

Solutions to Exercises

Winfried Just¹, Hannah Callender² and M. Drew LaMar³

¹Department of Mathematics, Ohio University, Athens, OH, USA, ²University of Portland, Portland, OR, USA,

³The College of William and Mary, Williamsburg, VA, USA

Exercise 9.1. For what kind of populations would the uniform mixing assumption be nearly satisfied? When would it not be realistic?

Sample Solution: Mixing may be nearly uniform if hosts move around a lot relative to the size of their habitat, encounter each other rarely, and there is no social structure. \square

Exercise 9.2. What could we say about the state ($|\mathbf{S}(t^*)|$, $|\mathbf{I}(t^*)|$, $|\mathbf{R}(t^*)|$) in a continuous-time model? Would it still always be true that $|\mathbf{I}(t^*)|$ is equal to the number of secondary infections caused by the index case?

Sample Solution: In a continuous-time model, some host i that got infected by the index case could infect another host i_1 before host j^* gets removed, or host i could recover earlier than the index case so that $T_i^R < t^*$. Thus $|\mathbf{I}(t^*)|$ might be larger or smaller than the number of secondary infections caused by the index case. The only thing about the state at time t^* that we can say with certainty is that $|\mathbf{R}(t^*)| \geq 1$. \square

Exercise 9.3.

- (a) Assume a next-generation *SIR*-model and an initial state that corresponds to the introduction of one index case into an otherwise susceptible population. Find $(S(1), I(1), R(1))$.
- (b) How would the result change if we were considering an *SIS*-model instead?

Sample Solution:

- (a) You have already seen that $|\mathbf{R}(1)| = 1$, which implies that the mean value $R(1)$ must also be equal to 1. Because $|\mathbf{S}(1)|$ is the actual number of secondary cases caused by the index case, we must have $I(1) = R_0$ by definition. Because we assumed constant population size, it follows that $S(1) = N - R_0 - 1$.
- (b) In an *SIS*-model, we don't have an **R**-compartment. We could interpret this as $R(t) = 0$ at all times t . We still have $I(1) = R_0$, but upon recovery, the index case moves back into the **S**-compartment. It follows that $S(1) = N - R_0$. \square

Exercise 9.4. Explain why the conclusion of the theorem will not in general hold for *SI*- or *SEIR*-models.

Sample Solution: In an *SI*-model, no infectious host ever recovers, and we will always have $I(t^+) \geq I(t)$.

In an *SEIR*-model, there may be a large number of exposed hosts at time t in addition to a few infectious ones. While hosts who are already infectious at time t would be expected to have caused few secondary infections by time t^+ , we would also expect that a large number of hosts will have moved from the **E**-compartment to the **I**-compartment by time t^+ . This may give a possible net increase of I between times t and t^+ . \square

Exercise 9.5.

- (a) Use IONTW to simulate a compartment-level next-generation *SIR*-model with **infection-prob** = 0.09, population size $N = 10$, and one index case. Find R_0 . Run 8 repetitions in slow motion. For detailed instructions, see Module 9.1 of [1].
- (b) Describe your results in terms of $|\mathbf{I}(t)|$ for $t = 0, 1, 2, \dots$. Do you always observe that $|\mathbf{I}(t+1)| < |\mathbf{I}(t)|$? Do your results confirm Theorem 8.9.1, or do they appear to conflict with it? How would you explain the relation of your observations to the theorem?

Sample Solution: For the suggested parameter settings, $R_0 = 0.81 < 1$.

Most likely, in some simulations you will see that $|\mathbf{I}(t)| = 0$ for all $t = 1, 2, \dots$. In other simulations you may see something like $|\mathbf{I}(0)| = |\mathbf{I}(1)| = |\mathbf{I}(2)| = 1$ and $|\mathbf{I}(3)| = 0$. In yet other simulations you may see $|\mathbf{I}(0)| = 1 < |\mathbf{I}(1)|$, so that I occasionally *increases*.

At first glance, these outcomes appear to contradict Theorem 9.1. But the theorem tells us only something about the *expected values* $I(t)$ of $|\mathbf{I}(t)|$, while in simulations you observe *actual values* of the r.v.s $|\mathbf{I}(t)|$. So there is no conflict with the theorem.

However, the simulations do not directly confirm the theorem. Empirical verification of whether or not actual data conform to the predictions of the theorem would require running a much larger batch of simulations and computing averages of the observed outcomes. \square

Exercise 9.6. Use IONTW to simulate a compartment-level next-generation *SIR*-model with **infection-prob** = 0.005, population size $N = 100$, and one index case. Run 100 repetitions. Find the largest and the mean number of hosts who experienced infection in the 100 simulated outbreaks. For detailed instructions, see Module 9.1 of [1].

Sample Solution: We found $R_0 = 0.495$.

The number of hosts who had experienced infection during a given run was shown in the output file under the column header `count turtles with [removed?]`.

Out of our 100 repetitions, the simulation with the largest number of hosts who had experienced infection was 19. The mean number for all 100 simulations was 2.38. \square

Exercise 9.7. Use IONTW to examine the dependence of the final size on N when $R_0 < 1$ is kept constant. For detailed instructions, see Module 9.1 of [1]. How do the observed mean and maximum number and the mean and maximum final size appear to depend on the population size N ?

Sample Solution: Following the instructions in Module 9.1, we obtained the following results, shown in Table S9.1.

TABLE S9.1 Results from Exercise 9.7

N	R_0	Max #	Avg #	Max Final Size	Avg Final Size
100	0.495	19	2.38	0.19	0.0238
200	0.4975	15	2.06	0.075	0.0103
400	0.49875	8	1.9	0.02	0.00475

As N increases, we observe a decrease in the maximum and average numbers and final sizes. In particular, as N doubles, we see a decrease in the average final size by a factor of 2. \square

Exercise 9.8. Use IONTW to simulate *SIR*-compartment models with $R_0 > 1$ for population sizes of $N = 100, 200$ and analyze your results. For detailed instructions, see Module 9.1 of [1].

Sample Solution: Following the instructions given in online Module 9.1, we obtained the results shown in Tables S9.2 and S9.3.

TABLE S9.2 Results from Exercise 9.8 for Minor Outbreaks

N	R_0	Minor Outbreak Prop.	Max #	Avg #	Max Final Size	Avg Final Size
100	1.98	0.19	7	2.05	0.07	0.021
200	1.99	0.12	5	1.75	0.025	0.009
100	1.485	0.36	18	4.11	0.18	0.041
200	1.493	0.23	12	2.36	0.06	0.012

TABLE S9.3 Results from Exercise 9.8 for Major Outbreaks

N	R_0	Max #	Avg #	Max Final Size	Avg Final Size
100	1.98	91	80.99	0.91	0.810
200	1.99	180	160.42	0.90	0.802
100	1.485	83	60.56	0.83	0.606
200	1.493	153	108.98	0.77	0.545

One immediate observation is that the proportion of minor outbreaks is significantly affected by R_0 . The larger R_0 , the fewer minor outbreaks, as we would expect. For $R_0 > 1$ but close to 1, we will observe larger proportions of minor outbreaks, and the cutoff to distinguish minor from major outbreaks becomes less clear. For minor outbreaks, as N increases from 100 to 200, there is a slight but probably insignificant decrease in the maximum number of infections, but there is a noticeable decrease in the maximum final size and mean final size. This difference becomes more significant for smaller values of R_0 .

For major outbreaks, for the larger R_0 value, as N increases, the maximum and average final sizes are almost identical, as we would expect. So the larger the R_0 , the smaller N 's effect on the final size. For major outbreaks in general, we see that R_0 is the controlling factor for both maximum and average final size. \square

Exercise 9.9. Based on the above reading of Theorem 9.3, try to estimate z_∞ and $r(\infty)$ for $R_0 = 1.5, 2$ based on the data you collected in Exercise 9.8. Do your findings appear to confirm the predictions of Theorem 9.3?

Sample Solution: For R_0 near 2, we could estimate z_∞ to be around 0.19 for $N = 100$ and 0.12 for $N = 200$. For R_0 near 1.5, we could estimate z_∞ to be around 0.36 for $N = 100$ and near 0.23 for $N = 200$. Again, it decreases as N increases. Also, as R_0 increases, we see a decrease of z_∞ , as the theorem predicts. We see that R_0 affects z_∞ more significantly than N affects z_∞ , at least for these choices of R_0 and N . While we observed a slight decrease in the estimated values of z_∞ as N increases, this observation is due to finite size effects. Had we been able to compare, for example, population sizes $N = 10,000$ and $N = 20,000$ for the same values of R_0 , we would most likely not see a significant difference in the observed values of z_∞ .

