6.207/14.15: Networks
Lecture 8: Diffusion through Networks

Daron Acemoglu and Asu Ozdaglar
MIT

October 7, 2009

Outline

- Spread of epidemics in networks
- Models of diffusion without network structure
  - Bass model
- Models of diffusion that explicitly incorporate network structure
  - Diffusion with immune nodes
  - SIR model (susceptible, infected, removed)
  - SIS model (susceptible, infected, susceptible)

Reading:

- Jackson, Chapter 7, Sections 7.1,7.2.
- EK, Chapter 21.

Introduction

- The study of epidemic disease has always been a topic where biological issues mix with the social ones.
- The patterns by which epidemics spread through a society is determined not just by the properties of the pathogen carrying it (including its contagiousness, the length of its infectious period, and severity), but also by the network structure within the population.
  - Opportunities for a disease to spread from one person to another is given by the contact network, indicating who has contact with whom on a regular basis.
- We are interested in the following questions:
  - Under what conditions will an initial outbreak spread to a nontrivial portion of the population?
  - What percentage of the population will eventually become infected?
  - What is the effect of immunization policies?
- The problem is relevant not only to disease transmission, but also to diffusion through a network of information, opinions, and adoption of new technologies or behaviors.

Bass Model–1

- An early model of diffusion is the Bass model.
- Although it does not capture any explicit social network structure, it still incorporates imitation.
- The model is built on two parameters: p captures the rate at which agents spontaneously get infections (in response to outside stimuli); and q captures the rate at which agents get infected through others (secondary infections).
  - In the context of adoption of technologies, p can be interpreted as the rate of innovation and q as the rate of imitation due to social effects.
- Consider a discrete-time model and let $F(t)$ be the fraction of agents infected at time t.
- The Bass model is described by the difference equation:

$$F(t) = F(t-1) + p(1 - F(t-1)) + q(1 - F(t-1))F(t-1).$$

- The term $p(1 - F(t-1))$ is the infection rate times the fraction of uninfected agents. The term $q(1 - F(t-1))F(t-1)$ is the contagion rate times the frequency of encounters between healthy and infected agents.

Bass Model–2

- A continuous time version of this model is described by

$$\frac{d\,F(t)}{dt} = (p + qF(t))(1 - F(t)),$$

  with $F(0) = 0$.

- This is a nonlinear differential equation, but admits a closed form solution

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}.$$

- Note that the levels of p and q scale time, the ratio of q to p determines the overall shape of the curve.

Bass model–3

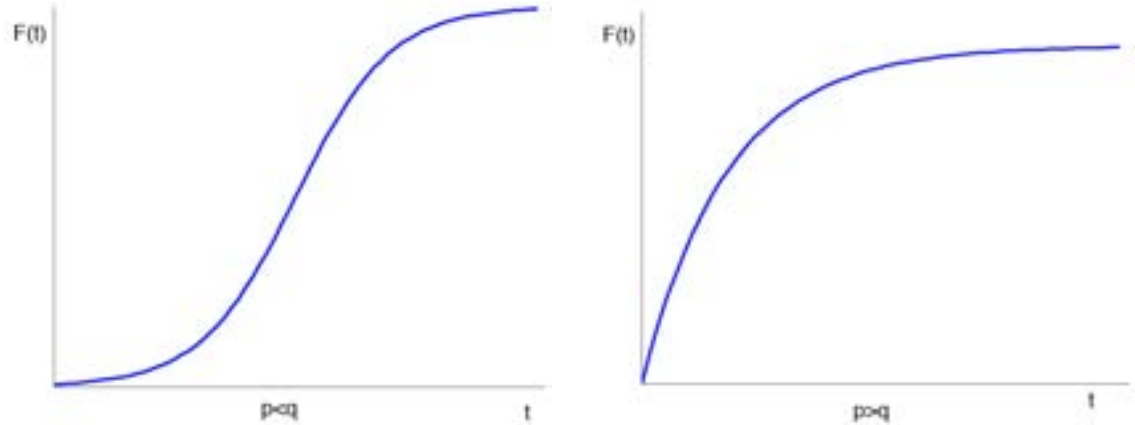


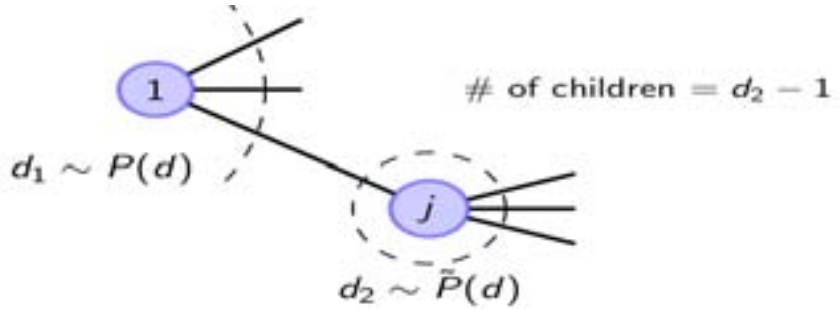
Figure: Diffusion curves: left is for $p < q$ and right is for $p > q$.

