

1.022 Introduction to Network Models

Amir Ajorlou

Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

Lectures 18 and 19

- ▶ Bayes' Rule was a crucial conceptual ingredient in the first generation of e-mail spam filters,
- ▶ Suppose that you receive a piece of e-mail whose subject line contains **“check this out”** (a popular phrase among spammers)
- ▶ What is the chance the message is spam? (without looking at the sender or the message content)

- ▶ This is a question about conditional probability:

$$\Pr(\text{spam} | \text{"check this out"})$$

- ▶ To determine this value, we need to know some facts:
 - ▶ 40% of all your e-mail is spam
 - ▶ 1% of all spam messages contain the phrase "check this out"
 - ▶ 0.4% of all non-spam messages contain this phrase

- ▶ Writing these in terms of probabilities:

$$\Pr(\text{spam}) = 0.4$$

$$\Pr(\text{"check this out"}|\text{spam}) = 0.01$$

$$\Pr(\text{"check this out"}|\text{not spam}) = 0.004$$

$$\begin{aligned}\Pr(\text{"check this out"}) &= \Pr(\text{spam}) \cdot \Pr(\text{"check this out"}|\text{spam}) + \\ &\quad \Pr(\text{not spam}) \cdot \Pr(\text{"check this out"}|\text{not spam}) \\ &= 0.4 \times 0.01 + 0.6 \times 0.004 = 0.0064\end{aligned}$$

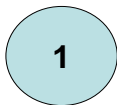
⇒ Use Bayes' Rule to write

$$\begin{aligned}\Pr(\text{spam}|\text{"check this out"}) &= \frac{\Pr(\text{"check this out"}|\text{spam})\Pr(\text{spam})}{\Pr(\text{"check this out"})}\Pr(\text{spam}) \\ &= \frac{0.004}{0.0064} = \frac{5}{8} = 0.625\end{aligned}$$

- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{\textit{Chinese}, \textit{Indian}\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others

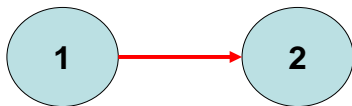
- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{\textit{Chinese}, \textit{Indian}\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others
- ▶ **Realization:** Assume $\theta = \textit{Indian}$

- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{\textit{Chinese}, \textit{Indian}\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others
- ▶ **Realization:** Assume $\theta = \textit{Indian}$
 - ▶ Agent 1 arrives. Her signal indicates 'Chinese'.
 - ▶ She chooses to have a Chinese dinner



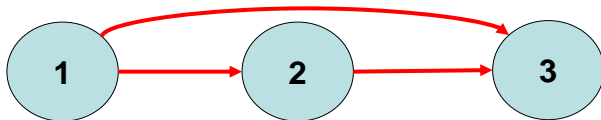
Signal = 'Chinese'
Decision = 'Chinese'

- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{\text{Chinese}, \text{Indian}\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others
- ▶ **Realization:** Assume $\theta = \text{Indian}$
 - ▶ Agent 2 arrives. His signal indicates 'Chinese'.
 - ▶ He also chooses to eat Chinese food



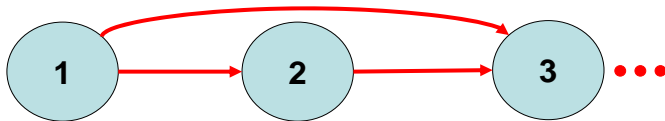
Decision = 'Chinese' **Signal = 'Chinese'**
Decision = 'Chinese'

- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{Chinese, Indian\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others
- ▶ **Realization:** Assume $\theta = Indian$
 - ▶ Agent 3 arrives. Her signal indicates 'Indian'.
 - ▶ She disregards her signal and copies the decisions of agents 1 and 2.



Decision = 'Chinese' Decision = 'Chinese' Signal = 'Indian' Decision = 'Chinese'

- ▶ Individuals (hereafter called agents) arrive in town sequentially and choose to dine in an Indian or in a Chinese restaurant.
- ▶ One restaurant strictly better, underlying state $\theta \in \{\text{Chinese}, \text{Indian}\}$
 - ⇒ is determined by initial random event that the agents can't observe
- ▶ Agents have independent binary private signals (s_i)
 - ⇒ Signals indicate the better option with probability $p > 1/2$.
- ▶ Agents observe prior decisions, but not the signals of others
- ▶ **Realization:** Assume $\theta = \text{Indian}$
 - ▶ If the first two agents choose Chinese, everyone else selects Chinese.
 - ▶ People do not converge on the better restaurant



Decision = 'Chinese' Decision = 'Chinese' Decision = 'Chinese'

- ▶ Agent i should guess "Chinese" if

$$\Pr(\theta = \text{"Chinese"} \mid \text{what Agent } i \text{ has observed} + \text{her private signal}) > \frac{1}{2}$$

and guess "Indian" otherwise.