For R_0 near 2, we could estimate $r(\infty)$ to be around 0.81 for $N = 100$ and 0.80 for $N = 200$. We see no notable difference for different values of N . For R_0 near 1.5, we could estimate $r(\infty)$ to be around 0.61 for $N = 100$ and near 0.54 for $N = 200$. Again, it decreases as N increases, but by very little. This is also due to finite size effects. More important, as R_0 increases, we see an increase in $r(\infty)$, which is in agreement with the theorem. We see that R_0 affects $r(\infty)$ more significantly than N affects $r(\infty)$, at least for these choices of R_0 and N . \square

Exercise 9.10. Can you think of some examples where the assumption of symmetric transmission probabilities of pathogens would not be justified?

Sample Solution: In some STIs such as HIV, male-to-female transmission during a given contact is more likely than female-to-male transmission [2].

If host i has a weaker immune system than host j , more infectious agents will need to be transferred from i to j than from j to i to make the contact effective. This translates into different transmission probabilities, depending on who is infectious. The latter asymmetry is actually rather common, but we cannot study its effect if we assume homogeneity of hosts. \square

Exercise 9.11. Find the sets $V(G_1)$ and $E(G_1)$.

Sample Solution: $V(G_1) = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E(G_1) = \{\{1, 2\}, \{1, 5\}, \{1, 7\}, \{2, 5\}, \{4, 5\}, \{6, 8\}\}$. \square

Exercise 9.12. Mathematicians have their own version of the Kevin Bacon game, where the network is the graph of scientific collaborations. Two nodes (mathematicians) are connected by an edge if they are coauthors of a joint scientific publication. The *Erdős number* is the distance of a node from Paul Erdős, who (co)authored around 1525 papers.

What is the Erdős number of Paul Erdős himself? What does it mean that the Erdős number of W. Just is 2? What can you deduce from this information about the Erdős numbers of the other authors of this chapter?

Sample Solution: Because there is a path of length 0 from every node to itself, the Erdős number of Paul Erdős himself must be 0.

The information about W. Just implies that he never coauthored a paper with Paul Erdős; otherwise his Erdős number would be 1. But there must have been a coauthor i_1 of Erdős who also coauthored at least one paper with W. Just, so that (i_0, i_1, i_2) is a path in the collaboration network, with i_0 denoting P. Erdős and i_2 denoting W. Just.

If i_3 denotes any of the other coauthors, then (i_0, i_1, i_2, i_3) will be a path of length 3 in the collaboration network after this chapter appears. At that time, the Erdős number of i_3 will be *at most* 3. However, you cannot conclude from the given information that it will be equal to 3, as there may be a shorter path from i_0 to i_3 in the collaboration network. \square

Exercise 9.13. Is it true that a graph G is complete if, and only if, $\text{diam}(G) = 1$?

Sample Solution: Almost. If $\text{diam}(G) = 1$, then for every pair (i, j) of nodes there must be a path of length ≤ 1 from i to j in G . If $i \neq j$, then $P = (i, j)$ is the only possibility. But $P = (i, j)$ can only be a path if $\{i, j\}$ is an edge. Thus $\text{diam}(G) = 1$ implies that G is complete.

Conversely, in a complete graph, each pair (i, j) for $i \neq j$ is a path of length 1 between i and j , so the diameter of a complete graph is always ≤ 1 . But a graph with just one node and no edges (such as the connected component of node 3 in G_1) is also complete: Because we ruled out loops, all possible edges are present. For this special case of a complete graph the diameter is 0; all complete graphs with at least two nodes do have diameter 1. \square

Exercise 9.14. Find the degrees k_i of all nodes in G_1 and the mean degree $\langle k \rangle$.

Sample Solution: $k_1 = k_5 = 3$, $k_2 = 2$, $k_4 = k_6 = k_7 = k_8 = 1$, $k_3 = 0$. The mean degree is equal to

$$\langle k \rangle = \frac{3 + 2 + 0 + 1 + 3 + 1 + 1 + 1}{8} = 1.5.$$

\square

Exercise 9.15. Find a formula for the degree of each node and the total number of edges of a complete graph K_N .

Sample Solution: Each node i is adjacent to all *other* nodes; thus $k_i = N - 1$. In other words, the graph is $(N - 1)$ -regular. The set of edges is the set of all two-element subsets of $V(K_N) = \{1, \dots, N\}$. You may recall from your introductory probability or statistics course that this is the set of all *combinations* of 2 out of N elements. It has size $\binom{N}{2} = \frac{N(N-1)}{2}$. \square

Exercise 9.16. Find all subgraphs and induced subgraphs of the complete graph K_2 .

Sample Solution: We describe the subgraphs in mathematical terminology here; the reader may want to draw the corresponding pictures.

$K_2 = (\{1, 2\}, \{\{1, 2\}\})$ itself is a subgraph, because removal of any edges or nodes is optional, not mandatory.

All other subgraphs are *proper subgraphs*. If we remove only the edge, we get $G_{1,2}^\emptyset = (\{1, 2\}, \emptyset)$, where the set of edges is empty. This is the only subgraph of K_2 that is *not* induced by a set of vertices.

If we remove one vertex, we need to remove the edge as well, as we cannot have edges dangling without one endpoint. This gives two more subgraphs: $G_1^\emptyset = (\{1\}, \emptyset)$ and $G_2^\emptyset = (\{2\}, \emptyset)$.

In this particular example, all proper subgraphs are *empty graphs* that have no edges whatsoever. These structures are different from the *empty set* that we would obtain by removing all vertices. The empty set is not considered a subgraph, as we specifically required in our definition of a graph that the vertex set V be nonempty. \square

Exercise 9.17. Show that a graph G is a tree if, and only if, for every pair of nodes $i \neq j$ there exists exactly one simple path from i to j in G .

Sample Solution: A graph G is connected if, and only if, for every pair of nodes $i \neq j$ there exists a path from i to j in G . If such a path exists, it can be pruned to a simple path. Trees are connected graphs without cycles. So it suffices to show that a graph contains a cycle if, and only if, for some nodes i, j there exist two distinct simple paths $P_1 = (i = i_1, i_2, \dots, i_m = j)$ and $P_2 = (i = i'_1, i'_2, \dots, i'_m = j)$ from i to j in G . In general, the nodes $i_1, i_2, \dots, i_{m-1}, i'_2, \dots, i'_{m-1}, j$ do not need to be all distinct, but if we choose i, j with two distinct simple paths between them so that $m - 1 = d(i, j)$ is as small as possible, they will be. By moving along the edges of P_1 from i to j and then from j along the edges of P_2 back to i , we traverse a cycle $(i, i_1, \dots, j, i'_{m-1}, \dots, i'_2, i)$ in G .

Conversely, if (i, i_2, \dots, i_m, i) is a cycle, then $P_1 = (i, i_2)$ and $P_2 = (i, i_m, i_{m-1}, \dots, i_2)$ are two distinct simple paths from i to i_2 in G . \square

Exercise 9.18. Which of these scenarios could actually be observed in a network-based model for this disease with contact network G_1 ?

Sample Solution: Scenario 1 could be observed if the index case, host 1, infects hosts 2 and 5, and host 5 in turn infects host 4. Note that all of the required transmissions of disease agents happen along edges of G_1 .

Scenario 2 could not be observed, as the infection would need to travel from host 1 along a path in G_1 to host 3, but host 3 is isolated in this network. In fact, host 3 could only be infected from outside the population. Isolated nodes can never experience infection in these models, unless they themselves happen to be the index case. The hosts represented by these nodes may feel lonely, but this loneliness protects them from infection and protects all other nodes from infection by isolated ones. Similarly, the infection cannot reach node 8 from nodes outside the connected component of 8.

Scenario 3 could not be observed either, because in order to reach node 4 from node 1, the infection would need to travel along a path in G_1 , and each such path must include node 5. Thus in an immunizing disease, node 5 can no longer be susceptible at time 2, when node 4 is infectious.

Scenario 4 could be observed if node 5 becomes infectious after time 1 and ceases to be infectious before time 2. This could definitely happen in continuous-time models and in the real world. However, this scenario would not be allowed in next-generation models.

In Scenario 5, host 7 must have been infected by host 1, because 1 is the only node adjacent to node 7 in G_1 . At time 1, host 1 is no longer infectious, but host 7 is still susceptible, so this scenario cannot happen in any of our models. In a real-world situation, it might be the case that transmission of the infection from 1 to 7 occurred at a slightly earlier time than time 1, and host 7 would still not be infectious or show any symptoms at time 1. In this case, the observable data may indicate a scenario that is empirically indistinguishable from Scenario 5. In terms of our models though, we would consider host 7 to be exposed at time 1 and would write $St(1) = (RISSESSES)$ instead. \square

Exercise 9.19. Construct the contact network G_{IV} that, according to the story, seems most appropriate for transmission of diseases in this community. Spell out the assumptions that you made in constructing this network.

Sample Solution: If we assume that contacts occur *only* during the meetings of the research groups, the graph shown in Figure S9.1 represents the resulting contact network. \square

Exercise 9.20.

- (a) Find $\langle k \rangle$ for the networks $G_{RR}(3, 5)$, $G_{RD}(3, 5)$, and G_{IV} that were introduced in Exercise 9.19.
 (b) Many species of Gibbons are believed to be sexually entirely monogamous. What would this imply about $\langle k \rangle$ for the sexual contact network in a population of Gibbons?

