1 Spreading Processes on Complex Networks

Spreading processes take place on networks are all around us. Diseases spread through contact networks. Gossip and memes spread through social networks. Avalanches of excitation spread through the mammalian cortex. Research topics spread through conference or hiring networks. Blackouts cascade through electrical grids.

These are all examples of *dynamical processes* in which nodes interact with their network neighbors, and through interactions, node behaviors or states can changes. This therefore places spreading processes in a category that some refer to as dynamics *on* networks. This contrasts with dynamics *of* networks, in which the network itself is changing in time, and not the states of its nodes.

In this lecture, we will introduce some of the basic ideas and models for spreading processes. However, there are *so many* different models for spreading processes, that here, our treatment may feel more like a survey of ideas, rather than a series of deep dives. Before getting started, we'll first introduce a few different dichotomies that help to categorize spreading processes.

Continuous time vs. discrete time dynamics. Just as in other dynamical systems, some models of spreading take place in continuous time, like standard epidemic models, while others take place in discrete time, like simplified models of neuronal networks [1].

Simple vs. complex contagions. A simple contagion is one in which a single exposure can cause transmission of the contagion across a network link. Measles is a good example of such a contagion, since a single exposure is all it takes for transmission to occur. A complex contagion, on the other hand, is one in which multiple exposure are required before the contagion is transmitted. Fashion trends or choices about technology purchases are often of this type, as an individual is unlikely to change his or her clothing styles unless (or until) many peers have made the change.

2 Epidemiological spreading processes

We will start by studying the spreading of an epidemic on a network, where the nodes can be persons (or computers, in the case of computer viruses; twitter accounts in the case of retweets; etc.), and two nodes are linked if they interact in a way that the disease can spread.

Note that the definition above has an interesting implication. If two different diseases have two different spreading mechanisms, then the "network" may be different. For instance, the influenza virus can be transmitted through the air while the common cold is transmitted via contact with

[1] When we speak of networks of actual, literal, biological neurons, we'll explicitly call them neuronal networks, to differentiate them from the machine learning community's neural nets.

surfaces. This means that in a room where no one touches anything, but people are coughing, links exist in the influenza transmission network while none exist in the common cold transmission networks.

We can model both the network of interactions in different levels of detail. For example, on one extreme, we may have a time resolved, agent-based simulations of individuals' movements and their interactions. On the other extreme, we may have a so-called "fully mixed" model in which all individuals can interact with all others. Note that a model of a disease on a specific complex network falls more toward the former, while the same model applied to an *ensemble* or *model* of complex networks (like the configuration model) falls slightly more toward the latter.

Similarly, we can model the disease in different ways. The simplest models consider different discrete states of disease or no-disease status, and the possible transitions between them. Sometimes the states are called *compartments* leading to the more general term *compartmental model*. Transitions between compartments may be stochastic, stochastic conditional on neighboring nodes, or deterministic, depending on the model. They may be modeled in discrete time or in continuous time. Therefore, we may think of compartmental models as discrete-state models of disease spread.

Prior to any derivations, as a brief tip of the hat and thanks, the notes on SIS models are based largely on lecture notes of Prof. Juan G. Restrepo, University of Colorado, Department of Applied Math.

2.1 SI-X Models

The simplest compartmental model is called the SI model because it has two states: S and I, representing nodes that are Susceptible and Infected. In the SI model, nodes are susceptible to infection until such time as they become infected, at which point they may transmit the infection to susceptibles. This model is called SI because the transition is assumed to be one-way, meaning that one can never become uninfected. The state diagram fos the SI model is then

$$S \to I.$$

The SIS model is the natural extension of the SI model. Here, nodes may become susceptible again after they are infected. This adds a layer of realism, since convalescence returns us to the susceptible state from the infected state. Of course, that transition assumes that the convalescent individual has no lasting immunity. The state diagram for the SIS model is then

$$S \to I \to S.$$

The SIR model represents an SI model that now has immunity included, where R stands for recovered or removed. One can also think of the R state as representing someone who has recovered

and has developed an immunity. Therefore, for the purposes of modeling, that person is effectively no longer part of the system at all. The state diagram for the SIR model is then

$$S \to I \to R.$$

Finally, many other flavors exist. One can imagine a model in which recovery or immunity is only temporary. Such a model would have states

$$S \to I \to R \to S.$$

Furthermore, some disease are infectious to others before they are detectable in the individual who is infected, suggesting that the I compartment ought to be split into "Transmitting" and "Infected" where one becomes visibly infected after a deterministic amount of time in the transmitting state.

$$S \to T \to I \to \dots$$

Compartmental models are even used to model vector-borne diseases like dengue and malaria, in which there are states for humans, and mosquitos, separately, which interact and affect each others' transition rates between states. Age-structured models are also common, which involve the separate modeling of individuals of different ages as different "metapopulations." The literature here is vast, exciting, and ranges from the physicist's approximations to the epidemiologist's agent-based detailed model.

