6.207J/14.15J Networks, Spring 2022

Problems on Social Media Networks

Problem 1: Sheep or Narcissist?

Consider the network manipulation model of Mostagir, Ozdaglar, and Siderius (2022). In this exercise, we look at how a DeGroot learner might calibrate the weight to assign to her own news as opposed to the news and opinions of those in her social media network.

Suppose there is a single knowledgeable agent (node 1) and two DeGroot agents (nodes 2 and 3). Recall that the knowledgeable agent has belief $\pi_{1,t} = 0$ for all t of the incorrect state, whereas the two DeGroot agents update their beliefs $\pi_{i,t+1}$ of the incorrect state according to:

$$\begin{cases} \pi_{2,t+1} = \theta_2\gamma_2 + (1-\theta_2)\pi_{3,t}/2 \\ \pi_{3,t+1} = \theta_3\gamma_3 + (1-\theta_3)\pi_{2,t}/2 \end{cases} \tag{1}$$

where $\gamma_i = 1$ if the principal spams agent i with disinformation and $\gamma_i = 0$ if not. These network learning dynamics correspond to a complete network of three nodes where all other agents' beliefs are weighted equally. Recall that social learning accounts for $1-\theta_i$ proportion of the learning whereas learning from one's own news accounts for θ_i proportion.

We consider an extension of this model where DeGroot agents are still boundedly rational and must comply with Equation (1), but can strategically choose θ_i to try and avoid manipulation. Formally, consider the following game:

1. At $\tau = 1$, DeGroot agents strategically choose θ_2 and θ_3 simultaneously;
2. At $\tau = 2$, the principal elects to target agent 2 and/or agent 3 (i.e., γ_2 and γ_3), for which the principal pays $\varepsilon(\gamma_2 + \gamma_3)$ (i.e., the principal pays ε for each agent he targets);
3. At $\tau = 3$, belief updating occurs over an infinite horizon as in the baseline model;
4. At $\tau = 4$, DeGroot agent i is manipulated if $\lim_{t\to\infty} \pi_{i,t} > 0.1$, in which case the agent receives utility 0 and the principal receives utility 1. If $\lim_{t\to\infty} \pi_{i,t} < 0.1$, the agent receives utility 1 and the principal receives utility 0.

a. Taking $\boldsymbol{\theta} \equiv (\theta_2, \theta_3)$ and $\boldsymbol{\gamma} \equiv (\gamma_2, \gamma_3)$ as given, solve for $\lim_{t\to\infty} \pi_{2,t}$ and $\lim_{t\to\infty} \pi_{3,t}$.

b. Show that $\boldsymbol{\theta} = \mathbf{0}$ is a pure-strategy Nash equilibrium where no agent is manipulated.

c. Suppose $1 < \varepsilon < 2$. Show that $\boldsymbol{\theta} = \mathbf{1}$ is also a pure-strategy equilibrium where no agents are manipulated, but that $\boldsymbol{\theta} = (1, 0)$ and $\boldsymbol{\theta} = (0, 1)$ are not equilibria. Conclude that neither $\theta_i = 0$ nor $\theta_i = 1$ are (weakly) dominant strategies.

d. Under the same conditions of (c), show that other pure-strategy Nash equilibria exist for intermediate values of $\theta^* \in (0, 1)$ where $\boldsymbol{\theta} = (\theta^*, \theta^*)$, but where both agents are manipulated.

e. Interpret the results of (b), (c), and (d) in terms of how DeGroots should weight their own news feeds as opposed to the opinions of their peers on social media. Are all such equilibria equally resistant to manipulation?

Problem 2: Censorship Policy Backfire

Consider the online misinformation model of Acemoglu, Ozdaglar, and Siderius (2022). In this exercise, we formally show that a regulator who censors only a fraction of misinformation may make the overall spread of misinformation worse.

Let us consider two islands of equal size $N/2$: a left-wing island where all agents have prior belief $b_i = 1/3$ (that $\theta = R$) and a right-wing island where all agents have prior belief $b_i = 2/3$ (that $\theta = R$). There is an article of reliability score r with likelihood of being truthful $\phi(r)$ and likelihood of being misinformation $1-\phi(r)$. Recall that an article that is truthful produces a message $m = \theta$ with probability $p > 1/2$ and an article that contains misinformation produces a message $m = \theta$ with probability $q \leq 1/2$. Let π_i be the posterior belief of agent i that the article is truthful (does not contain misinformation) given reliability score r and right-wing message $m = R$.

a. Write the expression for π_i.

   Suppose each agent i decides between taking the action $a_i = \mathcal{S}$ ("share") and $a_i = \mathcal{D}$ ("dislike"). Take the payoff from sharing to be $U_i = (2\pi_i - 1) + (S_i - D_i)/(2N)$, where S_i is the number of re-shares after agent i's share and D_i is the number of dislikes after agent i's share. Take the payoff from disliking to be $1 - \pi_i$.[1] Moreover, assume all articles are equally likely to be truthful or contain misinformation ex-ante (i.e., $\phi(r) = 1 - \phi(r) = 1/2$) and that $p = 1$ and $q = 0$.

b. Argue that the optimal platform algorithm (to maximize total shares) is a two-island homophily model with $p_s = 1$ and $p_d = 0$ and where the article is initially recommended to a (seed) agent on the right-wing island.

c. Now, suppose the regulator removes δ fraction of the misinformation in circulation (in other words, if an article contains misinformation it is removed with probability δ and remains with probability $1 - \delta$). Write the expression for π_i as in (a), assuming this censorship policy is common knowledge.

d. Show that if $1/3 < \delta < 1/2$, the optimal platform algorithm is the complete network (in other words, a single-island model where all links exist).

e. Conclude that misinformation spreads more (on average) under a censorship policy with $1/3 < \delta < 1/2$ than under no censorship policy ($\delta = 0$).

[1] Observe that under this formulation, "ignore" is always dominated by "dislike" so indeed the agent only decides between "share" and "dislike".