Sample Solution:

- (a) For the network $G_{RR}(3, 5)$, there are four nodes of degree 2 (the corners), eight nodes of degree 3 (along the edges), and three nodes of degree 4 (in the interior). Thus, we have

$$\langle k \rangle = \frac{4 \times 2 + 8 \times 3 + 3 \times 4}{15} \approx 2.93.$$

For the network $G_{RD}(3, 5)$, there are four nodes of degree 3 (the corners), eight nodes of degree 5 (along the edges), and three nodes of degree 8 (in the interior). This gives

$$\langle k \rangle = \frac{4 \times 3 + 8 \times 5 + 3 \times 8}{15} \approx 5.07.$$

Finally, the sequence of degrees of the graph G_{IV} is given by $\bar{k} = (2, 4, 2, 1, 4, 3, 5, 5, 3, 2, 5)$, which leads to $\langle k \rangle \approx 3.27$.

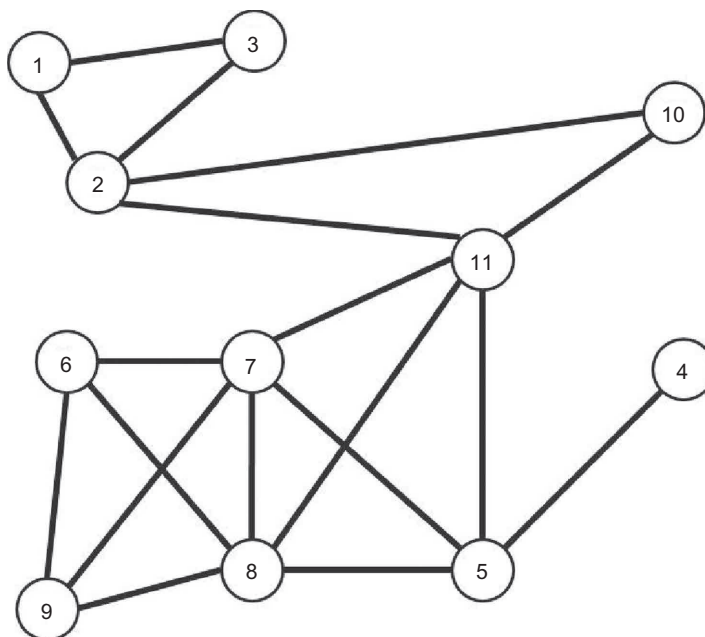


FIGURE S9.1 The contact network G_{IV} at the Ivory Tower Research Institute.

- (b) Because Gibbons are believed to be sexually entirely monogamous, this implies the degree k_i of each node i in the sexual contact network is at most 1. It follows from Eq. (9.9) that the mean degree $\langle k \rangle = \frac{1Q_1 + 0Q_0}{N} \leq 1$. The mean degree will be equal to 1 if, and only if, each member of the population is sexually active so that $Q_1 = N$ and $Q_0 = 0$. □

Exercise 9.21.

- (a) Find the neighborhoods $\mathcal{N}_1(3)$, $\mathcal{N}_{\leq 1}(3)$, $\mathcal{N}_2(3)$, and $\mathcal{N}_{\leq 2}(3)$ in the graph $G_{\text{NN}}^1(9, 2)$.
 (b) How are the Erdős numbers that were mentioned in Section 9.3.1 related to the neighborhoods in the scientific collaboration graph?
 (c) Suppose G is a graph with vertex set $\{1, \dots, N\}$. Show that for every node i the strongly connected component of i in G is equal to the set $\mathcal{N}_{\leq N-1}(i)$.

Sample Solution:

- (a) $\mathcal{N}_1(3) = \{1, 2, 4, 5\}$, $\mathcal{N}_{\leq 1}(3) = \{1, 2, 3, 4, 5\}$, $\mathcal{N}_2(3) = \{6, 7, 8, 9\}$, and $\mathcal{N}_{\leq 2}(3)$ comprises all nodes. Note: We assume here that nodes are numbered from 1 to 9. IONTW will assign label 0 to node 9.
 (b) If node i_0 represents Paul Erdős, then \mathcal{N}_ℓ consists of all researchers with Erdős number ℓ , and $\mathcal{N}_{\leq \ell}$ consists of all researchers with Erdős number $\leq \ell$.
 (c) For every ℓ , all nodes in $\mathcal{N}_{\leq \ell}(i)$ are at a finite distance from i in G . This implies that each $\mathcal{N}_{\leq \ell}(i)$, and in particular $\mathcal{N}_{\leq N-1}(i)$, is a subset of the connected component of i in G . Conversely, every node in the connected component of i in G must by definition belong to $\mathcal{N}_{\leq \ell}(i)$ for some finite ℓ . It suffices to show that the smallest such ℓ can never exceed $N - 1$. But if $d(i, j) = \ell$, then there must exist a simple path $P = (i = i_1, i_2, \dots, i_\ell, j)$ with $\ell + 1$ distinct nodes. Because the entire vertex set contains only N distinct nodes, it follows that $\ell \leq N - 1$. □

Exercise 9.22. How large or small does the reduction factor r need to be so as to prevent all major outbreaks? What does this imply for a successful policy of imposing quarantine?

Sample Solution: In order to prevent major outbreaks from happening, we need to reduce R_0 to $R_0^- \leq 1$. If we consider a continuous-time model, it follows from Eq. (9.4) that

$$R_0 \approx \beta \langle \tau^I \rangle N \quad \text{and} \quad R_0^- \approx \beta \langle \tau^I \rangle^- N = \beta r \langle \tau^I \rangle N \approx r R_0. \quad (\text{S9.1})$$

By setting $R_0^- = 1$ and solving (S9.1) for r we find that we need $r \leq \frac{1}{R_0}$.

In the context of the 2009 H1N1 infection, this means that we need $r = \frac{\langle \tau^I \rangle^-}{\langle \tau^I \rangle} \leq 0.57$. If the policy of quarantine is sufficiently strictly enforced so that it cuts $\langle \tau^I \rangle$ into half, it will most likely prevent major outbreaks from occurring. □

Exercise 9.23.

- (a) Explain how these types of behavior modification will influence the following quantities that we introduced in this chapter: b, c, v, R_0 . What is their likely effect on the epidemic curve?
 (b) Does adoption of such measures also make sense if they are not sufficient to prevent all major outbreaks?

Sample Solution: Refraining from certain types of socializing will reduce the probability c of making contact during a unit time interval. Changing behavior during contacts or using hand sanitizer more often will reduce v , the conditional probability that at least one of the contacts that occur during a given time interval will be effective. In view of Eq. (9.2), both types of control measure will reduce the probability b of at least one effective contact during a unit time interval. Thus by Eq. (9.4), both types of behavior modification will reduce R_0 to some smaller value $R_0^- < R_0$.

If $R_0^- > 1$, behavior modification alone will not prevent all major outbreaks. But if, for example, these measures result in a reduction of $R_0 = 1.75$ to $R_0^- = 1.25$, the predicted final size would become 0.37, about half of what was predicted in the absence of any control measures. The effect on the epidemic curve will be similar to the one illustrated in the right panel of Figure 8.2 of Chapter 8. \square

Exercise 9.24. How could culling be translated into our mathematical framework?

Sample Solution: If culling takes place before an anticipated outbreak, the mathematical interpretation would be exactly the same as for vaccination. From the point of view of the animals there is a world of difference between being culled or vaccinated, but the predictions for the course of a possible outbreak will be exactly the same.

However, while vaccinations usually are not effective immediately, culling is. For this reason, culling is sometimes implemented as the control measure of last resort when the outbreak has already started and vaccination is no longer an option. \square

Exercise 9.25.

- (a) Use Eq. (9.13) to derive a formula for HIT in terms of R_0 .
- (b) Find HIT for the 2009 H1N1 infection.

Sample Solution:

- (a) According to Eq. (9.13), the threshold value for K that will give herd immunity satisfies $R_0 \frac{N-K}{N} = 1$, or equivalently,

$$\text{HIT} = \frac{K}{N} = 1 - \frac{1}{R_0}. \quad (\text{S9.2})$$

- (b) Substituting the value $R_0 = 1.75$ into Eq. (S9.2), we get $\text{HIT} = 0.43$. Thus, achievement of herd immunity would require vaccinating at least 43% of the population. \square

Exercise 9.26. Outline the pros and cons of immunizing the whole population versus only the fraction needed for achieving herd immunity. Discuss possible obstacles to implementing your favored policy. Debate the issues with a partner or in a group of students.

Sample Solution: This is a very open-ended discussion question for which giving a sample solution would be inappropriate. \square

Exercise 9.27. Calculate the herd immunity threshold (HIT) for this model and derive a prediction about how many hosts need to be vaccinated to achieve herd immunity.

