Epidemics

Kiran Kumar(04CS1011)

April 25, 2008

1 Introduction

An epidemic by definition is generally a widespread disease that affects many individuals in a population. An Epidemic study on networks address the challenge to study how information flows on the network.

1. Computer Worms and viruses
2. Information flow in social networks, say news or rumors
3. Study of Asymptotic behavior of the population by Theoretical Biologists.

2 Models to Study Epidemics:

The basic idea is:

1. Group individuals into categories or **States**.
2. Calculate the rates at which various events occur (*infection, recovery, death, etc.*)
3. Move individuals between categories/change the population sizes.
4. Write down equations based on these rates. Use the equations to describe the system outcomes (analytically, statistically, or computationally).

2.1 Box Models

In this model we divide the host population up into boxes according to epidemiological status.
The most important categories are:

1. **S**usceptible: hosts that are susceptible to infection.
2. **I**nfective: hosts that are infected, and capable of infecting other hosts.

   In addition, we might have categories such as:

3. **R**ecovered (or Removed): hosts that have acquired immunity to the disease, or been quarantined, or vaccinated, or otherwise been made unavailable for infection.
4. **E**xposed (or Latent): hosts that have been infected but are incubating the disease, not yet Infective.

We will refer to these as the **State Variables** of our system.

The list of categories or states we have in a particular model depends on the biology of a particular disease:

- Is there an incubation period?
- Do any hosts recover?
- If so do they acquire immunity or do they become susceptible again?

We often refer to these models by their acronyms:

1. **SI** (also known as the "simple epidemic") is an epidemic where hosts never recover,
2. **SIS** one where they recover but become susceptible again (e.g. gonorrhea),
3. **SIR** where they recover and gain immunity (e.g. measles),
4. **SEIR** where there is an incubation period, etc

Now we have to consider the transitions: how do individuals move from one box to another?

- birth (entering the S box from outside, or possibly entering I if there is vertical transmission so that some hosts are born infected);
- "natural" (non-disease-related) death (leaving any box);
- disease-induced mortality (leaving I);
- infection (moving from S to I);
- recovery (moving from I to R)

Quantify the transitions: how fast do individuals get infected, die, and how do these transition rates depend on the numbers or densities of hosts in different categories? These are some pretty standard assumptions.

1. birth may occur at a constant per capita rate and may come from just the susceptible population or from both susceptible and infective hosts: $b(S + I)$
2. natural death often occurs at a constant per capita rate in each category: dS, dI
3. disease-induced mortality occurs at a constant per capita rate: aI
4. infection is often assumed to be proportional to both S and I. This assumption implies that we treat individuals as though they were gas molecules, which is not a bad first approximation. If we double the number of susceptibles or the number of infectives, we double the overall infection rate: $\beta(S/N)I$

Since the spread of a disease is a dynamic process we will want to keep track of the spread of a disease over time. Thus we will denote the time period by the subscript t on each of our state variables.

3 SIS Model

Let us start with a simple model of an epidemic.

3.1 The First Model

Labels used:

N_t = Total size of population in period t

S_t = Number of Susceptible in period t

I_t = Number of Infected in period t

Individuals potentially move from the susceptible to the infected group when a susceptible person comes in contact with an infected person. What counts as a contact varies with the disease. Once someone is infected it takes some amount of time, or some number of periods in our model for the infected individual to move back to the susceptible pool.

Assume that $N_t = N$ for all t. In other words there is a constant population size. Further let us assume that there is a constant unchanging population (We ignore births and deaths).

Let us begin by looking at the transition from period 0 to period 1. We begin period 0 with I_0 infected individuals. We can call this the state of the system at time 0. New individuals get infected by coming in contact with members of the infected population. Let us assume that each infected person contacts γ non-infected people in each period. Thus the number of possible new infections is γI_0. But not all of the contacts result in an infection. Suppose that only α percent of contacts result in an infection. Thus each infected individual results in $\alpha\gamma$ new infections in each period. Further to keep things simple let us assume that each infected person in period t moves back to the susceptible pool in the next period $t + 1$. Let us now write out an equation that describes this process:

$$I_1 = I_0 - \gamma\alpha I_0 + I_0$$

Note: We assumed that each infected person recovers in the next period.

$$I_1 = \gamma\alpha I_0$$
$$I_2 = \gamma\alpha I_1$$
$$I_2 = (\gamma\alpha)^2 I_0$$
$$I_t = (\gamma\alpha)^t I_0$$

The number of people currently infected in period t is the product of the contact and transmission parameters raised to the power t multiplied by the initial size of the infected population.

