Lecture 9: Diffusion through Networks and Societies

Alexander Wolitzky

MIT

6.207/14.15: Networks, Spring 2022

Plan

Having covered basics of graph theory and random networks, now ready to start analyzing processes and behavior on networks.

Start with classic models of "diffusion" (of a disease, idea, or product) through society.

These are called **compartmental models**, because at each point in time each individual is in one of several states/"compartments."

- **Susceptible (S):** Individual has not yet been infected, and is susceptible to infection from others.
- **Infectious (I):** Individual is currently infected, and can pass the infection to others.
- **Removed (R):** Individual is no longer infected and cannot be infected again, either due to immunity or death.

These models are very important in fields like epidemiology and marketing. As we'll see, there are also tight (and somewhat surprising) connections with the ER and configuration models.

Different Compartmental Models: Preview

By including different compartments in the model, we can capture different types of phenomena. We'll cover three different compartmental models.

SI Model: Once you're infectious, stay infectious forever.

- As long as the network is connected or there's any mixing in society, the infection eventually reaches everone.
- This doesn't describe most diseases, but does describe the spread of a superior idea or product that is eventually adopted by everyone.
- Such ideas or products can also be discovered through processes other than infection by others.
- The **Bass model** is a generalization of the SI model that allows such "innovation" in addition to "imitation"/"infection" by others.

Different Compartmental Models: Preview (cntd.)

SIR Model: Once infected, you're infectious for a while, and then recover forever.

- A good model of diseases you can only catch once, like chicken pox, measles, or a particular strain of influenza or Covid.
- Now the disease typically doesn't reach everyone, because eventually the number of susceptible people is low enough that most infectious people recover without infecting anyone else. (More precisely, reproduction number R_t drops below 1.)
- The point where this happens is the **herd immunity threshold**, which plays a key role in the model.
- Eventually, everyone is recovered, and the epidemic ends.

Different Compartmental Models: Preview (cntd.)

SIS Model: Like SIR, but once you recover you can get infected again.

- A good model of diseases you can catch many times, like the cold.
- Now the epidemic never ends, and instead a **steady state** is reached, where at each point in time the number of people who recover equals the number who get infected.
  (That is, $R_t = 1$.)

Population Heterogeneity

The simplest versions of all these models assume a **homogeneous population:** everyone is equally likely to meet everyone else.

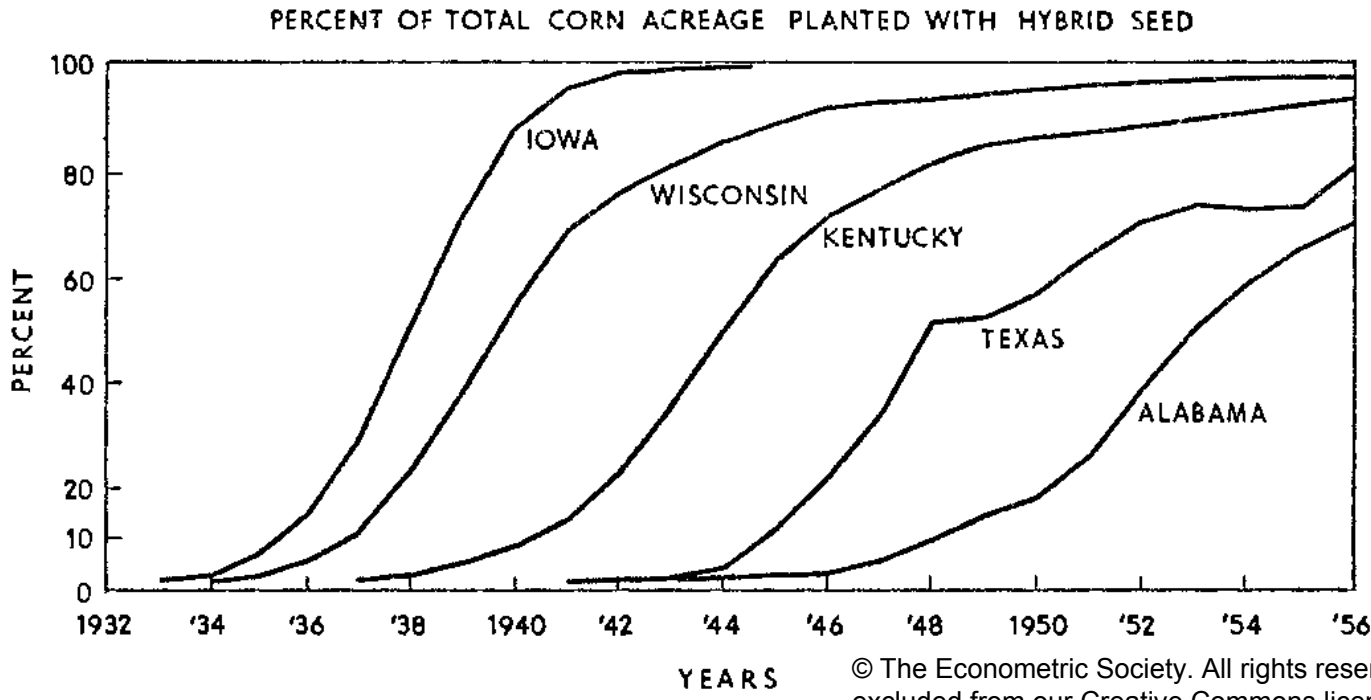
A tractable, realistic, and important extension of the model is to instead assume **heterogeneous contact rates:** some people are more active than others (and hence have more contacts each period).

