Social and Economic Networks

Diffusion through networks

Xiang Sun

November 1–8, 2017

Outline

1. Diffusion
2. Bass model
3. Diffusion on random networks
   - Giant component
   - Contagion with immunity and link failure
4. SIS model
   - Uniformly random infection
   - Degree-based random meeting model

Section 1

Diffusion

Networks and behavior

How does network structure impact behavior?

- Simple infections, contagion: diffusion.
  - mechanical way where it goes from one node to another.
- Opinions, information: learning.
  - how do people process information? how does information flow?
- Choices, decisions: games on network.

Diffusion

- 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.
- Not only disease transmission, but also diffusion through a network of information, opinions, and adoption of new technologies or behaviors.

S-shape adoption

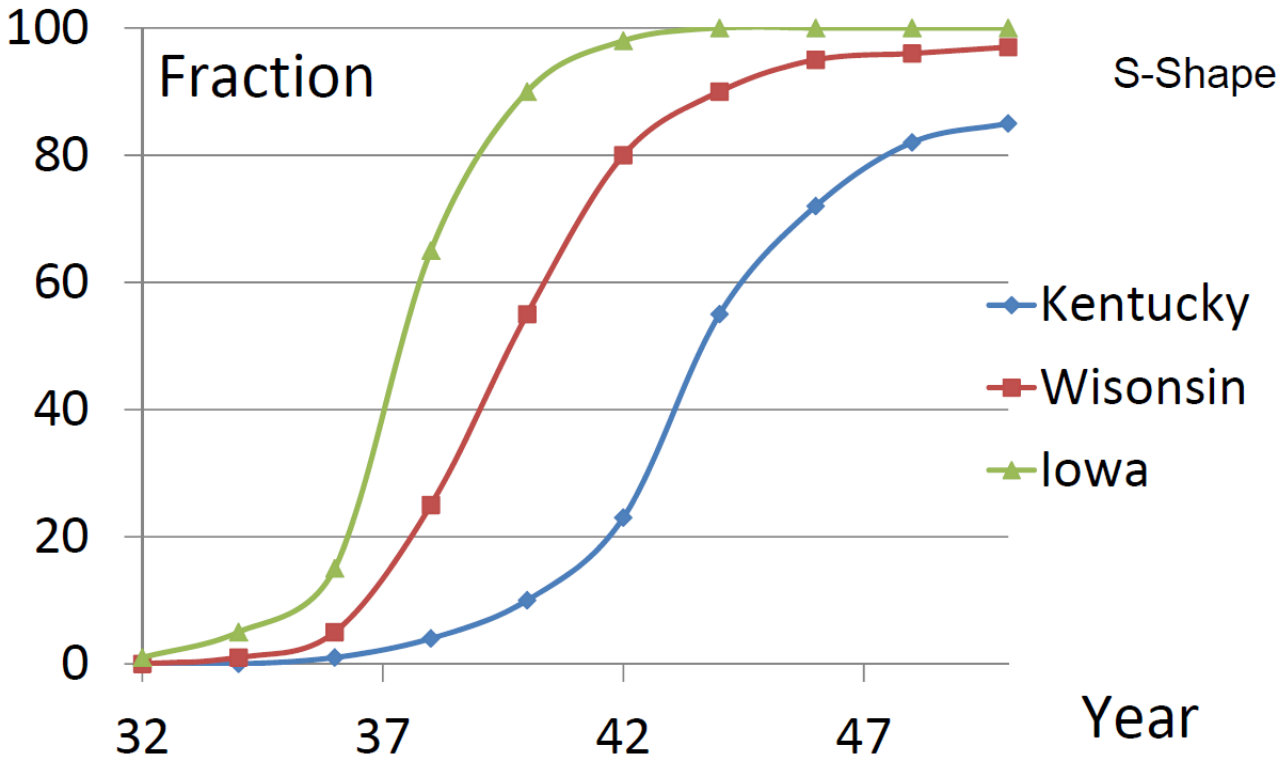
- Diffusion over time and space
- Initial adopters
  - Who are they? High degree? Innovators?
- Increase in speed
  - Word of mouth, observations of neighbors
- Eventual slowdown
  - Saturation

Diffusion: Illustration 1

Griliches (1957): Hybrid corn diffusion

- The hybrid corn is somewhere between 15% to 25% higher than the existing corn strains.
- Iowa is a state which is basically is mostly corn (a very good climate for growing corn).
- There is more variety of things grown in Wisconsin.
- In Kentucky, even less hospitable to corn, and has other things going on in terms of what it's growing.

