Transmission of infections on contact networks II: Network Structure *vs.* disease dynamics

Winfried Just
Department of Mathematics, Ohio University

July 20, 2018
2^{nd} Portuguese Meeting in Biomathematics
University of Aveiro

Review: The kind of models covered here

We study the spread of Infectious diseases in fixed population of hosts (without demographics).

It is assumed that transmission of the disease can occur only during an effective contact between a susceptible and an infectious host who are connected by an edge in a given contact network.

We will use the simulation program IONTW that was written by Drew LaMar for some educational materials that we developed together with Hannah Callender Highlander and Natalia Toporikova. It is freely available for downloading at

`http://www.ohio.edu/people/just/IONTW/`

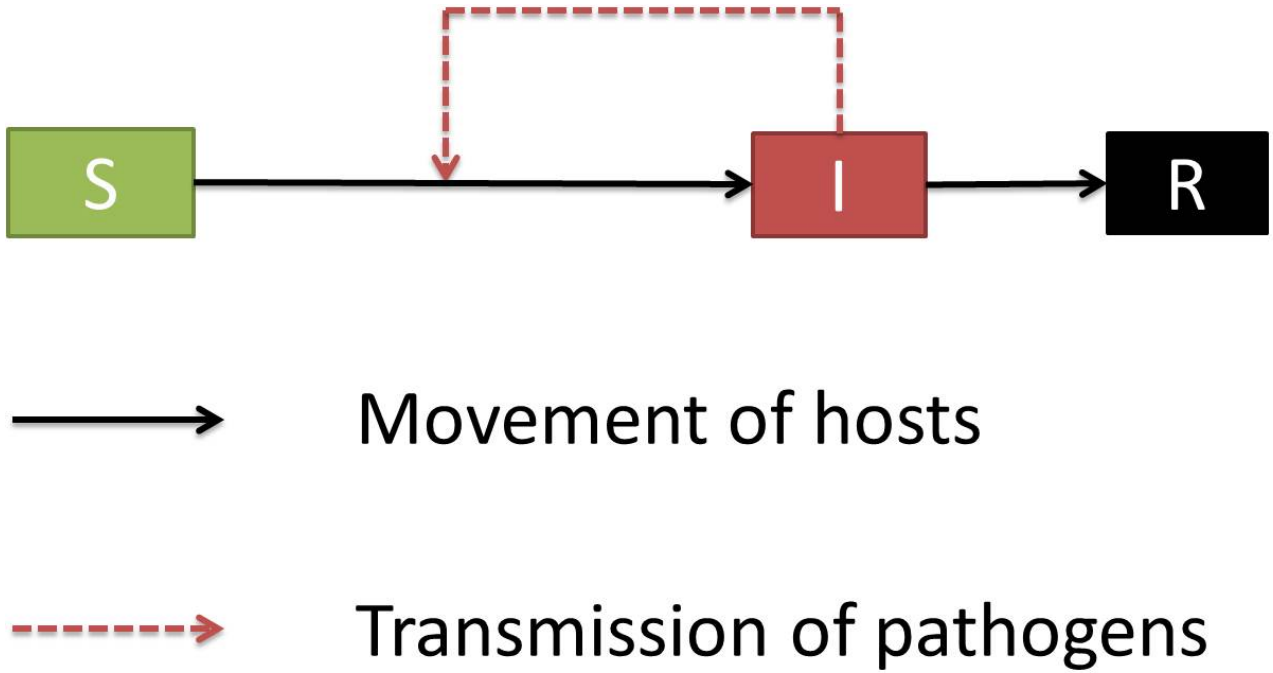
or for running directly on the web at

`https://qubeshub.org/iontw`

Review: *SIR*-models

Models of **type** *SIR* may be suitable for **immunizing infections** like measles or chicken pox where recovery confers permanent immunity.

