Week 2 Session 2: Contact Network Epidemiology

Joel Miller & Tom Hladish

<https://scholar.harvard.edu/joelmiller/teaching>

<http://tjhladish.github.io/sismid/>

<http://epidemicsonnetworks.readthedocs.io/en/latest/>

LECTURE 1: DISEASE SPREAD

Two major features

* Relation between mode of transmission and population structure
* How does the immune system respond to exposure

What are we interested in?

* P: probability of epidemic
* A: attack rate (the fraction infected, or attack ratio because not really a rate)
* R0: the average number of infections caused by those infected early in the epidemic
* I(t): the time course of the epidemic
* For SIS:
  + I(inf): equilibrium level of infection
  + P, R0, I(t) too

Modeling approaches

* Trade-off between complexity and understanding (transparency)
* Easy but not real = compartmental models
* Complex and real = agent-based models
* Somewhere in the middle = network models

**Simple compartmental models**

Continuous or discrete time / SIR or SIS

Assumptions

* Every individual is average
* Every interaction of u is with a randomly chosen other individual
* The probability an interaction is with a susceptible and infected = S/n\*I/n

**Networks, in general**

* Collection of individuals who are joined together based on interactions that may spread the disease in question
* Connections (edges) may be
  + Transient = brief or temporally distinct
  + Weighted = closer level of contact
  + Clustered = if mutual edges also have edges
  + Heterogeneously distributed = more variation in degree distribution
  + Directed = who is infecting whom
* Examples of important networks
  + Airlines
  + Communities
  + Livestock

Definitions

* Network is a collection of nodes joined by edges
* Two nodes joined together are neighbors/partners
* Number of neighbors a given node has is its degree (k)
* No real difference between network and graph
* Contact can mean either the actual physical contact, or a node in the same network. Try to avoid using
* Degree distribution: P(k) = proportion of nodes with degree k
* Partnership duration = changing contacts through time, individuals enter/leave population
  + Changing partnership reduces the effect of the local “susceptible depletion” meaning each of one node’s infections depend on # of contacts and their length
* Clustering = frequency of short cycling
* Edge weights = some edges may have higher transmission probabilities than others
* Assortativity = individuals may actively select similar partners (and similar degree)
  + Can change the dynamics of disease spread:
    - High transmission rates 🡪 less spread in network (high degree infected)
    - Low transmission rate 🡪 more spread in network (enough time for low degree to become infected)
* Modularity = some parts of network are more densely connected than others (existence of sub-communities)

Random network models

* “Null” model to compare your model to 🡪 are they statistically different?
* Way to look at particular properties in a network
* Common random network models (more description on the lecture slides):
  + Erdos-Renyi (Gn,p)
  + Configuration
  + Chung-Lu
  + Exponential random graph model (ERGM)
* Connections between neighbors: if we choose a random individual in a configuration model network, its expected degree IS HIGHER than the expected degree of its random partner!
  + \*Your friends have more friends than you! (In configuration models)
  + Although, after that one step, there is not a greater probability of higher degree
* Theory: high degree people get infected sooner and will transmit to more people

Real world networks not always contact networks for disease

* Typically need physical interactions
* Often highly clustered
* Appropriate for respiratory diseases
* Sexual networks = low clustering, highly heterogeneous, transient partnerships play large role
* Location-location networks = city-city, farm-farm, habitat-habitat, etc.

**Empirical networks**

* “In nature” may mean animals or humans in “natural” settings

[9] Rodrigo Hamede, Jim Bashford, Menna Jones, and Hamish McCallum.

Simulating devil facial tumour disease outbreaks across empirically derived contact networks.

Journal of Applied Ecology, 49(2):447{456, 2012.

[10] Meggan E Craft, Erik Volz, Craig Packer, and Lauren Ancel Meyers.

Disease transmission in territorial populations: the small-world network of serengeti lions.

Journal of the Royal Society Interface, 8(59):776{786, 2011.

