

# The replacement number\*

Winfried Just<sup>†</sup> Hannah Callender<sup>‡</sup> M. Drew LaMar<sup>§</sup> Ying Xin<sup>¶</sup>

December 23, 2015

In this module we introduce the important notion of the *replacement number*, which generalizes the basic reproductive number  $R_0$ . We investigate how this number behaves near the start of an outbreak in two types of models: The first type is based on the uniform mixing assumption and the second type assumes a contact network that is a random  $k$ -regular graph with small  $k$ . We also illustrate a method for estimating the value of  $R_0$  from epidemiological data.

## 1 The replacement number in models based on the uniform mixing assumption

### 1.1 Definition of the replacement number and its basic properties in compartment-level models

Did it strike you as a bit peculiar that the notation  $R_0$  for the most important parameter in disease modeling involves a subscript 0? The subscript 0 in this notation does double duty of reminding us what  $R_0$  stands for: the mean number of secondary infections that are caused by a single index case that is introduced at time  $t = 0$  into a population with 0 other infectious, exposed, or removed hosts.

As the outbreak develops, more hosts may become infectious, while other hosts may be removed. Thus at later times  $t$ , the mean number of secondary infections that is caused by hosts who are infectious at time  $t$  will in general be *different* from  $R_0$ . In the literature it is sometimes denoted by  $R$  and called the *replacement number* (see, for example, [1]). In contrast to  $R_0$  that is a fixed constant, the replacement number changes over time and depends on the current state  $st$  of the system. It is somewhat unfortunate that the letter  $R$  is also suggestive of the **R**-compartment. We already used the notation  $R(t)$  for the number of removed hosts at time  $t$  in ODE-models. In order to avoid possible confusion, we will use the notation  $R_t^{st}$  for the value of the replacement number at time  $t$  when the system is

---

\*©Winfried Just, Hannah Callender, M. Drew LaMar, Ying Xin 2015

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

<sup>‡</sup>University of Portland E-mail: callende@up.edu

<sup>§</sup>The College of William and Mary E-mail: drew.lamar@gmail.com

<sup>¶</sup>Department of Mathematics, Ohio University, Athens, OH 45701 E-mail: yx123812@ohio.edu

in state  $st$ . Whenever we use this notation, we will implicitly assume that there is at least one infectious host in state  $st$ . More precisely, we will use the following definition:

**Definition 1** *The replacement number  $R_t^{st}$  is the mean number of secondary infections caused by a host  $i$  who is infectious at time  $t$  when the system is in state  $st$  over the time interval  $[t, T_i^R)$ , where  $T_i^R$  denotes the time of removal of host  $i$  after  $t$ .*

These definitions of  $R_t^{st}$ ,  $R_t$ , and in particular  $R_0$ , implicitly assume that infectious hosts will eventually become removed. This will not be the case, for example, in  $SI$ -models without demographics, where  $T_i^R$  in effect is assumed to be infinite for each host. For these models one can either consider  $R_t^{st}$  to be undefined or  $R_t^{st} = \infty$ . Moreover, in models where removal does not confer permanent immunity, such as  $SIS$ -models, it is important that  $T_i^R$  in Definition 1 is interpreted as the time when the host will recover from *this particular bout of infection*, but not during subsequent bouts that may result from re-infection after recovery. We will sidestep these complications here by focusing only on  $SIR$ -models in the remainder of this module, although some of the results carry over to other types of models.

In the models that we consider here, the mean length of the interval during which a host  $j$ , who is infectious at time  $t$ , will remain infectious after time  $t$  does not depend on how long this host already has been infectious. While somewhat unrealistic, the assumption greatly simplifies modeling and is embodied in the code of IONTW; see Section 8.6 of the online appendix to [2] for a discussion. This assumption is also implicit in the classical ODE-based compartment-level models that were discussed in our module *Differential equation models of disease transmission* at this web site<sup>1</sup>. Thus in the models that we are considering here the number  $R_t^{st}$  depends *only* on the state of the system at time  $t$ . The subscript  $t$  is, strictly speaking, redundant. But it will be useful when we want to study how the replacement number changes over time. In deterministic models such as ODE models, the state at time  $t$  is entirely determined by the initial state of the system. Under the assumptions of homogeneity of hosts and uniform mixing, in such models  $R_t^{st}$  depends only on  $t$  and the numbers of hosts that were initially infectious, susceptible, removed, or exposed. In this case we will use the symbol  $R_t$  for the replacement number at time  $t$  under the assumption that the outbreak started at time  $t = 0$  with 1 infectious host in an otherwise susceptible population. For time  $t = 0$  we then get exactly the quantity that was previously defined as  $R_0$ .