Diffusion: Illustration 1 (Cont.)



- Iowa was the state which adopted this earliest on.
- Wisconsin was a little later adoption.
- Kentucky was even later than Wisconsin.

Diffusion: Illustration 1 (Cont.)

- The adoptions start out fairly slowly where it takes.

  It took quite awhile before it really starts accelerating.
- Eventually peak.
- We get this very nice s shape which is actually observed in a number of different.

Diffusion: Illustration 1 (Cont.)

- Why does it start out slowly?
- why does it start accelerating in this manner?
- It eventually has to asymptote and slow down.
* It cannot go above 100% so it's got to slow down eventually. It can't just keep going forever.

Diffusion: Illustration 2

- Adoption of new drug by doctors: Coleman, Katz, and Menzel (1966).
- Before they started this they surveyed the doctors: ask which other doctors would you go to for advice.

$\Rightarrow$ This gives a network of doctors.

- Result:

Fraction adopting by (months)	6	8	10	17
named by 0 others (36)	0.31	0.42	0.47	0.83
named by 1 or 2 others (56)	0.52	0.66	0.70	0.84
named by 3+ others (33)	0.70	0.91	0.94	0.97

  - More connected are earlier adopters.
  - The diffusion process actually differed based on the position of the doctors in the network.

Questions

- Under what conditions will an initial outbreak spread to a nontrivial portion of the population?
- What is the extent of a diffusion?
- How does it depend on the particulars of a network structure?
- Can we say something about these time patterns, where does the S-shape come in?
- Can we say something about welfare analyses?
- If you want to accelerate a diffusion, how would you want to do it?
- What is the effect of immunization policies?

Section 2

Bass model

Bass model

- A benchmark model with no explicit social structure.
- It incorporates imitations.
- It produces things like the S-shape.
- Two states/actions/behaviors 0 and 1.
  - 0: you are not infected/you are not adopting the new product/….
  - 1: you are infected/you are adopting the new product/….
- Initial state/action/behavior: 0.
- Move from 0 to 1, and do not move back.
- $F(t)$: fraction of the population who have adopted action 1 at time t.

Bass model (Cont.)

- p: rate of spontaneous adoption/innovation.
- q: rate of imitation of adoption.
- $$\frac{\mathrm{d}}{\mathrm{d}t}F(t) = p \cdot (1 - F(t)) + qF(t) \cdot (1 - F(t)).$$
  - The first term: spontaneous infection rate times the fraction of uninfected agents.
  - The second term: the contagion rate times the frequency of encounters between healthy and infected agents.
- Initially, $F(0) = 0$.

