Differential equation models of disease transmission*

Winfried Just† and Hannah Callender‡

December 23, 2015

In this module we explore ODE models of disease transmission and compare some of their predictions with those of agent-based models. Part of this material will be referenced in later modules.

1 Construction of basic ODE models for disease transmission

Most readers will already have encountered at least one ODE model of disease transmission. Here we briefly review the construction of such models and some of their predictions. For an in-depth review of these models the reader may want to consult the survey article [1].

All ODE models of disease transmission are continuous-time compartment-level models. Here we focus on models of type SI, SIS, SIR, and $SEIR$. The ODE models of these types have variables S, I, and, if applicable, R, E. They can be interpreted as actual numbers, expected numbers, or (actual or expected) proportions of hosts in the **S**-, **I**-, **R**, and **E**-compartments. These variables will change over time, and $S(t), I(t), R(t), E(t)$ denote their values at a given time t. The state of the system at time t is the vector $(S(t), I(t))$ for SI- and SIS-models, the vector $(S(t), I(t), R(t))$ for SIR-models, and the vector $(S(t), I(t), R(t), E(t)$ for $SEIR$-models.

We will ignore demographics, that is, births, immigration, emigration, and deaths from causes that are unrelated to the disease. Mathematically this means that at all times the equality

$$S(t) + I(t) + R(t) + E(t) = N \tag{1}$$

holds, where N is a constant that represents the size of the host population. If the variables are interpreted as proportions, then we will need to choose $N = 1$. Equation (1) is an example of what mathematicians call a *conservation law*. It says that the value of some quantity, total population size in this case, remains constant. Conservation laws allow us to reduce the number of variables by expressing one of them in terms of the others and

†Department of Mathematics, Ohio University, Athens, OH 45701 E-mail: mathjust@gmail.com

‡University of Portland E-mail: callende@up.edu

the conserved constant. This is often a useful trick. We will use it later for analyzing our models, but for constructing them it is usually easier to work with a full set of variables.

ODE models are based on derivatives of the variables with respect to time. For this to make sense, the variables should be able to take any value in some interval of the real line. This is strictly speaking impossible for the interpretation of the variables as actual sizes of the compartments (try to think of 3.14 susceptible humans). For large N, we can alleviate the problem to some extent by assuming that the variables count hosts in batches of thousands or even millions. In this interpretation, $S = 3.14$ would make sense, but $S = \pi$ still would not. Thus ODE models need to be taken with a grain of salt. One can think of them as useful approximations for very, very large populations. But for small populations their predictions may be misleading.

1.1 SI models

Let us begin with the easiest case of an SI model. Here we have only two variables: S and I. A host can only move from the **S**-compartment to the **I**-compartment as the result of an effective contact with an infectious host. If this happens, $I(t)$ will increase, and $S(t)$ will decrease. Since new infections are the only process that occurs in this model, the rate $\frac{dI}{dt}$ at which $I(t)$ changes will be the rate at which hosts move from **S**-compartment to the **I**-compartment. In other words, $\frac{dI}{dt}$ will be equal to the rate at which effective contacts between infectious and susceptible hosts occur.

In order to derive a formula for this rate, recall that compartment-based models are based on the uniform mixing assumption. This means that for all possible pairs (i, j) of an infectious host i and a susceptible host j the rate β at which these particular hosts come into effective contact is the same. Thus the total rate of effective contacts between infectious and susceptible hosts will be β times the product of S and I, which represents the number of all such pairs (i, j). You may recall from your chemistry classes that this is exactly analogous to the law of mass action kinetics for a chemical reaction with two reactants[1] and gives:

$$\frac{dI}{dt} = \beta SI. \tag{2}$$

Since the only process that we are considering in this model are new infections, the rate at which I increases must be the same as the rate at which S decreases. In other words, $\frac{dS}{dt} = -\frac{dI}{dt}$. This gives the following system:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI. \end{aligned} \tag{3}$$

[1] In this interpretation, (3) would describe the rate of change of the reactants and reaction products for a single chemical reaction $S + I \to 2I$ with rate constant β.