Consider an outbreak that is started by a single index case in a population of size  $N$ . Under the assumptions of uniform mixing and homogeneity of hosts, we will have

$$\begin{aligned} R_t^{st} &\leq R_0 \text{ at all times } t \geq 0 \text{ and for all states } st, \\ \forall t \forall \varepsilon > 0 \quad \lim_{N \rightarrow \infty} P(R_t^{st} > R_0 - \varepsilon) &= 1. \end{aligned} \tag{1}$$

Note that unless we are studying a deterministic model, in the second line of (1) we can only make an estimate of a probability as the state  $st$  at time  $t$  will usually not be known with certainty.

---

<sup>1</sup><http://www.ohio.edu/people/just/IONTW>

To see why (1) will be true, consider a host  $j$  who is infectious at time  $t$ . If  $j$  is the only infectious host in state  $st$  and all other hosts are susceptible, then  $R_t^{st}$  is the mean number of secondary infections that will be caused by this host and must be equal to  $R_0$  by the definition of  $R_0$ . Host  $j$  can infect other hosts only by making effective contact with them. Under the assumptions of uniform mixing and homogeneity of hosts, it does not matter which host has number  $j$ . Thus the mean number of other hosts with whom  $j$  will make effective contact during  $j$ 's interval of infectiousness is always the same. But if in state  $st$  there are other hosts who are no longer susceptible, then a larger proportion of these effective contacts will not be successful in terms of causing secondary infections. Thus unless  $j$  is the only infectious in state  $st$  while all other hosts are still susceptible, the strict inequality  $R_t^{st} < R_0$  will hold. This argument proves the first line of (1).

The second line of (1) essentially says that for any fixed  $t$  and sufficiently large population sizes  $N$ , with very high probability the replacement number  $R_t^{st}$  will remain very close to  $R_0$ . To see why this is true, consider a very large population and a time  $t$  near the start of an outbreak. Then already some other hosts may no longer be susceptible in state  $st$  time  $t$ , but, with probability very close to 1, their proportion will be very small. Thus *nearly all* effective contacts of an infectious host  $j$  will be with susceptible hosts, which is very similar to the situation for the index case at time 0. Under these additional assumptions,  $R_t^{st}$  should be only insignificantly smaller than  $R_0$ .

We want to emphasize that our little proof of (1) works only under suitable assumptions. We explicitly mentioned the uniform mixing assumption and homogeneity of hosts. We will see in Section 2 that the uniform mixing assumption is needed for this result. Homogeneity of hosts is also needed, as it implies that the mean duration of infectiousness  $\langle \tau_j^I \rangle$  is the same for all hosts  $j$ . Our argument also relies on some hidden assumptions that are usually made but not always clearly spelled out in the literature. One is that for any given pair of hosts  $\{i, j\}$  the transmission probabilities  $b_{i,j}$  or transmission rates  $\beta_{i,j}$  do not change over time<sup>2</sup>. A second one is that for different pairs of hosts  $\{i, j\} \neq \{i', j'\}$  transmission of pathogens from host  $j$  to host  $i$  and from host  $j'$  to host  $i'$  are independent events.

## 1.2 Initial nearly exponential growth in next-generation *SIR*-models

Let us consider an agent-based next-generation *SIR* model with a very large population size  $N$ . Assume that a given outbreak is started by a randomly chosen index case in an otherwise susceptible population. Moreover, let us assume that (1) holds in this model. More specifically, let us assume that for some  $T > 0$  and small  $\varepsilon, q > 0$  we will have  $P(R_t^{st} > R_0 - \varepsilon) \geq 1 - q$  for all times  $t$  with  $0 \leq t < T$ .

