Network Epidemiology : Structure

Aaron Clauset, CSCI 3352 Biological Networks
5 & 7 November 2019

Modeling an epidemic : the compartmental model

Infectious diseases are illnesses that spread from one organism to another. The details can be enormously complicated. Do we consider the molecular mechanisms of infection? Do we consider what happens to individual cells? Do we account for how the immune system responds? Most epidemiology models that use networks focuses instead on understanding how a disease spreads through a population: how many are sick, how many could get sick, and how many have recovered.

In this case, we are modeling the population dynamics of the disease. When we overlay such a model on a network of nodes and edges, we get something called **network epidemiology**, where the shape of the network can influence the spread of the disease.

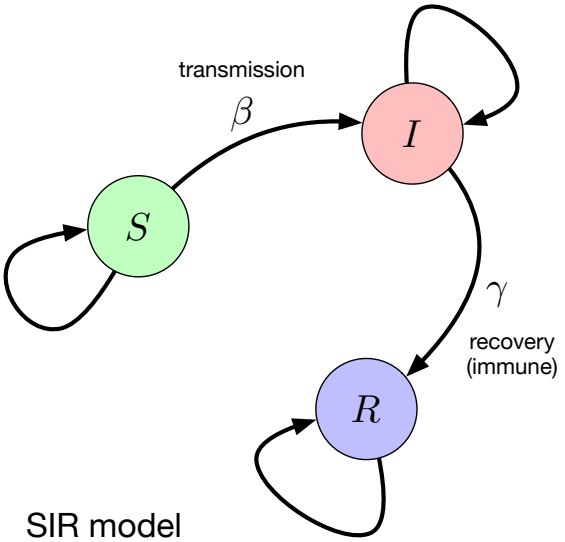
To begin, we'll first describe the basic **compartmental model** of epidemics, which you can imagine unfolding on a fully connected network, i.e., one in which every person is connected to every other person. In this setting, every individual can transmit an infection to, or receive an infection from, any other individual in the population. (This idea is equivalent to saying the population is "well mixed.")

Epidemiology is another example of *dynamics on a network*. Here, the network structure remains fixed, and instead each node has a **state variable** $x_i \in \{S, I, R\}$ or $x_i \in \{S, I\}$ that tells us which "compartment" a node is currently in, along with a set of "transition" rules describe how a node's state switches from one compartment to another. The basic compartmental model defines one compartment S for "susceptible" individuals, one compartment I for "infected" individuals, and, optionally, one compartment R for "recovered" individuals. Transitions among these compartments then define events like a susceptible individual becoming infected $S \to I$, an infected individual recovering with $I \to R$ or without $I \to S$ immunity.

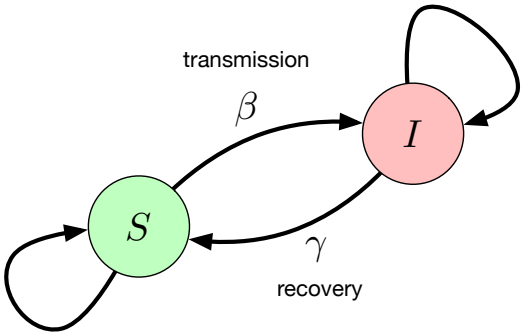
The particular set of states and transition rules we choose defines our model of the spread of a disease. For instance, we might model the infection stage of an illness by using a sequence of "infected" compartments, e.g., $I_1 \to I_2 \to I_3$, which would represent a progression of a disease from one distinct stage to another. Similarly, there might be multiple types of susceptible states, capturing different kinds of susceptibility. This kind of flexibility means that there are *many different models* of disease spread, e.g., a model for how influenza spreads should be different from a model of ebola because these diseases spread through different means, and have different consequences on the hosts.

The most basic epidemiology models are the **Susceptible-Infected-Recovered** or *SIR* model and the **Susceptible-Infected-Susceptible** or *SIS* model, which have the

following states and transitions:



SIR model



SIS model

- **SIS** can model diseases like influenza or the common cold, where recovery does not imply immunity (because there are other strains circulating).
- **SIR** is more realistic for diseases like chickenpox and measles, where recovery does imply immunity (because there's only one strain in circulation over the course of a human lifetime).

Two kinds of networks

Let's answer the two fundamental questions for using a network $G = (V, E)$ in this setting: *What are nodes? And, what are edges?*

Typically, nodes are people. (Or, if we're modeling disease spread in an animal population, nodes might be parakeets, or cats, etc.) Then, edges are potential exposures. If two nodes are connected, it means that one can expose the other to an illness. If that's not possible, then there cannot be an edge between those nodes. Hence, different modes of exposure define different networks.

