School of Information Sciences
University of Pittsburgh

TELCOM2125: Network Science and Analysis

Konstantinos Pelechrinis

Spring 2015

Figures/Materials are taken from:
M. O. Jackson, “Social and Economic Networks”

Part 10: Diffusion

The problem

- **Different “commodities” can disseminate through a population**
  - Information
  - Product adoption
  - Diseases
  - …

- **The vast majority of the times network structure plays a crucial role in the way this diffusion takes place**

Bass model

- **Bass model is a very basic model for describing the dynamics of product adoption**
  - No network structure involved

- **Two states are possible for every node/person**
  - 0: product has not been adopted
  - 1: product has been adopted

- **Let's assume that F(t) is the fraction of the population that have adopted the product at time t**

Bass model

- **A person can adopt the product through two mechanisms:**
  - Innovation – which happens with a rate p
  - Imitation – which happens with a rate q

- **Hence, the rate with which the fraction F(t) changes is given by:**

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

- **Solving the above equation we get:**

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

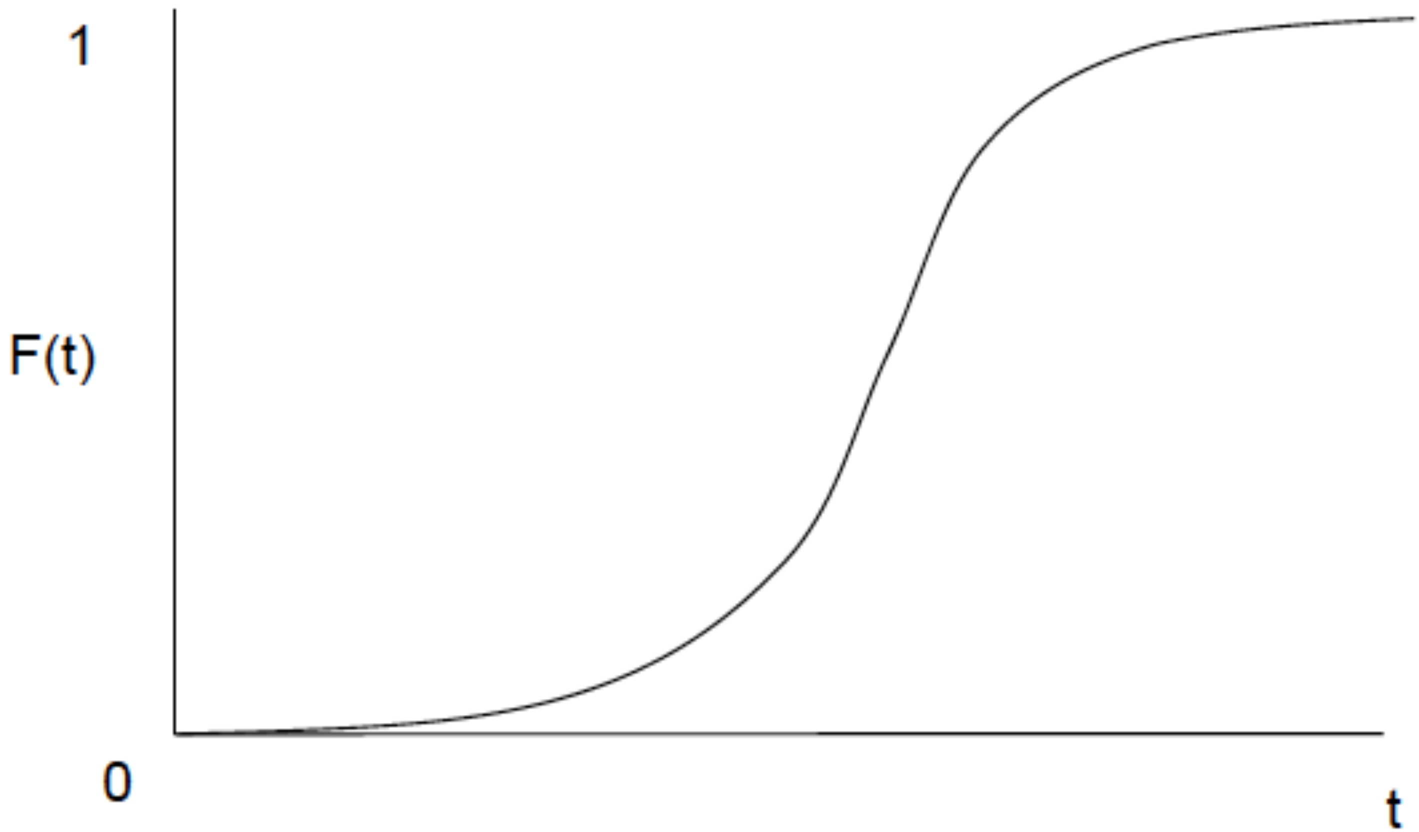
Bass model and the S-curve

- **When F(t) is almost 1 the rate with which the fraction of the population that has adopted the product increases is 0**