[11] Meggan E Craft and Damien Caillaud.

Network models: an underutilized tool in wildlife epidemiology?

Interdisciplinary perspectives on infectious diseases, 2011, 2011.

**LECTURE 2: EPIDEMIC IN NETWORKS**

Kermack-McKendrick continuous time model

* Transmission = B\*(K)\*SI/N where (K) is the average degree and B is the transmission rate per partnership and SI/N is the proportion of S and I in the population

**Stochastic simulation with SIR**

* As n increases, the final epidemic size becomes more similar (less variation) except that a proportion always does not have the epidemic take off
* In large populations, small outbreaks affect the same number of people (or, population size doesn’t matter. If disease never spreads to enough people to take off, it will fail and only infect small amount ~0)
* In large populations, epidemics affect approximately the same PROPORTION of the population! Meaning, epidemics in (large) populations will infect a predictable proportion of the population when they take off
  + Thus, equilibrium depends on population size but proportion infected does not
* With smaller populations, there are many “in between” epidemic sizes…and as population increases, you either get a failed epidemic or a “predictable size” outbreak
* Notation: dot over a system state (SIR) means rate of change
  + The equations are the same as typical SIR but with \*average degree added to S and I
  + “Kermack-McKendrick equations”
  + m = # of other nodes one node infects
* R0
  + In “average” population, no one in particular is getting infected first
  + In heterogeneous, you would expect those that are more highly connected to get infected first, and thus R0 WOULD change with time!
  + R0 = B(K)/gamma
* Probability of an epidemic
  + The movement of individuals from one compartment to another is always proportional
  + Meaning you can always find the value of one class by the volume in other classes
  + All movements are independent and “memoryless” (just based on where you are at the current time)
  + Probability of no epidemic = f(0) (probability of one person becoming infected and not transmitting)
  + We often don’t care if a few people get infected, but we want to know that the infection goes no further than 1 generation
* Epidemic size
  + Want to find equilibrium level of infection, or S-dot and I-dot are 0
  + I.e. recovery and transmission rates are equal!
  + If Ieq < 0 OR gamma > B(K), then R0 is < 1 and disease will die out
  + If gamma < B(K), then RO >1 and disease can have 2 equilibriums (disease free and endemic)
* Alternates to SIR equations?
  + Lots of work of rearranging equations
  + Result: as R 🡪 infinity = 1 – e^-R0\*r(inf)
  + Where r = R/N (proportion recovered)
  + Start with guess of r=1 to solve, and it will always work!
  + Psi-hat: the expected # of transmissions a random individual has received by time t
  + Essentially we are just adding a variable so you can work at the node level by its degree

**LECTURE 3: EPIDEMICS IN NETWORKS**

Sample stochastic models

* As population increases, homogeneous populations (in (K)) look like clear epidemic or not
* As population increases, heterogeneous population still has a lot of noise but there are trends in epidemic or not
* Heterogeneity in degree distribution increases R0\*\*

**Impacts of network properties**

* Holding (K) fixed, degree heterogeneity increases R0 because of size-biased infection
* With small transmission rate, diseases are less likely to start with homogeneous populations
* BUT with large transmission rate, bimodal degree populations have smaller outbreaks with time because the epidemic can die out with low degree individuals
* If there is assortative mixing, more high degree people will become infected first and spread to other high degree people 🡪 larger R0 at first but epidemic may decrease slightly over time
* Longer partnership duration decreases R0 because repeated transmissions are “wasted” on those that are/were infected
  + Concurrency: idea that disease transmission occurs in 2 directions: a🡪b, b🡪c. Partnerships do not wait to form for when others end. If you do wait for one partnership to end, then transmission is slowed by the population structure
  + With SIS, there is a difference because the long lasting partnerships ~guarantee transmission. Then individual can become reinfected and re-spread to partners
* Clustering: people who infected are likely to have partners who were infected by others. For SIR, decreases R0. For SIS, high degree nodes have higher chance of becoming infected but who they will infect also are probably infected by someone else (complex)
  + With higher transmission rates, clustering has more of an effect on epidemic size
  + Yet for lower transmission rates, there may be an increase in successive epidemic sizes as those clusters are somewhat more realized and S pool grows (whereas high transmission will not so much because the chance someone you infected is already infected by someone you also interact with is high)
  + For calculating R0, don’t need to typically consider clustering
  + Size of epidemic is not really affected by clustering