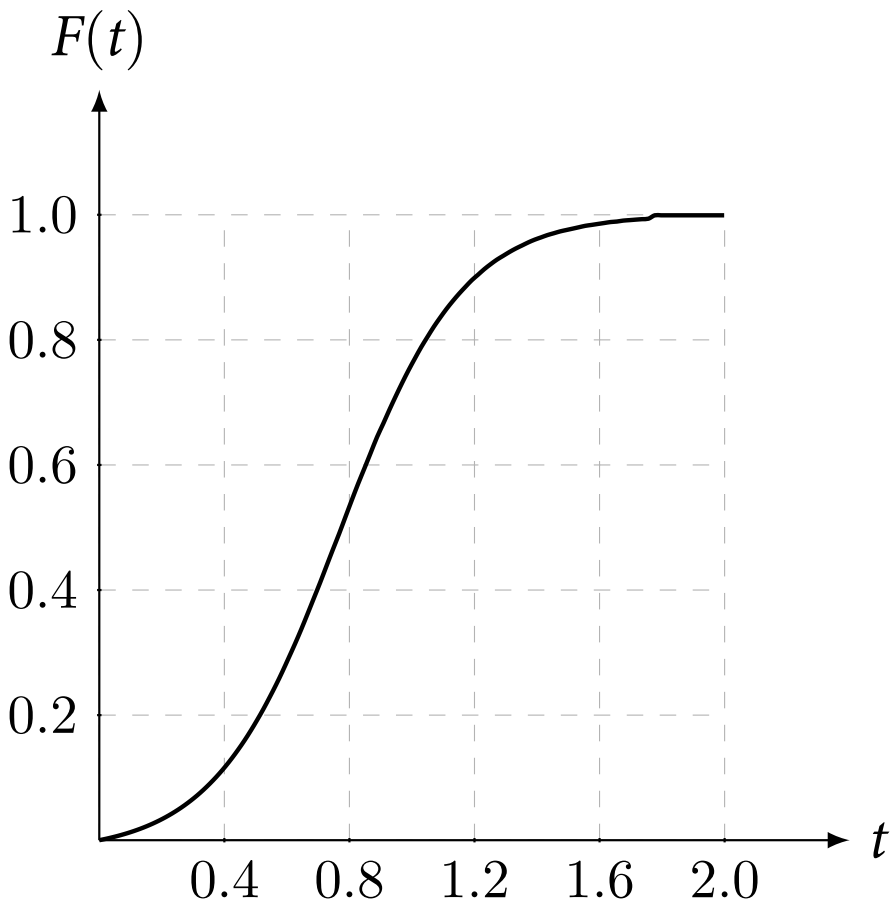
Getting the S-shape

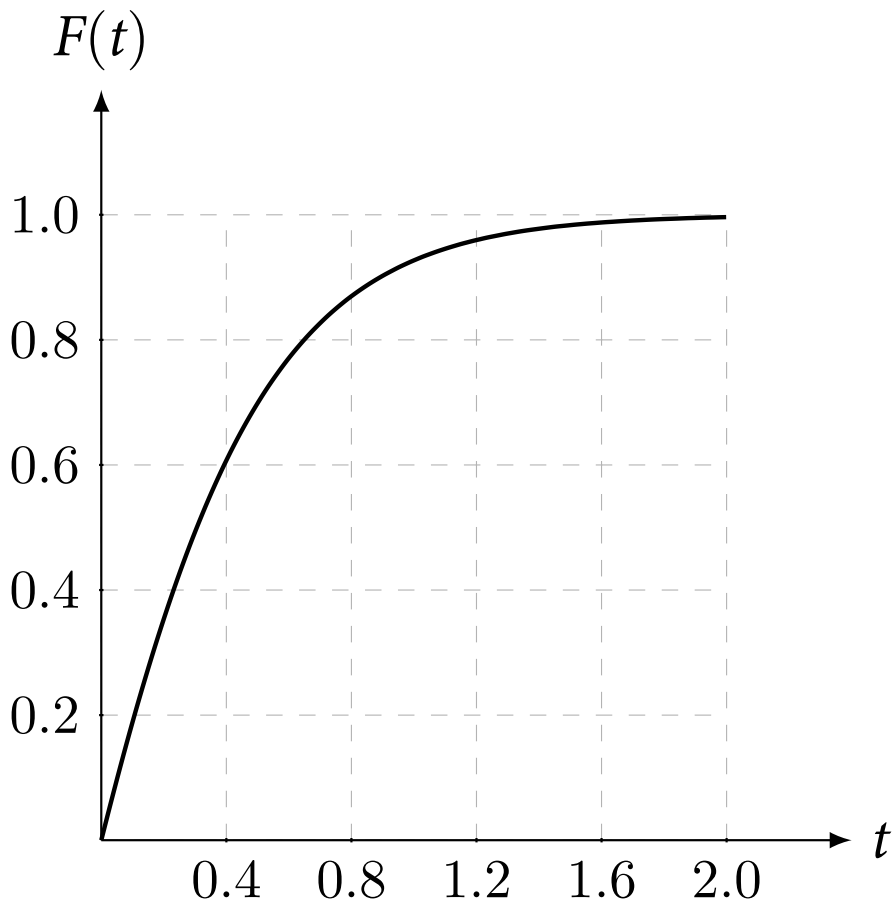
$$\frac{\mathrm{d}}{\mathrm{d}t}F(t) = (p + qF(t)) \cdot (1 - F(t)).$$

- When $F(t) = 0$, $\frac{\mathrm{d}}{\mathrm{d}t}F(t) = p$.
  - everything is just happening from the spontaneous adopters.
  - no imitation going on because there's nobody to imitate.
- When $F(t) = \epsilon$, $\frac{\mathrm{d}}{\mathrm{d}t}F(t) = (p + q\epsilon)(1 - \epsilon)$.
  - to get initial convexity, need $(p + q\epsilon)(1 - \epsilon) > p$.

  $\Rightarrow$ $q > p$.
- When $F(t)$ nears 1, $\frac{\mathrm{d}}{\mathrm{d}t}F(t) = 0$.
  - the fraction of people who haven't adopted yet becomes small, and that's what gives you the last part of the curve.

Getting the S-shape (Cont.)





Left: $q > p$; Right: $q < p$.

Getting the S-shape (Cont.)