Sample Solution: For $R_0 = 2$ we get $\text{HIT} = 1 - \frac{1}{R_0} \approx 0.5$. So one would need to vaccinate at least 100 hosts to achieve herd immunity. \square

Exercise 9.28. Sort and analyze the column of your output file that represents the number of hosts who experienced infection in each simulation. Calculate the corresponding proportions of unvaccinated hosts who experienced infection.

Do the results indicate that the vaccination prevented all major outbreaks among the $200 - K$ hosts who did not get vaccinated? *Hint:* Keep in mind that the numbers in this column always include all vaccinated hosts.

Sample Solution: In our experiment we saw no clear gap that would allow us to unambiguously distinguish between minor or major outbreaks. In 65 of our simulations, at most 3 of the 100 hosts who did not get vaccinated experienced infection, which is much better than the approximately 20% of minor outbreaks that one would observe without control measures. However, in one of our simulations more than 45% of hosts who did not receive the vaccine experienced infection, so we could not say with confidence that the vaccination strategy prevented all major outbreaks. \square

Exercise 9.29.

- (a) Redo Exercise 9.28 for vaccinating $1.2K$ and $0.8K$ hosts, respectively.
- (b) Compare your results with those of Exercise 9.28 in terms of the mean and maximum proportions of unvaccinated hosts who experienced infection and in terms of the proportion of simulations in which two or fewer hosts experienced infection.
- (c) Is the effect of increasing vaccination coverage from $0.8HIT$ to HIT the same as increasing vaccination coverage from HIT to $1.2HIT$? Discuss implications for public health policy.

Sample Solution: In a population of $N = 200$ hosts, a herd immunity threshold of $HIT = 0.5$ corresponds to vaccinating $K = 100$ hosts. We obtained the following results, shown in Table S9.4.

K	N-K	Max # Inf	Avg # Inf	(Max # Inf)/(N-K)	(Avg # Inf)/(N-K)
80	120	78	18.68	0.65	0.16
100	100	50	6.92	0.5	0.07
120	80	29	3.17	0.36	0.04

In our simulations, the most significant jump appears from $K = 80$ to $K = 100$ in terms of the average number and average proportion of infections. This suggests that the major effect of increasing vaccination coverage from HIT to $1.2HIT$ might be to reduce the actual number of hosts that will experience infections in minor outbreaks. \square

Exercise 9.30. Suppose we have only enough vaccine available to achieve a 10% coverage. Would this have much of an effect in terms of protecting the remaining 90% of the population? Explain how you would approach this question with the help of simulations and discuss your findings in terms of implications for public health policy.

Sample Solution: We could investigate this scenario by setting $K = 20$ (10% of $N = 200$) and running a batch of 100 simulations, just as we did in the previous exercises, and comparing these results to those when no vaccinations were given ($K = 0$). Our results are shown in Table S9.5.

K	N-K	Max # Inf	Avg # Inf	(Max # Inf)/(N-K)	(Avg # Inf)/(N-K)
20	180	154	90.36	0.86	0.5
0	200	174	125.8	0.87	0.63

From our results we see that vaccinating only 10% of the population does very little to the maximum proportion of infections. It appears to slightly decrease the average proportion. Further, when $K=20$, in 68% of our simulations, more than half of the population experienced infection, compared to 79% when no individuals were vaccinated. While vaccination does have some effect, we would have no confidence in our ability to prevent a major outbreak by only vaccinating 10% of the population. Our recommendations for public health policy would therefore be to examine other vaccination strategies that use the available doses of vaccine for targeting certain groups of hosts. Exploring such strategies will require relaxing the uniform mixing assumption. We will investigate one such targeted vaccination strategy in Module 9.4 and additional examples in the modules at our website [3]. \square

Exercise 9.31. Based on the parameter values above, for both measles and diphtheria calculate α and β from Eq. (9.4) of Section 9.2.2 and γ as we explained in Section 8.6 of [4]. Assume a population size of $N = 200$ and use as midpoints the parameters that are given only in the form of ranges. Round all answers to 4 decimal places.

Sample Solution: Measles: $R_0=16$, $\alpha = 1/6.5$, $\gamma = 1/7.5$, and $\beta = [(1/6.5)*16]/200$.

Diphtheria: $R_0=6.6$, $\alpha = 1/17.5$, $\gamma = 1/3.5$, and $\beta = [(1/17.5)*6.6]/200$. □

Exercise 9.32. Open the output file and sort the data by the column labeled `count turtles with [removed?]`. Try to find a clear jump in the number of removed hosts during the simulations. Based on this jump, group your data into those corresponding to minor and major outbreaks. Record the maximum durations for all outbreaks that you would classify as minor and major, according to this observation, as well as the proportion of simulations in which you observed a minor outbreak. Calculate the mean duration and mean final size for the groups of minor and major outbreaks separately.

Sample Solution: For our set of 200 runs, we had a minor outbreak proportion of 0.13. The mean duration for minor outbreaks was 3.1971 with mean final size of 0.0056. For major outbreaks the mean duration was 140.4698 with mean final size 0.9985. □

Exercise 9.33. Repeat Exercise 9.32 for the case of an *SIR*-model for diphtheria. This can be done by first changing **latent-period** to **Off** before defining a new batch processing experiment. Alternatively, you can edit the previous experiments by setting [“latent-period” false] in the **Experiment** window of **BehaviorSpace**.

Sample Solution: For our set of 200 runs, we had a minor outbreak proportion of 0.15. The mean duration for minor outbreaks was 3.5057 with mean final size of 0.0058. For major outbreaks the mean duration was 122.4569 with mean final size 0.9986. □

Exercise 9.34. Repeat Exercises 9.32 and 9.33 for measles instead of diphtheria.

Sample Solution: For our set of 200 runs of an *SEIR*-model, we had a minor outbreak proportion of 0.05. The mean duration for minor outbreaks was 0.9472 with mean final size of 0.0065. For major outbreaks the mean duration was 68.3527 with mean final size 1.

For our set of 200 runs of an *SIR*-model, we had a minor outbreak proportion of 0.075. The mean duration for minor outbreaks was 0.3315 with mean final size of 0.0057. For major outbreaks the mean duration was 41.1494 with mean final size 1. □

Exercise 9.35. Find the sets $V(G)$ and $E(G)$ for the graph that is currently in the **World** window.

Sample Solution: For the graph G shown in Figure S9.2 we have $V(G) = \{0, 1, 2, 3, 4\}$ and $E(G) = \{\{0, 1\}, \{2, 3\}, \{3, 4\}\}$. □

Exercise 9.36. Have you seen a graph like this before? *Hint:* Think about shifting the numbering of the nodes by one so that it ranges from 1 to 8 instead of from 0 to 7.

Sample Solution: After changing the labels as in the hint, the graph in the **World** window would become the same mathematical object as the graph G_1 of Figure 9.2. It is drawn by IONTW in a different way, but it would be the *same mathematical object* as G_1 after we change the labels, as the vertex set and the edge set would become the same. However, since $V(G) \neq V(G_1)$ and $E(G) \neq E(G_1)$, the graph G whose representation you actually see in the **World** window is not the same mathematical object as G_1 . Mathematicians call graphs that differ only by the labelling of the nodes *isomorphic*. □

Exercise 9.37.

- (a) Create another plain text file `triangle.txt` that will give you a picture like Figure 9.1 in the **World** window after you press **Load**.
- (b) Is there a less tedious way of making this graph show up?

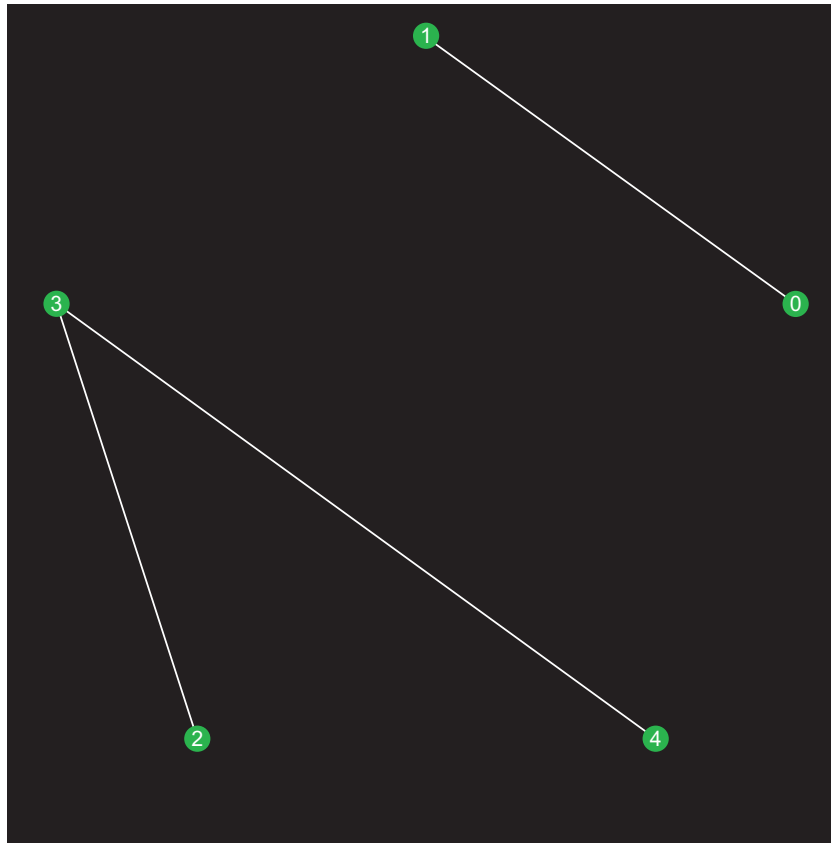


FIGURE S9.2 A random graph G with five nodes.