- E.g., in the SIR model, the herd immunity threshold can be much lower with heterogeneous contact rates, because high-activity people get sick first and then recover.
- Also affects how we can fight epidemics: e.g., should we prioritize high-contact people like essential workers for vaccination?

We'll consider versions of the SIR and SIS models with both homogeneous and heterogeneous populations.

- Next week: more on targeting and other strategic aspects of diffusions.

Diffusion of Innovations



Percentage of total corn acreage planted with hybrid seed (Ryan and Gross, 1943; Griliches, 1957).

- S-shaped adoption curves indicative of social learning / "word-of-mouth".
- What kind of formal models generate this pattern?
- When is adoption S-shaped vs. concave?

Bass Model

The Bass model simply tracks how a product or technology spreads through a large population when individuals can adopt the innovation as a result of **innovation** (discovering it on their own) or **imitation** (discovering it from others).

- Let $F(t)$ be the fraction of the population who have adopted the product by time t.
- Let p be the "innovation rate".
- Let q be the "imitation rate".

The **Bass model** is given by the difference equation

$$F(t+1) - F(t) = (1 - F(t))(p + qF(t))$$

- Of fraction $1 - F(t)$ who haven't yet adoped, fraction p innovate (adopt on their own) and fraction $qF(t)$ meet an adopter ($F(t)$) and imitate (q; copy the adopter).
- Assume $p + q \leq 1$, so $F(t+1)$ stays below 1.
- The **SI model** is the special case where $p = 0$.

Analysis

Easier and more common to analyze the continuous-time version of the model:

$$\dot{F}(t) = (1 - F(t))(p + qF(t)),$$

with $F(0) = 0$
(where dot denotes time derivative, $\dot{F}(t) = dF(t)/dt$).

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

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

- The resulting curve $F(t)$ is the **adoption curve** of the product.
- Note: the sum $p + q$ scales time (how "fast" the innovation diffuses), while the ratio $\frac{q}{p}$ determines the shape of the curve.
- If $p > q$ the adoption curve is **concave**; if $p < q$ the adoption curve is **S-shaped**.

Concave or S-Shaped Adoption?

Whether adoption curve is concave or S-shaped depends on whether it is concave or convex at low adoption level (it's always concave at high adoption levels, as the market becomes saturated).

Whether curve is concave or convex at low adoption levels is determined by a race between two forces:

- Increasing adoption speeds up adoption as there are more adopters to imitate.
  This force is stronger when q is higher.
- Increasing adoption slows down adoption as there are fewer non-adopters who can innovate.
  This force is stronger when p is higher.

An easy calculation shows that if $p < q$ then the first force "wins" and adoption is S-shaped; while if $p > q$ then the second force wins and adoption is concave.

