{n}\Big),$$

implying a small spectral gap and therefore large value of the bound.

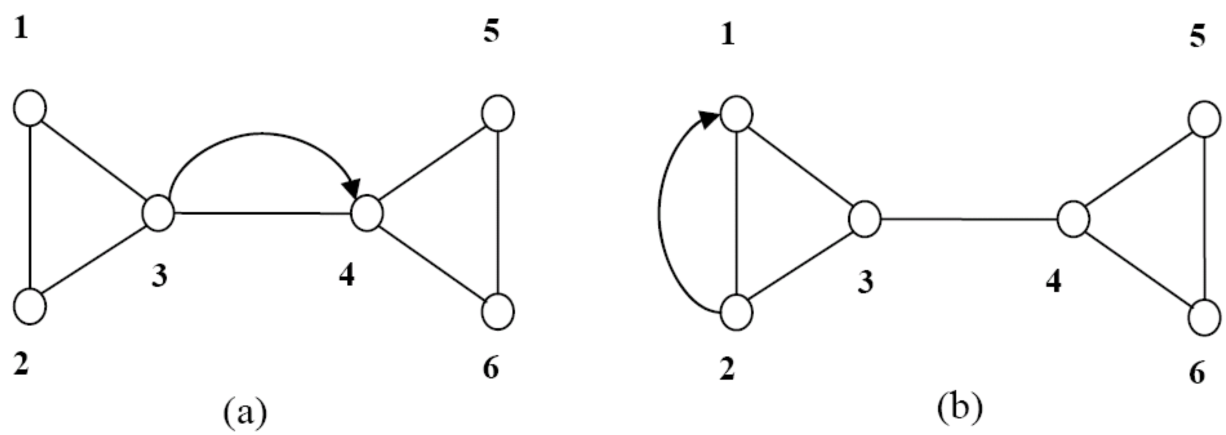
- Suggests considering cuts associated with the forceful and regular agents

Location of Forceful Agents II

- Bounds so far characterize the variation of the stationary distribution in terms of total influence of forceful agents, $\sum_{i,j} p_{ij}\alpha_{ij}$, and the second largest eigenvalue of matrix T
- Bounds do not depend on the location of the forceful agents

Example: Consider 6 agents connected with undirected graph induced by T and two different misinformation scenarios:

- forceful agent over a bottleneck and forceful agent inside a cluster



The stationary distribution for each case is given by

$$\pi_a = \frac{1}{6}(1.25, 1.25, 1.25, 0.75, 0.75, 0.75)', \quad \pi_b = \frac{1}{6}(0.82, 1.18, 1, 1, 1, 1)'.$$

Main Theorems (IV): Exact Characterization of Stationary Distribution

Theorem: Let π denote the unique stationary distribution of $\tilde{W}$. Then,

$$\pi_k - \frac{1}{n} = \sum_{i,j} \frac{p_{ij}\alpha_{ij}}{2n^2}((1-2\epsilon)\pi_i + \pi_j)\big(m_{ik} - m_{jk}\big) \quad \text{for all } k,$$

where m_{ij} is the mean first passage time from state i to state j of a Markov chain $(X_t, t = 0, 1, 2, \ldots)$ with transition matrix T, i.e.,

$$m_{ij} = \mathbb{E}[T_j \mid X_0 = i],$$

where $T_i = \min\{t \geq 0 \mid X_t = i\}$.

- Proof relies on using Schweitzer's exact perturbation result and relating the mean first passage times to the fundamental matrix of the Markov chain
- Implies that the sensitivity of each agent to influence links depend on the relative distance of that agent to the forceful and the forced agent
  - Explains the insensitivity of the agents in the right cluster in the previous example, part (b)

Main Theorems (V): Bounds in terms of Relative Min-Cut

Theorem: Let π denote the unique stationary distribution of $\tilde{W}$. Then

$$\left\|\pi - \frac{1}{n}e\right\|_\infty \leq \sum_{\{i,j\}\in\mathcal{A}} \frac{|p_{ij}\alpha_{ij} - p_{ji}\alpha_{ji}|}{2c_{ij}},$$

where c_{ij} is the minimum $i-j$ cut on the graph induced by matrix T, i.e.,

$$c_{ij} = \min_{\substack{S\subset\mathcal{N} \\ i\in S, j\notin S}} \left\{\sum_{k\in S}\sum_{l\in S^c} w_{kl}\right\}.$$

- **Return to Barbell Graph:** The minimum $i-j$ cut of the barbell graph is

$$c_{ij} \approx n^2 \times \frac{1}{n} = O(n),$$

implying misinformation of locally forceful agents in both clusters will not spread.

Proof Outline

- For all k, we have the relation

$$\begin{aligned}\left|\pi_k - \frac{1}{n}\right| &\leq \sum_{\{i,j\}\in\mathcal{A}} \frac{|p_{ij}\alpha_{ij} - p_{ji}\alpha_{ji}|}{2n^2}|m_{ik} - m_{jk}| \\ &\leq \sum_{\{i,j\}\in\mathcal{A}} \frac{|p_{ij}\alpha_{ij} - p_{ji}\alpha_{ji}|}{2n^2} \max\{m_{ij}, m_{ji}\},\end{aligned}$$

  where the second inequality follows from $m_{ik} \leq m_{ij} + m_{jk}$ & $m_{jk} \leq m_{ji} + m_{ik}$.

- Use Max flow-Min cut Theorem (from linear network optimization theory) to relate the mean passage time m_{ij} to cuts between i and j.
  - Max flow-min cut theorem states that the maximum amount of flow between any two nodes is equal to the capacity of the minimum cut, i.e., is dictated by its bottleneck

Conclusions

- Framework for the analysis of spread of misinformation in a society represented by a general social network.
- Under a minimal set of assumptions, misinformation does not prevent convergence to consensus.
- However, consensus can be on an undesirable set of beliefs, reflecting those of forceful agents (possibly spreading misinformation).
- Bounds on the effect of misinformation and influence of forceful agents.
- Under various benchmark assumptions, this influence is limited.
- **Future Work:**
  - Bounds on other moments of the belief distribution.
  - What happens when the society is not connected?
  - Worst-case analysis: robustness against adversarial behavior.