Sample Solution:

- (a) Your file might look like this:

```
network
3
0 1
1 2
0 2
```

- (b) Choose **network-type** → **Complete Graph**, set **num-nodes** to 3, and press **New**. □

Exercise 9.38.

- (a) Find all simple paths from node 1 to node 3 and the distance $d(1, 3)$.
 (b) Are there any paths of length 4 from node 1 to node 3? If so, how many? If not, why can there be no such paths?
 (c) Are there any paths of length 3 from node 1 to node 3? If so, how many? If not, why can there be no such paths?

Sample Solution:

- (a) $(1, 2, 3)$ is the only simple path from node 1 to node 3. Because this path has length 2, it follows that $d(1, 3) = 2$.
 (b) The following 4 paths from node 1 to node 3 have length 4: $(1, 0, 1, 2, 3)$, $(1, 2, 1, 2, 3)$, $(1, 2, 3, 2, 3)$, $(1, 2, 3, 4, 3)$.
 (c) There are no paths of length 3 from 1 to 3 in this graph. To see this, note that every edge has an even-numbered and an odd-numbered endpoint. Thus, any path of length 3 that starts at the odd-numbered node like node 1 must end at an even-numbered node. □

Exercise 9.39.

- (a) Find all simple paths from node 1 to node 4 and the distance $d(1, 4)$.
 (b) Change **d** from 1 to 2 and press **New**. What happens to the number of simple paths from node 1 to node 4 and to $d(1, 4)$ if you add new edges in this way?
 (c) Does the addition of new edges in point (b) affect the distances between *all* pairs of nodes? If not, for which pairs do the distances remain the same?

Sample Solution:

- (a) There are 3 simple paths from node 1 to node 4 in this graph: $(1, 0, 2, 4)$, $(1, 3, 2, 4)$, $(1, 3, 5, 4)$. All of them have length 3 and it follows that $d(1, 4) = 3$.
 (b) We get many additional simple paths from 1 to 4, including $(1, 2, 4)$, $(1, 3, 4)$, $(1, 0, 3, 4)$, $(1, 0, 3, 2, 4)$, $(1, 0, 3, 2, 5, 4)$. The shortest simple paths have length 2, and it follows that $d(1, 4) = 2$.
 (c) No. The distance between nodes that are already connected by an edge in $G_{\text{NN}}^1(6, 1)$ remains 1 in $G_{\text{NN}}^1(6, 2)$ and the distance from each node to itself remains 0. \square

Exercise 9.40. Consider the graph in the **World** window.

- (a) Find the degrees of nodes 1 and 3, as well as the mean degree $\langle k \rangle$ using Eq. (9.5).
 (b) How do the results of point (a) change if you first increase **d** from 1 to 2 and then from 2 to 3? *Hint:* Sometimes edges overlap and you cannot clearly distinguish them in the **World** window.
 (c) Are the graphs that you considered in this exercise regular, that is, k -regular for some k ?

Sample Solution:

- (a) The degrees of nodes 1 and 3 are $k_1 = 2$ and $k_3 = 3$. The mean degree is $\langle k \rangle = \frac{2+2+3+3+2+2}{6} = \frac{14}{6} \approx 2.33$.
 (b) For **d** = 2: The degrees of nodes 1 and 3 are $k_1 = 3$ and $k_3 = 5$.
 The mean degree is $\langle k \rangle = \frac{3+3+5+5+3+3}{6} = \frac{22}{6} \approx 3.67$.
 For **d** = 3 all nodes have degree $k_i = 5$ and therefore also $\langle k \rangle = 5$.
 (c) For **d** = 1, 2 the graphs that we considered are not regular. For **d** = 3 we get a regular graph; more specifically, the 5-regular graph $G_{\text{NN}}^2(6, 3) = K_6$. \square

Exercise 9.41.

- (a) Find the degree of each node for the graph in the **World** window as well as the mean degree $\langle k \rangle$ using Eq. (9.5).
 (b) Is this graph regular?

Sample Solution:

- (a) The degree k_i of each node is 0 and hence also $\langle k \rangle = 0$.
 (b) Yes. \square

Exercise 9.42. Are these graphs actually different as mathematical objects or just the same graph drawn in different ways? Explain your method for obtaining the answer.

Sample Solution: Most of the time you will see a graph that is a single cycle, but drawn in different ways. Mathematicians would say that these graphs are *isomorphic*. But usually they will be different mathematical objects, as the edges will connect different pairs of nodes. To see this, carefully check which nodes are adjacent to node 0 each time you press **New**. If node 0 is adjacent to different nodes in two graphs, these graphs cannot be the same.

Occasionally you may see a graph that has two connected components of size 3 each. These graphs are not even isomorphic to the graphs that consist of a single cycle. \square

Exercise 9.43.

- (a) Do you see (many) regular graphs?
 (b) Does the mean degree $\langle k \rangle$ appear always to be the same whenever you press **New**?
 (c) Closely examine the degrees in a couple of graphs that you create in this way. How are these degrees related to the information that is plotted in **Network Metrics**?

Sample Solution:

- (a) No.
- (b) No. Sometimes we get more and sometimes fewer edges. It follows from Eq. (9.6) that the mean degree will be smaller for graphs with fewer edges and the same number of nodes.
- (c) The heights of the bars give the number of nodes with a given degree. □

Exercise 9.44. Set **num-nodes** to 10 and **lambda** to 2 and press **New**.

- (a) Calculate the mean degree $\langle k \rangle$ using Eq. (9.5).
- (b) Set **num-nodes** to 20 and again calculate the mean degree $\langle k \rangle$. If this is getting tedious, try to find a shortcut that uses the information plotted in **Network Metrics**.
- (c) Repeat part (b) by increasing **num-nodes** to 50. For your calculations, use the shortcut that you found in part (b).
- (d) Does the mean degree appear to be reasonably close to **lambda** for large network sizes?

Sample Solution:

- (a) Results will vary depending on your graph. We found a graph with $\langle k \rangle = 2.6$.
- (b) We found a graph with 4 nodes of degree 0, 10 nodes with degree 1, 4 nodes with degree 2, and 2 nodes with degree 3. Because each node i contributes a term k_i to the numerator of Eq. (9.5), we can use this information to calculate $\langle k \rangle = \frac{(0)(4)+(1)(10)+(2)(4)+(3)(2)}{20} = 1.2$.
- (c) We found a graph with 5 nodes of degree 0, 18 nodes with degree 1, 14 nodes with degree 2, 7 nodes with degree 3, and 3 nodes each with degrees 4 and 5. By using the shortcut we found in part (b) we can calculate $\langle k \rangle = \frac{(0)(5)+(1)(18)+(2)(14)+(3)(7)+(4)(3)+(5)(3)}{50} = 1.88$.
- (d) Yes, the result we got for 50 nodes is reasonably close to **lambda** = 2. □

Exercise 9.45. For each of the following parameter settings, create a network, and determine whether the network is a tree or forest. Record this information in a table, together with the numbers of nodes and edges for each of the graphs.

- (a) **Empty Graph**; **num-nodes** = 1, 2
- (b) **Nearest-neighbor 2**; **num-nodes** = 5, 6, 7; **d** = 1
- (c) **Preferential Attachment**; **num-nodes** = 8; **lambda** = 1; **d** = 1
- (d) **Random Regular**; **num-nodes** = 10; **lambda** = 1, 2

Sample Solution: See Table S9.6. □

Exercise 9.46. Now look at the data that you collected for the previous exercise and form a conjecture on how the number of edges in a tree is related to the number of nodes. Try to confirm your conjecture for the tree G_T of Figure 9.3. Can you prove your conjecture?

Sample Solution: In any tree, the number of nodes exceeds the number of edges by exactly 1. In fact, the following mathematical theorem gives what mathematicians would call a *characterization of trees*. It shows that we could have defined trees in terms of the numbers of nodes and edges instead of the absence of cycles.

Theorem 9.1. *A connected graph G is a tree if, and only if, $|V(G)| = |E(G)| + 1$, where $|V(G)|$ and $|E(G)|$ denote the numbers of nodes and edges of G .*

One can prove the “only if” part of this theorem by *mathematical induction* as follows: If G has exactly 1 node, then G has no edges, and thus $|V(G)| = 1 = 0 + 1 = |E(G)| + 1$.