In this type of model, time is incremented in integer steps that are scaled so that  $\langle \tau^I \rangle = 1$ , and each infectious host gets removed at the next time step. What can we say about the numbers<sup>3</sup>  $|\mathbf{I}(t)|$  of infectious hosts at times  $t = 1, 2, \dots, T$  in this model? Since our agent-based models are stochastic, these numbers are random variables (r.v.s) and we

---

<sup>2</sup>See *Network-based models of transmission of infectious diseases: a brief overview* at this web site <http://www.ohio.edu/people/just/IONTW/>.

<sup>3</sup>Recall that we are using the notation  $|A|$  for the size of a set  $A$ .

cannot predict them with certainty. But we can predict their expected values  $\langle |I(t)| \rangle$ , which will for convenience be denoted here by  $I(t)$ .

By definition of  $R_0$  and since the index case will have been removed by time step 1, we must have  $I(1) = R_0$ .

We can have no certainty about the state of the system at time  $t = 1$ , but our assumption implies that with probability  $\geq 1 - q$  we will have  $R_1^{st} > R_0 - \varepsilon$ . In this case, each host who is infectious at time step 1 will cause on average  $> R_0 - \varepsilon$  secondary infections. It follows that

$$I(2) \geq (1 - q)I(1)(R_0 - \varepsilon) = (1 - q)R_0(R_0 - \varepsilon). \quad (2)$$

A similar argument applies at subsequent time steps. Since the probability that at least one of  $t - 1$  events that have probability  $\leq q$  each occurs is bounded from above by  $(t - 1)q$ , regardless of independence, this gives by induction:

$$I(t) \geq (1 - (t - 1)q)R_0(R_0 - \varepsilon)^{t-1}. \quad (3)$$

A similar argument based on the first line of (1) gives the upper bound

$$I(t) \leq R_0^t. \quad (4)$$

If  $p$  is any given probability with  $p < 1$  and  $0 < R_0^- < R_0$ , then for all sufficiently large  $N$ , we will be able to choose  $q, \varepsilon$  small enough so that  $(1 - (t - 1)q) > p$  and  $R_0 - \varepsilon > R_0^-$ . For these choices, (3) together with (4) predict that

$$\forall t \leq T \quad (R_0^-)^t \leq I(t) \leq R_0^t. \quad (5)$$

If  $R_0^-$  is chosen very close to  $R_0$  and if  $R_0 > 1$ , (5) predicts *nearly exponential growth* (with base  $R_0$ ) of the expected number  $I(t)$  of infectious hosts for times  $t \leq T$ , that is, in the initial stages of an outbreak.

In next-generation models hosts get removed after exactly one time step. Thus in these models,  $|I(t)|$  gives an estimate both of the expected *incidence* (the rate at which new infections occur) and the expected *prevalence* (the total number of infectious hosts) at a given time. Other types of models allow us to make a distinction between these two notions. We will examine this distinction for ODE models in the next subsection.

**Exercise 1** *Determine whether a next-generation SIR-model for which (1) holds predicts initial near-exponential growth with base  $R_0$  for the expected number  $R(t)$  of removed hosts at time  $t$ .*

In the literature, nearly exponential growth of the incidence and the prevalence in the initial stages of an outbreak is sometimes considered the defining property of an *epidemic*. This definition is not without problems. As we will see in later modules, not for all types of contact networks should we expect initial nearly exponential growth for major outbreaks.

### 1.3 Initial nearly exponential growth in ODE-based *SIR*-models

This subsection assumes some familiarity with ODE models of disease transmission as covered in our module *Differential equation models of disease transmission* at this web site<sup>4</sup>. Readers who want to focus exclusively on agent-based network models can skip it.