- Solution:

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

- The ratio of q to p determines the overall shape of the curve.
- Initially only p matters, then q takes over.

Applications

- This model has been used extensively in forecasting by trying to estimate from initial take up.
- As this process takes off, you do not need much data to begin to analyze and form estimates of p and q.
- Once you have got the estimates of p and q, you can get estimates of what the rest of the process is going to look like.
  - how many people go see the movie in the first week?
  - how many people go see the movie in the second week?
  - based on that first week and second week, you may predict something.

Section 3

Diffusion on random networks

Questions

- When do we get diffusion?
- What is the extent of diffusion?
- How does it depend on the particulars of the process as well as the network?
- Who is likely to be infected earliest?

Component structure

- Reach of contagion is determined by the component structure.
- The component structure of a network naturally partitions a society into separate groups who do not interact or communicate with each other.
- The component structure serves as a natural first limit on the extent of diffusion or contagion when we examine diffusion through a social network.
- The component structure is also important to understand with respect to navigation, as if one can only follow paths in the network, then components are again natural barriers.

Extent of diffusion

- Get nontrivial diffusion if someone in the giant component is infected/adopts.
- Size of the giant component determines likelihood of diffusion and its extent.
- Poisson random network models allow for giant component calculations.

Subsection 1

Giant component

Giant component

- How big is the giant component when there is one?
- If $p < \frac{1}{n}$, then all are isolated.
- If $p > \frac{\log n}{n}$, then all path connected.

Size of the giant component

- q: fraction of nodes in largest component.
- For any node, the chance it is in the giant component is q.
- Observation: Chance that this node is outside of the giant component is the chance that all of its neighbors are outside of the giant component.

Size of the giant component (Cont.)

- Probability that a node is outside of the giant component: $1 - q$.
- It is equal to the probability that all of its neighbors are outside.

$\Rightarrow$ It is equal to $(1 - q)^d$, where d is the node's degree.

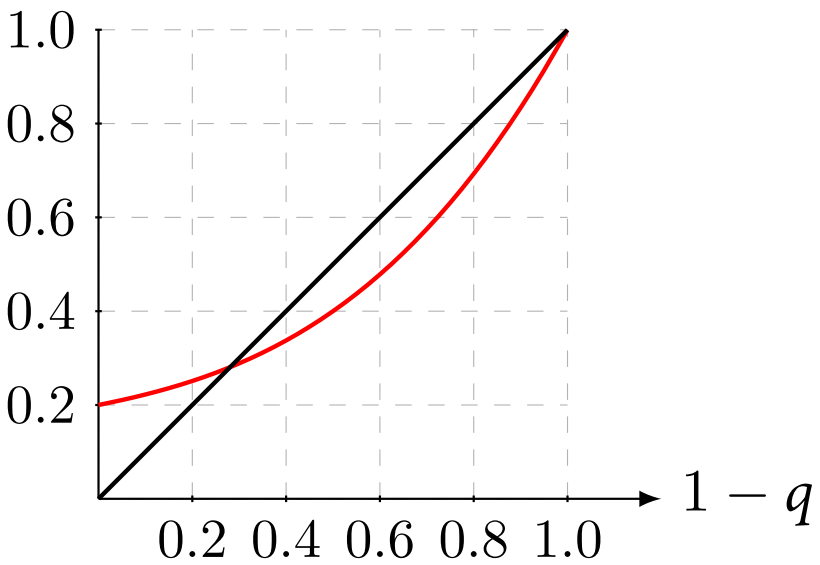
- So, probability $1 - q$ that a node is outside of the giant component is