In Reality, p is Usually Less Than q

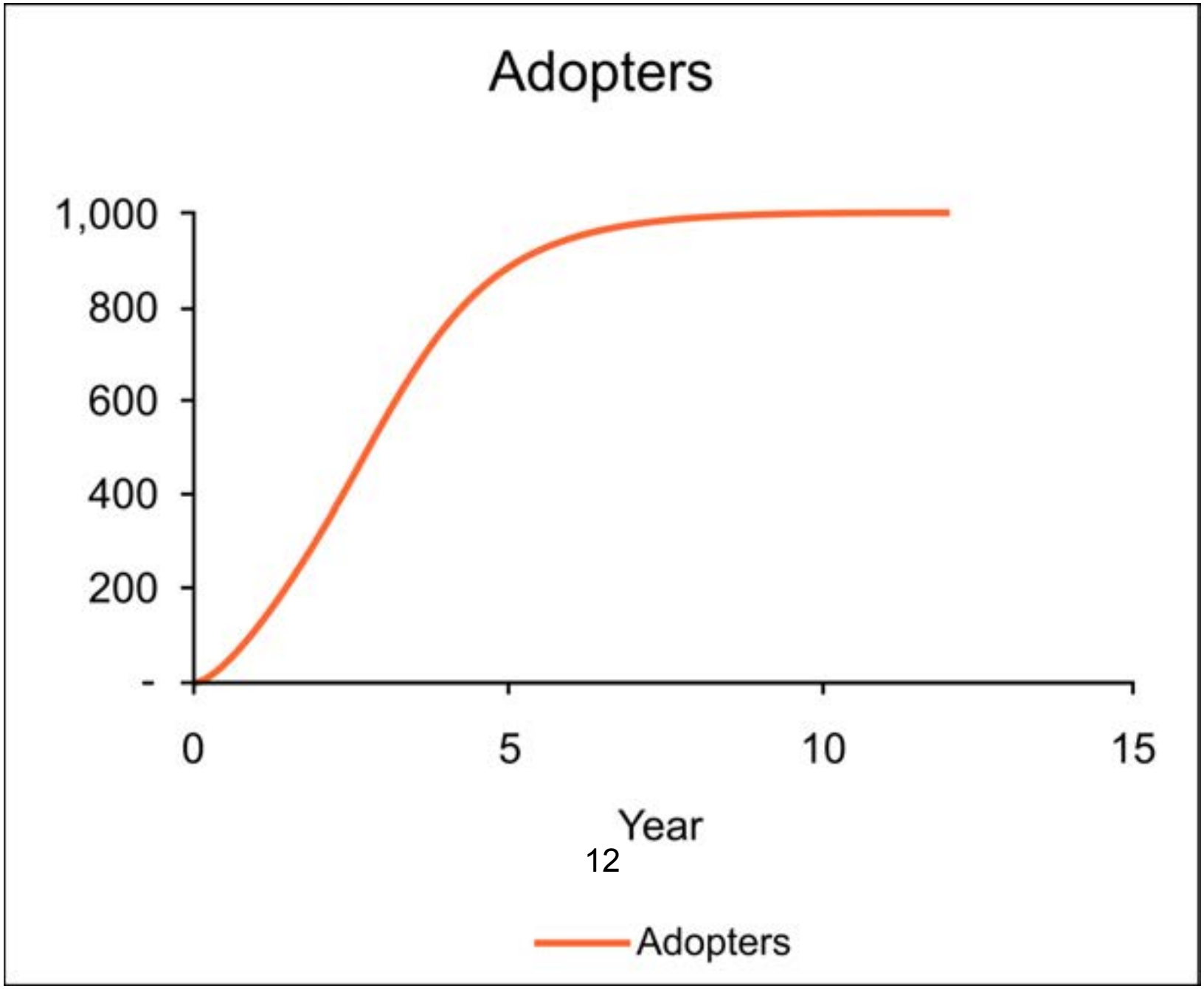
Marketing researchers have attempted to estimate the parameters of the Bass model for a wide range of products.

- Typical estimates find $p \approx 0.03$ and $q \approx 0.4$ (where t is measured in years).
- These numbers should be taken with a large grain of salt, but the finding that the imitation rate is often an order of magnitude greater than the innovation rate appears to be robust.
- This provides an explanation for the prevalence of S-shaped adoption curves, such as the the diffusion of hybrid corn in the US Midwest.

