Local non-Bayesian social learning with stubborn agents

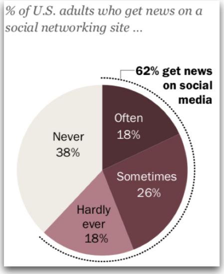
Daniel Vial, Vijay Subramanian

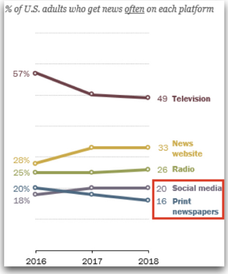
ECE Department, University of Michigan

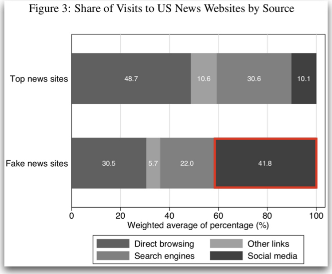
Motivation

Social learning in the presence of malicious agents

Most prominent example: fake news on social networks







[Shearer, Gottfried 2017] [Shearer 2018] [Allcott, Gentzkow 2017]

Overview

Salient features:

1. Simultaneous consumption/discussion of news
2. Legitimate news partially reveals "truth"
3. Fake news more likely in "echo chambers"

We analyze model incorporating these features:

1. Agents receive signals/share beliefs about true state θ
2. Regular agents: signals $=$ noisy observations of θ
3. Stubborn agents: signals uncorrelated with θ; ignore others' beliefs

Main questions:

- Do stubborn agents prevent regular agents from learning θ?
- How can stubborn agents maximize influence?

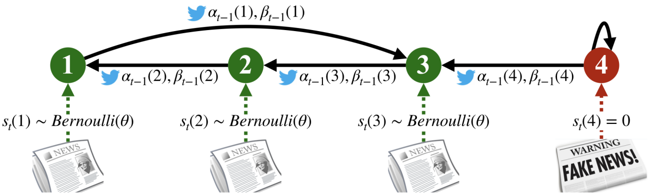
Learning model (basic ingredients)

True state $\theta \in (0, 1)$, (regular) agents A, stubborn agents/bots B

Signals at time t: $s_t(i) \sim \text{Bernoulli}(\theta)$ for $i \in A$, $s_t(i) = 0$ for $i \in B$

Beliefs at time t: $\text{Beta}(\alpha_t(i), \beta_t(i))$ for $i \in A \cup B$

If $j \to i$ in graph, i observes $\alpha_{t-1}(j), \beta_{t-1}(j)$ at t; $i \in B$ has only self-loop



Learning model (belief updates)

How should i use signal $s_t(i)$ + neighbor parameters $\{\alpha_{t-1}(j), \beta_{t-1}(j) : j \to i\}$?

We adopt non-Bayesian model similar to [Jadbabaie et al. 2012]

Bayesian update using signal, then average with neighbors in graph:

$$\alpha_t(i) = (1-\eta)(\alpha_{t-1}(i) + s_t(i)) + \frac{\eta}{d_{in}(i)} \sum_{j \in A \cup B : j \to i} \alpha_{t-1}(j)$$

$$\beta_t(i) = (1-\eta)(\beta_{t-1}(i) + 1 - s_t(i)) + \frac{\eta}{d_{in}(i)} \sum_{j \in A \cup B : j \to i} \beta_{t-1}(j)$$

Quantity of interest:

$$\theta_t(i) = \mathbb{E}\left[\mathrm{Beta}(\alpha_t(i), \beta_t(i))\right] = \frac{\alpha_t(i)}{\alpha_t(i) + \beta_t(i)}$$

(View as summary statistic of i's belief/opinion at t)

Learning horizon

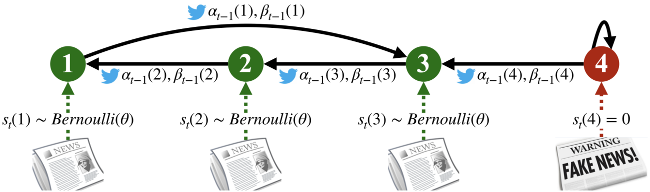
As learning horizon (i.e. number belief updates) grows ...