We have built our first ODE model! Let us see what the model predicts. First note that $\frac{d(S+T)}{dt} = \frac{dS}{dt} + \frac{dI}{dt} = -\beta SI + \beta SI = 0$. This is exactly as it should be. In an SI-model we have $R(t) = E(t) = 0$ at all times t, and (1) tells us that $S(t) + I(t) = N$ is a constant. We can now substitute $S = N - I$ and rewrite (3) in a single variable I as follows:

$$\frac{dI}{dt} = \beta(N - I)I. \tag{4}$$

The ODE (4) has two *fixed points* or *equilibria,* $I^* = 0$ and $I^{**} = N$, where $\frac{dI}{dt} = 0$ and the state of the system does not change. As long as $0 < I < N$, the derivative $\frac{dI}{dt}$ is positive, and I will increase. Thus in all solutions that start with an initial condition $I(0) > 0$ the number of infectious individuals will keep increasing towards the value $I = N$. Technically, this means that the equilibrium $I^* = 0$ is *unstable* and the equilibrium $I^{**} = N$ is *locally asymptotically stable.*

Exercise 1 *Does equation* (4) *look familiar? Have you seen this equation before in the context of population dynamics, with different names of the variables and constants? Do all biologically realistic solutions of that other model also make biological sense in the SI-model when we reinterpret the meaning of the variables?*

If we substitute $S = N - I$ back into (3) we find that the original system has two equilibria: $(S^*, I^*) = (N, 0)$ is called a *disease-free* equilibrium (as there are no infected hosts). It is unstable. $(S^{**}, I^{**}) = (0, N)$ is called an *endemic* equilibrium (as at least some hosts are infected). It is locally asymptotically stable. Moreover, all biologically realistic solutions of (3) that start with $(S(0), I(0)) \neq (S^*, I^*)$ will approach the endemic equilibrium.

Exercise 2 *Is it possible to derive from* (3) *the theoretical prediction that if there is no recovery from the disease and some hosts are infectious, the whole population will eventually become infectious?*

1.2 SIR models

Now let us build an ODE version of the SIR model. Here we need three variables, S, I, R, and we need to consider simultaneously two processes: new infections and removal of hosts.

The first of these processes is the only one that affects susceptible hosts. As in the SI model, we get $\frac{dS}{dt} = -\beta SI$. This process also contributes a term βSI to $\frac{dI}{dt}$.

Removal of hosts is the only process that affects R. It proceeds at rate $\frac{dR}{dt}$. Only hosts in I are candidates for removal, and in an SIR-model removal occurs as a result of recovery from the disease with permanent immunity or of death from the disease. In none of these scenarios can hosts ever leave the **R**-compartment. Thus $\frac{dR}{dt}$ is proportional to the number of infectious hosts, that is, $\frac{dR}{dt} = \alpha I$, where α is the rate of removal. Since removal of hosts decreases I, the process also contributes a term $-\alpha I$ to $\frac{dI}{dt}$.

We obtain the following model:

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \alpha I \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \tag{5}$$

This is the famous model of Kermack and McKendrick that was published in [3], one of the very first papers on mathematical modeling of disease transmission dynamics. The predictions of the SIR-model (5) will be discussed in some detail in Section 2.

Exercise 3 *Show that* (5) *satisfies the conservation law* $S(t) + I(t) + R(t) = N$.

1.3 $SEIR$ models

If we add an **E**-compartment to the SIR-model, we need four variables, S, E, I, R. Effective contacts with infectious hosts move susceptible hosts out of the **S**-compartment at rate $\frac{dS}{dt} = -\beta SI$ and into the **E**-compartment at rate βSI. Hosts in the **E**-compartment move into the **I**-compartment at rate γE, and removal works exactly as in the SIR-model.

We obtain the following ODE-version of the $SEIR$-model:

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS \\ \frac{dE}{dt} &= \beta IS - \gamma E \\ \frac{dI}{dt} &= \gamma E - \alpha I \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \tag{6}$$

The analogous argument as in the solution of Exercise 3 shows that (6) satisfies the conservation law (1). Some predictions of the $SEIR$-model (6) will be discussed in Section 2.

1.4 SIS models

The reader may already be familiar with the following ODE version (7) of SIS-models. It can be derived similarly to the ODE versions of SI-, $SEIR$-, and SIR-models.

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS + \alpha I \\ \frac{dI}{dt} &= \beta IS - \alpha I. \end{aligned} \tag{7}$$

Since $S(t) + I(t) = N$ for all t, by letting $S = N - I$, this model can be simplified to

$$\frac{dI}{dt} = \beta I(N - I) - \alpha I = \beta I(N - \frac{\alpha}{\beta} - I). \tag{8}$$

This looks somewhat familiar.

Exercise 4 *Where have you previously seen ODE models like* (8)*? What do models* (8) *and* (7) *predict?* Hint: *Compare with Exercise 1 and the subsequent analysis of the SI-model.*

2 The basic reproductive ratio R_0 and the final size

2.1 Calculating R_0

The basic reproductive ratio R_0 is the expected number of secondary infections that will be caused by an average index case that is introduced into a large and entirely susceptible population. In the types of ODE models that we consider here uniform mixing is assumed and every host is considered "average." A more detailed discussion of this assumption can be found in Section 9.2 of [2]. However, since the variables may correspond to counting hosts in large batches, such as thousands or millions, it is not immediately clear which initial state would correspond to a single index in a large and otherwise susceptible population. In the context of an SIR-model, we can think of such an initial state as $(S(0), I(0), R(0)) = (N - \varepsilon, \varepsilon, 0)$, where ε is a very small but positive number.