For example:

- close proximity (influenza, common cold)
- physical contact with object harboring live pathogen (pink eye)
- physical contact with fluids (STDs
- contact with an animal vector (mosquitos and malaria)

Not all potential exposures will result in a *transmission*, and hence we can define a second network $G_T = (V, E_T)$ that captures only the actual transmissions that occur. In this network, edges are *directed* transmissions $T \subseteq E$, which are a subset of potential exposures. These edges are directed because an infected person receives their infection from a unique infected individual. Hence, G_T is a tree.

The mean out-degree $\langle k^{\text{out}} \rangle$ of the transmission network is equivalent to R_0, the **basic reproduction number** of a disease, which is the average number of secondary infections produced by each infected individual. This quantity is the first-order measure of how "spreadable" a disease is.

Specifying a SIR or SIS model

Both SIR and SIS models have parameters β, γ. These represent, respectively, the rate or probability of becoming infected, given an exposure, and the rate or likelihood of recovery, given an infection. The higher the value of β, the more likely an exposure leads to an infection. For instance, a highly contagious disease like measles will have a very high value of β, while something like the common cold will have a much lower value. Similarly, the higher the value of γ, the faster an infected individual recovers to become non-infectious again. The ratio β/γ provides a simple quantitative measure of the relative flow into the I compartment, and is equivalent to R_0.

Epidemic dynamics in SIS

The dynamics of the basic SIS compartmental model provide a reference point for our intuition about epidemic dynamics on a network. We can mathematically model the number of individuals in each of the two S, I compartment as follows. This model is a *continuous time* model, while in practice, network models are typically *discrete time* (with synchronous updates). More on that below.

- Let N be the total population (number of nodes), which we assume to be constant.
- An infected person has a probability $\bar{\beta}$ per unit time to pass the infection to some susceptible person.
- An infected person also has a probability γ per unit time to recover and become susceptible again.
- Let $S(t)$ and $I(t)$ denote the number of susceptible and number of infected individuals, and note that $N = S(t) + I(t)$.

Now, consider how $S(t)$ changes over a small interval of time:

$$S(t + \Delta t) = S(t) - \left[S(t) \times \bar{\beta} \Delta t I(t)\right] + \left[I(t) \times \gamma \Delta t\right]$$

- The first term is just the number of susceptibles we began with.
- The second term measures the flow of individuals out of the susceptible S compartment and into the infected I compartment, due to new infections. Crucially, this term is proportional to the product $S(t)I(t)$, because each susceptible is connected to every infected, and we assume that transmissions

are independent.

- Finally, the third term measures the flow of individuals back into the susceptible S compartment out of the infected I compartment, due to recovery. (In the SIR model, this term would flow into the R compartment, instead.)

We can convert this difference equation into a differential equation by dividing S, I, and $\bar{\beta}$ by N and Δt, yielding

$$s = S/N \qquad i = I/N \qquad \beta = \bar{\beta}N$$

and

$$\frac{\mathrm{d}i}{\mathrm{d}t} = -\beta i(1-i) - \gamma i$$

We can use this equation to answer a simple question: will a new epidemic spread?

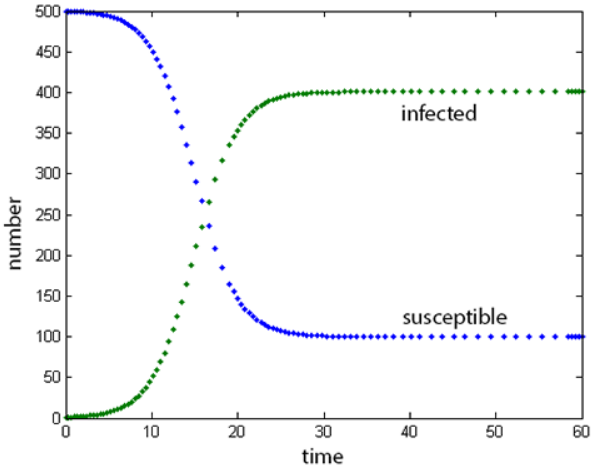
In this case, i will be very small, and we want to know how it will grow. Under this assumption, we make the approximation $(1-i) \approx 1$, which yields

$$\frac{\mathrm{d}i}{\mathrm{d}t} \approx i(\beta - \gamma)$$

Now, we can see clearly that $\beta > \gamma$, an epidemic will tend to spread. Here, now, we can see the mathematical connection between β/γ and R_0. In addition, we can see that there's a *critical value* at $\beta/\gamma = 1$, above which the epidemic tends to grow, while below which it tends to die out quickly. This value is called the **epidemic threshold**.

