Learning in Networks with a Development Application

May 8, 2012

Introduction

Overview

i. Overview of the Question
   a. Development Puzzles
   b. Network Learning
ii. Networks
   a. Representing Networks
   b. Bayesian Learning
   c. DeGroot Learning
iii. Networks in Development
   a. Theory
   b. Experiment

Puzzles

- ⋆ Productive investments remain unmade, even when the returns from the technology are easily observable and high, e.g.,
  - Irrigation pumps in India (Rosenzweig/Wolpin, *JPE*, 1993)
  - Pineapple and fertilizer use (Conley/Udry, *AER*, 2010)
  - De-worming pills unpurchased (Miguel/Kremer, *Econometrica* 2004)
  - Anti-malarial mosquito nets unpurchased (Banerjee/Duflo, 2011)
  - Water-purifying chlorine pills unpurchased (ibid.)
- ⋆ cf. Conley and Udry, "Social Learning Through Networks." People don't seem to talk to one another. Can we contribute to the solution of the above problems with a very simple intervention?
- ⋆ I suggest that we can do empirical studies of network learning in studying the development applications *and* that we can learn something about these development problems by studying the learning and information sharing which happen in a network context.

Network Learning Puzzles

The General Problem: Agents in a network must figure out the unknown state of the world based on their own signals and the actions of their neighbors.

Two basic frameworks for learning on networks:

i. Bayesian - Agents on graphs make optimal decisions about the unobserved state of the world by observing the payoffs received by their neighbors.

ii. DeGroot - From Morris DeGroot's (1974) "Reaching a Consensus" (J. Am. Stat. Assoc.). Agents have fixed weights on the information they observe from their neighbors and from themselves (details to follow). They update sequentially, but keep the weights on the agents they observe constant.

Basic Network Facts

A Very, Very Brief Introduction

Definition

A *network* $\Gamma = (N, E)$ is a collection of nodes $N = \{1, 2, \ldots, m\}$ connected by edges $E = \{e_{ij} | i, j = 1, \ldots, m\}$ where we say that i is connected to j if $e_{ij} > 0$. [1]

Remark

There is a basic distinction between a directed and undirected network, in that we might have $e_{ij} > 0 \Rightarrow e_{ji} > 0$ (undirected) or not (directed).

Remark

The structure of the graph can be summarized in a $m \times m$ matrix $\mathbf{T}$, where $[\mathbf{T}_{ij}] = e_{ij}$.

[1] This is basically just a graph.

Basic Network Facts

Walks, Paths and Cycles[2]

Definition

A walk in $\mathbf{\Gamma}$ is a sequence of nodes $i_1, i_2, \ldots, i_k$ not necessarily distinct such that $\mathbf{T}_{i_k, i_{k+1}} > 0$ for all $k \in \{1, 2, \ldots, K-1\}$. A path $\mathbf{P}$ has distinct nodes.

Definition

A cycle is a walk $i_1, \ldots, i_k$ such that $i_1 = i_k$. The length of a cycle is $k-1$. A cycle is simple if the only node appearing twice is i_1.

Definition

A matrix $\mathbf{T}$ is strongly connected if there is a path $\mathbf{P}$ in $\mathbf{T}$ from any node to any other node.

[2] Golub/Jackson 2010

Basic Network Facts

Paths, Connectedness, &c.

Definition

$\mathbf{T}$ is aperiodic if the greatest common divisor of all *directed* cycle lengths is 1

Other useful network properties:

i. Degree of a node: number of links to i
ii. Components: $\Gamma' = (N', E')$ such that $N' \subset N, E' \subset E$ and Γ' is connected and $i \in N', e_{ij} \neq 0 \Rightarrow j \in N'$ and $e_{ij} \in E'$
iii. Centrality: Degree, Closeness, Eigenvalue, Betweenness
iv. Closed: There is no link from any agent in $C \subset N$ to any agent not in C. That is, $\nexists i \in C, j \notin C$ with $\mathbf{T}_{ij} > 0$

Learning Models

Bayesian[3]

Network learning in a Bayesian context - *General features*:

i. **Strategic Considerations**: assuming representative agents at each node in the network (i.e., strategic considerations ignored).

ii. **Uncertainty**: represented as $(\Omega, \mathcal{F}, P)$ where Ω is a compact metric space, $\mathcal{F}$ is a σ-field and P is a probability measure.

iii. **Actions**: $\mathbf{A} \subset \mathbb{R}$

iv. **Signals**: Each agent i receives a signal $\sigma_i(\omega)$, where $\omega \in \Omega$.

v. **Time**: $t = 1, 2, \ldots \in \mathbb{N}$

vi. **Network**: Agents $i = 1, 2, \ldots, n$ represented by a family of sets $N_i, i = 1, \ldots, n$ such that $N_i \subseteq \{1, 2, \ldots, i-1, i+1, \ldots, n\}$ (neighbors observed by i).