Here if $\gamma\alpha < 1$ the number of infected individuals decreases to 0 very rapidly;

the disease disappears. If $\gamma\alpha > 1$ the number of infected individuals keeps increasing; the disease spreads throughout the population. This is sometimes called the epidemic threshold. If the number of contacts times the transmission rate is less than one this means that each infected person infects less than one person on average. So, the number of infected individuals will decrease. Just like our model when the average number of people infected is greater than one; the disease continues to spread to a larger and larger fraction of the population. Thus we reach the epidemic threshold whenever greater than one person is infected by each infected person.

3.2 Full SIS Model

In the last section we assumed that each contact of an infected person was with a non-infected, or susceptible, person. It is more realistic to assume that the number of susceptible contacts is a function of the number of susceptible persons in the population. Also, we assumed that each infected person was fully recovered after one time period. This may be true if we are measuring time in weeks or months, but it probably isn't true for some diseases if we are measuring time in days. Thus we would like our model to allow for the possibility that it takes multiple time periods for someone to move from the infected group back to the susceptible group.
First let us define two new state variables that will measure the fraction of the total population that are susceptible and infected. Let $i_t = I_t/N_t$ be the fraction of the population that is currently infected. Define $s_t = S_t/N_t$ as the fraction of the population that is currently susceptible. We also will define a new parameter κ that measures the percent of the population that recovers from a disease each period.

$$I_{t+1} = I_t - \kappa I_t + \alpha\gamma s_t I_t$$
$$S_{t+1} = S_t + \kappa I_t - \alpha\gamma s_t I_t$$

We have κI_t individuals who recover each period and thus leave the infected group and re-enter the susceptible group. And we have $\alpha\gamma s_t I_t$ individuals who enter the infected group and leave the susceptible group. Notice that the change in the infected group always equals the change in the susceptible group when we have a constant population.
Equivalently we can write these equations using our proportion state variables:

$$i_{t+1} = i_t - \kappa i_t + \alpha\gamma s_t i_t$$

$$s_{t+1} = s_t + \kappa i_t - \alpha\gamma s_t i_t$$

Now, with these equations written we can use them to understand the epidemic threshold in the SIS model. If more people flow into the infective state than flow out this means that the level of infectives is increasing. If the opposite is true (more flow out than in) the level of infectives is decreasing. Thus our epidemic threshold is determined by whether $\kappa i_t > \alpha\gamma s_t i_t$ or $\kappa i_t < \alpha\gamma s_t i_t$.
If $\kappa > \alpha\gamma s_t$ this means that more people leave the infective state than enter it. Thus the level of the disease is decreasing. If $\kappa < \alpha\gamma s_t$ this means that more people enter the infective state than leave it. Thus the level of the disease is increasing. Now, another way to write this inequality is: $\alpha\gamma s_t/\kappa$. This is the epidemic threshold for the SIS model. If the fraction is greater than one the level of infectives increases (more than one person is infected by each infective) And if it is less than one the level of infectives decreases (fewer than one person is infected by each infective.)
Equilibrium is reached when the number of susceptibles and infectives is in steady state; both are constant proportions of the population. This would be the case if $\alpha\gamma s_t/\kappa = 1$. This would define a steady state of the system where s_t and i_t are constant in all periods.

3.3 Dynamics

In equilibrium, variables I_t and S_t , do not change from period to period; they are constant. To find the steady state we need to find a solution to our system of equation above where $S_t = S_t + 1$ and $I_t = I_t + 1$. One way to do this is to drop the time subscript on our equations above and solve for s and i. Thus we are looking for a solution to the following set of equations:

$$i = i - \kappa i + \alpha\gamma si$$
$$s = s + \kappa i - \alpha\gamma si$$
$$\Rightarrow s = \kappa/\alpha\gamma$$

and also,

$$s + i = 1 \qquad \text{(as sum of these fractions is 1)}$$
$$\Rightarrow i = 1 - \kappa/\alpha\gamma$$

Note the intuitive properties of these equations. As the number of contacts

of an infected person or the transmission probability increases the number of susceptible individuals decreases (and the number of infected individuals increases). And as the time to recover increases (meaning κ decreases) the number of susceptible individuals decreases (the number of infected people increases).

4 SIR Model

We are interested in the spread of an infectious disease where individuals may be susceptible to the disease, may be currently infected with the disease, or may be recovered and immune from the disease. Thus we have three groups or states in which we can place individuals. In addition we see that our data is a time series where we have a number of infected individuals at each point in time. Similarly we also have a number of susceptible and recovered individuals at each point in time. Let us begin with the notations.

S_t = the number of susceptible individuals in the population at time t.
I_t = the number of infected individuals in the population at time t.
R_t = the number of recovered individuals in the population at time t.
N = the population size.
Correspondingly define the three groups as fractions of the total population N in lower case.
$s_t = S_t/N$ (the susceptible fraction of the population at time t.)
$i_t = I_t/N$(the infected fraction of the population at time t.)
$r_t = R_t/N$(the recovered fraction of the population at time t.)
We will work with both of these sets of notation in the modelling process. Note that each individual in the population is in one of the three groups. Thus $S_t + I_t + R_t = N$ and $s_t + i_t + r_t = 1$.