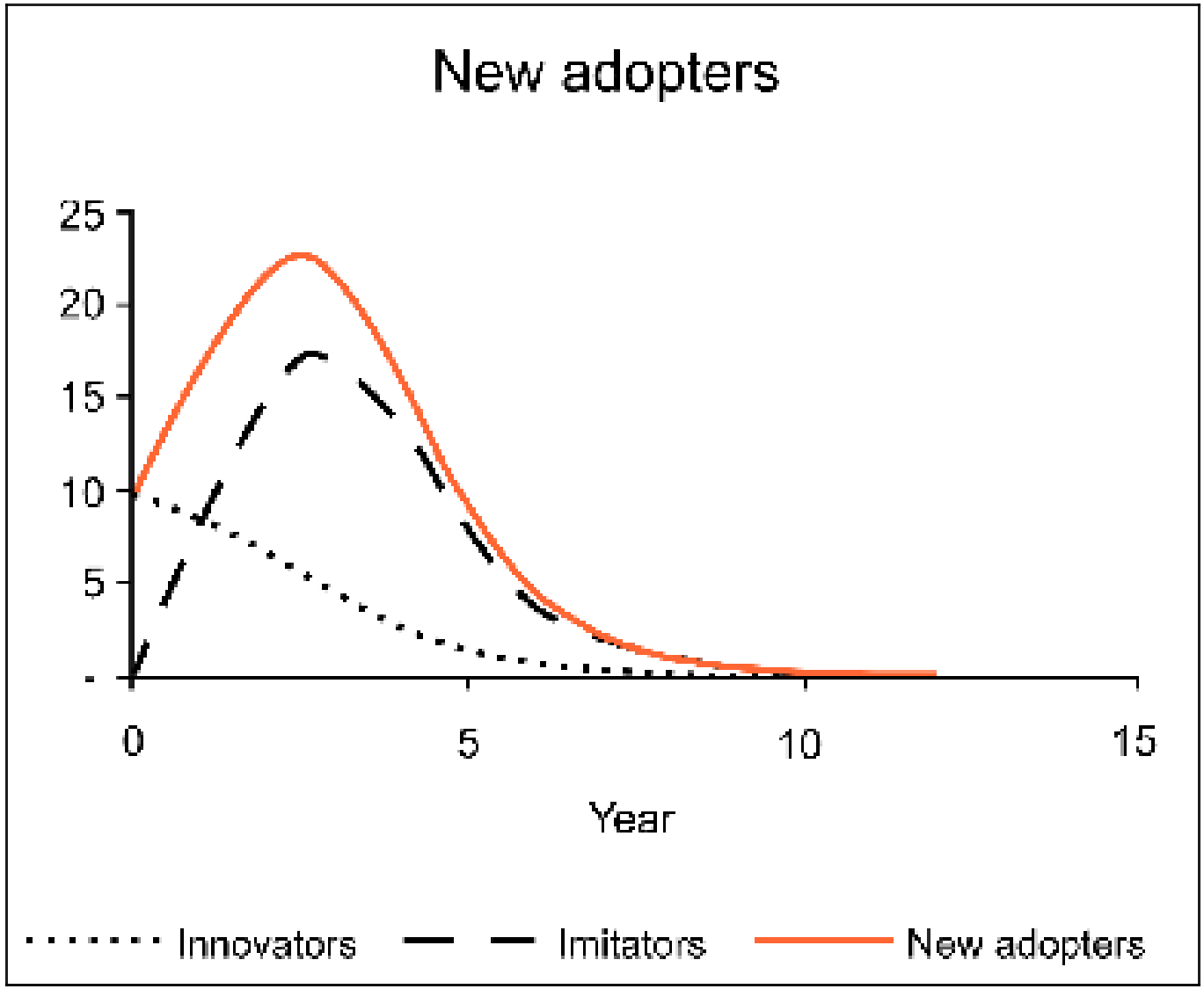
In addition to explaining S-shaped adoption curves, the Bass model can also be used to predict the future path of adoption based on the early adoption pattern.

- Same thing happens in the SIR model. This is a main way of predicting the path of an epidemic.

Adoption Curve (p=0.03, q=0.38)



New Adopters (p=0.03, q=0.38)



The SIR Model

Let's turn to the classic SIR epidemiology model.

- Like Bass model with "adopter"="infectious", eventual recovery from infection, and no innovators.
- The SIR model was the main model used early on to predict the path of Covid. We will see what it's strengths and weaknesses are in this regard.

At each point in time t, the population is divided into three compartments:

- Fraction $S(t)$ are **susceptible**
- Fraction $I(t)$ are **infectious**
- Fraction $R(t)$ are **recovered** or **removed**
- For each t, we have $S(t) + I(t) + R(t) = 1$.