The system (5) cannot be solved in closed form. For the purpose of studying R_0 we may consider a related simpler system. First note that in a very large population the proportion of susceptible hosts will not change much during the interval of infectiousness of the index case. So let us for simplicity assume that $S(t) = N$ during this interval, and that $\varepsilon = 1$. We can think of an ODE model where $I(0) = 1$ represents a single index case as describing how fractions of this index case cause secondary infections and get removed over time. This does not strictly speaking make biological sense, but works well for analyzing the ODEs. Now let $X(t)$ denote the part of the index case that has not yet been removed until time t. Moreover, let $G(t)$ denote the number of hosts that get infected by the index case until time t. In other words, $G(t)$ is the number of hosts in the first generation $Gen(1)$ of the infection until time t.

It will be convenient to consider the following Initial Value Problem:

$$\begin{aligned} \frac{dX}{dt} &= -\alpha X \\ \frac{dG}{dt} &= \beta X N \\ X(0) &= 1 \\ G(0) &= 0. \end{aligned} \tag{9}$$

One can think of the DE in (9) as describing a modified version of (5) where an additional susceptible host is introduced at the start of the outbreak and each host who becomes infected by the index case immediately gets removed from the system and replaced by a new susceptible host. The IVP (9) can easily be solved analytically. We get

$$\begin{aligned} X(t) &= e^{-\alpha t}, \\ G(t) &= \frac{\beta N}{\alpha} - \frac{\beta N}{\alpha} e^{-\alpha t}. \end{aligned} \tag{10}$$

Note that the second line of (10) implies that

$$\lim_{t \to \infty} G(t) = \frac{\beta N}{\alpha}, \tag{11}$$

which can be interpreted as the expected total number of secondary infections caused by the index case in the system described by (9).

When N is very large, the difference between (9) and the corresponding IVP for (5) becomes negligible. Since we have assumed that $X(0) = 1$ corresponds to 1 infectious host, it follows that the right-hand side of (11) will give the following approximation for R_0 in the the SIR-model (5) that becomes arbitrarily good when $N \to \infty$:

$$R_0 = \frac{\beta N}{\alpha}. \tag{12}$$

This is the same expression that we get for continuous-time network-based models where the contact network is the complete graph K_N and N is very large. Note that while (7) of the brief overview of network-based models of transmission of infectious diseases at this web site[2] gives only approximate equality, the simplifying assumptions of the ODE model turn it into an actual equality.

Exercise 5 *How would you modify the above derivation of R_0 for ODE models of type $SI, SEIR$ and SIS? What expressions do you obtain and how can they be interpreted?*

2.2 The predicted final size in SIR- and $SEIR$-models

Now let us explore some important predictions of the SIR-model (5). If $I(t) = 0$, then $\frac{dS}{dt} = \frac{dI}{dt} = \frac{dR}{dt} = 0$. Thus each of the vectors $(S, 0, N - S)$ for $0 \leq S \leq N$ is an equilibrium, and these are the only biologically meaningful equilibria of (5). All of these equilibria are disease-free.

As long as there are infectious hosts ($I(t) > 0$), the variable R increases. If there are both susceptible and infectious hosts, that is, if $I(t)S(t) > 0$, then the variable S decreases.

The situation is more nuanced for I: The equation for $\frac{dI}{dt}$ shows that as long as $I > 0$, this variable will increase if $S > \frac{\alpha}{\beta}$ and decrease if $S < \frac{\alpha}{\beta}$. If we count hosts in large batches, then we can assume that for an initial state with one infectious host in an otherwise susceptible

[2] `http://www.ohio.edu/people/just/IONTW`

population we have $S(0) = N - \varepsilon \approx N$. Thus when $R_0 = \frac{\beta N}{\alpha} \leq 1$, then the inequality $S \leq N \leq \frac{\alpha}{\beta}$ will hold at all times, and hence the model predicts $\frac{dI}{dt} \leq 0$ and a decrease of I at all times. In contrast, when $R_0 = \frac{\beta N}{\alpha} > 1$, the number of infectious hosts will initially increase, attain its maximum at a time when $S(t) = \frac{\alpha}{\beta} = \frac{N}{R_0}$, and subsequently decrease.