2.2 The fully-mixed SIS model

In this model, everyone interacts with everyone *in the same way*. You may consider this to be a model without a network, or, you may consider this to be a model in which the connections are all-to-all.

Due to the fact that everyone in a particular state will have identical interaction patters with all others, we will track only the total numbers of individuals in each state. Let N be the total population size (or number of nodes). We will assume that this is constant. An infected person has a probability $\bar{\beta}$ per unit time to pass the infection to a susceptible person. An infected person has a probability γ per unit time to heal and become susceptible again.

For accounting purposes, let $S(t)$ be the expected number of susceptibles at time t, and let $I(t)$ be the expected number of infected at time t. Note that $N = S + I$, so we can (and will) write $I(t) = N - S(t)$.

Now consider how $S(t)$ changes in a small amount of time Δt.

$$S(t + \Delta t) = S(t) - S(t) \cdot \bar{\beta}\Delta t I(t) + I(t) \cdot \gamma \Delta t \ . \tag{1}$$

The term in red represents the probability that one susceptible becomes infected, while the term in blue represents the probability that one infected will become susceptible. We now divide by N and Δt, and let

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

Note that when we do this to β, it changes its interpretation—we can now think of it as the expected number of individuals that a single infected individual infects per unit time. Now, try making these substitutions yourself, and take the limit as $\Delta t \to 0$. Equation 1 becomes

$$\frac{di}{dt} = \beta i(1-i) - \gamma i \tag{2}$$

We now have an equation for the dynamics of the infected population of our system. What should we do with it?

One of the most fundamental questions to ask is: *Will there be an epidemic at all?* Put differently, if a small fraction individuals are infected, will the infection tend to spread or will it die out? We can answer this question by analyzing the dynamics of Eq. (2) for small values of i. When i is small, i^2 is very small, so when we linearize Eq. (2) and discard the very small terms, we get

$$\frac{di}{dt} \approx \beta i - \gamma i = (\beta - \gamma)i \ ,$$

which means that whenever $\beta > \gamma$ (or whenever $\beta > \gamma > 1$), an epidemic will tend to grow (exponentially, in the small-i regime). Otherwise, an epidemic will decay (exponentially) to zero. For this reason, the point

$$\frac{\beta}{\gamma} = 1$$

is often referred to as the *epidemic threshold*.

We can also draw conclusions about the steady-state fraction of infected individuals in the network. Set Eq. (2) to 0, and solve to find that $i_\infty = 1 - \frac{\gamma}{\beta}$.

2.3 SIS on a network

How does a network of connections change the dynamics that we observe? One place to start is the epidemic threshold. For an all-to-all network, we found, above, that an epidemic would occur whenever $\bar{\beta}N/\gamma > 1$. Now, we consider the same process as above on an adjacency matrix, A, in which $A_{nm} = 1$ if and only if nodes n and m interact. We will still define $\bar{\beta}\Delta t$ as the probability that an infected node will pass the infection to an uninfected neighbor in a time Δt, and let $\gamma\Delta t$ be the probability that an infected node will spontaneously recover in time Δt.

To begin a network-based analysis, let $i_n(t)$ be the probability that node n is infected at time t. Then

$$\begin{aligned} i_n(t+\Delta t) &= P(\text{node } n \text{ is infected at time } t \text{ and does not heal.}) \\ &\quad + P(\text{node } n \text{ is susceptible at time } t \text{ and gets infected by a neighbor}). \\ &= i_n(t)[1-\gamma\Delta t] + [1-i_n(t)] \sum_{m=1}^{N} A_{nm} i_m(t) \bar{\beta} \Delta t \end{aligned}$$

and this can be rearranged, as in the all-to-all case, to find that

$$\frac{di_n}{dt} = -\gamma i_n + (1-i_n)\bar{\beta} \sum_{m=1}^{N} A_{nm} i_m \ . \tag{3}$$