Consider the ODE version of an *SIR*-model with population size  $N$ :

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

This model is deterministic, so that the state  $st$  at time  $t$  will be uniquely determined by the initial condition, that is, the state at time 0. If we assume that at time  $t = 0$  there is exactly one infectious host while all other hosts are susceptible, we can use the simplified notation  $R_t$  instead of  $R_t^{st}$  here. In particular, the probability  $P(R_t^{st} > R_0 - \varepsilon)$  in the second line of (1) will be either 0 or 1 depending on  $t$ , and this line simply translates into the assertion that

$$R_t > R_0 - \varepsilon \text{ for all sufficiently small } t.\tag{7}$$

Of course, (7) is vague about the meaning of “sufficiently small  $t$ .” Let us consider here all times  $t$  at which a proportion of at least  $1 - \varepsilon$  of all hosts is still susceptible. It depends on the parameters  $\alpha$  and  $\beta$  how small  $t$  actually needs to be in terms of physical time units. But we can see that all times “sufficiently close to 0” would qualify, as long as  $N$  is large enough so that 1 initially infectious host corresponds to a proportion of less than  $\varepsilon$  of the population. This follows from the fact that the solutions of ODEs are continuous functions. Readers who are interested in more precise estimates may want to do the following exercise:

**Exercise 2** Suppose  $I(0) = 1$ ,  $S(0) = N - 1$ , and  $\frac{\beta N}{\alpha} > \frac{1}{1-\varepsilon}$ . Show that if  $S(0) = N - 1 > (1-\varepsilon)N$ , then  $S(t) \geq (1-\varepsilon)N$  for all  $t \leq T$ , where  $T = \frac{1}{\beta(N-1)-\alpha} \ln \left( \frac{(\varepsilon N - 1)[\beta(N-1) - \alpha]}{\beta(N-1)} + 1 \right)$ .

In ODE-models,  $I(t)$  is the prevalence of the infection, and the incidence is the rate at which new infections occur. In (6), the latter is the term  $\beta IS$ . Now notice that for all times  $t$  as above we have:

$$(\beta N - \alpha)I(t) \geq \frac{dI}{dt} = \beta I(t)S(t) - \alpha I(t) \geq ((1 - \varepsilon)\beta N - \alpha)I(t).\tag{8}$$

Recall that  $R_0 = \frac{\beta N}{\alpha}$  for model (6). Thus as long as  $R_0 > \frac{1}{1-\varepsilon} > 1$  and  $I(t) > 0$ , both estimates of  $\frac{dI}{dt}$  in (8) are positive.

---

<sup>4</sup><http://www.ohio.edu/people/just/IONTW>

Now consider the linear ODEs

$$\frac{dI}{dt} = (\beta N - \alpha)I \tag{9}$$

and

$$\frac{dI}{dt} = ((1 - \varepsilon)\beta N - \alpha)I. \tag{10}$$

Their solutions are the exponential functions  $I(t) = I(0)e^{(\beta N - \alpha)t}$  and  $I(t) = I(0)e^{((1 - \varepsilon)\beta N - \alpha)t}$ . It follows from (8) that the solution  $I(t)$  of (6) and the corresponding incidence function  $\beta I(t)S(t)$  satisfy the following inequalities for all  $t$  as above:

$$\begin{aligned} I(0)e^{(\beta N - \alpha)t} &\geq I(t) \geq I(0)e^{((1 - \varepsilon)\beta N - \alpha)t}, \\ \beta NI(0)e^{(\beta N - \alpha)t} &\geq \beta I(t)S(t) \geq (1 - \varepsilon)\beta NI(0)e^{((1 - \varepsilon)\beta N - \alpha)t}. \end{aligned} \tag{11}$$

Thus both the prevalence function  $I(t)$  and the incidence function  $\beta I(t)S(t)$  are closely sandwiched between two exponential functions for all sufficiently small  $t$ . We can see that when  $R_0 > 1$ , both of these functions initially grow nearly exponentially with base  $e^{\beta N - \alpha}$ .