The model has two parameters:

- β =transmission rate
- γ =recovery rate

SIR Model: Equations

The SIR model is defined by the following three equations:

$$\dot{S}(t) = -\beta S(t) I(t) \tag{1}$$

- Each of $S(t)$ susceptible people meets an infectious person w/ prob $I(t)$. When this happened, gets infected w/ prob β.
- $S(t)$ decreases over time.

$$\dot{I}(t) = \beta S(t) I(t) - \gamma I(t) \tag{2}$$

- Inflow into state I: $\beta S(t) I(t)$ susceptible people get infected.
- Outflow from state I: $\gamma I(t)$ infected people recover.
- Over time, $I(t)$ first increases and then decreases.

$$\dot{R}(t) = \gamma I(t) \tag{3}$$

- The $\gamma I(t)$ infected people who recover go to state R.

The Basic Reproduction Number

With transmission rate β and recovery rate γ, on average an infectious person will infect β/γ others before she recovers, if everyone else is susceptible.

- The number β/γ is called the **basic reproduction number** of the disease, and is denoted by R_0.

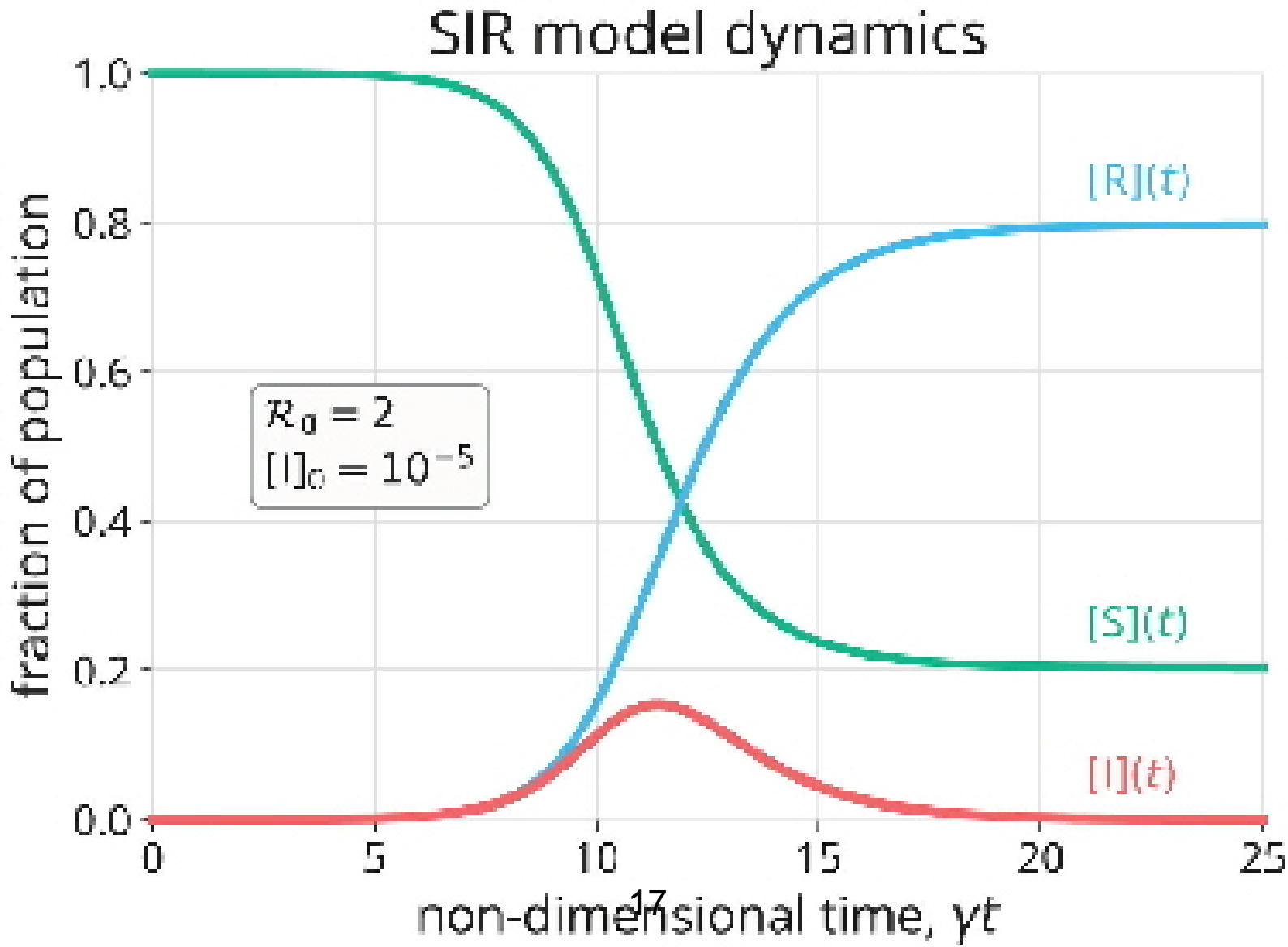
Since $\beta = \gamma R_0$, we can rewrite the SIR equations as