For any given initial state $(S(0), I(0), R(0))$, the system will approach one of the equilibria $(S, 0, N - S)$. But which which one? Let us use the following notation:

$$\begin{aligned} S(\infty) = \lim_{t\to\infty} S(t), \qquad I(\infty) = \lim_{t\to\infty} I(t), \qquad R(\infty) = \lim_{t\to\infty} R(t), \\ s(t) = \frac{S(t)}{N} \quad \text{and} \quad r(t) = \frac{R(t)}{N} \quad \text{for } r \in [0, \infty]. \end{aligned} \tag{13}$$

In this terminology, SIR-models predict that $I(\infty) = 0$. For an outbreak that starts with an initial state with $R(0) = 0$, the fraction $r(\infty) = 1 - s(\infty)$ represents its final size, that is, the proportion of all individuals that will eventually experience infection. When $R(0) = 0$, that is, $S(0) + I(0) = N$, the value of $s(\infty)$ depends on $s(0)$ and R_0 as follows:

$$s(\infty) = s(0)e^{R_0(s(\infty)-1)} \tag{14}$$

We will show in Subsection 2.3 how this equation can be derived. It cannot be solved explicitly, but one can analytically investigate the behavior of its solutions. The following lemma summarizes these properties.

Lemma 1 *(a) Suppose $0 < s(0) < 1$ is fixed. Then*

(a1) For every $R_0 > 0$ Equation (14) has exactly one solution $s(\infty) = s(R_0, \infty)$ in the interval $[0, 1]$. This solution satisfies the inequalities $0 < s(\infty) < 1$.

(a2) The function $s(R_0, \infty)$ is decreasing, that is,
$R_0^- < R_0^+$ implies $s(R_0^-, \infty) > s(R_0^+, \infty)$.

(a3) $\lim_{R_0\to 0^+} s(R_0, \infty) = s(0)$.

(a4) $s(R_0, \infty) < \frac{s(0)}{R_0}$. In particular, $\lim_{R_0\to\infty} s(R_0, \infty) = 0$.

(b) Suppose $s(0) = 1$. Then

(b1) For every $R_0 > 0$ the value $s(\infty) = 1$ is a solution of (14).

(b2) If $R_0 \leq 1$, then (14) has no solutions in the interval $(0, 1)$.

(b3) If $R_0 > 1$, then (14) has exactly one solution $s(R_0, \infty)$ in the interval $(0, 1)$.

(b4) The function $s(R_0, \infty)$ is decreasing, that is,
$1 < R_0^- < R_0^+$ implies $s(R_0^-, \infty) > s(R_0^+, \infty)$.

(b5) $\lim_{R_0\to 1^+} s(R_0, \infty) = 1$.

(b6) $s(R_0, \infty) < \frac{1}{R_0}$. In particular, $\lim_{R_0\to\infty} s(R_0, \infty) = 0$.

The following exercise takes some effort, but only requires knowledge of basic calculus.

Exercise 6 *Prove Lemma 1.* Hint: *Consider a function that expresses the difference between the left-hand side and the right-hand side of* (14) *and investigate its behavior at the endpoints of the interval* $[0, 1]$ *as well as its first and second derivatives. For points (a2) and (b4) use implicit differentiation.*

Introduction of a single infectious host into a large otherwise susceptible population corresponds to a situation where $s(0) \approx 1$. For $R_0 \leq 1$, Lemma 1 implies that in this case the value of $s(\infty)$ will be very close to 1. In biological terms, introduction of one or very few infectious hosts into an entirely susceptible population will result in a minor outbreak, with $r(\infty)$ very close to zero. On the other hand, if $R_0 > 1$ and $s(0) = 1$, then (14) also has a second solution in the interval $(0, 1)$. If $s(0)$ is slightly less than 1, then $s(\infty)$ will be near this second solution. In biological terms, introduction of one or very few infectious hosts into an entirely susceptible population will result in an major outbreak, with $r(\infty)$ comprising a fixed positive fraction of the population.

It is interesting to compare the predictions of $SEIR$-models and SIR-models with the same parameters α and β. The latter are a simplification of the former, based on the biologically unrealistic assumption that hosts become infectious immediately at the time of exposure. One would like to know how much this simplification would influence the predictions of the models.