Notes on deriving equations

* With SIR, etc. equations, we are assuming all individuals are “equal” in their S and I classes properties or characteristics
  + In a network, it DOES matter exactly who is infected
* Star example more like wolf network, but star networks are only useable for smaller populations with symmetry
* This is annoying, but it is a way to look at of the possibilities
* Overall, once you upscale to large networks…all of this is ~proportional to # S and I in the population, and changes by (K)
  + They make the Kermack-McKendrick
  + Basically make all regular SI models underestimates of true network
  + Not always “best” model
* Instead, we want a model that accounts for degree correlation (“assortative mixing”)
  + Include: pairs with their degree and their “status” (SIR class)
* Heterogeneous population model:
  + What is the probability that a susceptible individual with degree k is interacting with an infected node? = new probability of transmission and infection
  + Early on, the most connected people will become infected.
  + Disease can persist in any power-law network because the chain of transmission may hit a well-connected node and then re-start the transmission chain
  + There is no epidemic threshold for SIS disease, even if (K^2) is finite because there is always a pool of I around high degree nodes, which re-infects high degree nodes, and “restarts” outbreaks. This is contradictory to our prediction!

Revisiting probability of an epidemic

* Example: 6 sided dice 0-5
* Whatever the value (n) you roll, you roll the dice that value times
* What is the probability you stop rolling the dice?
* Once you have momentum, it will take a long time to stop OR it won’t stop in foreseeable future
* In dynamic network, you are always assuming that I is meeting new individuals all of the time. In static network, you need to reduce # of S, I, and k by 1 each time step
* Must take into account both initial infections and secondary infections, and that anyone in population can have a certain degree (based on degree distribution of the population)

SIS endemic equilibrium

* Adding and changing equations to make sense? Theta has no physical meaning

Scale free

* This is network in which (K) is finite, but (K^2) is infinite

**LAB 1: EPIFIRE TUTORIAL**

All labs are in discrete time and finite populations

Why network models?

* Compromise between complexity and comprehension
* But really consider WHY this type of model as opposed to other types of modeling. You MUST have of one of two features in your population to use networks:
  + Structure is important!
  + You have the data to build network!

**Mechanisms of network models**

Percolation

* Transmissibility: probability of transmission from I node to S node
* Node 🡪 node by transmissibility

Chain binomial

* Only difference, nodes can remain infectious for n time steps
* Node 🡪 node by transmissibility\*infectious period
* Cohorts are partially overlapping in that not all are becoming infected/recovering during same time steps

epiFire

* Node state evolution shows you a random sample of nodes and how they changed through the simulation
* Blue = S, red = I, yellow = R

Starting at section 3

**Poisson vs. Power-law**

* R0 tells us how quickly the epidemic spreads at first
* R0 does NOT tell us:
  + Final outbreak size
  + Patterns of spread
  + Stochasticity 🡪 proportion of events leading to epidemics or not
* When we keep transmissibility the same, there are larger outbreaks in more homogenous populations (Poisson)
* Poisson is a better approximation of mass-action models because contacts are more normally distributed

**Urban network for flu**

* How does changing the # initially infected change the epidemic size?
  + Because you are forcing those to be infected instead of them getting infected by some small probability
  + It also makes the epidemic end sooner
* Keep R0 = 1.1 but change around the distribution…what happens?
  + Poisson: only small outbreaks, no huge epidemics
  + Exponential: a few small outbreaks, no big
  + Power law: same as above
  + Urban: same as above
  + Constant: epidemics occur! In this case, mass action predicts fewer outbreaks than occur from network
  + Small world: more like a Poisson distribution of outbreak sizes, different from others
  + Because you are forcing contacts between individuals in the constant and small world, those networks have more outbreaks in general
    - Whereas Poisson, exponential, power, and urban, if you don’t infect many well-connected nodes, then you don’t get an epidemic