- **When F(t) is 0, the rate with which the fraction of the population that has adopted the product increases is p**
  - Initially only innovators matter since there is no one to imitate

- **When F(t) is a small positive number ε:** $\frac{dF(t)}{dt} = (p + q\varepsilon)(1 - \varepsilon)$
  - To get initial convexity we want: $(p + q\varepsilon)(1 - \varepsilon) > p \Rightarrow q > p$
  - After a while imitation (i.e., q) takes over

Bass model and the S-curve



Networks and diffusion

- **Information, products, ideas, computer viruses, diseases etc. spread via network connections**

- **“Infections” can spread only through connections**

- **Questions**
  - Will an “infection” take hold ?
  - How many nodes will it reach ?
  - When do we get diffusion ?
  - How does it depend on the particulars of the diffusion process as well as the network structure ?
  - Who is likely to be infected first ?

Diffusion in random networks

- **The reach of diffusion is dictated by the structure of the components**
- **Non-trivial diffusion if someone in the giant component is “infected”/adopts**
  - Size of the giant component determines the likelihood of diffusion and its extent
- **We will use our results for Erdos-Renyi graphs**

Diffusion in random networks

- **We have seen that a giant component emerges when:**

$$p > \frac{1}{n}$$

  - We have also calculated in this case the size of the giant component S

- **Probability of being at the giant component increases with the degree of a node**
  - More connected, more likely to be infected

Extensions

- **Until now we have assumed that all edges facilitate diffusion**

- **Immunity: Delete a fraction of nodes that are immune to diffusion and study the giant component on the remaining nodes**

- **Probabilistic diffusion: Some links might not facilitate diffusion**
  - Just lower p

SIS model

- **SIS is extensively studied in epidemiology**
- **Allows nodes to change behavior back and forth between two states**
  - Susceptible
  - Infected
- **It can be used for modeling of catching some recurring disease, who votes for whom, etc.**

SIS model

- **A susceptible node can be infected through *contact* with its already infected neighbors**
  - Probability that it gets infected is proportional to number of infected neighbors with rate $v > 0$, plus a spontaneous ε

- **An infected node gets well (i.e., becomes susceptible again) at a rate $\delta > 0$**

- **Let us denote the fraction of infected nodes with ρ**

Benchmark

- **Let's start with a benchmark where all nodes mix with even probabilities**
  - Randomly meet/contact an individual at each period
  - Large Markov chain

- **Steady-state:**

$$\frac{d\rho}{dt} = 0$$

$$\frac{d\rho}{dt} = (1-\rho)(v\rho+\varepsilon) - \rho\delta = 0$$

Mean-field

- **Drop ε** $$\frac{d\rho}{dt} = (1-\rho)v\rho - \rho\delta = 0$$
- **Two solutions:** $$\rho = 1 - \frac{\delta}{v} \quad (if > 0)$$ $$\rho = 0$$
- **If δ>v then recover faster than get sick**
  - No infection stays
- **Otherwise, infection stays at some level**
  - For low recovery rates can lead to large infections
    - ✓ $\rho \rightarrow 1$

Where's the network?

- **So far we have considered random interaction**

- **Missing heterogeneity in degree**
  - We will address it in next slides

- **Missing local patterns**

Effect of degree distribution

- **Node i has degree d_i**
  - Random matching with d_i matces

- **Let's denote with $\rho(d)$ the fraction of nodes of degree d infected**

- **Let θ be the fraction of randomly chosen neighbors infected**

Probability of meeting an infected node

- **P(d) fraction of nodes have d meetings**
  - Meeting = degree
- **More likely to meet someone with high degree d**
- **Likelihood of meeting node of degree d is:**

$$\frac{P(d)d}{E[d]}$$

Why?