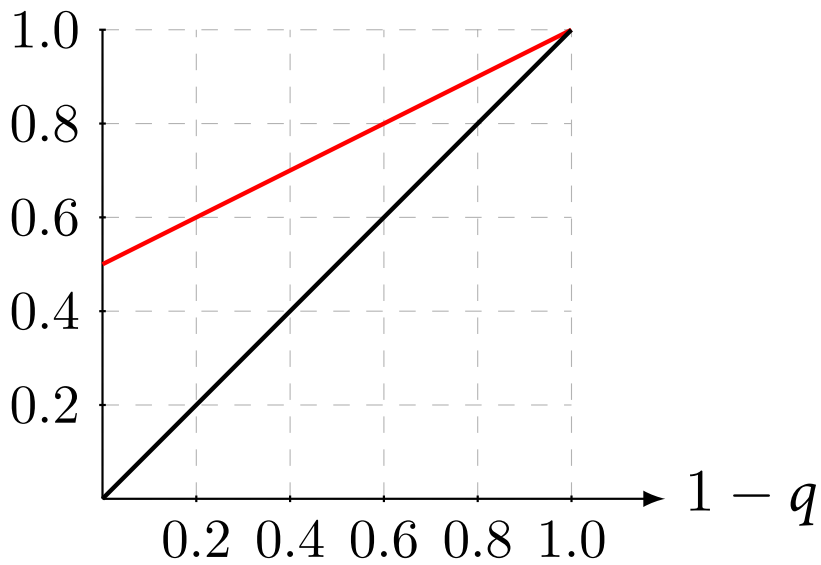
$$1 - q = \sum_d (1 - q)^d P(d),$$

where $P(d)$ is the chance that the node has d neighbors.

Size of the giant component (Cont.)

$$1 - q = \sum_d (1 - q)^d P(d).$$





Size of the giant component (Cont.)

- In a Poisson random network:

$$P(d) = \frac{(n-1)^d p^d e^{-(n-1)p}}{d!}.$$

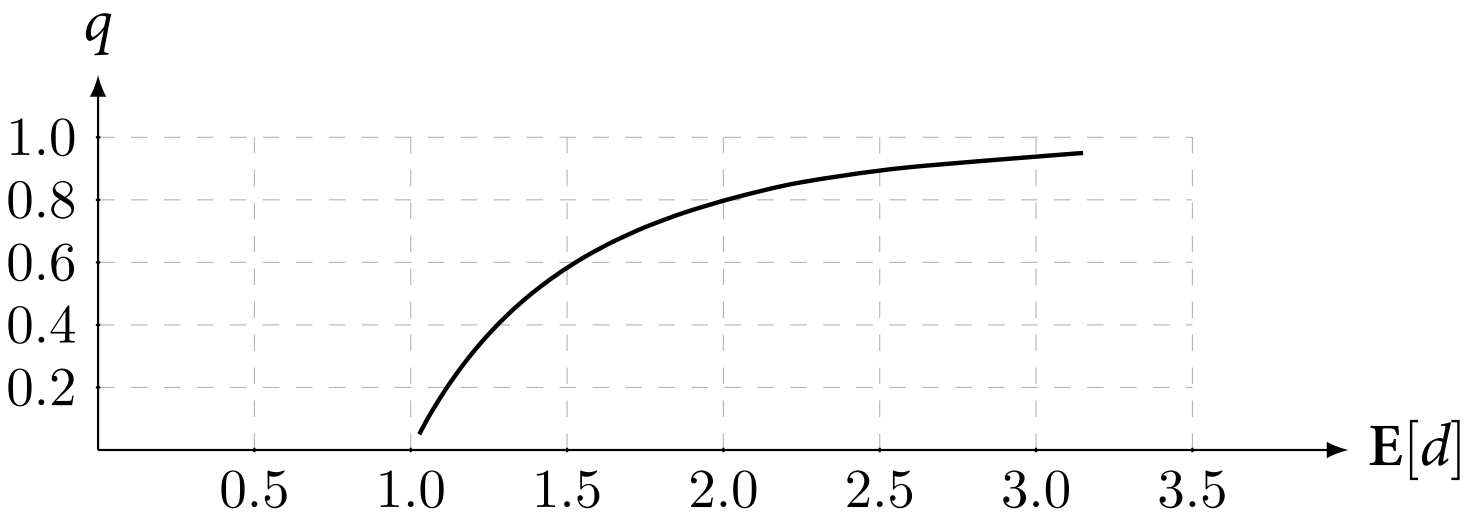
- So

$$1 - q = e^{-(n-1)p} \sum_d \frac{[(1-q)(n-1)p]^d}{d!}$$
$$\approx e^{-(n-1)p} e^{(1-q)(n-1)p} = e^{-pq(n-1)}.$$

$$\Rightarrow \frac{\log(1-q)}{q} = (n-1)p = \mathbf{E}[d].$$

Size of the giant component (Cont.)

$$\frac{\log(1-q)}{q} = (n-1)p = \mathbf{E}[d].$$



Who is infected?

- Probability of being in the giant component:

$$1 - (1 - q)^d.$$

- It is increasing in d.

$\Rightarrow$ More connected, more likely to be infected.

Subsection 2

Contagion with immunity and link failure

Extensions

- Immunity: delete a fraction of nodes and study the giant component on remaining nodes.
- Probabilistic infection.
  - Random infection: have some links fail, just lower p

Contagion with immunity and link failure

- Some node is initially exposed to infection.
- π of the nodes are immune naturally.
- only some links result in contagion: fraction f.
- What is the extent of the infection?