How does the epidemic threshold change when we consider the dynamics of Eq. (3)? We will linearize as before, neglecting the products of $i_n i_m$, which will be very small when the typical i is small. This results in the equation:

$$\frac{di_n}{dt} = -\gamma i_n + \bar{\beta} \sum_{m=1}^{N} A_{nm} i_m \tag{4}$$

which in vector form is

$$\frac{di}{dt} = \left(-\gamma \mathcal{I} + \bar{\beta} A\right) i \tag{5}$$

and which is solved by

$$i(t) = u e^{(\bar{\beta}\lambda - \gamma)t} \ . \tag{6}$$

One can check (you should!) that this is a solution provided that u is an eigenvector of A and λ is its corresponding eigenvalue. Which eigenvalue-eigenvector pair should we consider? As in previous lectures, this turns out to be the principal (largest) eigenvalue. For a network A which is irreducible (i.e., connected) this largest eigenvalue is unique and positive, and is referred to as the Perron-Frobenius eigenvalue.

What is the eigenvector u? Recall that if λ is the Perron-Frobenius eigenvalue, then u is the corresponding eigenvector which makes its entries the *eigenvector centralities of the nodes*! Therefore, nodes with higher eigenvector centrality will tend to have a higher probability of being infected than nodes with a lower eigenvector centrality. Can you build an intuitive understanding of this? We've now solved the equation above, but have not computed an epidemic threshold. Based on the solution, the exponential term will grow whenever $\bar{\beta}\lambda - \gamma > 0$, meaning that the epidemic threshold here is

$$\frac{\bar{\beta}}{\gamma} = \frac{1}{\lambda} \ .$$

Recall that in the all-to-all case, we decided to change from $\bar{\beta}$ to β. We rescaled by N so that β was interpretable as the expected number of new infections resulting from an infected node in an uninfected population, per unit time. Previously, this meant that $\beta = \bar{\beta}N$, but here, to maintain the same interpretation, we set $\beta = \langle k \rangle \bar{\beta}$ since typically, each node connects to $\langle k \rangle$ other nodes. Therefore, our epidemic threshold becomes

$$\frac{\beta}{\gamma} = \frac{\langle k \rangle}{\lambda} \ . \tag{7}$$

What should we make of this equation? It looks just like our previous equation for the well-mixed case, except for that the right hand side now reads as a ratio of the mean degree to the principal eigenvalue instead of one. What does that mean?

2.3.1 A few notes on the largest eigenvalue

It turns out that the Perron-Frobenius eigenvalue comes up a lot. As a consequence, people have been very interested to know how it can be approximated from the values. For a network without degree correlations, it turns out that

$$\lambda \approx \frac{\langle k^2 \rangle}{\langle k \rangle}.$$

You may recognize that value on the right hand side from the work on the friendship paradox! Recall that $\langle k^2 \rangle \geq \langle k \rangle^2$ with equality only if the network is degree-regular. This means that $\lambda \geq \langle k \rangle$. Plugging the approximation above into Eq (7), we find that the epidemic threshold can be approximated as

$$\frac{\beta}{\gamma} \frac{\langle k^2 \rangle}{\langle k \rangle^2} = 1 \ . \tag{8}$$

In networks where some nodes have exceptionally large degree, the approximation above is no longer good. Specifically,

$$\lambda \approx \sqrt{k_{\max}} \qquad \text{if } \sqrt{k_{\max}} > \frac{\langle k^2 \rangle}{\langle k \rangle} \log^2 N.$$

These results are due to Chung and Lu, whose names you'll recognize from the configuration model lectures. They assume that there are no degree correlations. However, when there are degree correlations, correction terms are needed. For more details, please see *Approximating the largest eigenvalue of network adjacency matrices* by Restrepo, Ott, and Hunt.

2.4 Vaccination and Herd Immunity

Suppose that you vaccinate a fraction q of the nodes, chosen uniformly at random. How large does q need to be in order to prevent the disease from spreading? If we assume that the vaccine is

perfectly protective, then the probability that an infected node passes the infection to an uninfected node is $\beta(1-q)$, since the disease transmits only if the recipient is not vaccinated. Therefore, to understand the role of vaccination, we need only replaced β in Eq. (9) with $\beta(1-q)$,

$$\frac{\beta(1-q)}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle^2} = 1 \ . \tag{9}$$

A little algebra shows us that

$$\frac{\beta(1-q)}{\gamma}\frac{\langle k^2\rangle}{\langle k\rangle^2} \leq 1 \iff q \geq 1 - \frac{\gamma\langle k\rangle^2}{\beta\langle k^2\rangle} \ . \tag{10}$$

When a fraction q of nodes are vaccinated satisfying Eq (10), it pushes the SIS dynamics into the subcritical regime in which epidemics do not spread to infect a large fraction of the network. Put differently, we need only vaccinate qN nodes to protect the population of N nodes from epidemics. This phenomenon is called *herd immunity* and it is an important idea in epidemiology. Simply due to the intrinsic dynamics of a disease, a vaccine for some can provide protection for all.