- Many empirical studies have found diffusion patterns that are S-shaped (e.g. adoption of hybrid corn seeds among Iowa farmers).
- Let us interpret this in the context of adoption of new technologies.
- First adopters are almost entirely those who adopt from their spontaneous innovation (when $F(t)$ is close to 0, $\dot{F}(t) = p$).
- As process progresses, there are more agents to be imitated leading to an increase in the rate of diffusion, which eventually slows down since there are fewer agents to do the imitating.

Diffusion in a Network with Immune Nodes

- The problems of modeling contagion or the spread of information through a society involve determining "when paths exist that connect different nodes", i.e., understanding the component structure.
- Let us consider the following problem.
- There is a society of n individuals. Initially one of them is infected with a disease. Each individual is immune with probability π.
- The question of whether the disease can spread to a nontrivial fraction of the population amounts to whether the infected individual lies in a "giant component" of the network with the immune nodes removed.
- We have studied this problem when the underlying network is an Erdös-Renyi network.
- We will now generalize to arbitrary degree distributions.

Diffusion with General Degree Distributions



- Recall that the degree distribution of a neighboring node is given by

$$\tilde{P}(d) = \frac{dP(d)}{\langle d \rangle}.$$

- From this, we showed that the expected number of children is given by

$$\tilde{\mathbb{E}}[\text{number of children}] = \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle}.$$

- Hence the expected number of infected children (basic reproductive number of the disease) is

$$\lambda \equiv (1 - \pi) \frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle}.$$

Diffusion with General Degree Distributions

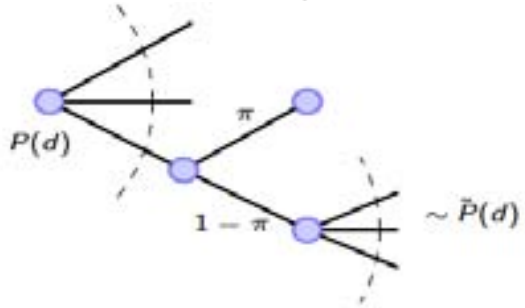
- Recall the branching process analysis:
  If $\lambda < 1$, then with probability one, the disease dies out after a finite number of stages.
  If $\lambda > 1$, then with positive probability, the disease persists by infecting a large portion of the population.
- This yields the following threshold for the probability of immune π: a giant component will emerge if
$$\pi < \frac{\langle d^2 \rangle - 2\langle d \rangle}{\langle d^2 \rangle - \langle d \rangle},$$
  i.e., if the fraction of immune nodes is below this threshold.
- For example, for a regular network (each node with degree $\bar{d}$), this leads to
$$\pi = \frac{\bar{d} - 2}{\bar{d} - 1}.$$
  - If $\bar{d} = 2$, giant component never emerges.
  - If $\bar{d} = 3$, giant component emerges if $<$ half the population is immune.
- For the Erdös-Renyi graph, we have $\langle d^2 \rangle = \langle d \rangle^2 + \langle d \rangle$ and $\langle d \rangle = (n-1)p$.
- This yields the threshold $\pi = 1 - \frac{1}{(n-1)p}$ or $p(n-1)(1-\pi) = 1$, as before.

Diffusion with General Degree Distributions

- For a power-law degree distribution (or a scale-free network) with $P(d) \sim d^{-\gamma}$, $\gamma < 3$, we have that the $\langle d^2 \rangle$ is diverging in n.
- Therefore, the contagion threshold for this case is $\pi = 1$, i.e., all nodes have to be immune before the giant component of susceptible nodes disappears.
  - Under such degree distributions, there are enough very high degree nodes that many nodes are connected to and the network has a giant component even when many nodes are eliminated uniformly at random.
- Immunized nodes can be viewed as nodes that are removed from the system.
- We have seen that Internet has a power-law distribution with exponent $\sim 2.1 - 2.7$. The preceding shows that Internet is robust: remove 98% of the nodes, you still have connectivity.
- However, a targeted removal of highest-degree nodes implies a much lower threshold:
  - If $\gamma = 2.5$, then $\pi = 0.056$!! (removing 5% of the nodes disconnects the network).
- Leads to the catchy phrase "Internet is robust, yet fragile."

Size of the Infected Population

- Compute size of giant component, gives the size of infected population.
- Consider a node and the event that this node is in the giant component, or equivalently the event that the branching process does not die out.



- Let $\tilde{q}$ denote the probability that the branching process does not die out starting from a neighboring node:

$$1 - \tilde{q} = \pi + (1 - \pi) \sum_{d=1}^{\infty} \tilde{P}(d)(1 - \tilde{q})^{d-1}.$$

- Let q denote the probability that the branching process does not die out:

$$1 - q = \sum_{d=0}^{\infty} P(d)(1 - \tilde{q})^{d}.$$

- The size of the giant component is given by qn.