[3] Gale/Kariv 2003

vii. **Utility**: Given by $U : \mathbf{A} \times \Omega \to \mathbb{R}$, bounded and measurable

viii. **Stategy**: Maximize $U(a_{it}, \omega)$ by choice of a_{it}. More formally, strategy is a random variable $X_{it}(\omega), \mathcal{F}_{it}/\mathbf{A}$ measurable

ix. **Information**: $\mathcal{F}_{it} = \sigma \langle \sigma_i, \{X_{js}, s \le t-1\} \rangle$

Definition

A weak perfect Bayesian equilibrium consists of a sequence of random variables $\{X_{it}\}$ and σ-fields $\{\mathcal{F}_{it}\}$ such that for each $i = 1, \ldots, n$ and $t = 1, 2, \ldots$ we have

a. $X_{it} : \Omega \to \mathbf{A}$ is $\mathcal{F}_{it}$ measurable.

b. $\mathcal{F}_{it} = \sigma \langle \sigma_i, \{X_{js}, s \le t-1\} \rangle$

c. $\mathbb{E}[U(x(\omega), \omega)] \le \mathbb{E}[U(X_{it}(\omega), \omega)]$ for any $\mathcal{F}_{it}$ measurable function $x : \Omega \to \mathbf{A}$

[4] Gale/Kariv 2003

Learning Models

Bayesian[5]

Theorem

Let $\{X_{it}, \mathcal{F}_{it}, i = 1, \ldots, n, t = 1, 2, \ldots\}$ be an equilibrium. For each i, define $V_{it}^* : \Omega \to \mathbb{R}$ by

$$V_{it}^* = \mathbb{E}[U(X_{it}, \bullet)|\mathcal{F}_{it}]$$

Then V_{it}^* is a submartingale with respect to $\mathcal{F}_{it}$ and there exists a random variable $V_{i,\infty}^*$ such that V_{it}^* converges to $V_{i,\infty}^*$ almost surely.

Corollary

The Imitation Principle - Let $\{X_{it}, \mathcal{F}_{it}\}$ be as above and let V_{it}^* be the equilibrium payoff. Then for any $j \in N_i$ and any t, $V_{it}^* \geq \mathbb{E}[V_{j,t-1}|\mathcal{F}_{it}]$. Futhermore, in the limit, $V_{i,\infty}^* \geq \mathbb{E}[V_{j,\infty}^*|\mathcal{F}_{i,\infty}]$, where $\mathcal{F}_{i,\infty}$ is the σ-field generated by $\cup_{t=1}^{\infty}\mathcal{F}_{it}$

Corollary

Let $\{X_{it}, \mathcal{F}_{it}\}$ be the equilibrium above and let V_{it}^* be the equilibrium payoffs. Then if $j \in N_i$ and j is connected to i, then $V_{i,\infty}^* = \mathbb{E}[V_{j,\infty}^*|\mathcal{F}_{i,\infty}]$

[5] Gale/Kariv 2003

Learning Models

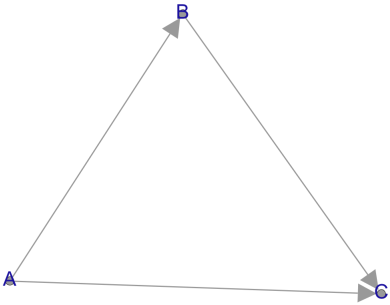
Bayesian Example

Comments

i. Less *analytically* tractable in larger networks than in smaller networks

ii. Why? Suppose we have the following network

⋆ $A \to B \to C$ and $A \to C$.