4.1 Dynamics

If we think about the process of a disease that fits the SIR framework we have a flow of individuals from the susceptible group to the infected group and then to the recovered or removed group.

Susceptible $\longrightarrow$ Infected $\longrightarrow$ Recovered

An individual potentially moves from the susceptible to the infected group

when she comes in contact with an infected person.Suppose that each infected person contacts γ individuals in each period of time on average. Now each contact may not result in transmission of the disease. Perhaps only αpercent of the contacts result in transmission. Thus the potential number of transmissions may be at most $\alpha * \gamma$. Let us define this value as $\beta = \alpha * \gamma$. β is the average number of transmissions possible from a given infected person in each period.

If we assume that individuals are mixed randomly then each potential transmission may be from an infected person to a susceptible person which results in a new infected person. Or a transmission may occur from an infected person to another infected person which results in nothing happening since the person are already infected. Or the potential transmission may occur from an infected person to a recovered or immune person. In this case again nothing changes. Since only st percent of the population is susceptible each infected person generates only β st new infections each period. Each infected person recovers (or is removed/ dies) at some rate. Let the fraction of the infected group that recovers be κ.

Given the current state of the population in period t described by S_t , It and R_t we can write a series of difference equations that describe the motion of the system. First lets describe the susceptible population. We begin period t with S_t individuals in the susceptible population. From this population we lose on average $\beta s_t I_t$ from the population. Thus in period $t+1$ we have:

$$S_{t+1} = S_t - \beta s_t I_t \text{ — (1)}$$

Through similar reasoning we see that:

$$R_{t+1} = R_t + \kappa I_t \text{ — (2)}$$

and

$$I_{t+1} = I_t + \beta s_t I_t - \kappa I_t = I_t(1 + \beta s_t - \kappa) \text{ — (3)}$$

Similarly we could write each of these in terms of the population fractions:

$$s_{t+1} = s_t - \beta s_t i_t \text{ — (4)}$$

Through similar reasoning we see that:

$$r_{t+1} = r_t + \kappa i_t \text{ — (5)}$$

and

$$i_{t+1} = i_t(1 + \beta s_t - \kappa) \text{ — (6)}$$

If we add up these equations we will find that $s_{t+1}+i_{t+1}+r_{t+1} = s_t+i_t+r_t = 1$.

4.2 Steady States

Unlike the SIS model there do not exist positive steady state levels of infected individuals in the SIR model if there is a constant population. Let us see why this is the case. First suppose that it is equal to 0. Then from Equation 6 you can see that $I_t + 1 = 0$. Thus if we start with no people infected we stay there. This was true in the SIS model too. There has to be an initial infected person in order for there to be more infected people. But in the SIS model there were other steady states as well. Let us see why those do not exist here. Suppose that $i_t > 0$. I_{t+1} will be either greater than or less than but never equal to (except in a special case we going to see.)
So, in most cases it will either be increasing or decreasing. We see from Equation 6 that $i_t + 1 = i_t(1 + \beta s_t - \kappa)$. Let $\rho_t = (1 + \beta s_t - \kappa)$. This is the epidemic threshold for the SIR model with a constant population. If $\rho_t is greater than 1 then we are multiplying it by a number greater than 1 so i_{t+1} > i_t$. The number of infected individuals is increasing. But if ρ_t is less than 1 we are multiplying it by a number less than 1 so $i_{t+1} < i_t$. The number of infected individuals is decreasing. Clearly any time $\rho_t > 1$ or $\rho_t < 1$ we are not at a steady state.
Now let us consider the possibility of $\rho_t = 1$ and show why this can not persist. If $\rho_t = 1$ then either $i_t = 0$ or $\rho_{t+1} < \rho_t$. The first case is not interesting. This just says that there are no infected individuals so there is no possibility of anyone else being infected. The second case is more tricky. Let us again look at the definition of ρ_t, $\rho_t = (1 + \beta s_t - \kappa)$ We see that there are three constants in the equation 1, β, and κ as well as one state variable s_t. Let us take another look at the equation for s_t given by Equation 4. We see that this equation is decreasing whenever $i_t > 0$. Thus if $i_t > 0$, ρ_t is decreasing for any $\beta > 0$. Thus there can never be a steady state where $\rho_t = 1$ and $i_t > 0$. Further if it is increasing it is doing so at a decreasing rate. And if it is decreasing it is doing so at an increasing rate. Combining these results indicates that the infected population will always disappear in the long run if we have a constant population.
Now let us take a brief look at s_t and r_t . First let us look at s_t . We know that s_t is decreasing for $i_t > 0$. And we also know that it equals 0 in steady state. So, from this we can deduce that s_t will also reach a steady state value since it is constant if i_t is equal to 0 (see Equation 4) Now, r_t is increasing for $i_t > 0$ and at steady state when $i_t = 0$. Like s_t r_t will also reach a steady state value.