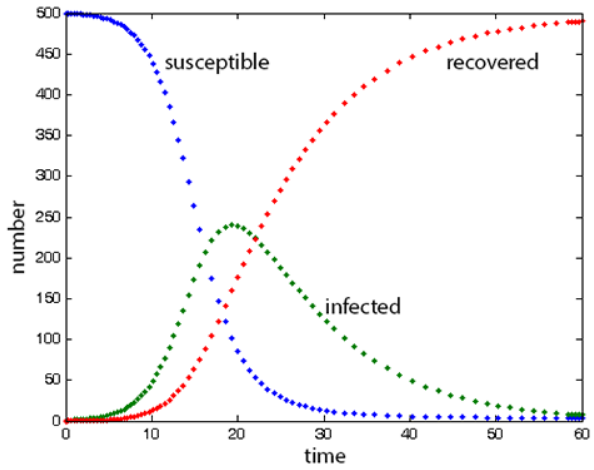
SIS and SIR dynamics

The SIS model produces dynamics like the following. Because individuals can go back and forth between the two compartments, the relative proportion of individuals in each stabilizes after some time, and is determined by the ratio β/γ.



[Adapted from: https://en.wikiversity.org/wiki/File:Sissys.png]

The SIR model produces dynamics like the following. Can you reason about the what value was chosen for the ratio β/γ here, based on the dynamics you see?

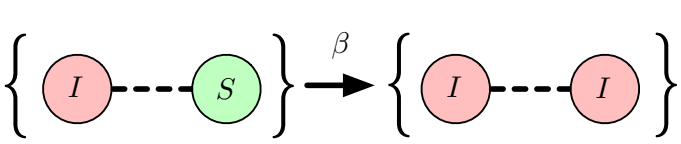


[Adapted from: https://en.wikiversity.org/wiki/File:Sirsys-p9.png]

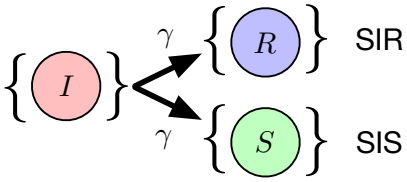
Epidemics on a network

In network epidemiology, not every susceptible connects to every infected, and as a result, the structure of the potential exposure network shapes both **which** nodes can be infected, and **when** a node gets infected.

In these models, time usually progresses in discrete steps. To do the accounting of transitions between the compartments, we define a set of update rules for the network, one for the edges, and one for the nodes:



edge rule



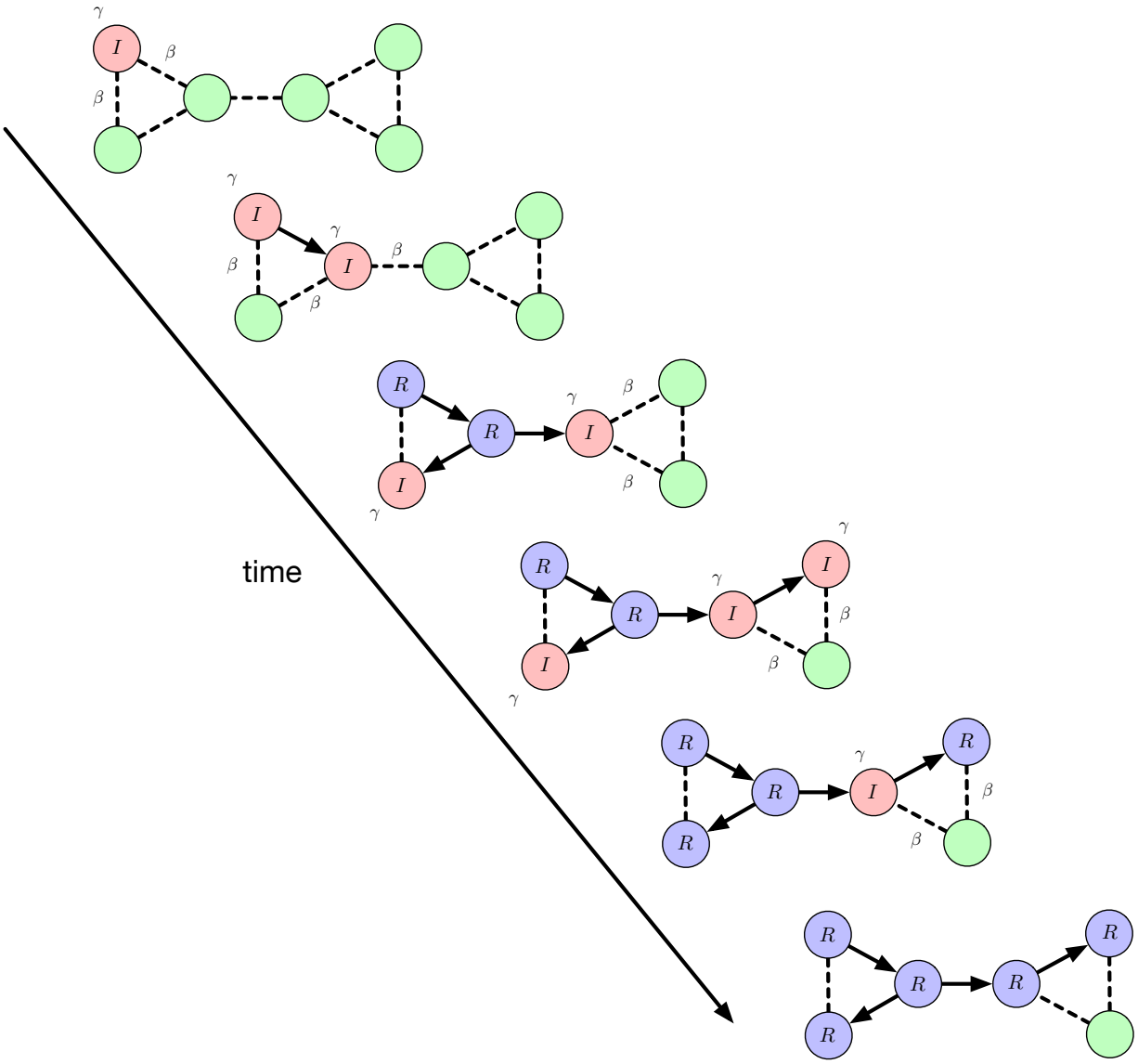
node rule

When we find a motif of the form on the left, we transform it into the motif on the right with probability β for an edge update (transmission) and γ for a node update (recovery).

An example on a network

Given a network where some node is the initial infection (the "patient 0"), we can apply the rules to simulate the spread of the infection across the network. The simulation halts when there are no infected nodes.

In the example below, edges in the potential exposure network G are dashed, while edges in the transmission network G_T are solid.

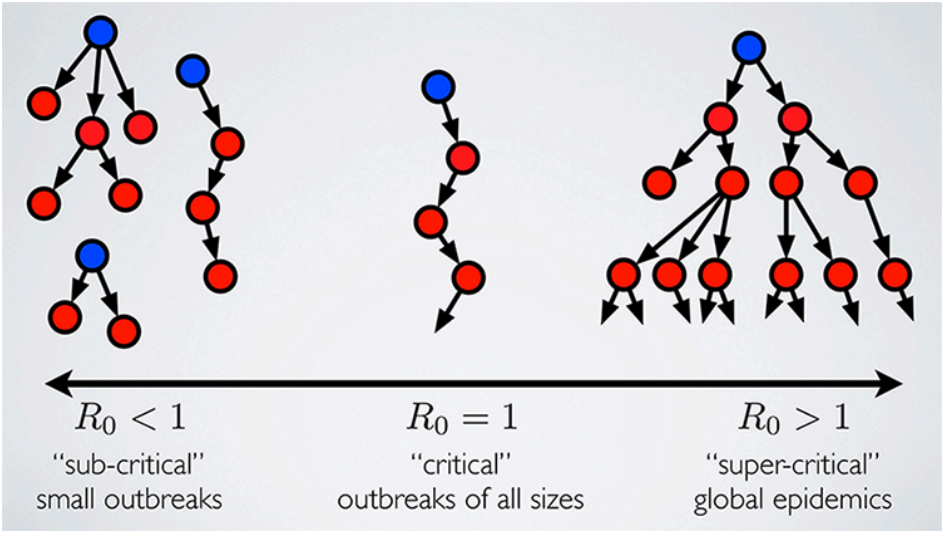


At the end of this epidemic, there is only one node (bottom right) that escaped infection entirely. The bottom-left node could have been infected by either of its neighbors, but only one of those edges is included in the final transmission network.