- ...agents receive more unbiased observations
- ...influence of bots spreads

Learning horizon plays important, but non-obvious role

Difficult to analyze finite horizon for fixed graph

- Will consider sequence $\{G_n\}_{n \in \mathbb{N}}$ of random graphs, where G_n has n agents
- Will consider horizon $T_n \in \mathbb{N}$ for G_n (finite for each finite n)

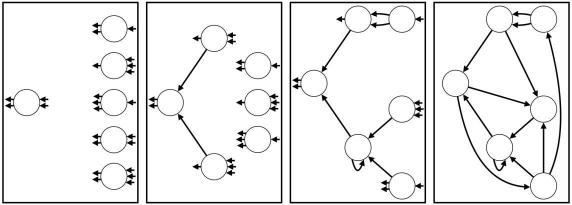


Graph model

1. Realize $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i=1}^n$ satisfying

$$d_{out}(i) \in \mathbb{N},\ d_{in}^A(i) \in \mathbb{N},\ d_{in}^B(i) \in \mathbb{Z}_+,\ \sum_{i=1}^n d_{out}(i) = \sum_{i=1}^n d_{in}^A(i)\ a.s.$$

2. From $\{d_{out}(i), d_{in}^A(i)\}_{i=1}^n$, construct sub-graph with nodes $A = \{1, \dots, n\}$ via *directed configuration model* [Chen, Olvera-Cravioto 2013]



3. Connect $d_{in}^B(i)$ bots (with only self-loop) to each $i \in A$

Here bot connections $\{d_{in}^B(i)\}_{i=1}^n$ given; later, will consider optimal connections

Assumptions

Key random variable: "density" of (regular) agents, measured as

$$\tilde{p}_n = \sum_{i=1}^{n} \underbrace{\frac{d_{in}^A(i)}{d_{in}^A(i) + d_{in}^B(i)}}_{\text{Fraction in-neighbors trying to learn}} \times \underbrace{\frac{d_{out}(i)}{\sum_{j=1}^{n} d_{out}(j)}}_{\text{Sample w.r.t. out-degree distribution}}$$

Assumption 1 (for belief convergence):

- $\lim_{n\to\infty} \mathbb{P}(|\tilde{p}_n - p_n| > \delta_n) = 0$ for some $\{p_n\}_{n\in\mathbb{N}}, \{\delta_n\}_{n\in\mathbb{N}} \subset (0,1)$ s.t. $\lim_{n\to\infty} \delta_n = 0$
- $\lim_{n\to\infty} T_n = \infty$

Assumption 2 (for branching process approximation):

- Sparse degrees (finite mean/variance) with high probability
- $T_n = O(\log n)$

$\Rightarrow$ Guarantees $\theta_{T_n}(i)$ depends on $o(n)$ other agents ("local" learning)

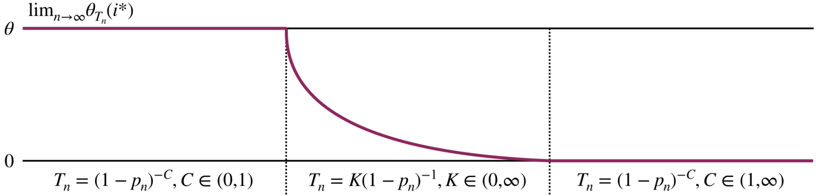
Main result

Theorem

Given assumptions, we have for $i^* \sim \{1, \ldots, n\}$ uniformly,

$$\theta_{T_n}(i^*) \xrightarrow[n\to\infty]{\mathbb{P}} \begin{cases} \theta, & T_n(1-p_n) \xrightarrow[n\to\infty]{} 0 \\ \theta(1-e^{-K\eta})/(K\eta), & T_n(1-p_n) \xrightarrow[n\to\infty]{} K \in (0,\infty) \\ 0, & T_n(1-p_n) \xrightarrow[n\to\infty]{} \infty \end{cases}.$$

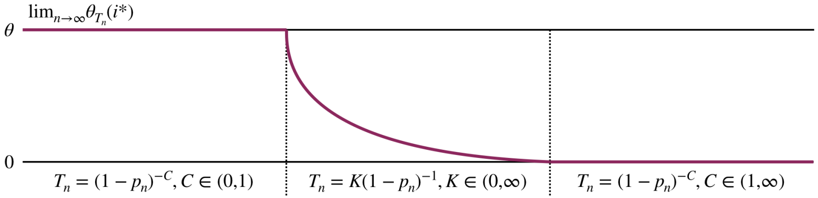
Illustration, assuming T_n, p_n related as $T_n \propto (1-p_n)^{-C}$:



Remarks on main result

Again assuming T_n, p_n related as $T_n \propto (1 - p_n)^{-C}$:

1. *Phase transition* occurs (small change to $C \approx 1 \Rightarrow$ big change belief)
2. For fixed p_n, agents initially (at small T_n) learn, later (at large T_n) forget!
3. For fixed $T_n \propto (1 - p_n)^{-1}$, bots experience "diminishing returns"
4. When $T_n(1 - p_n) \to K \in (0, \infty)$, limiting belief $= \theta(1 - e^{-K\eta})/(K\eta)$:
   - As $\eta \to 0$, agents ignore network, belief $\to \theta$
   - As $\eta \to 1$, belief $\to \theta(1 - e^{-K})/K$ (not $\to 0$, "discontinuity")