- **Hence, likelihood of meeting infected node is:**

$$\theta = \sum_{d} \frac{\rho(d)P(d)d}{E[d]}$$

Solving the model

- **Steady state for each d:**

$$\frac{d\rho(d)}{dt} = (1-\rho(d))v\theta d - \delta\rho(d) = 0$$

$$\rho(d) = \frac{\lambda\theta d}{\lambda\theta d + 1}, \quad where \quad \lambda = \frac{v}{\delta}$$

- **Substituting, we have for the likelihood of meeting an infected neighbor at random:**

$$\theta = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d + 1)E[d]}$$

- **<u>Steady state infection rate of people you meet</u> is the solution of:**

$$\theta = H(\theta) = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d + 1)E[d]}$$

Solving the model

- What can we say about how this depends on the network structure?
- How does the infection rate of neighbors θ depend on P(d) and E[d]?
- We need to see how H(θ) looks like and how it depends on P(d) and E[d]

Properties of H(θ)

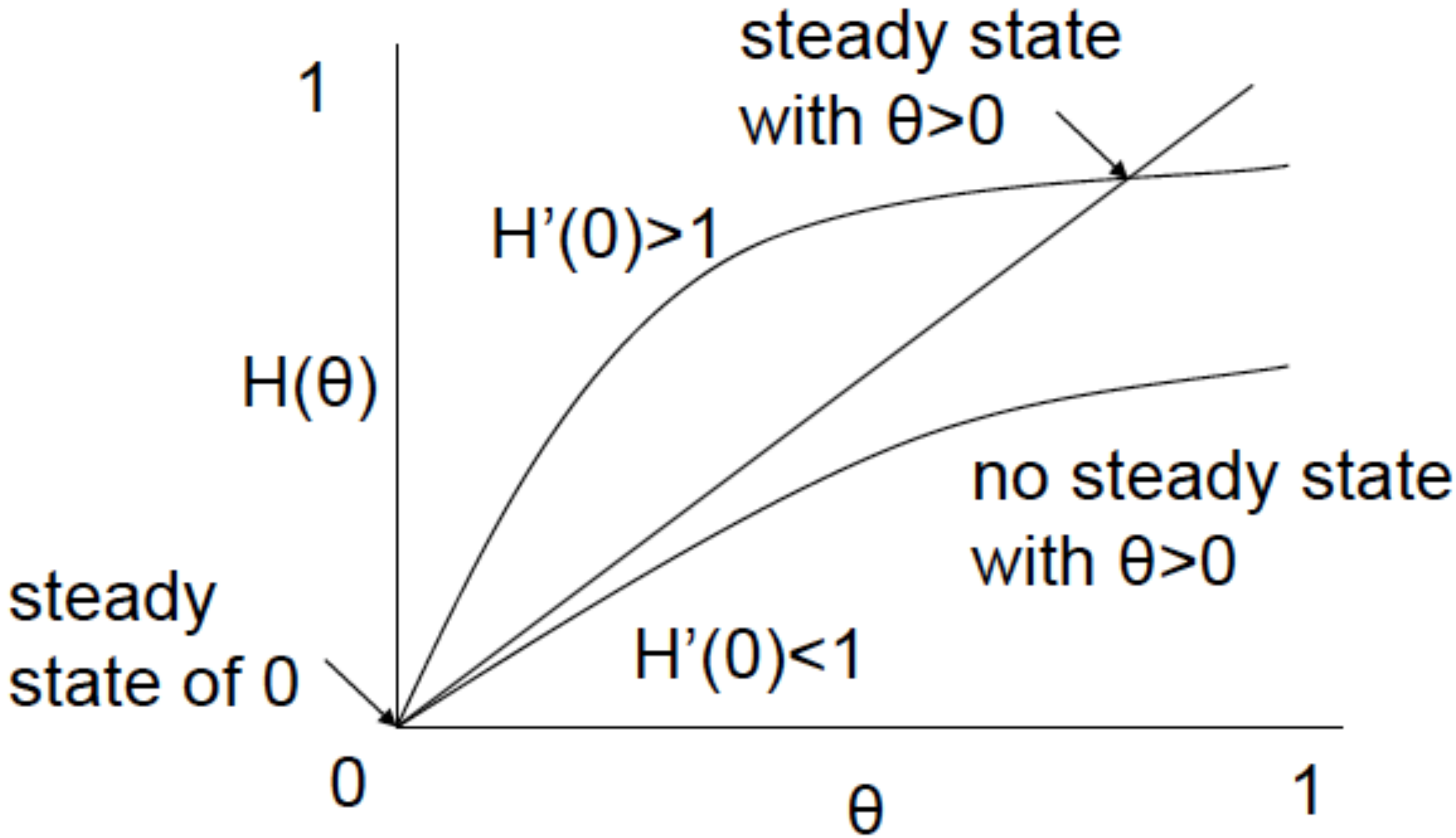
- **H is increasing**

$$\frac{dH(\theta)}{d\theta} = \sum_d \frac{P(d)\lambda d^2}{(\lambda\theta d + 1)^2 E[d]} > 0$$