$$\begin{aligned} \dot{S}(t) &= -\gamma R_0 S(t) I(t) \\ \dot{I}(t) &= \gamma R_0 S(t) I(t) - \gamma I(t) \\ \dot{R}(t) &= \gamma I(t). \end{aligned}$$

Note that $\dot{S}$, $\dot{I}$ and $\dot{R}$ are all multiplied by γ.

- This means that, once we fix $R_0 = \beta/\gamma$, γ just changes the timescale of the epidemic process.
- Therefore, quantities that don't directly reference time (e.g., herd immunity threshold, how many people will ultimately become infected) depend only on R_0.

SIR Model: Illustration



SIR Model: Observations

The SIR model cannot be completely solved in closed form, but it's easy to evaluate numerically.

We can also make some important analytical observations:

- The infectious share of the population is maximized when $\dot{I}(t) = 0$, or
$$R_0 S(t) = 1.$$
- This is the **herd immunity threshold:** the point where $S(t)$ has dropped far enough that an infected person infects on average 1 other person before she recovers.
- But "herd immunity" is a bit of a misnomer: people keep getting infected after herd immunity is reached, albeit at a decreasing rate.
- $R(t)$ keeps increasing and asymptotes to a value $R(\infty) = 1 - S(\infty)$.
- The difference between the herd immunity threshold and $R(\infty)$ is the degree of **overshooting**.

SIR Model: History

The SIR model was developed by Kermack and McKendrick in 1927, building on work by Ross and Hudson in 1917.

Ronald Ross: British doctor who won Nobel prize for discovering that mosquitos spread malaria. Also an amateur novelist, poet, artist, and mathematician.

Hilda Hudson: English Algebraic geometer, first woman to speak at the International Congress of Mathematicians. Also wrote on the relationships between math and religion.

William Ogilvy Kermack: Scottish biochemist. Blind for almost his whole career due a chemical explosion in his lab.

Anderson Gray McKendrick: Scottish doctor, epidemiologist, and amateur mathematician. Several contributions to models of epidemics and population growth.

How Many People Ever Get Sick?

Solving for $R(\infty)$ (share of people who ever get sick):

- Since $\dot{S}(t) = -\gamma R_0 S(t) I(t)$ and $\dot{R}(t) = \gamma I(t)$, we have

$$\frac{\dot{S}(t)}{S(t)} = -R_0 \dot{R}(t).$$

- The solution to this differential equation with $S(0) = 1$ is

$$S(t) = e^{-R_0 R(t)}.$$

- Since $I(t)$ eventually goes to 0, we have

$$1 = S(\infty) + R(\infty) = e^{-R_0 R(\infty)} + R(\infty).$$

- Therefore,

$$R(\infty) = 1 - e^{-R_0 R(\infty)}.$$

How Many People Ever Get Sick? (cntd.)

We showed that the share of people who ever get sick is

$$R(\infty) = 1 - e^{-R_0 R(\infty)}.$$

Intuition:

- $R(\infty)$ people ultimately get sick.
- Each has on average R_0 meetings that would cause infection with someone who is susceptible.
- The distribution of the number of such meetings is Poisson (since meetings are independent).
- Therefore, probability of **not** getting infected is $e^{-R_0 R(\infty)}$.

Interestingly, $R(\infty)$ is exactly the size of the giant component in the ER model with $\lambda = R_0$!

This is useful: for example, we know what $R(\infty)$ looks like as a function of R_0.

- $R(\infty) = 0$ whenever $R_0 \leq 1$, then increases rapidly as a concave function of R_0.