- ▶ The prior probabilities are

$$\Pr(\theta = \text{"Chinese"}) = \frac{1}{2} \quad \& \quad \Pr(\theta = \text{"Indian"}) = \frac{1}{2}$$

- ▶ We assume private signals indicate the better option with probability $p > \frac{1}{2}$

$$\Pr(s_i = \text{"Chinese"} \mid \theta = \text{"Chinese"}) = p > \frac{1}{2}$$

$$\Pr(s_i = \text{"Indian"} \mid \theta = \text{"Indian"}) = p > \frac{1}{2}$$

- ▶ Can use Bayes' Rule to calculate

$$\Pr(\theta = \text{"Chinese"} \mid s_i = \text{"Chinese"}) = \frac{\Pr(\theta = \text{"Chinese"}) \cdot \Pr(s_i = \text{"Chinese"} \mid \theta = \text{"Chinese"})}{\Pr(s_i = \text{"Chinese"})}$$

- ▶ The numerator is $\frac{1}{2} \cdot p$
- ▶ For the denominator

$$\begin{aligned} \Pr(s_i = \text{"Chinese"}) &= \Pr(\theta = \text{"Chinese"}) \cdot \Pr(s_i = \text{"Chinese"} \mid \theta = \text{"Chinese"}) \\ &+ \Pr(\theta = \text{"Indian"}) \cdot \Pr(s_i = \text{"Chinese"} \mid \theta = \text{"Indian"}) \\ &= \frac{1}{2}p + \frac{1}{2}(1 - p) = \frac{1}{2} \end{aligned}$$

$\frac{1}{2}$ makes sense – the roles of “Chinese” and “Indian” in this experiment are completely symmetric.

$$\Rightarrow \Pr(\theta = \text{"Chinese"} \mid s_i = \text{"Chinese"}) = \frac{\frac{1}{2} \cdot p}{\frac{1}{2}} = p > \frac{1}{2}$$

- ▶ We get the intuitive result that the first agent should select “Chinese” when her private signal $s_1 = \text{“Chinese”}$.
- ▶ Bayes’ Rule also provides a probability, namely p , that the guess will be correct.
- ▶ All individuals trust their own observations when there is a tie between two choices.
- ▶ The calculation is similar for the second agent, and we skip it here to move on to the calculation for the third agent, where a cascade begins
 - ⇒ $\Pr(\theta = \text{“Chinese”} \mid s_1 = \text{“Chinese”}, s_2 = \text{“Chinese”}, s_3 = \text{“Indian”})$
- ▶ Using Bayes’ Rule

$$\Pr(\theta = \text{“C”} \mid s_1 = \text{“C”}, s_2 = \text{“C”}, s_3 = \text{“I”}) = \frac{\Pr(\theta = \text{“C”}) \cdot \Pr(s_1 = \text{“C”}, s_2 = \text{“C”}, s_3 = \text{“I”} \mid \theta = \text{“C”})}{\Pr(s_1 = \text{“C”}, s_2 = \text{“C”}, s_3 = \text{“I”})}$$

$$\Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"} \mid \theta = \text{"C"}) = p^2(1 - p)$$

$$\begin{aligned} \Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"}) &= \Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"} \mid \theta = \text{"C"}) \cdot \Pr(\theta = \text{"C"}) \\ &\quad + \Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"} \mid \theta = \text{"I"}) \cdot \Pr(\theta = \text{"I"}) \\ &= \frac{1}{2}p^2(1 - p) + \frac{1}{2}p(1 - p)^2 \end{aligned}$$

- ▶ Plugging all this back into

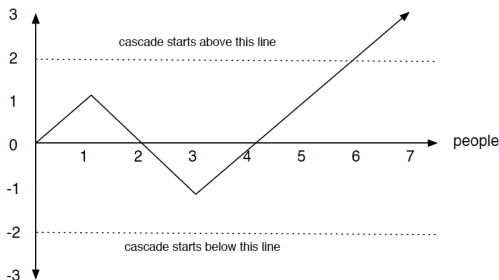
$$\begin{aligned} \Pr(\theta = \text{"C"} \mid s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"}) &= \frac{\Pr(\theta = \text{"C"}) \cdot \Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"} \mid \theta = \text{"C"})}{\Pr(s_1 = \text{"C"}, s_2 = \text{"C"}, s_3 = \text{"I"})} \\ &= \frac{\frac{1}{2} \cdot p^2(1 - p)}{\frac{1}{2} \cdot p^2(1 - p) + \frac{1}{2} \cdot p(1 - p)^2} \\ &= p > \frac{1}{2} \end{aligned}$$