Note also that the fraction q can be nearly 1 whenever: (i) recovery is slow (γ is small), (ii) the disease is highly transmissible (β is large), or (iii) the network has an especially broad degree distribution $\langle k^2\rangle \gg \langle k\rangle^2$. In fact, for some networks, $\langle k^2\rangle$ diverges, meaning that no amount of vaccination can overcome the structural tendency toward epidemics!

To illustrate this latter point, suppose that you have degrees with a power-law degree distribution, $(k) = ck^{-a}$ where c is the normalization constant. Suppose for convenience that $k \in [k_{min}, \infty)$. Then, approximating the discrete power law with its continuous version,

$$\begin{aligned}\langle k^2\rangle &= \int_{k_{min}}^{\infty} k^2 P(k) dk \\ &= \int_{k_{min}}^{\infty} k^2 c k^{-\alpha} dk \\ &= c\int_{k_{min}}^{\infty} k^{2-\alpha} dk \end{aligned} \tag{11}$$

This integral will diverge when $\alpha \leq 3$, which means that for sufficiently broad power-law degree distributions, the epidemic threshold is always satisfied, regardless of vaccination, β, or γ.

3 Excitable cascades on networks

The Greenberg-Hastings cellular automaton is a simple model for excitable dynamics. In its simplest form, it is a discrete-state, discrete-time system that models excitation with three states:

resting, excited, and refractory. When resting, such a node can accept an incoming stimulus, in which case it will be in the excited state at the next time step. When excited, the node will attempt to excite its network neighbors, and it will then, deterministically, enter a non-negative integer number of refractory states, through which it will proceed deterministically until eventually returning to the resting state.

By analogy, this automaton is a discrete-time version of SIRS dynamics, with a twist: a node only stays in the excited (infected) state for one time step, before proceeding deterministically through a variable number of R states until returning back to S.

In a departure from most of the coursework so far, we now consider a weighted network, in which A_{nm} is a value in $[0, 1]$ that represents the probability that an excited node n causes a resting node m to become excited in the next time step.

3.1 An equation for the network's dynamics

Let $p_m(t)$ be the probability that node m is in the excited state at time t, and for simplicity, assume that nodes have zero refractory states, so that an excited node will proceed deterministically from the excited state back to the resting state. In this case, the probability that a node is in the excited state at time $t + 1$ is given by the probability that the node is in the resting state, $1 - p_m(t)$, and that the node receives a stimulus from at least one of its network neighbors,

$$p_m(t+1) = \big(1 - p_m(t)\big) \left(1 - \prod_n (1 - p_n(t) A_{nm})\right) \ . \tag{12}$$

Note that here, we have written the probability that a node receives a stimulus from at least one of its network neighbors as one minus the probability that the node does *not* receive a stimulus from any of its neighbors.

3.2 Fixed point and stability

Equation (13) has a fixed point that is easy to check: $p_m = 0$ for all nodes m. As with the case of SIS epidemics, we can ask whether $p = 0$ is stable by considering small perturbations away from it. Suppose that all the values of p are small, such that any quadratic terms in p are very small and can be ignored. In this case, Eq. (13) can be linearized—try this yourself—resulting in

$$p_m(t+1) = \sum_n p_n(t) A_{nm} \ . \tag{13}$$

Using the ansatz $p_m(t) = u_m \lambda^t$, and plugging in, we get $\lambda u_m = \sum_n u_m(t) A_{nm}$ which written as a matrix-vector equation is

$$\lambda u = uA \tag{14}$$

which, just as with SIS dynamics, is solved when λ is the Perron-Frobenius eigenvalue of A, and u is the (left) eigenvector. Note that under this definition, the left eigenvector corresponds to the generalized in-degree centrality of nodes. Thus, we can interpret the equation above as stating that the stability of dynamics will be given by whether λ is greater than or less than one, and the relative probabilities for nodes to be excited are proportional to their left-eigenvector centralities. Additional analysis of this type of system has been done and may be of interest to some. [2]

4 Cascade models, opinion, economics, and influencers

Rather than reproduce the beautiful book chapter of Jon Kleinberg here, please see
`http://www.cs.cornell.edu/home/kleinber/agtbook-ch24.pdf`

[2] See, e.g., *Predicting criticality and dynamic range in complex networks: effects of topology* by Larremore, Shew, and Restrepo. Physical Review Letters 106, p. 058101 (2011).