Now assume that the theorem is true for all connected graphs G with $|V(G)| = N$ nodes. This is called the *inductive assumption*. We have already shown that it is true for $N = 1$.

TABLE S9.6 The Data for Exercise 9.45

Network Type	# of Nodes	# of Edges	Tree/Forest/Neither
Empty Graph	1	0	Tree
Empty Graph	2	0	Forest
Nearest-neighbor 2 $d = 1$	5	4	Tree
Nearest-neighbor 2 $d = 1$	6	7	Neither
Nearest-neighbor 2 $d = 1$	7	6	Tree
Preferential Attachment $\lambda = 1, d = 1$	8	7	Tree
Random Regular $\lambda = 1$	10	5	Forest
Random Regular $\lambda = 2$	10	10	Neither

At this stage we would need to prove an auxiliary result, called a *lemma*, that shows that every tree G with $N > 1$ nodes has at least one leaf, and if we remove this leaf, we obtain a subgraph G^- that is a tree with one fewer node.

Remember that there cannot be any dangling edges in a subgraph. Thus the subtree G^- must also have one fewer edge than G , because we must have removed the edge that connects the removed leaf with the remainder of the tree together with the leaf itself.

We could now prove the *inductive step* as follows: Consider a tree G with $N > 1$ nodes. Remove one leaf. The resulting subgraph G^- is also a tree, with N nodes, and one fewer edge than G . By inductive assumption, the equality $|V(G^-)| = |E(G^-)| + 1$ holds. But then we get

$$|V(G)| = N + 1 = |V(G^-)| + 1 = (|E(G^-)| + 1) + 1 = |E(G)| + 1,$$

as required.

For the proofs of the lemma and the “if” direction of Theorem 9.1, the reader may want to consult an introductory textbook on graph theory. \square

Exercise 9.47. For each of the following parameter settings, create a regular tree and find the sets of nodes at the levels $\mathcal{N}_\ell(0)$ for $0 \leq \ell \leq h(T)$:

- (a) $\lambda = 1, d = 8$. This type of tree is called a *star tree*.
- (b) $\lambda = 4, d = 2$. This type of tree is called a *binary tree*.
- (c) How, exactly, is the parameter d related to the degrees of the nodes?

Sample Solution:

- (a) $\mathcal{N}_1(0) = \{1, 2, 3, 4, 5, 6, 7, 8\}$
- (b) $\mathcal{N}_1(0) = \{1, 2\}$, $\mathcal{N}_2(0) = \{3, 4, 5, 6\}$, $\mathcal{N}_3(0) = \{7, 8, 9, 10, 11, 12, 13, 14\}$, and $\mathcal{N}_4(0)$ comprises all other nodes.
- (c) The degree of the root 0 is always equal to d . The degree of all nodes other than the root and the leaves is $d + 1$.

\square

Exercise 9.48. Show that the diameter $\text{diam}(T)$ of a tree satisfies the inequalities

$$h(T) \leq \text{diam}(T) \leq 2h(T). \quad (\text{S9.3})$$

Find sufficient and necessary conditions for these inequalities to be strict.

Sample Solution: The distance of any node in $\mathcal{N}_{h(T)}$ from node i is by definition equal to $h(T)$. This gives the first inequality. The second inequality follows from the observation that we can construct a path P from any node j_1 to any node j_2 by first walking from j_1 to i along a simple path of length $d(j_1, i) \leq h(T)$ and then walking from i to j_2 along a simple path of length $d(i, j_2) \leq h(T)$. This shows that

$$d(j_1, j_2) \leq d(j_1, i) + d(i, j_2) \leq h(T). \quad (\text{S9.4})$$

The first inequality in (S9.4) is called the *triangle inequality*. It holds for any kind of legitimate distance function $d(\cdot, \cdot)$. For trees it turns into an equality if, and only if, the path P that we described above is simple. This path does not need to be simple, but it will be if j_1 and j_2 are different nodes in the same level of the tree. If this level is the highest one, then the second inequality in Eq. (S9.4) also turns into an equality. Thus, a necessary condition for the second inequality in Eq. (S9.3) to be strict is that the highest level $\mathcal{N}_{h(T)}(i)$ contains exactly one node. Upon some further reflection we can see that this condition is also sufficient.

Similarly, if $j \in \mathcal{N}_{h(T)}(i)$ and if $\mathcal{N}_1(i)$ contains two distinct nodes i_1, i_2 , then the simple path from j to i can pass through at most one of these nodes, say i_1 . But then the simple path from j to i_2 must pass through i . It follows that the existence of two distinct nodes at level $\ell = 1$ is a sufficient condition for the first inequality in (S9.3) to be strict. This condition is not necessary, though. Consider Figure S9.3. For the tree on the left, $h(T) = 3 < 4 = \text{diam}(T)$. While the root i has only one neighbor, the strict inequality is caused by the branching at node b .

Clearly, a necessary condition for $h(T) < \text{diam}(T)$ is that the tree T has at least one branching node. However, as the tree on the right of Figure S9.3 shows, this condition is not sufficient. The figure indicates that a condition for $h(T) < \text{diam}(T)$ that will simultaneously be necessary and sufficient must involve the relation between the levels of the branching nodes and the heights of the subtrees above these nodes. It is not too hard but a bit tedious to precisely formulate and derive such a condition. We omit the details here. \square

Exercise 9.49. Now some of the edges with red endpoints are red and some white. Does this make sense in terms of the interpretation of edge color that we gave above?

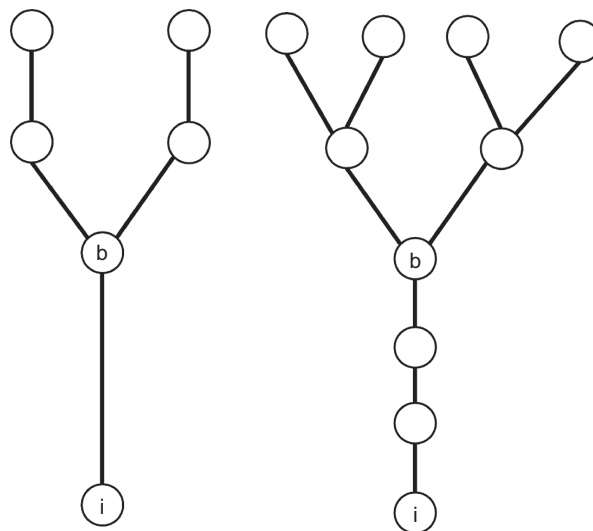


FIGURE S9.3 For the tree on the left, $h(T) = 3 < 4 = \text{diam}(T)$. For the tree on the right, $h(T) = 5 = \text{diam}(T)$.

Sample Solution: Yes. For all white edges with at least one red endpoint, *both* endpoints are red, which signifies infectious hosts. An effective contact between two hosts who are already infectious cannot be a successful one from the point of view of the pathogens. \square

Exercise 9.50. How could you use these capabilities to create an initial state in which nodes 0, 1, 2, 3 are infectious and nodes 4, 5, 6 are removed, while all other nodes are susceptible?

Sample Solution: In the current settings, after pressing **Reset**, press **Set**, and enter in the dialogue box the vector

[0 1 2 3]

Next we could change **set-state-to** \rightarrow **Removed**, press **Set** again, and enter in the dialogue box [4 5 6]

Note that this procedure will make user-specified sets of nodes red and gray. In Module 9.4, you will learn how to make two disjoint sets of specified sizes of randomly chosen nodes infectious and removed. \square

Exercise 9.51.

- (a) Can you always figure out who infected whom?
- (b) How is what you are seeing related to the notions of *path* and *distance*?

Sample Solution:

- (a) Yes.
- (b) The infection travels from the root to the leaves along the unique simple path between the root and the given node. At each step of the simulation, the set of nodes that are in the exposed or infectious state comprises all nodes that are at a fixed distance from the root. \square

Exercise 9.52. How do the observations differ from the ones in the previous simulation? How would you explain the observed differences?

Sample Solution: The infection still travels from the root to the leaves along the unique simple paths, but no longer at the same speed along each branch. The reason is that with the new value of **end-latency-prob**, there is now some randomness in the times when exposed (yellow) nodes become infectious (turn red). \square

Exercise 9.53.

- (a) How do the observations differ from the ones in the previous simulation? How would you explain the observed differences?
- (b) The subgraphs that are induced by the green nodes may contain several connected components. In the original tree, there will always be a path between any two such components of the subgraph. What property in terms of disease transmission do all these paths have in common?

Sample Solution:

- (a) It is no longer the case that all nodes will always eventually turn gray. Because **infection-prob** has been reduced from 1 to 0.5, it will no longer be the case that an effective contact occurs along each edge that turns red at some time during the simulation. The subgraph that is induced by the nodes that remain green at the end will usually have several connected components.
- (b) Each such path must contain a node that turned gray at the end of the simulation and an edge that turned red at some step of the simulation when no effective contact between the two hosts represented by its endpoints occurred. \square

Exercise 9.54.