- ▶ If the first two agents choose “Chinese”, then agent 3 is better off choosing “Chinese” even if her signal indicates “Indian” (since two signals are stronger than one, and the behavior of the first two agents indicates that their signals were pointing to “Chinese”).
- ▶ This is because

$$\Pr[\theta = \text{“Chinese”} \mid s_1 = s_2 = \text{“Chinese” and } s_3 = \text{“Indian”}] > \frac{1}{2}$$

- ▶ This reasoning applies later in the sequence (though agents rationally understand that those herding is not revealing information).

- ▶ If the first two people choose “Chinese” then everyone in order will choose the Chinese restaurant as well
- ▶ Of course, the same thing happens if the first two people choose the Indian restaurant



Easley, David and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010. © Cambridge University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

- ▶ We refer to this phenomenon as **herding**, since agents after a certain number “herd” on the behavior of the earlier agents.
- ▶ Bikchandani, Hirshleifer and Welch refer to this phenomenon as **informational cascade**.
- ▶ Information cascades can lead to non-optimal outcomes

- ▶ In general networks, this gets too complicated very quickly!
- ▶ In fact we proved that the problem of learning is NP Hard, i.e., even computers cannot solve it efficiently (in the worst case).
- ▶ We have also developed simpler approaches
- ▶ Let us turn to a simple model of (rule-of-thumb) learning that also incorporate **network structure**.

- ▶ First introduced by DeGroot (1974) and more recently analyzed by Golub and Jackson (2007).
- ▶ Beliefs updated by taking weighted averages of neighbors' beliefs
- ▶ A finite set $\{1, \dots, n\}$ of agents
- ▶ Interactions captured by an $n \times n$ nonnegative **interaction matrix** T
 - ▶ $T_{ij} > 0$ indicates the trust or weight that i puts on j
 - ▶ T is a stochastic matrix (row sum=1)
- ▶ There is an underlying state of the world $\theta \in \mathbb{R}$
- ▶ Each agent has initial belief $x_i(0)$; we assume $\theta = 1/n \sum_{i=1}^n x_i(0)$
- ▶ Each agent at time k updates his belief $x_i(k)$ according to

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

- ▶ For all i , $T_{ij} > 0$ only if $j \in N(i)$, i.e., $[T_{ij}]_j$ captures the neighborhood of agent i .

- ▶ Each agent is updating his or her beliefs as an average of the neighbors' beliefs.
- ▶ Reasonable in the context of one shot interaction.
- ▶ Is it reasonable when agents do this repeatedly?

Definition

T is a (row) stochastic matrix, if the sum of the elements in each row is equal to 1, i.e.,

$$\sum_j T_{ij} = 1 \text{ for all } i.$$

Definition

T is a doubly stochastic matrix, if the sum of the elements in each row and each column is equal to 1, i.e.,

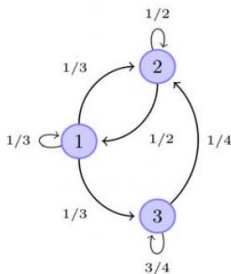
$$\sum_j T_{ij} = 1 \text{ for all } i \text{ and } \sum_i T_{ij} = 1 \text{ for all } j.$$

- ▶ Throughout, assume that T is a stochastic matrix. Why is this reasonable?

- ▶ Consider the following example

$$T = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

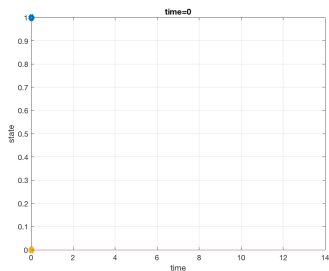
- ▶ The interaction matrix induces a weighted directed graph: there is an edge from i to j (with weight T_{ij}) if $T_{ij} > 0$.