Contagion with immunity and link failure (Cont.)

- Consider a random network on n nodes.
- Delete fraction π of the nodes.
- Delete fraction $1 - f$ of the links.
- If starts at a node in giant component of the remaining network, then the giant component of that network is the extent of the infection; otherwise negligible.

Contagion with immunity and link failure (Cont.)

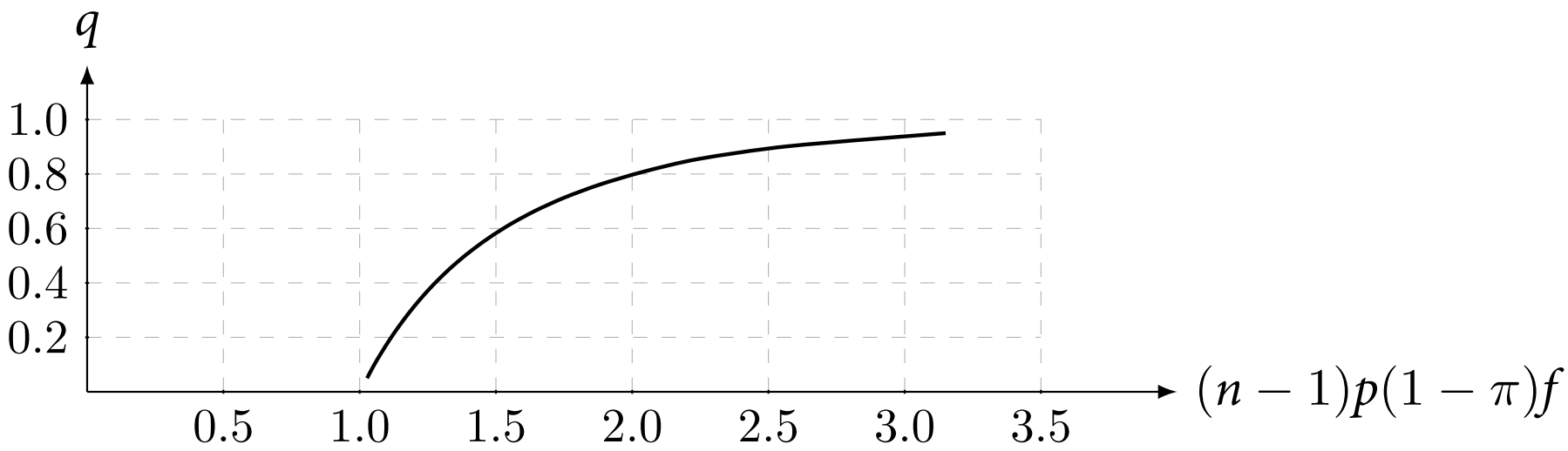
- Let q be the fraction of nodes of the remaining network in its giant component.
- $q(1-\pi)$ is the probability of a nontrivial contagion.
- Conditional on a contagion it infects $q(1-\pi)$ of the original nodes.

$\Rightarrow$ q solves

$$-\frac{\log(1-q)}{q} = (n-1)p(1-\pi)f.$$

Contagion with immunity and link failure (Cont.)

$$-\frac{\log(1-q)}{q} = (n-1)p(1-\pi)f.$$



Implications

- Infection can fail if π is high enough or f or p are low enough.
- High π: immunization, low virulence.
- Low f: low contagiousness.
- Low p: low contact among population.

Section 4

SIS model

SIS model

- In the SIS model, a node can be in one of 2 states:
  - Susceptible (易感): Before the node has caught the disease, it is susceptible to infection from the neighbors.
  - Infected (感染): Once the node has caught the disease, it is infectious and has some probability of infecting each of its susceptible neighbors.
    After recovering from infected, the nodes become susceptible again (rather than being removed).
- Models diseases such as certain variations of the common cold.

SIS model: Comments

- It is a simple model of diffusion.
- It is highly stylized and not directly applicable to a lot of things.
- But it is a useful model because it gives us some basic intuitions of how things work.
- A lot of the insights that will come out of this model will be quite useful, even if the model is a little bit too simple and stark to actually match a lot of things.

SIS model (Cont.)

- Probability that get infected is proportional to number of infected neighbors with rate $\nu > 0$, plus spontaneous ϵ.
- Get well randomly in any period at rate $\delta > 0$.
- Let $\rho = \rho(t)$ be the percent infected.