Recall that when time is scaled so that the mean duration of infectiousness  $\langle \tau^I \rangle = 1$ , then  $\alpha = 1$  and  $\beta N \approx R_0$ . Thus in this case the ODE version of the SIR model predicts initial nearly exponential growth with base  $e^{R_0 - 1}$ . For  $R_0 > 1$  this value will be larger than  $R_0$  (due to a phenomenon akin to compounding interest), but when  $R_0$  is only slightly larger than 1, then  $e^{R_0 - 1} \approx R_0$ . For example, if  $R_0 = 1.2$ , then  $e^{R_0 - 1} = 1.2214$ .

#### 1.4 Estimating $R_0$ from data

The results of this section give us a possible method of estimating  $R_0$  from data on the prevalence or incidence function in the initial states of an outbreak. Suppose we observe an outbreak of an immunizing infection with  $\langle \tau^I \rangle = 1$  week, and Table 1 gives the numbers of reported new infections for the first 10 weeks.

<b>Week:</b>	0	1	2	3	4	5	6	7	8	9
<b>New cases:</b>	1	3	6	12	17	26	38	59	57	46

Table 1: Number of reported new cases per week.

Clearly, the incidence did not grow nearly exponentially during the whole time period. It decreases after week 8, and the growth seems to markedly slow down after week 5. But when we focus on the data for the first 5 weeks, we may see a nearly exponential increase, and the base of the exponential function might be a good estimate of  $R_0$ .

**Exercise 3** Try to estimate  $R_0$  based on the given data for the first 5 weeks.

(a) Assume that a next-generation model would adequately predict the initial stages of the outbreak for which Table 1 gives data.

(b) How would the result change if we were to base the method on an ODE version of the SIR-model with time scaled to  $\langle \tau^I \rangle = 1$  instead?

The method we suggested for Exercise 3 works well for giving ball-park figures, but it has its problems. First of all, it can work only if the number of reported cases gives a fairly reliable estimate of the actual incidence function. For real outbreaks in human populations, this may not be a realistic assumption; see [2] for a discussion. Moreover, the method is based on a particular model; as we have seen, it may give slightly different values for continuous-time and next-generation models. For either type of model, the method assumes that the bounds on the replacement number that we derived in this section are valid in the initial stages of an outbreak. These were derived under the uniform mixing assumption. The next section will give you a glimpse into how the situation may change in models that assume spread of the infection on a given contact network.

## 2 The replacement number in network-based models

The results of Section 1 crucially depended on (1), which was derived under the uniform mixing assumption. In network-based models, the uniform mixing assumption usually fails, and we might conjecture that (1) may fail as well. We will show here that this can indeed happen for the second line of (1); in later modules we will show that the first line of (1) may also fail in some network-based models.

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

**time-step:** 1  
**infection-prob:** 0.75  
**end-infection-prob:** 1  
**network-type** → **Random Regular**  
**num-nodes:** 15  
**lambda:** 2  
**auto-set:** **On**

Click **New** to create an instance of  $G_{Reg}(15, 2)$  with one index case. Click **Metrics** and verify that  $R_0 = 1.5$  in this model.

Press **Labels** to see the numbers associated with each node, and record the number assigned to the initially infectious node. Now set the speed slider to a very slow setting; adjust for comfortable viewing as needed. Record the numbers of the nodes that are infectious after each **tick**; that is after each time step of the simulation, and try to make out who infected whom.

In our run, node 8 was initially infectious and infected nodes 0 and 6; we recorded this as

8 → 0, 6.

For the next time steps of the infection we observed

$$\begin{aligned} 0 &\rightarrow 10 & 6 &\rightarrow 13 \\ 10 &\rightarrow 4 \\ 4 &\rightarrow 13 \\ 13 &\rightarrow 11. \end{aligned} \tag{12}$$

This was the end of the outbreak.