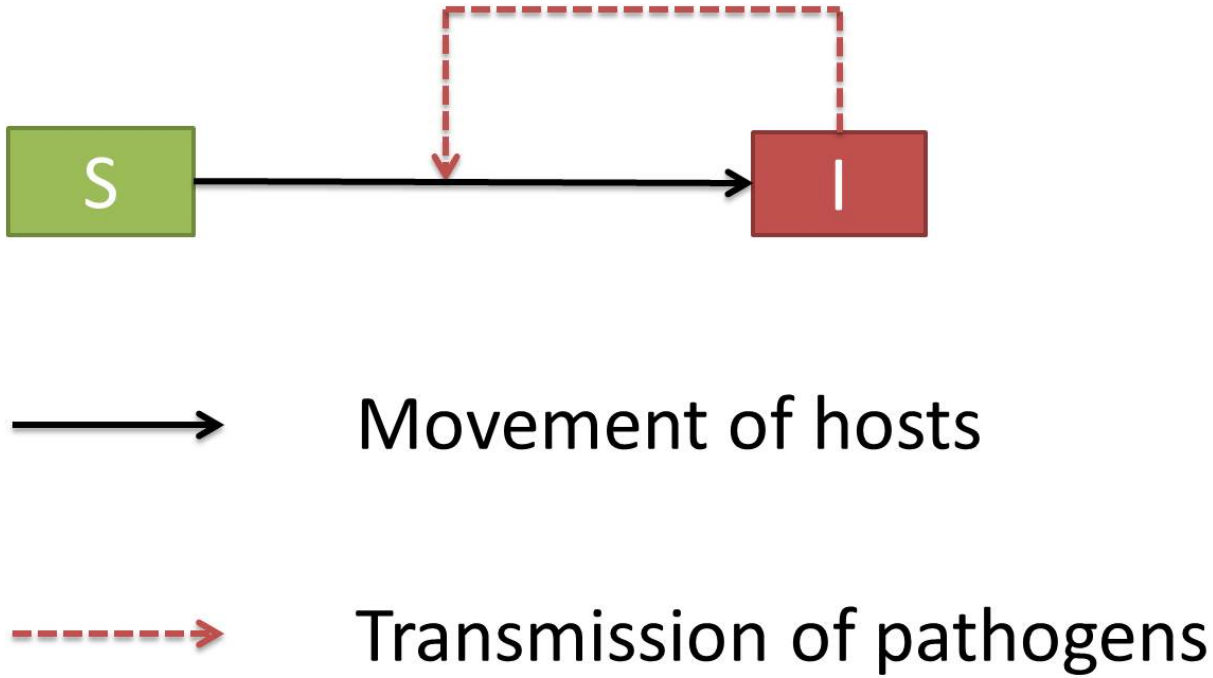
Schematically, such a model can be represented as:



SI-models

Models of **type** *SI* may be suitable for infections such as HIV/AIDS where there is no recovery. .

Schematically, such a model can be represented as:



Review: Some questions that we are trying to answer

- If one **index case** is introduced into an entirely susceptible population, will a **major outbreak** result? That is, will a significant fraction of the population eventually get infected?
- If a major outbreak does occur, what proportion of hosts will experience infection? This proportion is called the **final size** (of the outbreak).
- Which **control measures** are most effective in either preventing a major outbreak or reducing its final size?

Review: For agent-based models of type *SIR* or *SI* we need

- A population of N agents that represent hosts.
- A time line $t \in \mathbb{N}$ for discrete-time models or $t \in \mathbf{R}$ for continuous-time models.
- States. At any given time, an agent can be either in state S (susceptible), I (infectious), or R (removed).
- For discrete-time models:
  - A transmission probability b_{ij} that agent i, if currently infectious, will transmit a critical number of pathogens to agent j during the current time step.
  - A removal probability a_i that agent i, if currently infectious, will transition into state R by the next time step.
- For continuous-time models:
  - A transmission rate β_{ij} at which events of infections of agent j by agent i occur, if agent i is currently infectious and agent j is currently susceptible.
  - A removal rate α_i at which agent i, if currently infectious, will transition into state R.

Some special assumptions

In IONTW we always make the assumption of homogeneity of hosts. Mathematically speaking this meas that there are single constants a, α such that $a_i = a$ and $\alpha_i = \alpha$ for all i.