Subsection 1

Uniformly random infection

Uniformly random infection

- Randomly meet an individual each period.
- $$\frac{\mathrm{d}}{\mathrm{d}t}\rho = (1-\rho)(\nu\rho + \epsilon) - \rho\delta.$$
- Steady state: $\frac{\mathrm{d}}{\mathrm{d}t}\rho = 0$.

$\Rightarrow$

$$\rho = \frac{(\nu - \delta - \epsilon) + \sqrt{(\nu - \delta - \epsilon)^2 + 4\epsilon\nu}}{2\nu}.$$

Without ϵ

- $$\frac{\mathrm{d}}{\mathrm{d}t}\rho = (1-\rho)\nu\rho - \rho\delta.$$
- $\frac{\mathrm{d}}{\mathrm{d}t}\rho = 0$ has two solutions:
  - $\rho = 0$.
  - $\rho = 1 - \frac{\delta}{\nu}$ (if it is positive).

Implications

- $\rho = 1 - \frac{\delta}{\nu}$.
- If $\delta > \nu$, then recover faster than get sick.

$\Rightarrow$ No infection stays.

- If $\delta < \nu$, infection stays at some level.

* Low recovery rates can lead to large infections.

Limitations

- Missing heterogeneity in degree.
- Missing local patterns.

Subsection 2

Degree-based random meeting model

Degree-based random meeting model

- Nodes interact randomly according to their degree d_i.

  Let $P(d)$ be the degree distribution in the society.
- ν: the transmission rate of infection.
- δ: the recovery rate of an infected individual.
- $\rho(d) = \rho(d, t)$: fraction of nodes of degree d infected.
- $\theta = \theta(t)$: probability that a given meeting is with an infected individual.

Degree-based random meeting model (Cont.)

- $P(d)$: fraction of nodes that have d meetings.
- The probability that a meeting of node i is with a degree-d node is $\frac{P(d)d}{\mathbf{E}[d]}$.
- It is essential to keep track of nodes degrees since nodes with different degrees tend to have different infection rates.
- So likelihood of meeting infected node is:

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

Steady state

- Steady state: for each d,

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\rho(d) = (1-\rho(d))\nu\theta d - \rho(d)\delta.$$

- $\rho(d) = \frac{\lambda\theta d}{\lambda\theta d+1}$, where $\lambda = \frac{\nu}{\delta}$.
  $\rho(d)$ is increasing in d.

$\Rightarrow$

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

θ for regular degree distributions

- Consider regular networks: each node has a degree $\mathbf{E}[d]$.
- $$\theta = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d + 1)\mathbf{E}[d]} = \frac{\lambda\theta\mathbf{E}[d]}{\lambda\theta\mathbf{E}[d] + 1}.$$
- It is positive only if $\mathbf{E}[d] > \frac{1}{\lambda} = \frac{\delta}{\nu}$.
- Intuition: If the number of meetings is large enough relative to the relative recovery/infection rate, then the infection can be sustained. Otherwise, any infection will die out.

θ for power law degree distributions

- $P(d) = 2d^{-3}$.
- $$\theta = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d + 1)\mathbf{E}[d]} = \sum_{d=1}^{\infty} \frac{2\lambda\theta}{\mathbf{E}[d](d^2\lambda\theta + d)}.$$

$\Rightarrow$ $1 = \frac{2\lambda}{\mathbf{E}[d]} \log(1 + \frac{1}{\lambda\theta})$.

- Note that $\mathbf{E}[d] = 2$.
- $\theta = \frac{1}{\lambda(e^{\frac{1}{\lambda}} - 1)} \in (0, 1)$.