Record such lists for about 5 outbreaks. Now record, for each time step at which any new infections occurred, the mean number of secondary infections caused by each of the currently infectious nodes. For our list we would get:

2 at step 1 (as node 8 infected two nodes)  
1 at step 2 (as each of the two infectious nodes caused 1 secondary infection)  
0.5 at step 3 (as only one of the two infectious nodes caused a secondary infection)  
1 at step 4  
1 at step 5  
0 at step 6

This is interesting. In this model we have  $R_0 = 1.5$ , which means that by step 1 on average 1.5 secondary infections will have been caused by the index case. We saw 2, which is as close as it gets to  $R_0$ . The means that we calculated for steps  $t = 1, 2, \dots, 6$  give very crude estimates of  $R_t^{st}$ , averaged over all possible states  $st$  at time  $t$ . According to (1), for small  $t$  we might expect averages close to 1.5, but for  $t > 0$  we *never* observed a value above 1, and we bet you didn't either. Why?

**Exercise 4** (a) Show that in a next-generation model with a 2-regular graph a host who is infectious at any time  $t > 0$  can cause at most 1 secondary infection.

(b) Let  $k$  be any positive integer and consider a next generation network-based SIR-model with probability  $b$  of effective contact over a unit time interval. Find  $R_0$ . Then generalize part (a) to derive an upper bound  $R_{ub}$  for  $R_t^{st}$  that applies to all  $t > 0$  if the contact network is  $k$ -regular.

(c) Test your prediction of point (b) with running some simulations in slow motion for  $b = 0.5$  and networks  $G_{Reg}(16, 3)$ .

If you solved part (b) of Exercise 4 correctly, then you will have found that for all  $t > 0$  we must have

- $R_{ub} \leq 0.75$  if  $b = 0.75$ , and the contact network is 2-regular,
- $R_{ub} \leq 1$  if  $b = 0.5$ , and the contact network is 3-regular.

Since  $R_0 = 1.5$  in both cases, we can see that the second line of (1) fails for 2-regular and 3-regular graphs, even when the number of nodes  $N$  is very large.

Now recall our derivation of the upper bound (4) for  $I(t)$ . The same argument applies here, and in view of your result for Exercise 4(b), for next-generation *SIR*-models on large  $k$ -regular graphs we can estimate

$$I(t) \leq R_0 R_{ub}^{t-1}. \quad (13)$$

The alert reader will have noticed that the complete graphs that we use as contact networks for studying the uniform mixing assumption are  $k$ -regular graphs with  $k = N - 1$ . Thus the upper bound (13) does apply to them. However, if  $N$  is very large, the difference between  $R_0$  and  $R_{ub}$  as calculated in your solution of Exercise 4(b) becomes negligible so that (13) does not invalidate the results that we derived in Subsection 1.2.

However, if  $k$  is as small as 2 or 3, the difference between  $R_0$  and  $R_{ub}$  may drastically alter the predictions of the model. For example, when  $b = 0.75$  and  $k = 2$ , then  $I(t) \leq 1.5(0.75)^{t-1}$ , so that  $I(t)$  quickly approaches 0 even when the network is very large. This explains what we saw in the explorations of Section 3 of our module *Exploring random regular graphs with IONTW* at this web site<sup>5</sup>! In our simulations for random 2-regular contact networks with these parameter settings, all observed outbreaks were minor, despite the fact that  $R_0 > 1$ . The results for parameter settings  $b = 0.5$  and  $k = 3$  were less clear-cut. As  $R_{ub} = 1$  in this case, we might expect that these outcomes would be similar to what one would see in compartment-level models with  $R_0 = 1$ . Mathematical theory predicts that all outbreaks will be minor for compartment-level models with  $R_0 = 1$ , but for moderately large population sizes this can still translate into a significant fraction of hosts who will experience infection.

**Exercise 5** Prove that the following version of (1) holds for next-generation *SIR*-models of disease transmission on contact networks of type  $G_{Reg}(N, k)$  for the value of  $R_{ub}$  that you found in Exercise 4(b):

$$\begin{aligned} R_t^{st} &\leq R_{ub} \text{ at all times } t > 0 \text{ and for all states } st, \\ \forall t \forall \varepsilon > 0 \quad \lim_{N \rightarrow \infty} P(R_t^{st} > R_{ub} - \varepsilon) &= 1. \end{aligned} \quad (14)$$