When $a = 0$ or $\alpha = 0$, then we obains SI-models.

Under the uniform mixing assumption there are single constants b, β such that $b_{ij} = b$ and $\beta_{ij} = \beta$ for all i, j with $i \neq j$.

Today we will not assume but work with network-based models of disease transmission.

In them we also have single parameters b, β, but assume that

- $b_{ij} = b, \beta_{ij} = \beta$ when i and j are connected in the contact network,
- $b_{ij} = 0, \beta_{ij} = 0$ otherwise.

Review: Contact networks as graphs

We can model contact networks as a **graph.**

The **nodes** of the graph represent hosts, the **edges** connect any two hosts who have reasonable frequent contact of the kind that may lead to the transmission of the particular infection.

Giving a precise meaning to "reasonably frequent" here is a decision that the modeler has to make.

The degree $deg(i)$ of a node i is the number of edges that connect i with other nodes.

Different contact networks can be relevant for different diseases.

The uniform mixing assumption can be modeled in this framework by assuming that the nework is the complete graph where all possible edges between the N nodes are included.

Paths and connected components in graphs

- A path from node i to node j in a graph is a sequence of nodes $(i = i_0, i_1, \ldots, i_\ell = j)$ such that $\{i_m, i_{m+1}\}$ is an edge for each $m < \ell$.
- The distance $d(i, j)$ is the length ℓ of the shortest path from i to j if such a path exists, and is ∞ otherwise.
- The connected component of a node i in a graph is the set of all nodes j with $d(i, j) < \infty$.
- The diameter of a connected component is the largest distance between any two nodes in this connected component.
- A graph is connected when it has exactly one connected component.

Paths, connected components, and disease transmission

- An infection can travel from the index case i to another node j only along a path of the contact network.
- Thus in an SIR-model, the set of nodes that will experience infection during an outbreak must be contained in the connected component of the index case.
- In an SI-model, the set of nodes that will experience infection during an outbreak will be equal to the connected component of the index case.

Random graphs

In reality, contact networks may not be well-structured. It is also difficult to collect enough reliable data on a contact network on an actual population.

But we can often estimate some network parameters like the mean degree of a node in the network, or, more generally, the degree distribution.

We can then assume that the actual network will be somewhat typical for networks that are drawn from a certain distribution (for a given population size N) and study the expected dynamics on such (instances of) random networks.

We say that (dynamics on) random graphs satisfy a given property (more formally: satisfy this property a.a.s — asymptotically almost surely if the probability that the property holds goes to 1 as the population size N for which we draw the random netwok goes to ∞.

Erdős-Rényi random graphs

Erdős-Rényi random graphs are the most basic example. Here each possible edge between the N nodes is included with probability p, independently for all pairs $\{i, j\}$ of distinct nodes $i \neq j$. They have the following properties (a.a.s.):

- The mean degree is close to $\lambda := p(N-1)$.
- The degree distribution is approximately Poisson with parameter λ and nodes with degrees much higher than λ are extremely unlikely.
- They are disconnected.
- For $\lambda < 1$, all connected components have size $O(\log N)$.
- For $\lambda > 1$, there exists exactly one giant component of size approximately ΘN, where Θ is the unique solution of $1 - x = e^{-\lambda x}$ in the interval $(0, 1)$. The diameter of this component will be of order $O(\log N)$.

Connection with the final size under uniform mixing

Consider a discrete-time SIR-model with $a = 1$ on a complete graph with N nodes.

When we simulate an outbreak, we would normally at each time step t and for each node i that is infectious at time t:

- Decide randomly and independently for each j that is still susceptible at time t whether there will be an effective contact between i and j at this time step (with probability b).
- Change the states of some nodes j to "infectious" according to these decisions.

Equivalently, we could:

- First draw an Erdős-Rényi random graph on N nodes with connection probability $p = b$.
- Then change the states of some nodes j at subsequent time steps of the simulation to "infectious" according to whether or not there is an edge in the Erdős-Rényi random graph.