Trajectories in $SEIR$-models (6) behave in qualitatively similar ways as trajectories of SIR-models (5). The number $S(t)$ of susceptible hosts always decreases. While $E(t)$ and $I(t)$ may initially increase, both of these quantities eventually decrease, and if we define $E(\infty) = \lim_{t\to\infty} E(t)$, then

$$E(\infty) = I(\infty) = 0. \tag{15}$$

In other words, each trajectory of (6) will eventually approach one of the disease-free equilibria $(S, 0, 0, N - S)$. Interestingly enough, as we will show in Subsection 2.3, for $SEIR$-models the exact same equation (14) as for SIR-models predicts which equilibrium will be approached from a given initial state with $R(0) = 0$. In particular, this equation predicts that the final size of an outbreak will be the same as for the corresponding SIR-model. This is the main reason why SIR-models are often adequate for deriving realistic predictions. While the simplifying assumption that hosts become infectious immediately at the time of exposure to the disease is biologically unrealistic, it does not significantly distort the predictions when the primary interest is in the final size of outbreaks.

2.3 Derivation of (14)

Let us now show how Equation (14) can be derived. First consider the SIR-model (5). Define a new quantity

$$C(t) = \frac{\alpha}{\beta} \ln S(t) - S(t) - I(t). \tag{16}$$

Exercise 7 *Show that the quantity $C(t)$ is conserved along all solutions of* (5)*, that is, show that for each solution of the system there exists a constant C such that $C(t) = C$ for all t.*

Recall from Section 2 that $S(\infty) = \lim_{t\to\infty} S(t)$ denotes the number of hosts that will escape infection and that $I(\infty) = \lim_{t\to\infty} I(t) = 0$. The conservation law for $C(t)$ implies that

$$\frac{\alpha}{\beta}\ln S(0) - S(0) - I(0) = \frac{\alpha}{\beta}\ln S(\infty) - S(\infty) - I(\infty). \tag{17}$$

Exercise 8 *Derive equation* (14) *from equation* (17)*.*

Exercise 9 *Prove that Equation* (14) *is also valid for the $SEIR$-model as defined by* (6)*.* Hint: *Define a quantity $D(t)$ that is satisfies an analogous conservation law for* (6) *as $C(t)$ does for* (5)*.*

3 Modes of transmission

So far we have assumed that the population size N is fixed. But it is not entirely clear whether the constant β will be the same for populations of different sizes. In order to reflect the possible dependence of β on N, the SIR-model (5) should really be written in the following form:

$$\begin{aligned}\frac{dS}{dt} &= -\beta(N)IS\\ \frac{dI}{dt} &= \beta(N)IS - \alpha I\\ \frac{dR}{dt} &= \alpha I.\end{aligned} \tag{18}$$

Under so-called *density-dependent transmission* the value of β does not depend on N. That is, we will have $\beta(N) = \beta$ for all N, which gives (5). Under so-called *frequency-dependent transmission* the value of $\beta(N)$ is inversely proportional to N so that $\beta(N) = \frac{\beta_1}{N}$ for some constant β_1. Since demographics are ignored here and N remains constant, we still get (5) with $\beta = \frac{\beta_1}{N}$. But if we want to estimate the relevant β for a population of size N_1 based on data that were collected from a population of a different size N_0, the difference between the two modes of transmission matters.

To get an intuition about the meaning of the phrases "density-dependent" and "frequency-dependent," think about Mr. Jones who socializes by taking an hour-long walk through his neighborhood each day and makes contact by shaking hands with each passer-by. What happens to the rate $\beta(N)N$ at which Mr. Jones shakes hands when the number of inhabitants of Mr. Jones's hometown Shakersville increases from $N = N_0$ to $N = N_1$?

If Shakersville is confined to a small valley surrounded by steep hillsides, then the population density of Shakersville will increase by a factor of $\frac{N_1}{N_0}$. Assuming that Mr. Jones lives in a typical neighborhood, this will result in a corresponding increase in the number of people in his neighborhood, which also increases the rate $\beta(N)N$ by the same factor. Thus $\beta(N_1)N_1 = \frac{N_1}{N_0}\beta(N_0)N_0$. It follows that $\beta(N)$ does not depend on N and is equal to a constant β. In other words, we get density-dependent transmission.

On the other hand, if Shakersville is a city in the great plains, then most likely the population growth will be absorbed by newly built suburbs and will not affect the mean number of people in a typical neighborhood. In particular, it will not alter the rate at which Mr. Jones shakes hands. Thus $\beta(N_1)N_1 = \beta(N_0)N_0$. It follows that $\beta(N)$ is inversely proportional to N, that is, $\beta(N) = \frac{\beta}{N}$ for some constant β. We get frequency-dependent transmission.

Many other possible forms for the function $\beta(N)$ have been considered in the literature; see [4] for a review. The ODE-models of Section 1 were implicitly constructed under the assumption of density-dependent transmission.