SIR Model–1

- In the SIR model, a node can be in one of 3 states:
  - Susceptible: Before the node has caught the disease, it is susceptible to infection from the neighbors.
  - Infected: Once the node has caught the disease, it is infectious and has some probability of infecting each of its susceptible neighbors.
  - Removed: After the disease has run its course, the node either dies or becomes completely immune (no longer susceptible).
- A good model for diseases such as chickenpox.
- Assume individuals are connected through a network generated under the configuration model with degree distribution $P(d)$.
- Suppose that the infection process is such that the probability that an infected node will infect a susceptible neighbor before the infected node is removed can be described by the probability of transmission t.
- Assume that the infection process is independent across links between susceptible and infected nodes.
  - The independence assumption is clearly violated in many cases.

SIR Model–2

- To analyze the reach of infection, we can remove links (in an independent and identical manner) with probability $1 - t$, and compute the resulting component size.
- The analysis then is analogous to the analysis of diffusion with immune nodes (with t in place of $1 - \pi$).
- How do we determine the transmission probability t?

  Model 1:
  - An infected node is removed within 1 time step (deterministic).
  - A node infects each of its susceptible neighbor i independently within time T_i that is exponentially distributed with parameter β.
  - We have $t = \mathbb{P}(T_i \leq 1) = 1 - e^{-\beta}$.

  Model 2:
  - An infected node is removed within time $T \sim \exp(\gamma)$.
  - A node infects each of its susceptible neighbor i independently within time $T_i \sim \exp(\beta)$.
  - We have $t = \mathbb{P}(T_i \leq T) = \frac{\beta}{\beta+\gamma}$.

SIS Model–1

- In the SIS model, susceptible nodes can become infected, and then recover in such a way that they become susceptible again (rather than being removed).
- Models diseases such as certain variations of the common cold.
- Consider a degree-based random meeting model: nodes interact randomly according to their degree d_i.
- Let $P(d)$ be the degree distribution in the society.
- The probability that a meeting of node i is with a degree d node is
$$\frac{P(d)d}{\langle d \rangle}.$$
- It is essential to keep track of nodes degrees since nodes with different degrees tend to have different infection rates.
- Let $\rho_d(t)$ denote the fraction of nodes of degree d infected at time t.
- Let $\theta(t)$ denote the probability that a given meeting is with an infected individual. Then:
$$\theta(t) = \frac{\sum P(d)\rho_d(t)d}{\langle d \rangle}.$$

SIS Model–2

- Let ν denote the transmission rate of infection and δ denote the recovery rate of an infected individual.
- We assume that the probability that a susceptible agent with degree d becomes infected in a period $[t, t+\epsilon)$ is $\epsilon\nu\theta(t)d$.
- Using a mean field analysis, we can write the evolution of $\rho_d(t)$:

$$\dot{\rho}_d(t) = (1-\rho_d(t))\nu\theta(t)d - \rho_d(t)\delta.$$

  The term $(1-\rho_d(t))\nu\theta(t)d$ represents the fraction of nodes of degree d that were susceptible and become infected and $\rho_d(t)\delta$ represents the fraction that recover to become susceptible again.
- Using this, we can characterize the steady state. Let $\theta(t) \longrightarrow \theta$ and $\rho_d(t) \longrightarrow \rho_d$, and $\lambda = \nu/\delta$:

$$\rho_d = \frac{\lambda\theta d}{\lambda\theta d + 1}, \qquad \text{and therefore,}$$

$$\theta = \sum_d \frac{P(d)\lambda\theta d^2}{\langle d\rangle(\lambda\theta d + 1)}.$$

Nonzero Steady State Infection Rate

- $\theta = 0$ is always a solution: if nobody is infected, the system stays that way.
- We next analyze when the steady state has a solution with $\theta > 0$.
- Assume the degree distribution is regular, all nodes have degree $\bar{d}$. Then:

$$\theta = \frac{\bar{d}\lambda\theta}{\bar{d}\lambda\theta + 1},$$

  implying a solution $\theta = 1 - \frac{1}{\lambda\bar{d}}$, which is positive only if $\bar{d} > 1/\lambda = \delta/\nu$.
  - If the number of meetings is sufficiently large compared to the relative recovery/infection rate, then the infection can be sustained.
- It can be shown that for power-law degree distributions, there is always a positive solution.

Nonzero Steady State Infection Rate

- In general, let $H(\theta)$ be
$$H(\theta) = \sum_d \frac{P(d)\lambda\theta d^2}{\langle d \rangle (\lambda\theta d + 1)}.$$
- We have $H(0) = 0$ and $H(\theta)$ is increasing and strictly concave in θ.
- Thus, for H to have a nonzero fixed point, we must have $H'(0) > 1$.
- Note that $H'(0) = \lambda \frac{\langle d^2 \rangle}{\langle d \rangle}$.
- Hence the condition for positive steady state infection is
$$\lambda > \frac{\langle d \rangle}{\langle d^2 \rangle}.$$
- For regular graphs, the threshold is $\lambda > 1/\bar{d}$, as before.
- For power law distributions (with $\gamma < 3$), $\langle d^2 \rangle$ is divergent, hence the above equation is satisfied for any positive λ.
  - Intuition: Individuals with high-degree nodes serve as conduits for infection. Even very low infection rates can lead them to become infected and infect many others.

14.15J / 6.207J Networks

Fall 2009