Special case

If $p_n \to p < 1$ (i.e. bots non-vanishing), stronger result holds:

Theorem

Suppose $p_n \to p \in (0,1)$, so that $\theta_{T_n}(i^*) \to 0$ in $\mathbb{P}$.
Then, under slightly stronger assumptions, and for any $\epsilon > 0$,

$$|\{i \in A : \theta_{T_n}(i) > \epsilon\}| = o(n) \textit{ with high probability as } n \to \infty.$$

"Slightly stronger assumptions":

- $T_n = \Omega(\log n)$ (instead of just $T_n \to \infty$)
- Minimum rates of convergence for "with high probability" statements

Key ideas of proof (1/2)

Recall parameter updates:

$$\alpha_t(i) = (1-\eta)(\alpha_{t-1}(i) + s_t(i)) + \frac{\eta}{d_{in}(i)} \sum_{j \in A \cup B : j \to i} \alpha_{t-1}(j) \tag{1}$$

$$\beta_t(i) = (1-\eta)(\beta_{t-1}(i) + 1 - s_t(i)) + \frac{\eta}{d_{in}(i)} \sum_{j \in A \cup B : j \to i} \beta_{t-1}(j) \tag{2}$$

Assume $\alpha_0(j) = \beta_0(j) = o(T_n)\ \forall\ j$ and define

- $P =$ column-normalized adjacency matrix
- $e_i =$ unit vector in i-th direction

Then iterating (1)-(2) yields

$$\theta_{T_n}(i) = \frac{1}{T_n} \sum_{\tau=0}^{t-1} s_{t-\tau} \left(\eta P + (1-\eta) I\right)^{\tau} e_i + o(1)$$

Interpretation: take Uniform($\{1, \ldots, T_n\}$)-length lazy random walk from i, sample signal of node reached

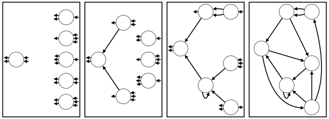
Key ideas of proof (2/2)

Previous slide: interpret $\theta_{T_n}(i)$ in terms of lazy random walk (LRW)

Bots are absorbing states on this LRW (owing to self-loops)

To analyze beliefs, analyze absorption probabilities

LRW and breadth-first-search graph construction can be done simultaneously



By $T_n = O(\log n)$ and sparsity, LRW explores tree-like sub-graph before horizon

Reduces random process on random graph to much simpler process
(simultaneous construction of tree / computation of absorption probabilities)

Formulation

Previously assumed $\{d_{out}(i), d_{in}^A(i), d_{in}^B(i)\}_{i=1}^n$ given

Now suppose $\{d_{out}(i), d_{in}^A(i)\}_{i=1}^n$ given, adversary chooses $\{d_{in}^B(i)\}_{i=1}^n$

By main result, adversary (with budget $b \in \mathbb{N}$) should solve

$$\min_{\{d_{in}^B(i)\}_{i=1}^n \in \mathbb{Z}_+^n} \underbrace{\sum_{i=1}^n \frac{d_{in}^A(i)}{d_{in}^A(i) + d_{in}^B(i)} \frac{d_{out}(i)}{\sum_{j=1}^n d_{out}(j)}}_{\text{Key random variable } \tilde{p}_n \text{ shown previously}} \text{ s.t. } \sum_{i=1}^n d_{in}^B(i) \leq b$$

Integer program (IP), so we devise approximation scheme

Approximation scheme

Independently attach each bot to i-th agent with probability proportional to

$$\max\left\{ d_{in}^{A}(i) \left(\sqrt{\lambda^* \frac{d_{out}(i)}{d_{in}^{A}(i)}} - 1 \right), 0 \right\} \qquad (3)$$

- (3) is solution to LP relaxation of IP; $\lambda^* > 0$ is efficiently computable
- Intuition: bots want to connect to i-th agent only if $\frac{d_{out}(i)}{d_{in}^{A}(i)} \geq \frac{1}{\lambda^*}$, i.e. only if i is influential ($d_{out}(i)$ large) + susceptible to influence ($d_{in}^{A}(i)$ small)

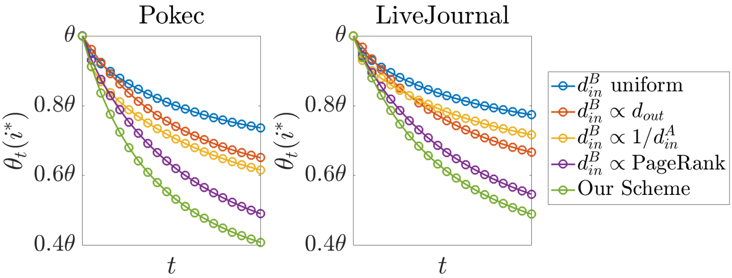
Theorem

For any $\delta > 0$, scheme gives $(2+\delta)$-approximation with high probability, i.e.