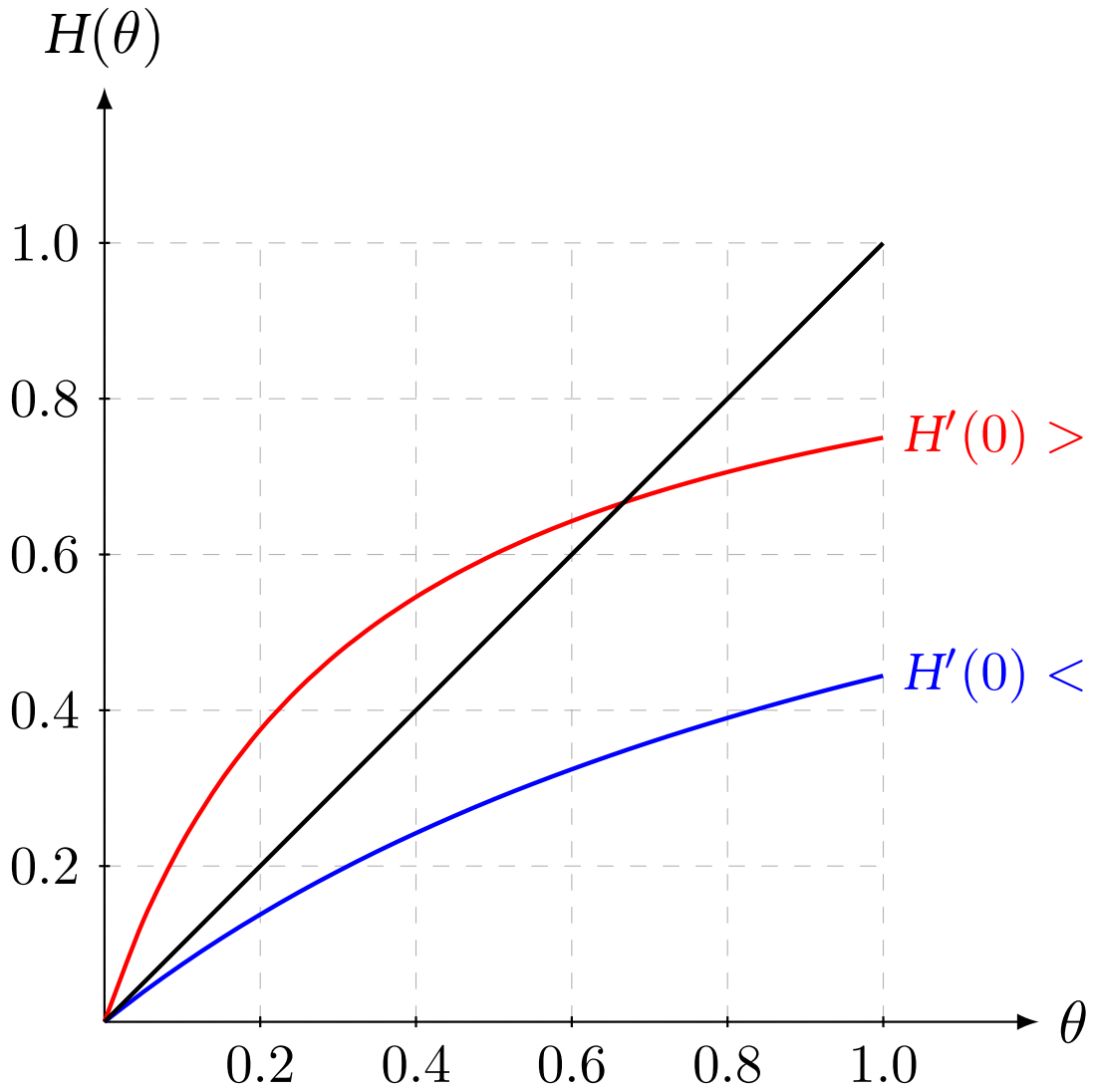
Properties of $H(\theta)$

- Let $H(\theta) = \sum_d \frac{P(d)\lambda\theta d^2}{(\lambda\theta d+1)\mathbf{E}[d]}$.
- $H(0) = 0$.
- $H(1) = \sum_d \frac{P(d)d}{\mathbf{E}[d]} \frac{\lambda d}{\lambda d+1} \leq \sum_d \frac{P(d)d}{\mathbf{E}[d]} = 1$.
- $H'(\theta) = \sum_d \frac{P(d)\lambda d^2}{(\lambda\theta d+1)^2\mathbf{E}[d]} > 0$.

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

  H is strictly concave.

Graph of $H(\theta)$



$H'(0) > 1$, positive steady state

$H'(0) < 1$, no positive steady state

Nonzero steady state

- $H'(0) = \lambda \frac{\mathbf{E}[d^2]}{\mathbf{E}[d]}$.
- There exists a nonzero steady state if and only if

$$\frac{\nu}{\delta} = \lambda > \frac{\mathbf{E}[d]}{\mathbf{E}[d^2]}.$$

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

Nonzero steady state (Cont.)

- Regular network: $\lambda > \frac{1}{\mathbf{E}[d]}$.
- Poisson random network: $\lambda > \frac{1}{1+\mathbf{E}[d]}$.
- Power law network: $\mathbf{E}[d^2]$ diverges $\Rightarrow$ always has a nonzero steady state.

Ideas

- High degree nodes are more prone to infection.
- 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.
- Higher variance, more such nodes to enable infection.