Learning Models

Bayesian Example

- ⇒ A learns from B and B learns from C. Thus in observing B, A must try to infer what B learned from C, whom A does not directly observe.
- ⇒ This is an issue of higher-order beliefs: "What does B believe about C?" and inference based on such beliefs
- ⇒ Gets worse with $A_1 \to A_2 \to \ldots \to A_n$

iii. There may be issues of computational complexity as well, but this is not my area: cf. Gregory Cooper, "The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks." *Artificial Intelligence* (1990)[6].

[6]cited in Chandrasekhar, Larreguy and Xandri (2011)

Learning Models

iv. Another complication - the varieties of informational incompleteness are numerous, e.g., let b denote the payoff of B, then

   a. A observes b exactly
   b. A observes $b + \epsilon$, for $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$
   c. A observes $b \geq \bar{c}$ or $b < \bar{c}$ for some constant $\bar{c}$
   d. We can imagine many more ways for A to learn from B.

   Some of these do not have solutions in closed form.

v. Furthermore, though we have convergence in payoffs in a connected network, we do not know what the *rate* of convergence is.

i. Each agent j (of m total agents) has a weight $p_{ji}, i = 1, \ldots, m$ where $\sum_{i=1}^{m} p_{ji} = 1$. We can think about this as the degree to which j trusts i, for example. These weights form a matrix:

$$T = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

ii. Each agent has some initial belief q_i^0 so we form the vector of initial beliefs $\mathbf{q^0} = (q_1, q_2, \ldots, q_m)'$.

iii. Beliefs evolve over time according to $\mathbf{q}^{(t+1)} = T\mathbf{q}^{(t)} = T^t\mathbf{q}^0$

i. Yes, $\mathbf{T}$ is a Markov transition matrix.

ii. This weighting scheme is not Bayesian, because the p_{ij} terms are constant. For example: agent i may learn from agent j in period t, but ignores what agent j may have learned from agent i in periods $t-k, k=1,2,\ldots$.

iii. This can lead to over-weighting of information which has already been processed.

iv. As a worst-case, let $\mathbf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then for any $\mathbf{q}^0 = (q_1^0, q_2^0)'$, we see that 1 and 2 just exchange beliefs forever.

Learning Models

DeGroot Mode Example

i. As another, more informative, example, consider the matrix $\mathbf{T}$:

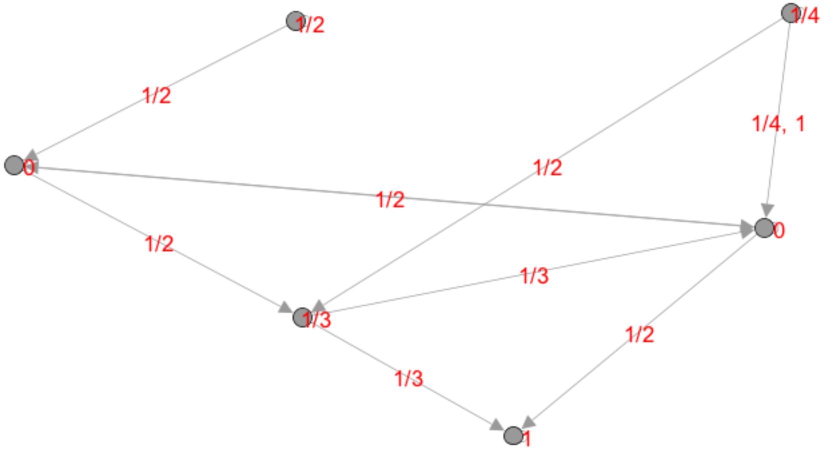
$$\begin{bmatrix} \text{Severine} \\ \text{Jan} \\ \text{Austin} \\ \text{Tobias} \\ \text{Joe} \\ \text{Maureen} \end{bmatrix} \overbrace{\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 1/4 & 0 & 1/4 \end{bmatrix}}^{\begin{bmatrix} Sev. & Jan & Aus. & Tob. & Joe & Mau. \end{bmatrix}}$$

ii. Some proxy for this information is not terribly hard to elicit in the field.

Learning Models

DeGroot Model Example Continued

With implied network structure:



Definition

A matrix $\mathbf{T}$ is convergent if $\lim_{t\to\infty} \mathbf{T}^t\mathbf{q}$ exists for all $\mathbf{q} \in [0,1]^m$

Definition

A group of agents $C \subset N$ reaches consensus for an initial vector of beliefs $\mathbf{q}^0$ if $\lim_{t\to\infty} q_i^t = \lim_{t\to\infty} q_j^t$ for all $i, j \in N$

Theorem

A matrix $\mathbf{T}$ is convergent if and only if every set of nodes that is strongly connected and closed is aperiodic.