**Identifying epidemics**

**LECTURE 4: PERCOLATION APPROACHES TO DISEASE SPREAD**

Now we are working in discrete time

* Instead of disease travelling on existing edges, you an equivalently decide in advance whether the edge will exist
* The sums of all options will equal the probability of the traveling along edges approach
  + Meaning, whether or not you get 2/4 S and 2/4 I in the same community, and the connections among those nodes differ, it is the SAME as the probability 2/4 S and 2/4 I in any order
* Below threshold value for epidemic outbreak, the size of the outbreaks does NOT depend on networks size. Above that threshold, however, epidemic size IS proportional to network size
  + \*\* The proportion of the population that becomes infected and the probability of an epidemic then becomes the same!

Now back to continuous time = > transmission events not quite as independent, rates used instead of probabilities

* Directed percolation analogy:
  + Not sure why you would do this?
  + Encodes dynamic process into static network
    - Easier to do computations on
* Final size = want to know the probability that a node remains S = probability that node started susceptible and remains susceptible until time t
  + Each node is weighed by their relative degree distribution (relative to average of pop)
  + What you are actually calculating is the probability a node is still S by time t (theta is time steps)

Use part 4 new info from slide 34 on

* Slide 35 shows us:
  + Degree distributions make different epidemics
  + Just knowing average degree doesn't tell you much
  + More hetero degree distribution = faster initial spread
* Pr(epidemic) = final size of epidemic with discrete time
  + Not as true in continuous time because infections are not independent (dependent on person infected's node degree, and length of their infectious period)
  + Need to calculate the pr(node) does NOT cause epidemic by not transmitting to neighbor or neighbors not transmitting beyond themselves
  + With discrete time, nothing is conditional on infectious period of node because there is same probability of transmission for each time step
  + QUESTION TO ASK: isn’t cont time the same as discrete with added variation? By “picking” transmission events in sims, isn’t that captured in B?
    - Continuous time is different in that the probability that someone transmits is depends on how long they are infectious! In discrete, this is fixed. In continuous, this changes depending on the distribution you chose I and R from.
    - Yes, you don’t really “pick” the event. You “pick” the B and the I for that individual from distribution set a priori
* You lose independence in SIS as well because the I node 🡪 infects neighbors (maybe) and then recovers and is S again…but it is has changed its local environment! So that initial I node in a grouping now has higher pr(trans) from its neighbors that it infected itself.
  + This SIS, simulations and generalizations ONLY work with constant I and B rates!!
  + Poisson distributions are "memoryless" 🡪 the times in I and R do NOT depend on who is infecting whom
  + \*\* The equilibrium size of an SIS epidemic with Poissonian transmission and recovery equals the probability that an epidemic occurs (i.e. one node causing epidemic, but need high transmission rate B for this to be true) (SAME result as percolation with SIR disease)

Simulations

* Epidemic dependent on if you include degree distribution
  + R0 may predict no epidemic, but when you include (K), then that may not be the case
* The effect of clusters can be small if degree is large and transmission probability is low; it is much more likely that the nodes of a cluster are getting infected by other nodes rather than each other
* **SIR epidemics in configuration model networks – relates most to me**
  + R0 calculation on there
  + Assume partnerships lasts “forever” or not at all (same as “random encounters”)

**LAB 3: MODELING NETWORKS IN PYTHON**

* Why use an edge list or an adjacency matrix?
  + Matrix ALSO represents non-connections (zeros) so they take up more space
  + Computations on connections are better in matrix (linear algebra)
  + May want the same pairing listed more than once, which may be easier on edge lists
* Network X

D