Other Lessons from the SIR Model

- $I(t)$ grows exponentially when $S(t) \approx 1$ and falls exponentially when $S(t) \approx 0$, so roughly normally distributed.
- If only temporary interventions to slow the disease are available (e.g., lockdowns but no vaccines), it is impossible to avoid reaching the herd immunity threshold, but overshooting can be reduced.
  - If impose lockdown as soon as herd immunity is reached, this brings $R(\infty)$ down to $1 - 1/R_0$.
  - So, can be substantial gains from lockdowns even if vaccines/improved treatments do not arrive, but greater gains from locking down until vaccines/treatments arrive (but also of course greater costs from longer lockdowns).
- Vaccinating fraction π of the population reduces $S(0)$ to $1 - \pi$ (and increases $R(0)$ to π), which can greatly reduce $R(\infty)$.

Next class, more such lessons in the context of Covid-19.

Heterogeneous-Agent SIR

The SIR equations can easily be generalized to allow K types of people with different β's, and hence different R_0's.

- For simplicity, assume $1/K$ people of each type.
- Assume meetings are uniform, so each meeting of a type-i agent is with a type-j agent with probability $R_{0j} / \sum_k R_{0k}$.

Heterogeneous-Agent SIR (cntd.)

This model is harder to analyze, but can show that the initial growth rate of the epidemic is now, rather than $R_0 - 1$ as in the homogeneous-agent model,

$$\bar{R}_0 - 1 = \frac{\mathbb{E}\left[R_{0i}^2\right]}{\mathbb{E}\left[R_{0i}\right]} - 1.$$

- $\bar{R}_0$ equals the reproduction number (conditional expected degree) in the configuration model with degree distribution given by the distribution of R_{0i}'s.

- Intuitively, variance in the R_{0i}'s accelerates the epidemic early on, because infections concentrate among the most active agents.
- However, same force implies that herd immunity is reached earlier (i.e., with more susceptible agents) in the heterogeneous-agent SIR model than in the homogeneous-agent model with $R_0 = \bar{R}_0$.

The SIS Model

The SIS model differs from the SIR model in that now an individual becomes susceptible again after recovering.

The equations of the SIS model are

$$\begin{aligned} \dot{S}(t) &= \gamma I(t) - \beta S(t) I(t), \\ \dot{I}(t) &= \beta S(t) I(t) - \gamma I(t), \end{aligned}$$

where again β is the transmission rate and γ is the recovery rate (and $S(t) + I(t) = 1$).

With $R_0 = \beta/\gamma$, can rewrite as

$$\begin{aligned} \dot{S}(t) &= \gamma I(t) - \gamma R_0 S(t) I(t), \\ \dot{I}(t) &= \gamma R_0 S(t) I(t) - \gamma I(t). \end{aligned}$$

Steady-State Infection Level

We could analyze the **initial spread** of the epidemic in the SIS model similarly to the SIR model, but the most important prediction of the SIS model is the **steady-state infection level**.

The steady-state infection level I is given by

$$\dot{S} = \gamma I - \gamma R_0 (1 - I) I = 0.$$

Solving for I gives

$$I = \begin{cases} 1 - \frac{1}{R_0} & \text{if } R_0 \geq 1 \\ 0 & \text{if } R_0 < 1 \end{cases}.$$

- Somewhat similar shape to size of giant component as a function of λ.

If vaccinate (remove) fraction π of the population then $S(t) + I(t) = 1 - \pi$, so the steady-state equation infection level is reduced to $1 - \pi - \frac{1}{R_0}$.

Heterogeneous-Agent SIS Model

Since the steady state of the SIS model is so simple, we can fully solve it even with heterogeneous agents.

- Preview: a main result will be that the steady-state infection level is positive if and only if basic reproduction number > 1, as in the earlier models.

- Let $P(d)$ be the share of the population with d meetings ("degree") each period.
- Assuming uniform matching: each meeting is with a degree-d individual with probability

$$\frac{P(d)\,d}{\langle d \rangle}.$$

- (Like heterogeneous-agent SIR, or configuration model.)

Heterogeneous-Agent SIS (cntd.)

- Let $I_d(t)$ be the share of degree-d nodes infected at time t.
- Then each **meeting** is with an infected individual with probability

$$\theta(t) = \frac{\sum_d P(d)\, d I_d(t)}{\langle d \rangle}.$$

  - This is a useful measure of the infection level.
  - **Note:** different from share of **people** who are infected, which is $\sum_d P(d)\, I_d(t)$.