Conclusions and a word of caution

Consider a discrete-time SIR-model with $a = 1$ on a complete graph with N nodes. Then for $\Theta > 0$ with $1 - \Theta = e^{-R_0\Theta}$:

- When $R_0 < 1$, then all outbreaks will be minor, with only $O(\log N)$ hosts experiencing infection.
- When $R_0 > 1$, then with positive probability, we will see major outbreaks of final size close to Θ.
- These predictions remain the same regardless of whether we consider exactly one or a small fixed positive number of index cases.
- We can estimate the probability that one index case causes a major outbreak as Θ.

Curiously enough, the last prediction will not hold if we instead consider a continuous-time model or a model with $a < 1$.

Why?

Clustering coefficients

Consider two randomly chosen people. Let p_r be the probability that they are friends.

Now consider a randomly chosen person and two friends of that randomly chosen person. Let p_f be the probability that they are friends.

Does it seem plausible that:

(a) $p_r > p_f$?

(b) $p_r = p_f$?

(c) $p_r < p_f$?

Real contact human networks will typically show some amount of clustering. This can be quantified by clustering coefficients. There are various definitions of these in the literature; here we will consider normalized clustering coefficients that essentially measure the ratio $\frac{p_f}{p_r}$.

The effect of clustering coefficients

Erdős-Rényi random networks are not realistic models of human contact networks because they have normalized clustering coefficients $= 1$.

For the same basic reproductive ratio R_0, final sizes tend to be smaller in networks with higher clustering coefficients.

Why?

The small-world property

Consider the following network of human contacts: For each pair of humans, connect them by an edge if they know each other well enough that they could ask each other for a small favor. It is generally believed that this graph is connected.

In 1967 the American social psychologist Stanley Milgram and his collaborators conducted a clever experiment to estimate the average distances between two people in this graph. The results indicate that these average distances are at most six. This is the famous six degrees of separation claim.

Why is this surprising?

Let us say that a class of graphs has the small-world property if some fixed positive percentile of the distances between nodes scales like $O(\log N)$.

Other examples of networks with the small-world property

Erdős-Rényi random networks are a.a.s. disconnected but still have the small-world property for connection probability $p > 1$ becuase the diameter of the giant connected component scales like $O(\log N)$.

Let us consider the graph whose nodes are all mathematicians. Connect two nodes by an edge if they have co-authored at least one paper. This graph has about 401 000 nodes, and a largest connected component of about 286,000 nodes. Mathematicians in this component have an Erdős number, which signifies the distance from Paul Erdős. These numbers range from 0 (Erdős himself) to 13, with a median of 5. There are 503 co-authors of Erdős with an Erdős number of 1, and 6 593 mathemaicians with an Erdős number of 2. `http://www.oakland.edu/enp/trivia/`

This graph also has a high clustering coefficient; it looks like a small-world network.

Recall our story: The Sisters of the Round Table

Consider the monastic order of the Sisters of the Round Table. The sisters spend most of their lives in their individual cells, where they devote themselves to prayer and meditation. The only time they have contact with each other is during meals that they take seated in a fixed order around a giant round table. Within this community, diseases can be transmitted only during mealtime.

Let us assume now that disease transmission can occur only between two sisters who have at most one other sister sitting in between them.

This network exhibits clustering, but does not have the small-world property.

But this is not the whole story ...

The Sisters of the Round Table who also will engage in conversation on their way to and from the table with their next-cell neighbors who may be seated across the table.

By adding the edges of Erdős-Rényi random graphs to the edges of nearest-neighbor graphs, we obtain **small-world models**. Such networks have both high clustering coefficients and the small-world property.

They might more realistically represent actual contact networks than either nearest-neighbor networks or Erdős-Rényi random graphs.

But: In these small-world models, the degree distribution is still very strongly concentrated around the mean value. Nodes with with extremely high degrees, that is with degrees significantly exceeding $\log N$ are still a.a.s. absent.

Some empirical observations

Empirically studied human contact networks tend to have degree distributions that roughly (very roughly!) conform to a power law with

$P(deg(i) = k) \sim k^{-\gamma}$,

where γ is some network-dependent constant with $2 < \gamma < 3$.