5 Kermack-McKendrick Model

The Kermack-McKendrick model is an SIR model for the number of people infected with a contagious illness in a closed population over time. It was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865). It assumes that the population size is fixed (i.e., no births, deaths due to disease, or deaths by natural causes), incubation period of the infectious agent is instantaneous, and duration of infectivity is same as length of the disease. It also assumes a completely homogeneous population with no age, spatial, or social structure.
The model consists of a system of three coupled nonlinear ordinary differential equations,

$$\frac{dS}{dt} = -\beta SI \text{ – (1)}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \text{ – (2)}$$

$$\frac{dR}{dt} = \gamma I \text{ – (3)}$$

where t is time, $S(t)$ is the number of susceptible people, $I(t)$ is the number of people infected, $R(t)$ is the number of people who have recovered and developed immunity to the infection, β is the infection rate, and γ is the recovery rate.
Here epidemiological threshold,

$R_0 = \beta S/\gamma$.

R_0 is defined as the number of secondary infections caused by a single primary infection; in other words, it determines the number of people infected by contact with a single infected person before his death or recovery.
When $R_0 < 1$, each person who contracts the disease will infect fewer than one person before dying or recovering, so the outbreak will peter out ($dI/dt < 0$). When $R_0 > 1$, each person who gets the disease will infect more than one person, so the epidemic will spread ($dI/dt > 0$). R_0 is probably the single most important quantity in epidemiology. Note that the result $R_0 = \beta S/\gamma$ derived above, applies only to the basic Kermack-McKendrick model, with alternative SIR models having different formulas for dI/dt and hence for R_0.

6 Eigenvector Centrality (EVC)

We wish to evaluate the nodes in a network in terms of their 'spreading power'. That is, we know that some nodes play an important role in spreading, while others play a less important one. One need only imagine the extreme case of a star: the centre of the star is absolutely crucial for the spreading of infection over the star, while the leaf nodes are entirely unimportant, having only the one aspect (common to every node in any network) that they can be infected.
A simple definition of centrality, which is certainly related to spreading power, is a node's degree centrality, i.e., its number of neighbors: k_i. This quantity is, however, not smooth at all - there is no necessary correlation between a given node's degree and that of its neighbors.
Seeking a smoother measure, we can give each node a centrality score which is simply the average of its neighbors' scores:

$$x_i = \frac{1}{k_i} \sum_{j=n(i)} x_j \text{ — (1)}$$

This choice gives a node importance weight which is too smooth i.e. normal solutions to equation (1) give equal scores x_i to every node i. Hence we discard this idea.
Eigenvector Centrality (EVC) involves one small change from equation 1: one defines the centrality of node i as being proportional to the sum of (but not the average of) i's neighbours' centralities:

$$e_i = \frac{1}{\lambda} \sum_{j=n(i)} e_j \text{ — (2)}$$

Equation 2, in contrast to node's degree k_i, makes a node's centrality dependent on its neighbors' centralities - hence giving a 'smooth' centrality measure - but it also gives non-trivial solutions (unlike equation 1) - because, since λ is the same for all nodes, equation 2 does not completely cancel out the boost in centrality from having many neighbors (as equation 1 does). Equation 2 can be rewritten as,

$$Ae = \lambda e \text{ — (3)}$$

where e is the vector of centrality scores, and A is the network's adjacency matrix.
Equation 3 reveals the motivation for the name 'eigenvector centrality': the EVC of e_i of node i is the i-th component of a chosen eigenvector e of the adjacency matrix A . To ensure that all centrality scores are positive, one

takes the principal eigenvector of A - that is, the one corresponding to the largest eigenvalue λ_{max}.
Thus we see that, because a node's EVC depends on that of its neighbors, the EVC values over a network may be thought of as 'smoothly varying' over the network. That is, a node with very high EVC cannot be surrounded by nodes with very low EVC. Of course, it is true that EVC tends to be positively correlated with a simpler measure of centrality, namely node degree. In fact, one might say that the principal difference between the two measures is that EVC is constrained by its definition to be smooth, while node degree centrality is not. This difference can, however, be nontrivial. For instance, a node with high degree, surrounded by many leaf nodes, and linked only tenuously to the bulk of a large and well-connected network, will have a low EVC, in spite of its high degree. The point is that EVC is sensitive to properties of neighborhoods, while node degree is not.