Heterogeneous-Agent SIS (cntd.)

For each degree d, $I_d(t)$ evolves according to

$$\dot{I}_d(t) = \beta \underbrace{d\left(1 - I_d(t)\right)\theta(t)}_{\text{susceptible}\to\text{infected}} - \underbrace{\gamma I_d(t)}_{\text{infected}\to\text{susceptible}}$$

Hence, at a steady-state, for each d we have

$$\beta d\left(1 - I_d(t)\right)\theta(t) = \gamma I_d(t),$$

or

$$I_d = \frac{dR\theta}{dR\theta + 1},$$

where $R = \beta/\gamma$.

- Note: higher-degree nodes have higher steady-state infection rates.

Steady-State Infection Level

We can now compute the steady-state infection level.

Since

$$I_d = \frac{dR\theta}{dR\theta + 1} \text{ for each } d$$

and

$$\theta = \frac{\sum_d P(d)\, dI_d}{\langle d \rangle},$$

we see that there is a steady state with infection level θ if and only if

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

This formula is reminiscent of the formula $d^2 / \langle d \rangle$ for expected neighbor's degree: both formulas account for oversampling of high-degree nodes; here we also have the feature that high-degree nodes have higher steady-state infection levels.

Steady-State Infection Level

There is a steady state with infection level θ if and only if

$$\theta = \sum_d \frac{P(d)\, d^2 R\theta}{\langle d \rangle (dR\theta + 1)}$$

- $\theta = 0$ is always a steady-state: if no one is infected, system stays that way.

When is there a solution with $\theta > 0$?

- We'll see that there is a positive steady-state infection level iff $R\frac{\langle d^2 \rangle}{\langle d \rangle} > 1$: that is, iff basic reproduction number exceeds 1.

Fixed Points

To see when there is a positive steady-state infection level θ, we use the mathematical idea of a **fixed point** of a function.

- A **fixed point** of a function $f(x)$ is a value x such that $f(x) = x$.
- That is, a point where $f(x)$ crosses the 45° line.

Let

$$H(\theta) = \sum_d \frac{P(d)\, d^2 R\theta}{\langle d \rangle (dR\theta + 1)}.$$

- If the current infection level is θ, the infection level drifts towards $H(\theta)$.
- The steady-state levels of θ are precisely the fixed points of H.
- There is a positive steady-state infection level iff H has a non-zero fixed point.

Steady-State Infection Level (cntd.)

$$H(\theta) = \sum_d \frac{P(d)\, d^2 R\theta}{\langle d \rangle (dR\theta + 1)}.$$

- $H(0) = 0$, $H(1) < 1$, and $H(\theta)$ is increasing and strictly concave in θ.
- Therefore, H has a non-zero fixed point iff $H'(0) > 1$.

Differentiate $H(\theta)$ to find

$$H'(\theta) = \sum_d \frac{P(d)\, d^2 R}{\langle d \rangle (dR\theta + 1)^2}.$$

Plug in $\theta = 0$ to find

$$H'(0) = \sum_d \frac{P(d)\, d^2 R}{\langle d \rangle} = R\frac{d^2}{\langle d \rangle}.$$

Hence, there is a positive steady-state infection level iff

$$R\frac{d^2}{\langle d \rangle} > 1.$$

Steady-State Infection Level (cntd.)

There is a positive steady-state infection level iff

$$R\frac{d^2}{\langle d\rangle} > 1.$$

That is, iff reproduction number exceeds 1.

Examples:

- When d is the same for everyone (homogeneous SIS), the threshold is $Rd > 1$, as we saw.
- For power-law distributions (with $\gamma < 3$), d^2 is divergent, so this condition holds for any positive R.

Summary

- The Bass model is a classic model of the diffusion of innovations via innovation and imitation, which generates S-shaped adoption curves.
- The SIR model is a classic model of diseases you can only get once. The disease eventually dies out. Key concepts are herd immunity and overshooting.
- The SIS model is a classic model of diseases you can get many times. The key concept is the steady-state infection level.
- Both the SIR and SIS models behave differently in homogeneous and heterogeneous populations. Understanding hetereogeneity is important for predicting the course of the epidemic and assessing possible interventions.

14.15 / 6.207 Networks

Spring 2022