Exercise 10 *How would you write a modified version of the SIS-model* (7) *that reflects an assumption of frequency-dependent transmission?*

Exercise 11 *Assume* $\beta = 2$ *and* $\alpha = 1$. *What happens in an SIR-model to* R_0 *and the final size as predicted by* (14) *when* $N \to \infty$ *and*

(a) transmission is density-dependent,

(b) transmission is frequency-dependent?

If you have worked through module *Exploring Erdős-Rényi random graphs with IONTW* at this web site[3] or through Module 6 of [2], the following exercise may be of interest to you.

Exercise 12 *Suppose you want to model the contact network in Shakersville by Erdős-Rényi random graphs* $G_{ER}(N, \lambda(N))$.

(a) How should $\lambda(N)$ *depend on* N *if Shakersville is located in a small valley surrounded by steep hillsides?*

(b) How should $\lambda(N)$ *depend on* N *if Shakersville is located in the great plains?*

4 Exploring the relation between ODE models and agent-based models

ODE models of disease transmission are relatively easy to study, but have an obvious drawback. Consider an initial state $(S(0), I(0), R(0))$ of (5) with one index case in an otherwise susceptible populations so that $S(0) \approx N$ and $R(0) = 0$. Assume that the value

[3] `http://www.ohio.edu/people/just/IONTW`

of $I(0)$ is positive but very small, so that it corresponds to one infectious host. If $R_0 > 1$, then the ODE model predicts that S will *always* decrease toward $S(\infty) = s(\infty)N$. Thus if taken too literally, the ODE version would predict that when $R_0 > 1$, introduction of an index case into an otherwise susceptible population will *always* result in a major outbreak. This is unrealistic. Even when $R_0 > 1$, there is a positive probability $z_\infty > 0$ that such an initial state will lead only to a minor outbreak. ODE models do not account for this phenomenon; in particular, they do not allow us to estimate z_∞.

The problem is that ODE models are *deterministic* and entirely ignore the randomness involved in disease transmission. In contrast, the agent-based models that are embodied in IONTW do account for the inherent variability between outbreaks. In this section we will explore how some predictions of ODE models relate to the outcomes that we may expect to observe when we simulate agent-based models with the same parameters using IONTW.

4.1 Exploring SI-models with IONTW

Recall from Exercise 2 that the ODE version of SI-models predicts that eventually the whole population will become infectious. Let us see how this prediction plays out in agent-based models.

Open IONTW, click **Defaults**, and work with the following parameter settings:

model-time: Discrete
time-step: 1
infection-prob: 0.001
end-infection-prob: 0
num-nodes: 60
auto-set: On

Click **New**, the **Go**. Sit back, relax, and watch the changes in the **Disease Prevalence Plot**. Since we are simulating an SI-model, when nothing seems to happen any more, you will need to terminate the simulation by clicking **Go** again. Eventually, the whole population will become infectious. Record the time when this seems to happen. Repeat a couple more times.

Next change

network-type $\rightarrow$ **Nearest-neighbor 1**
d: 1

Repeat the previous explorations. You may need to move the speed control slider to the right for comfortable viewing.

Now change

network-type $\rightarrow$ **Erdos-Renyi**
lambda: 2

Repeat the previous explorations.

Exercise 13 *ODE models are based on the uniform mixing assumption. The model of Subsection 1.1 predicts that eventually the whole population will become infectious.*

(a) Does the uniform mixing assumption appear to be necessary *for the prediction to carry over to network-based models? If not, formulate a* necessary and sufficient *condition on the contact network that will assure that the entire population will eventually become infectious.*

(b) If a network-based SI-model predicts that the whole population eventually becomes infectious, does the structure of the network influence the time that it takes until the whole population becomes infectious? Which properties of the network might best predict the time it takes until all nodes become infectious?

4.2 Taking into account demographics

While SI-models on fixed populations with connected contact networks predict that eventually the whole population will become infectious, in the explorations of Subsection 4.1 you have seen that it may take a very long time until this happens. If this time is of the same order of magnitude as the average life span of a host, the assumption of a fixed populations is no longer justified, and the prediction may become biologically unrealistic.

Now let us briefly explore what happens if we include demographics, that is, immigration, emigration, births, and mortality in the ODE version of SI-models. Mortality refers to deaths from causes that are unrelated to the disease that is being studied. Thus the per capita mortality rate μ should be the same for hosts in all compartments. The birth rate per host may or may not depend on whether the host is susceptible, infectious, or removed; births by infectious mothers may or may not transmit the disease to the newborn. The overall rates of immigration may not depend on the actual population size, while the overall rate of emigration will generally be higher if the environment is crowded. It is likely to depend on the population size $N(t)$ in complicated ways.

As you can see, demographics may enter the picture in many different and complex ways. We will restrict ourselves here to one particularly simple example, assuming a fixed rate B of immigration that does not depend on N and a fixed per capita mortality μ, with no births or emigration.