© Source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

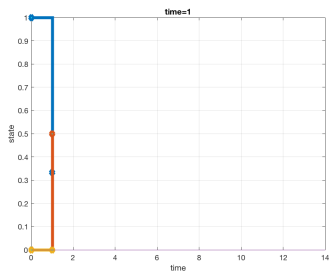
- ▶ Suppose that initial vector of beliefs is

$$x(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$



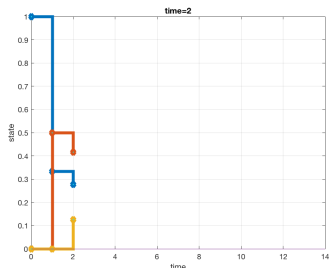
- ▶ Then updating gives

$$x(1) = Tx(0) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix}$$



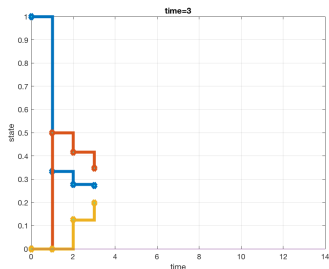
- ▶ In the next round, we have

$$\begin{aligned}
 x(2) = T_x(1) = T^2 x(0) &= \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} \\
 &= \begin{pmatrix} 5/18 \\ 5/12 \\ 1/8 \end{pmatrix}
 \end{aligned}$$



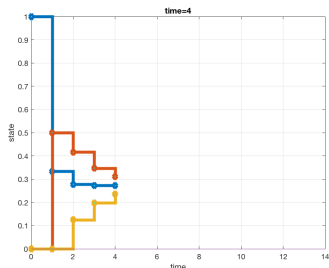
- ▶ In the next round, we have

$$\begin{aligned}
 x(2) = T_x(1) = T^2x(0) &= \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} \\
 &= \begin{pmatrix} 5/18 \\ 5/12 \\ 1/8 \end{pmatrix}
 \end{aligned}$$



- ▶ In the next round, we have

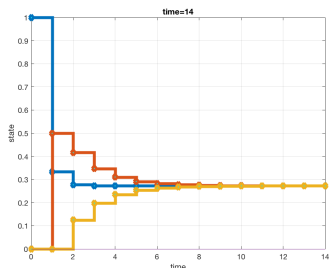
$$\begin{aligned}
 x(2) = T_x(1) = T^2 x(0) &= \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} \\
 &= \begin{pmatrix} 5/18 \\ 5/12 \\ 1/8 \end{pmatrix}
 \end{aligned}$$



- ▶ In the limit, we have

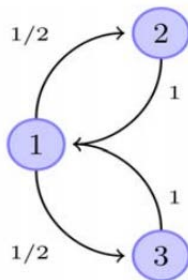
$$x(n) = T^n x(0) \rightarrow \begin{pmatrix} 3/11 & 3/11 & 5/11 \\ 3/11 & 3/11 & 5/11 \\ 3/11 & 3/11 & 5/11 \end{pmatrix} x(0) = \begin{pmatrix} 3/11 \\ 3/11 \\ 3/11 \end{pmatrix}.$$

- ▶ Note that the limit matrix, $T^* = \lim_{n \rightarrow \infty} T^n$ has identical rows.
- ▶ Is this kind of convergence general? Yes, but with some caveats.



- ▶ Consider instead

$$T = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$



© Source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

- ▶ In this case, we have

$$T^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \text{ for all } n \text{ even.}$$

$$T^n = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ for all } n \text{ odd.}$$

- ▶ Matrix oscillates and there is no convergence.

- ▶ Problem in the above example is **periodic** behavior.
- ▶ It is sufficient to assume that $T_{ij} > 0$ for all i to ensure aperiodicity. Then we have:

Theorem

Suppose that T induces a strongly connected network and $T_{ij} > 0$ for each i , then $\lim_n T^n = T^$ exists and is unique. Moreover, $T^* = ew_1^T$, where $e = [1, 1, \dots, 1]^T$ and w_1 is the left eigenvector corresponding to eigenvalue 1, normalized such that $w_1^T e = 1$.*

- ▶ In other words, T^* will have identical rows given by w_1 .

- ▶ An immediate corollary of the preceding result is:

Proposition

In the myopic learning model above, if the interaction matrix T defines a strongly connected network and $T_{ij} > 0$ for each i , then there will be consensus among the agents, i.e., $\lim_{n \rightarrow \infty} x_i(n) = x^$ for all i .*

- ▶ We can express x^* as

$$x^* = w_1^\top x(0) = \sum_i w_{1i} x_i(0).$$

- ▶ Hence, the limiting beliefs will be weighted averages of the initial beliefs, and the relative weights will represent the influences that the various agents have on the final consensus beliefs.
- ▶ This shows that the influences of the agents are given by the left eigenvector of matrix T corresponding to eigenvalue 1, providing a foundation for eigenvector based centrality measures.

MIT OpenCourseWare
<https://ocw.mit.edu/>

1.022 Introduction to Network Models
Fall 2018

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.