The role of R_0

The basic reproduction number plays a critical role in giving us a sense of what kind of an epidemic an infection will produce. When it's below the epidemic threshold, the transmission graphs (trees) tend to be small. Above the threshold, and the expected number of additional infections is greater than 1, which implies an exponential growth process (consider $R_0 = 2$; what kind of a tree does this create?) that will eventually

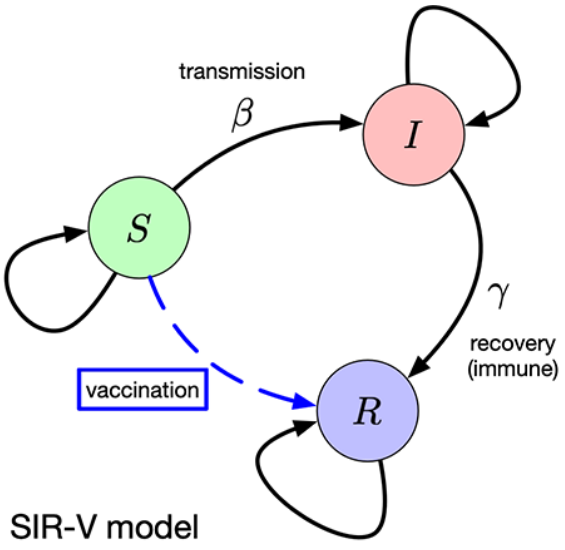
consume every node reachable from the initial infection. At the critical value of $R_0 = 1$, we get outbreaks of all different sizes.



In fact, we've seen (mathematically) exactly this behavior before! It was in the *phase transition* in the size of the giant component for the Erdos-Renyi random graph model $G(n, p)$ in Lecture 3. Look back at the figure that visualized ER graphs for different choices of mean degree $c = \{0.5, 1, 2, 4\}$. There, the mean degree c plays exactly the same role as R_0 here.

How to stop an epidemic on a network

Vaccination is a remarkably achievement of public health: for most people, it provides a means to transition from a susceptible state S to a "recovered" (immune) state R, without having to go through the experience of being infected I.



This is useful for individuals, but also useful for the population, because placing an individual into the R state effectively deletes their edges from the potential exposure graph G. And, deleting many edges from G necessarily reduces the mean out-degree R_0 in the transmission graph G_T, and, ideally, moves it below the critical threshold of $R_0 = 1$. That is, vaccinations break up the network into small, disconnected components, any one of which could host a small epidemic, but such an outbreak cannot spread to the other components. Vaccination thus protects the rest of the network, and this is called **herd immunity**.

The higher the value of R_0 for a disease, the greater a fraction of the population needs to be vaccinated in order to push the transmission graph below the critical threshold. Here are some examples of R_0 values:

- measles R_0=12-18 airborne data from 1912-1928 in US + 1944-1979 in UK
- chickenpox R_0=7-12
- polio R_0=5-7 fecal-oral route
- small pox R_0=1.5-20+ airborne droplet
- H1N1 flu R_0=1-3 airborne droplet
- ebola R_0=1.5-2.5 bodily fluids
- zika R_0=2

The data for the R_0 values is empirical, as the example of measles illustrates. That means that the value of R_0 for a particular disease depends on the current structure of human social organization and human behavior. The better the public health system, and the more isolated individuals are from each other, the lower the R_0 will be, even for the same disease.

Network strategies for targeted vaccination
Suppose we had only a limited supply of vaccines, but we still wanted to stop an epidemic. The limited number means we cannot rely on the classic vaccination strategy of "vaccinate everyone who isn't immuno-compromised". Instead, we can try to exploit the network structure itself, by vaccinating nodes that would either *slow down* the epidemic (making it take longer paths through the network) or *break up* the network into smaller components.

Here are some possible strategies:

- vaccinate by node attributes, e.g., the young and old (who are less likely to survive a bad infection)
- vaccinate randomly
- vaccinate higher-degree nodes first, e.g., teachers, doctors, service industry workers, etc.
- vaccinate "bridges" that link different social communities

Try your hand
Prof. Marcel Salathe's group at EPFL in Switzerland has created a web game that challenges you to precisely the above problem. Try it out here: VAX! https://vax.herokuapp.com

Further reading

van den Driessche, "Reproduction numbers of infectious disease models" *Infection Disease Modeling* **2**, 288-303 (2017).
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6002118/

Delamater et al., "Complexity of the Basic Reproduction Number R_0." *Emerging Infectious Diseases* **25**(1), 1-4 (2019)
https://www.ncbi.nlm.nih.gov/pubmed/30560777

Bansal, Grenfell and Meyers, "When individual behaviour matters: homogeneous and network models in epidemiology." *J. Royal Society Interface* **4**, 879-891 (2007)
https://royalsocietypublishing.org/doi/full/10.1098/rsif.2007.1100