- (a) Can you always figure out who infected whom?
- (b) Change **num-nodes** to 14, and create a new network and initial condition with one index case. Simulate an outbreak and answer the same question as in part (a).
- (c) Explain which property of the network was crucial for answering yes or no in parts (a) and (b).

Sample Solution:

- (a) Yes.
- (b) No.
- (c) In the second network, there are multiple paths of the same length between some pairs of nodes. □

Exercise 9.55.

- (a) Are the graphs $G_{\text{NN}}^1(N, d)$ regular?
- (b) How do the degrees of the nodes depend on the parameters **num-nodes** and **d**?

Sample Solution:

- (a) Yes.
- (b) As long as $N - 1 \geq 2d$, the degree of each node in $G_{\text{NN}}^1(N, d)$ is $2d$. Thus, as long as **num-nodes** is sufficiently large relative to **d**, the degree will not depend on **num-nodes**. Conversely, if **num-nodes** is sufficiently small relative to **d**, the degree will not depend on **d** and will be **num-nodes** - 1. In general, the degree will be the minimum of $2d$ and **num-nodes** - 1. □

Exercise 9.56.

- (a) For which of the following choices of m and n can we create the network $G_{\text{RR}}(m, n)$ with the above software capabilities?

$$m = 2, n = 10; \quad m = 12, n = 14; \quad m = 7, n = 33.$$

- (b) Find a necessary and sufficient condition on the pair (m, n) that allows for creation of $G_{\text{RR}}(m, n)$ as a **Nearest-neighbor 2** network.

Sample Solution:

- (a) We cannot create $G_{\text{RR}}(2, 10)$, because $(2)(10) = (4)(5)$ so that $G_{\text{NN}}^2(20, 1) = G_{\text{RR}}(4, 5)$. But the other two choices can be implemented as $G_{\text{RR}}(12, 14) = G_{\text{NN}}^2(168, 1)$ and $G_{\text{RR}}(7, 33) = G_{\text{NN}}^2(231, 1)$.
- (b) The inequality $m \leq n$ holds and the product $N = mn$ cannot be factored as $N = k\ell$ in such a way that k and ℓ are integers with $m < k \leq \ell < n$. □

Exercise 9.57. For the purpose of this exercise, let (i, j) denote the node in row i and column j of the grid. For which pairs of nodes (i_1, j_1) and (i_2, j_2) will IONTW put an edge into $G_{\text{NN}}^2(N, d)$ when $d > 1$? *Hint:* Sometimes edges overlap and you cannot clearly distinguish them in the **World** window.

Sample Solution: When $d > 1$, then $\{(i_1, j_1), (i_2, j_2)\}$ is an edge in $G_{\text{NN}}^2(N, d)$ if, and only if,

$$\max\{|i_1 - i_2|, |j_1 - j_2|\} < d.$$

□

Exercise 9.58.

- (a) Does the parameter **d** appear to influence the final size? Does it appear to influence the speed at which the disease spreads? *Hint:* Because we need to manually terminate the simulation by pressing **Go** again, we need to deduce the speed from features of the prevalence curves.
- (b) How would you explain the likely causes of the observed differences and similarities?

Sample Solution:

- (a) The parameter **d** has no influence on the final size: The whole population always becomes infectious. However, for smaller values of **d** it takes longer until all nodes are infectious.
- (b) We are simulating an *SI*-model with a positive value of **infection-prob**. Thus, infectious nodes will never recover and will eventually make effective contact with each adjacent node. Thus, each node that is adjacent to the index case will eventually become infectious, will infect each of its neighbors who are not already infectious, and so on, until all nodes are infectious. This will be true for all positive values of **d**. For smaller **d**, there are fewer edges along which the pathogens can be transmitted. This slows down the spread of the disease. □

Exercise 9.59. Based on your observations, form a conjecture about the set of hosts who will eventually experience infection. You may want to compare your answer with our sample solution before reading on.

Sample Solution: Not all nodes will eventually turn red; only the nodes in the connected component(s) that contain initially infectious hosts will. One might conjecture that the set of hosts who will experience infection is always equal to the connected component of the index case. We will see later how this conjecture, under suitable additional assumptions, can be proved. \square

Exercise 9.60. Can you always clearly see this connected component? If not, how can you get a little help from IONTW with the visualization?

Sample Solution: Press **Go** and see which nodes eventually turn red. The connected component of the initially infectious node will be the subgraph that is induced by these nodes. \square

Exercise 9.61. What do you observe? Does the conjecture that you formed in Exercise 9.59 hold up for *SIR*-models? If not, how would you modify it? Can you prove the modified conjecture?

Sample Solution: Not quite. It is no longer the case that the whole connected component of the index case will experience infection. One might still conjecture though that the set of nodes who will experience infection is always a subset of the connected component of the index case. If there is more than one initially infectious node, then the set of nodes that will experience infection should be a subset of the union of the connected components of the initially infectious nodes. \square

Exercise 9.62. What do you observe? How do these observations relate to the conjecture that you formed in Exercise 9.61? Do you perhaps want to revise your conjecture?

Sample Solution: The vaccinated nodes 0 and 10 act as barriers against the further spread of the infection. We might strengthen our conjecture to the effect that only nodes in the connected component of the index case *in the subgraph induced by the unvaccinated nodes* can experience infection. Sometimes all of these nodes will experience infection; sometimes only a proper subset of them will. \square

Exercise 9.63. We changed four parameters relative to the previous exercise. Which one is responsible for this difference?

Sample Solution: When we change \mathbf{d} from 1 to 2, we add edges of the form $\{i, i + 2\}$, in particular, an edge $\{9, 11\}$ and an edge $\{99, 101\}$. These edges allow the pathogen to be transmitted directly from host 9 to host 11 and directly from host 101 to host 99. Thus, vaccination of hosts 10 and 100 no longer creates insurmountable barriers against the spread of the pathogens. \square

Exercise 9.64. Analyze the column of your output file that represents the number of hosts who experienced infection in each simulation. Do the results confirm the predictions of the compartment-level model? Does the particular structure of this contact network appear to make it easier or more difficult for this particular disease to spread?

Sample Solution: In our experiment all hosts experienced infection in each run. This is worse than what the corresponding compartment-level model predicts. It appears that the structure of the network makes it easier for this particular disease to spread. \square

Exercise 9.65. Analyze the column of your output file that represents the number of hosts who experienced infection in each simulation. Do the results confirm the predictions of the compartment-level model? *Hint:* Remember that now we have 86 hosts who are removed from the outset, and the disease can spread only among the remaining 34 hosts.

Sample Solution: In our experiment the largest outbreak had 90 removed hosts at the end, which means that at most $90 - 86 = 4$ or a fraction of 0.12 of all unvaccinated hosts experienced infection during any of our simulated outbreaks. Vaccinating a fraction of HIT randomly chosen hosts does provide considerable protection of unvaccinated hosts in this model. \square

Exercise 9.66. Sort the column of your output file that represents the number of hosts who experienced infection in each simulation and analyze the observed outcomes. Do the results confirm the predictions of the compartment-level model?

Sample Solution: In our experiment, the smallest outbreak had 23 removed hosts at the end, and the largest had 120. The mean was 78.8. This means that between 10% and 100% of all 108 unvaccinated hosts experienced infection during our simulated outbreaks, with a mean percentage of 73% of all unvaccinated hosts. This is a lot less than the more than 90% that are predicted by the corresponding compartment-level model. Interestingly enough, after sorting the data we did not perceive a distinct gap between minor and major outbreaks.

Thus, the simulations indicate that the corresponding compartment-level model gives significantly distorted predictions about the effect of vaccinating a small percentage of randomly chosen hosts in the network-based model. \square

Exercise 9.67. How would you target vaccination to a particular set of 12 hosts so as to achieve the best possible result for the available amount of vaccine? First try to think of the best way to choose the hosts that will be vaccinated. Then run a batch of simulations similar to the ones of Exercise 9.66 for this choice of vaccinated hosts. Compare the results with the ones of the previous exercise in terms of minimum, maximum, and mean final sizes of the outbreak.

Sample Solution: The structure of the underlying network allows us to create barriers by vaccinating two adjacent hosts $i, i + 1$. The disease cannot cross this barrier from host $i - 1$ to host $i + 2$ or *vice versa*. With 12 doses of vaccine we can create up to 6 such barriers, and it seems best to space them evenly.

Thus we may want to try the following option for the .txt-file:

```
[ 1 2 21 22 41 42 61 62 81 82 101 102 ]
```

In our experiment, the minimum number of hosts who experienced infection was 15 and the maximum was 30. The value 30 was observed in 96 runs. The mean was 29.78. This translates into an average of close to 18, or 17% of all unvaccinated hosts. This was a lot less than what we observed in the simulated outbreaks where 12 randomly chosen hosts were vaccinated. The targeted vaccination strategy greatly reduced the expected and maximum numbers. \square

Exercise 9.68. Explain why Theorem 9.4 is true. Depending on your level of mathematical preparation, give either an informal argument or a formal mathematical proof.