Exercise 14 *Write down an equation for $\frac{dN}{dt}$ that describes the dynamics of the population size $N(t)$ under these assumptions. Find all equilibria of this ODE and their stability.*

Now let us explore what happens in an SI-model with this type of demographics. Assume that all new immigrants are susceptible. Since N can no longer be assumed constant, we need to be explicit about how β depends on N; for simplicity assume density-dependent transmission with $\beta(N) = \beta = const.$

Exercise 15 *Write down the analogue of the ODE model* (3) *with the demographics described as above. Explore its predictions and compare with the ones described for* (3) *in Subsection 1.1. Revisit the solution of Exercise 2 and compare with the predictions of the new model.*

4.3 Exploring the final size in SIR- and $SEIR$-models

Open IONTW, click **Defaults**, and change the following parameter settings:

model-time: Continuous
time-step: 1
infection-rate: 0.01
end-infection-rate: 1
num-nodes: 150
auto-set: On

For the chosen settings, (12) predicts $R_0 = \frac{(0.01)(150)}{1} = 1.5$. Click **New** to create a new network, and then **Metrics**. Look up the value of `R0` in the **Command Center**. You will see that it is slightly smaller than but very close to 1.5. The reason for the discrepancy is that ODE models implicitly assume that the actual population size N is very, very large, while our choice $N = 150$ is rather smallish in terms of this assumption. IONTW uses a different formula that is more accurate for small population sizes. It is given as Equation (5) in the brief overview of network-based models of disease transmission at this web site[4] and derived in Module 5 of the online appendix for [2]. For the purposes of this module we can work with the estimate $R_0 \approx 1.5$.

The model you have just set up is a continuous-time SIR-model on a complete graph. The initial state has 1 infectious host in an otherwise susceptible population so that $s(0) = \frac{149}{150} = 0.9933$. The corresponding ODE model would predict that $s(\infty)$ is the solution of (14) for this value of $s(0)$ and $R_0 = 1.5$. By numerically solving the equation, we find that we should expect

$$s(\infty) = 0.4099. \tag{19}$$

Let us see how the expectation works out. First run a few preliminary simulations by clicking **Go** and then **New** to initialize the next one. Adjust the speed control slider as needed for comfortable viewing. Watch the curves in the **Disease Prevalence** plots. After each run, move the cursor to the end of the green curve and record the approximate percentage of susceptible hosts at the end of each simulation that will show on the vertical axis of the plot. Repeat 20 times.

Exercise 16 *How do the results compare with the predictions of the ODE model? Do you always, or most of the time, see runs that terminate with about 41% of all hosts remaining susceptible? How would you explain the reasons for your observations?*

Let us get a larger and more reliable data set. With the current parameter settings, define and run a batch processing experiment by using the template given in the instructions on how to use the modules at this web site and the following specifications:

Define a **New** experiment.
Repetitions: 200

[4] `http://www.ohio.edu/people/just/IONTW`

Measure runs using these reporters:

`count turtles with [susceptible?]`

`ticks`

Setup commands:

`new-network`

Open the output file and sort the data in the column with the header `count turtles with [susceptible?]` from largest to smallest. Since R_0 is relatively close to 1 and the population size is relatively small, there may not be a large gap between major and minor outbreaks. Look for the topmost gap of at least 5 between consecutive entries in the column that you just sorted from largest to smallest and classify all outbreaks above this gap as minor and all outbreaks below this gap as major. Note that the column reports the numbers $S(\infty)$ of hosts who remained susceptible and did not experience infection.

Exercise 17 *(a) Compute the mean proportion of* $s(\infty) = \frac{S(\infty)}{150}$ *for all simulations and separately for the simulations that you classified as "major outbreaks."*

(b) Which of these means, if any, is close to the value predicted by (19)*? How should the prediction about the final size in the ODE version of the* SIR*-model* (5) *be interpreted in the context of agent-based models?*

(c) For future reference, also compute the overall mean of the column `ticks` *that reports the duration of the outbreaks and the mean of this column for all outbreaks that you classified as "major."*

As we mentioned in Subsection 2.2, (14) predicts the final size in both SIR- and $SEIR$-models with the same parameters α, β, and N. This is consistent with our explorations in Project 2 of the online appendix for [2]. There we compared the relation between SIR- and $SEIR$-models for measles and diphtheria. We found that these model types predict almost identical values for the final sizes, but somewhat longer durations of outbreaks in $SEIR$-models than in corresponding SIR-models. For both measles and diphtheria the values of R_0 are very high though, and it might be possible that this had caused the observed similarities and differences. Here we want to make an analogous comparison for models with a relatively small $R_0 \approx 1.5$.