Such networks are often called scale-free networks.

Such networks typically have some nodes called hubs with extremely high degrees, of order around $N^{1/\gamma}$.

Methods for generating random scale-free networks

- We can generate the as configuration models or generic models for a given legitimate degree distribution.

  There are several subtly different algorithms for generating such random graphs.

- We can obtain examples for $\gamma = 3$ by using the preferential attachment model of Barabasi and Albert. Here nodes are added one-by-one and attached with higher probability to nodes that already have high degree.

- Variants of the preferential attachment model that also incorporate such processes as random rewiring can give scale-free distributions with $\gamma < 3$.

Modeling vaccination

Vaccination can be modeled by moving a set of susceptible nodes into the **R**-compartment prior to any outbreak.

Under the uniform mixing assumption, it is predicted that when a proportion of at least

$V_{hit} = 1 - \frac{1}{R_0}$

of hosts is vaccinated, the we obtain herd immunity so that no major utbreaks will be observed.

The fraction V_{hit} is called the herd immunity threshold.

Vaccinating significantly more than V_{hit} hosts is not cost effective.

Vaccinating much fewer than V_{hit} hosts has little effect.

Can we do much better in network-based models?

Vaccinating the hubs—which hubs?

If the degree distribution is scale-free, then vaccinating a mach smaller proportion than V_{hit} of nodes with very high degrees, the hubs, may result in herd immunity.

But how do we find high-degree nodes to vaccinate if we don't know the network structure?

Why do your friens have more friends than you do?

In 1991, S. L. Feld published and article with that title in the *American Journal of Sociology.*

He showed that in all graphs of friendships with some heterogeneities of degrees most people have fewer friends than their friends have, on average.

This fact is called the Friendship Paradox.

Why would this be true?

Is this useful for vaccination strategies?

In acquaintance vaccination, we randomly pick a subset of hosts, ask each of them to name a random acquaintance, and then vaccinate that acquaintance.
But what if the acquaintance refuses to get vacinated?

Directions of future reseach

- When vaccination is voluntary, it can be studied with vaccination games. Together with Ying Xin and David Gerberry, we have beed studying such games where people base their decisions on imitation rather than rational calculations of costs. We obtained some interesting results that I will present next week in Lisboa under the uniform mixing assumption. This opens up a wide area of studying our version of imitation for network-based models of disease transmission.
- Even for optimal strategies of mandatory vaccination there are many open problems, especially on small-world networks.
- There has been surprisingly little work on versions of small-world models that are obtained by ading a scale-free random graph instead of an Erdős-Rényi random graph to a regular grid.
- Relatively little work has been done on network-based models with some heterogeneities of hosts.

Resources: Software

My web page `http://www.ohio.edu/people/just/IONTW/` and its mirror page `https://qubeshub.org/iontw` contains links to a number of background readings and modules for further exploration.

You can also download IONTW from this page, or run a web-based version at

`https://qubeshub.org/community/groups/iontw/iontwsimtool`

after a (free) sign-up procedure.

Resources: Book chapters

The following book chapters may be suitable as modules in and introductory mathematical biology course, for self-study, or as a basis for guided undergraduate research projects:

- Winfried Just, Hannah Callender, M. Drew LaMar, and Natalia Toporikova (2015); *Transmission of infectious diseases: Data, models, and simulations.* In Raina Robeva (ed.), *Algebraic and Discrete Mathematical Methods for Modern Biology,* Academic Press, 193–215.
- Winfried Just, Hannah Callender, and M. Drew LaMar (2015); *Disease transmission dynamics on networks: Network structure vs. disease dynamics.* In: Raina Robeva (ed.), *Algebraic and Discrete Mathematical Methods for Modern Biology,* Academic Press, 217–235.
- Winfried Just and Hannah Callender Highlander; *Vaccination strategies for small worlds.* In A. Wootton, V. Peterson, C. Lee, eds., *A Primer for Undergraduate Research: From Groups and Tiles to Frames and Vaccines,* Springer Verlag, 2018, 223–264.