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\textit{objective for approximation scheme}}{\textit{objective for optimal scheme}} > 2 + \delta \right) = 0.$$

Empirical performance

For real social networks, our approximation scheme outperforms heuristics, even those using network structure



(Networks from [*SNAP Datasets: Stanford Large Network Dataset Collection*])

Ultimately, new insights into vulnerabilities of social networks

Most similar models in literature

[Azzimonti, Fernandes 2018]

- (Almost) same belief update (minor differences to bot behavior)
- Only empirical results (allows for richer model, e.g. time-varying graph)

[Jadbabaie et al. 2012]

- Communicate distributions, not parameters, i.e.

$$\mu_t(i) = \eta_{ii} \mathrm{BU}(\mu_{t-1}(i), s_t(i)) + \sum_{j \neq i} \eta_{ji} \mu_{t-1}(j)$$

  where μ terms are distributions, $\sum_j \eta_{ji} = 1$, BU = "Bayesian update"
- Richer belief update, but stronger assumptions:
  1. Fixed, strongly-connected graph
  2. Infinite horizon
  3. No stubborn agents

Other relevant works

View our model as perturbation of classical deGroot model [DeGroot 1974]:

$$\theta_t = \theta_{t-1} W \text{ where } \theta_t, \theta_{t-1} \in \mathbb{R}^n \text{ and } W \text{ is column-stochastic}$$

Extensively studied, see surveys [Acemoglu, Ozdaglar 2011; Golub, Sadler 2017]

[Rahimian, Shahrampour, Jadbabaie 2015]: adopt belief of random neighbor, also relates to random walk (but need strong connectedness + infinite horizon)

[Acemoglu, Ozdaglar, ParandehGheibi 2010]: "forceful" but not fully-stubborn agents $\Rightarrow$ no absorbing states $\Rightarrow$ can use stationarity distribution

Stubborn agents have been considered in consensus setting, but infinite horizon typically assumed, e.g. [Acemoglu et al. 2011; Ghaderi, Srikant 2014]

References I

Acemoglu, Daron, Asuman Ozdaglar (2011). "Opinion dynamics and learning in social networks". In: *Dynamic Games and Applications* 1.1, pp. 3–49.

Acemoglu, Daron, Asuman Ozdaglar, Ali ParandehGheibi (2010). "Spread of (mis) information in social networks". In: *Games and Economic Behavior* 70.2, pp. 194–227.

Acemoglu, Daron et al. (2011). "Opinion fluctuations and persistent disagreement in social networks". In: *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, pp. 2347–2352.

Allcott, Hunt, Matthew Gentzkow (2017). "Social media and fake news in the 2016 election". In: *Journal of Economic Perspectives* 31.2, pp. 211–36.

Azzimonti, Marina, Marcos Fernandes (2018). *Social media networks, fake news, and polarization*. Tech. rep. National Bureau of Economic Research.

Chen, Ningyuan, Mariana Olvera-Cravioto (2013). "Directed random graphs with given degree distributions". In: *Stochastic Systems* 3.1, pp. 147–186.

DeGroot, Morris H (1974). "Reaching a consensus". In: *Journal of the American Statistical Association* 69.345, pp. 118–121.

Ghaderi, Javad, R Srikant (2014). "Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate". In: *Automatica* 50.12, pp. 3209–3215.

Golub, Benjamin, Evan Sadler (2017). "Learning in social networks". In:

Jadbabaie, Ali et al. (2012). "Non-Bayesian social learning". In: *Games and Economic Behavior* 76.1, pp. 210–225.

Leskovec, Jure, Andrej Krevl. *SNAP Datasets: Stanford Large Network Dataset Collection*. http://snap.stanford.edu/data.

References II

Rahimian, Mohammad Amin, Shahin Shahrampour, Ali Jadbabaie (2015). "Learning without recall by random walks on directed graphs". In: *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, pp. 5538–5543.

Shearer, Elisa (2018). "Social media outpaces print newspapers in the US as a news source". In: *Pew Research Center* 10.

Shearer, Elisa, Jeffrey Gottfried (2017). "News use across social media platforms 2017". In: *Pew Research Center, Journalism and Media*.