Change the following parameter settings:

end-latency-rate: 0.25

latent-period: On

This sets up an $SEIR$ model with the same R_0 as before. You may want to verify this by clicking **New**, then **Metrics**, and looking up the value of `R0` in the **Command Center**. In this model $\langle \tau^I \rangle = \frac{1}{\alpha} = 1$ and $\langle \tau^E \rangle = \frac{1}{\gamma} = 4$, so that the mean duration of latency is much longer than the mean duration of infectiousness. Intuitively one might expect significant differences between the predictions of this $SEIR$-model and the previous SIR-model where hosts are assumed to become infectious immediately at the time of exposure.

Set up and run a batch processing experiment with the current parameter settings and the same specifications as in the previous experiment. You can either define a **New** experiment or **Edit** the previous one by making the following changes:

Choose a new name for the output file
`["end-latency-rate" 0.25]`
`["latent-period" true]`

Open the output file and sort the data in the column with the header `count turtles with [susceptible?]` from largest to smallest. Look for the topmost gap of at least 5 between consecutive entries in the column that you just sorted from largest to smallest and classify all outbreaks above this gap as minor and all outbreaks below this gap as major.

Exercise 18 *(a) Compute the mean proportion of* $s(\infty) = \frac{S(\infty)}{150}$ *for all simulations and separately for the simulations that you classified as "major outbreaks."*

(b) Are these means close to the ones you found for the corresponding SIR-model? Does inclusion of an ***E****-compartment in the models appear to alter the predictions of the expected final size?*

(c) Compute the overall mean of the column `ticks` *that reports the duration of the outbreaks and the mean of this column for all outbreaks that you classified as "major." Compare these means with the ones you found for the corresponding SIR-model and explain the observed differences.*

4.4 Exploring SIS-models

Let us consider what happens if we scale up the SIS-model (8) to larger and larger populations. As we explained in Section 3, in general, β may depend on N and should be considered a function of N. Under density-dependent transmission we will have $\beta(N) = \beta = constant$ and under frequency-dependent transmission we will have $\beta(N) = \frac{\beta}{N}$ for some constant β.

Exercise 19 *Consider the equilibria that you found in your solution of Exercise 4. Where would the endemic equilibrium be located relative to overall population size for very large* N *if we assume*

(a) density-dependent transmission,

(b) frequency-dependent transmission?

Now let us explore some agent-based SIS-models. Open IONTW, click **Defaults**, and change the following parameter settings:

model-time: Continuous
infection-rate: 0.01
end-infection-rate: 1
gain-immunity: Off
num-nodes: 200
num/frac: 10

auto-set: On

This sets up an SIS-model with $\alpha = 1$ and $\beta = 0.01$. According to the solution of Exercise 4, the ODE model (7) predicts that all solution curves that start with $I(0) > 0$ will approach the equilibrium $(\frac{\alpha}{\beta}, N - \frac{\alpha}{\beta}) = (100, 100)$.

Let us see how this prediction plays out in the agent-based model. Click **New**, then **Go**. Observe what happens in the **Disease Prevalence** plot. Adjust the speed control slider for comfortable viewing as needed. Record your observations. Take some time to watch the movie. You may need to terminate the simulation by pressing **Go** again.

Now change the parameter settings as follows:

infection-rate: 0.02
num-nodes: 100

For this model the ODE system (7) predicts that all solution curves that start with $I(0) > 0$ will approach the equilibrium $(\frac{\alpha}{\beta}, N - \frac{\alpha}{\beta}) = (50, 50)$. Repeat the previous exploration and record any similarities and differences that you might observe.

Then repeat first with

infection-rate: 0.04
num-nodes: 50

and finally with

infection-rate: 0.1
num-nodes: 20

Exercise 20 *How do your observations differ from the predictions of the corresponding ODE models? In what sense are they similar? What happens if you decrease the population size? How can you explain these observations?*

References

[1] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[2] Winfried Just, Hannah Callender, and M Drew LaMar. Disease transmission dynamics on networks: Network structure *vs.* disease dynamics. In Raina Robeva, editor, *Algebraic and Discrete Mathematical Methods for Modern Biology*, pages 217–235. Academic Press, 2015.

[3] W. O. Kermack and A. G. McKendrick. A contributions to the mathematical theory of epidemics. *Proceedings of the Royal society of London. Series A*, 115:700–721, 1927.

[4] Hamish McCallum, Nigel Barlow, and Jim Hone. How should pathogen transmission be modelled? *Trends in Ecology & Evolution*, 16:295–17300, 2001.