- **H is strictly concave**

$$\frac{d^2 H(\theta)}{d\theta^2} = -2\sum_d \frac{P(d)\lambda^2 d^3}{(\lambda\theta d + 1)^3 E[d]} < 0$$

Graphical solution



Non-zero steady state

- **What about H'(0) ?**

- **Setting θ=0 to the formula above for the first derivative of H we have:**

$$\left.\frac{dH(\theta)}{d\theta}\right|_{\theta=0} = \sum_d \frac{P(d)\lambda d^2}{E[d]} = \lambda \frac{E[d^2]}{E[d]}, \quad \lambda = \frac{v}{\delta}$$

Theorem

- **Conditions for Steady State of Mean-Field SIS process**

- **There exists a non-zero steady state *if and only if*:**

$$\lambda > \frac{E[d]}{E[d^2]}$$

  - So need for infection/recovery rate to be high enough relative to average degree divided by the second moment of the degree distribution (roughly the variance)

Conditions for steady state

- **In a regular network we have:**

$$\lambda > \frac{1}{E[d]}$$

- **In an E-R network we have:**

$$\lambda > \frac{1}{1+E[d]}$$ **Why?**

- **In a power-law network, $E[d^2]$ diverges and hence, we always have a non-zero steady state**
  - In principle, diffusion is "easy" in a power-law network

Ideas/Take-aways

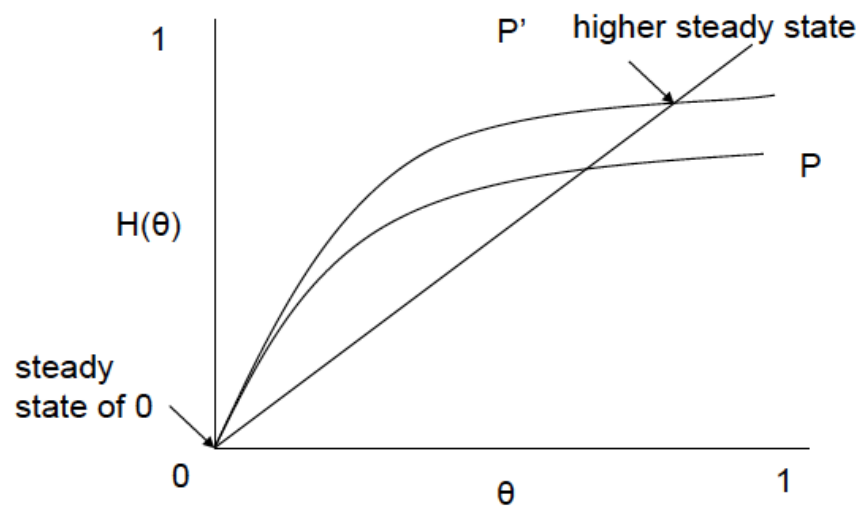
- High degree nodes are more prone to infection
- They serve as conduits for the infection
- Higher variance, more such nodes to enable infection

How does degree distribution affect solution

$$\theta = H(\theta) = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d + 1)E[d]}$$

- **How does H(θ) shift with changes at P(d)?**

- **It can be shown that for a new degree distribution P' such that:**
  - P' first order stochastically dominates P or
  - P' is a mean preserving spread of P
    - ✓ H(θ) increases at every θ



Intuitions

- **Mean preserving spread**
  - More high degree nodes and low degree nodes
- **Higher degree nodes are more prone to infection**
- **Neighbors are more likely to be high degree**
- **So, either first order stochastic dominance, or mean-preserving spreads in P increase θ**

What about average?

- **Until now we have examined the infection rate of the neighbors (θ)**

- **This is not the same with the infection rate of the population (ρ)!**

- **<u>Theorem:</u> If P’ is a mean preserving spread of P, then $\theta' > \theta$, but $\rho' > p$ if λ is *low*, while $\rho' < \rho$ if λ is *high***

Proof

- **For the steady state of nodes with degree d we have:**

$$\frac{d\rho(d)}{dt} = (1-\rho(d))v\theta d - \delta\rho(d) = 0$$

- **Taking the expectation over d we get:**

$$v\theta E[d] - \sum_d P(d)\rho(d)v\theta d - \rho\delta = 0 \Rightarrow$$

$$v\theta E[d] - v\theta^2 E[d] - \rho\delta = 0 \Rightarrow$$

$$\rho = \lambda\theta E[d](1-\theta)$$

- **ρ is increasing in θ iff θ<1/2**
  - However, θ is increasing in λ

SIS diffusion model

- **Simple and tractable**
- **Bring in relative meeting rates**
- **Can *order* infections by properties of the (contact) network**