Corollary

i. *A strongly connected and closed group reaches a consensus if and only if it is aperiodic.*

ii. *A consensus is reached if and only if there exists some t such that some column of $\mathbf{T}^t$ has all positive entries.*

[7] Golub/Jackson 2010 and Jackson 2008

Learning Models

i. Language is slightly different in the two cases, but the outcome is similar: convergence in payoff or convergence in beliefs *given* a connected network structure.

ii. In the development context, we should expect to see something similar, given a suitable network structure.

iii. So is there something wrong with the network structure? Other properties may matter also, but this is one working hypothesis (cf. Conley/Udry below).

iv. If connectedness is what is wrong, maybe it is easy to fix.

Development Puzzles

Conley and Udry

Conley and Udry, "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana." (Amer. J. Agricultural Econ.)

i. Farmers in Ghana switched from growing maize/cassava to pineapple (for export)
ii. We would hope that they would switch to the new technology after observing the payoffs to their neighbors from experimenting
iii. We would be wrong!
    a. Farmers don't know very much about the payoffs of their neighbors ("X had a y harvest" for $y \in \{$ great, average, terrible$\}$)
    b. Farmers know less about the inputs their neighbors use!
    c. Network is not connected in the relevant sense if people do not talk

Development Puzzles

A Field Experiment with Two Parts

For any of the puzzles listed above: high-yielding seeds, irrigation pumps, fertilizer use, de-worming pills, anti-malarial nets, chlorine pills, etc.

i. Identify the flows of information in the environment: what information is shared, with which individuals, etc.

ii. Then
   a. [D⇒N] Characterize the structure of the network and the flow of information through it, *or*
   b. [N⇒D] Discover that there is insufficient flow of information or the network lacks some important property

iii. Track how observable outcomes change over time as information flows:
   a. [D⇒N] How far is what is observed from what is predicted by either the Bayesian or DeGroot models?
   b. [N⇒D] How has changing the network structure impacted the development problem, if it has had any effect at all?

Development Puzzles

Progress

i. [Networks $\Rightarrow$ Development] Theory suggests that mere communication should help, given the right network environment: cf. the convergence-to-consensus condition for the DeGroot model. This is (potentially) a very cheap field experiment.

ii. [Development $\Rightarrow$ Networks] Departures from the theory in the distribution of payoffs over the long term can suggest where the theory differs from reality.

   a. Bayesian or DeGroot? Other? Or does it depend?
   b. As far as solving the puzzle, convergence happens in both cases, so we should improve the outcome by facilitating information sharing among neighbors, provided we get the network right.

Literature

On Network Learning

Basic Reference: Matthew Jackson, Social and Economic Networks. Princeton (Princeton, NJ). 2008.

a. Venkatesh Bala and Sanjeev Goyal, "Learning from Neighbors." *Review of Economic Studies*, 1998.

b. Manuel Mueller-Frank, "A General Framework for Rational Learning in Social Networks." *Oxford University working paper*, 2011.

c. Doulgas Gale and Shachar Kariv, "Bayesian Learning in Social Networks." *Games and Economic Behavior*, 2003

d. Daron Acemoglu, Munther Dahleh, Ilan Lobel, Asuman Ozdaglar, "Bayesian Learning in Social Networks." *Review of Economic Studies*, 2010.

e. Ben Golub and Matthew Jackson, "Naive Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal*, 2010. (DeGroot Model.)

f. Arun Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri, "Testing Models of Social Learning on Networks: Evidence from a Framed Field Experiment." *MIT working paper*, 2011.

Literature

On Development Puzzles

a. Mark Rosenzweig and Kenneth Wolpin, "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India." *Journal of Political Economy*, 1993 (On water pumps.)

b. Edward Miguel and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 2004 (On deworming pills.)

c. Abhijit Banerjee and Esther Duflo, Poor Economics. PublicAffairs (New York). 2011. (On bednets and chlorine pills, cf. Chapter 3.)

d. Timothy Conley and Christopher Udry, "Learning about a New Technology: Pineapple in Ghana." *AER*, 2010.

e. Andrew Foster and Mark Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *JPE*, 1995. (On the use of high-yielding varieties of seed.)