Now the same argument that we used to derive (3) under the assumption of uniform mixing gives the following analogue for next-generation *SIR*-models when the contact network is of type  $G_{Reg}(N, k)$ , both  $k$  and  $t$  are fixed, and the population size  $N$  is sufficiently large:

$$I(t) \geq (1 - (t - 1)q)R_0(R_{ub} - \varepsilon)^{t-1}. \quad (15)$$

As in (3), the value  $q$  can be chosen arbitrarily close to 0. When  $R_{ub} > 1$ , the bounds (13) and (15) on the mean number of infectious hosts  $I(t)$  in the initial stages of an outbreak in

---

<sup>5</sup><http://www.ohio.edu/people/just/IONTW>

a sufficiently large population predict initially nearly exponential growth of the number of infectious hosts, roughly like  $R_0 R_{ub}^{t-1}$ . When  $R_{ub} > 1$ , this will still give nearly exponential initial growth, but the base of the exponential function will no longer be  $\approx R_0$ , but  $\approx R_{ub} < R_0$ .

Let us try to confirm these predictions by running experiments. Change the current parameter settings in IONTW to:

**infection-prob:** 0.8  
**num-nodes:** 200  
**lambda:** 3

Set up and run a batch processing experiment for the current parameter settings with 100 repetitions by using our template with the following specifications:

**Measure runs using these reporters:**

count turtles with [infectious?]

**Setup commands:**

new-network

**Time limit:** 2

The output column will report the actual numbers of infectious hosts at time step 2.

Then repeat the experiment with

**Time limit:** 3

and run a third experiment with

**Time limit:** 4

Be sure to choose different names for each experiment so as not to overwrite the data in your output files.

**Exercise 6** *Compute mean values for the data in the last columns of your three output files. Do these outcomes give a good match for the prediction that  $I(t) \approx R_0 R_{ub}^{t-1}$  for  $t = 2, 3, 4$ ? If there seem to be significant discrepancies, how could you explain them?*

One might expect that the derivation of  $R_{ub}$  in Exercise 4(b) and the resulting estimates of  $I(t)$  for the initial stages of an outbreak will carry over to other types of large networks with small mean degrees, such as Erdős-Rényi networks  $G_{ER}(200, 3)$ . To verify this prediction, **Edit** your previous batch processing experiment by changing one entry in the input field **Vary variables as follows** from

["network-type" "Random Regular"]

to

["network-type" "Erdos-Renyi"]

and run first with

**Time limit:** 2

and then with

**Time limit:** 3

**Exercise 7** Compute mean values for the data in the last columns of your two output files. Do these outcomes give a good match for the prediction that  $I(t) \approx R_0 R_{ub}^{t-1}$  for  $t = 2, 3$  or are they closer to the prediction  $I(t) \approx R_0^t$  that we had derived under the uniform mixing assumption?

Recall that in Section 3 of our module *Exploring random regular graphs with IONTW* we discovered significant differences between simulations of outbreaks in models with contact networks  $G_{ER}(200, 3)$  and  $G_{Reg}(200, 3)$  with identical disease transmission parameters. The results of Exercise 7 go some way towards explaining these observations: It seems that initial growth in the former models will be nearly exponential with base  $R_0$ , while initial growth in the latter models will be nearly exponential with base  $R_{ub} < R_0$ . We can now understand why the base should be  $R_{ub}$  when the contact network is a random regular graph  $G_{Reg}(200, 3)$ .

But why should the base still be  $R_0$  for Erdős-Rényi contact networks  $G_{ER}(200, 3)$ ? After all, in both types of contact networks, nodes have small degrees and each host who becomes infectious at any time  $t > 0$  will already have at least one neighbor who is no longer susceptible.

Something else must be going on here. But what? To find out, work through our module *The friendship paradox* at this web site<sup>6</sup>.

## References

- [1] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.
- [2] Winfried Just, Hannah Callender, M Drew LaMar, and Natalia Toporikova. Transmission of infectious diseases: Data, models, and simulations. In Raina Robeva, editor, *Algebraic and Discrete Mathematical Methods for Modern Biology*, pages 193–215. Academic Press, 2015.

---

<sup>6</sup><http://www.ohio.edu/people/just/IONTW>