Sample Solution: We give here a version of the solution that is not an entirely formal proof but captures the essential ideas.

- (a) Suppose host i experienced infection during an outbreak. If i was not initially infectious, then there must have occurred an effective contact between i and another host j_1 at some time t_1 at which j_1 was infectious and i was susceptible. According to the model, this could only have happened if $\{i, j_1\}$ is an edge in G^- . The same applies to j_1 : Either j_1 was initially infectious, or there must have occurred an effective contact between j_1 and another host $j_2 \neq j_1$ at some time $t_2 < t_1$ at which j_2 was infectious and j_1 was susceptible. This requires that $\{j_1, j_2\} \in E(G^-)$. And so on.

Proceeding in this manner, we can construct a path (i, j_1, \dots, j_ℓ) in G^- . In an *SIR*- or *SEIR*-model where hosts gain permanent immunity upon cessation of infectiousness, the path must be simple; that is, all nodes in it must be distinct. Thus we must eventually run out of nodes and the construction must terminate. This can only happen when we reach a node j_ℓ that is among the initially infectious ones.

The case of *SIS*-models is a bit more subtle, as our description of the construction no longer implies that the path is simple. Roughly speaking, in this case we need to use the fact that $t_1 > t_2 > \dots$, so that we eventually must reach a time t_ℓ when only the index cases are infectious.

The existence of a path in G^- from i to an initially infectious node $j^* = j_\ell$ implies part (a) of the theorem.

- (b) Suppose that $P = (j^*, i_1, i_2, \dots, i_\ell = i)$ is a path in G^- from some initially infectious node j^* to i . Under the assumptions of *SI*- or *SEI*-models, node j^* will never cease to be infectious, and an effective contact between hosts j^* and i_1 will eventually occur. Thus, host i_1 will eventually become infectious, and will subsequently have an effective contact with host i_2 . And so on. Thus, host i will eventually become infectious.

Note that we could write *will* in the previous paragraph only because we had assumed that hosts never cease to be infectious. In other types of models, such as *SIR*, a host may recover or die before transmitting the infection to the next host along the path P , so that there will be only a positive probability, but no certainty, that host i will experience infection. \square

Exercise 9.69.

- (a) Repeat for **num-nodes** = 15, 21, 33.
- (b) Conjecture formulas for Q_k , q_k , and $\langle k \rangle$ for rectangular grids $G_{RR}(3, n)$.
- (c) Prove your conjectures of point (b).
- (d) Find $\lim_{n \rightarrow \infty} \langle k \rangle$ for the networks $G_{RR}(3, n)$.

Sample Solution:

- (a) For **num-nodes** = 12, 15, 21, 33, the mean degrees are 2.8333, 2.93333, 3.04762, 3.15152, respectively.
- (b) The mean degree in $G_{RR}(3, n)$ is given by $\langle k \rangle = \frac{10}{3} - \frac{2}{n}$.
- (c) There are 4 nodes of degree 2 (the corners), $2n - 2$ nodes of degree 4 (the nodes in the middle), and the remaining $2n - 2$ nodes have degree 4. Thus,

$$\langle k \rangle = \frac{8 + 4(n - 2) + 3(2n - 2)}{3n} = \frac{8 + 4n - 8 + 6n - 6}{3n} = \frac{10n - 6}{3n} = \frac{10}{3} - \frac{2}{n}.$$

 \square **Exercise 9.70.**

- (a) Explore the behavior of $\langle k \rangle$ for $G_{NN}^2(N^2, d)$ when $d = 1$ is fixed and $N \rightarrow \infty$. *Hint:* Don't try exploring networks of size larger than 1,000, as **Metrics** will take too long to complete its calculations.
- (b) Conjecture a formula for $\lim_{N \rightarrow \infty} \langle k \rangle$ for the networks $G_{NN}^2(N^2, 1)$.
- (c) Prove your conjecture of part (b).
- (d) Do the results that you observed in part (a) closely match your result for part (c)?

Sample Solution:

- (a) For $N = 10, 20, 30$ the mean degrees are $\langle k \rangle = 3.6, 3.8, 3.86667$, respectively.
- (b) $\lim_{N \rightarrow \infty} \langle k \rangle = 4$.
- (c) For $N \geq 3$, the network $G_{NN}^2(N^2, 1)$ contains 4 nodes with degree 2 (the corners), $4N - 8$ nodes with degree 3 (the other nodes on the edges), and $(N - 2)^2$ nodes with degree 4 (the nodes in the interior). In the terminology of Section 9.3.4.1 of the main text, this translates into $Q_2 = 4, Q_3 = 4N - 8, Q_4 = (N - 2)^2$. Because $Q_k = 0$ for all other values of k , we get from Eq. (9.9)

$$\begin{aligned} \lim_{N \rightarrow \infty} \langle k \rangle &= \lim_{N \rightarrow \infty} \frac{2Q_2 + 3Q_3 + 4Q_4}{N^2} = \lim_{N \rightarrow \infty} \frac{8 + 3(4N - 8) + 4(N - 2)^2}{N^2} \\ &= \lim_{N \rightarrow \infty} \frac{8}{N^2} + \lim_{N \rightarrow \infty} \frac{3(4N - 8)}{N^2} + \lim_{N \rightarrow \infty} 4 \left(\frac{N - 2}{N^2} \right)^2 = 0 + 0 + 4. \end{aligned}$$

- (d) For the particular networks we explored, we found values of the mean degree that were between 20% and 12% smaller than the limit we found in part (c). \square

Exercise 9.71. Mathematicians often talk about *finite size effects*. In our context this means that theoretical results that can be proved for large network sizes N do not show up clearly if N is small or moderate. Did you observe this phenomenon in your work on Exercises 9.69 and 9.70? If so, in which direction did finite-size effects distort the theoretical predictions that you derived for $N \rightarrow \infty$?

Sample Solution: For the particular networks that we explored, we always found values of the mean degree that were smaller than the theoretical limit for large network size. The discrepancy was more pronounced for smaller networks. We can attribute these discrepancies to finite size effects. \square

Exercise 9.72.

- (a) Press **Labels** to see how the nodes are numbered. Before starting the simulation, write out the neighborhoods $\mathcal{N}_\ell(j^*)$ and $\mathcal{N}_{\leq \ell}(j^*)$, where j^* is the initially infectious node and $\ell = 1, 2$. Keep these results in mind when you observe the simulation.
- (b) Set the speed control slider to about one-third of the slower range and adjust for comfortable viewing if needed. Press **Go** and observe what happens. Because this is an *SEI*-model, you will need to terminate the simulation by pressing **Go** again.
- (c) Explain how the color changes at each step are related to the neighborhoods that you found in part (a).

Sample Solution:

- (a) In our simulation we got $j^* = 12$. In this case $\mathcal{N}_1(j^*) = \{8, 13, 16\}$, $\mathcal{N}_2(j^*) = \{4, 9, 14, 17\}$, $\mathcal{N}_{\leq 1}(j^*) = \{8, 12, 13, 16\}$, $\mathcal{N}_{\leq 2}(j^*) = \{4, 8, 9, 12, 13, 14, 16, 17\}$.
- (c) The yellow nodes that appear in tick number $2\ell - 2$ comprise $\mathcal{N}_\ell(j^*)$; the red nodes in tick number $2\ell - 1$ comprise $\mathcal{N}_{\leq \ell}(j^*)$. □

Exercise 9.73. Repeat Exercise 9.72(b) for this model. Perform five runs. What is the most salient difference between the observations that you made for these networks and the observations you made in the previous exercise? How are these observations related to Exercise 9.21?

Sample Solution: In $G_{\text{NN}}^1(20, 2)$ all nodes will eventually turn red, which means that all hosts in the population eventually will experience infection. In contrast, not all nodes in the Erdős-Rényi network will eventually turn red, at least not in all simulations. Only the nodes in the connected component of the index case will. These observations complement the theoretical predictions of Exercise 9.21(c) for connected graphs. □

Exercise 9.74.

- (a) How many secondary infections would you *expect* the index case to cause?
- (b) Use **Last** to run about 10 simulations with this initial state, record the number of secondary infections for each simulation, and compute the mean. Do your results roughly conform with the expectation you formed in point (a)?

Sample Solution:

- (a) The mean number of infections caused by the index case should be 5.
- (b) The number of secondary infections is equal to the number of gray leaves of the tree. In our simulations the actual numbers ranged from 3 to 7, with a mean of 4.8. □

Exercise 9.75.

- (a) How would you need to modify your explorations to get a correct rough estimate for the model that we are currently exploring?
- (b) Perform 20 runs of the relevant simulations and compare your results with those of Exercise 9.74.

Sample Solution:

- (a) Change the following parameter settings:

set-state-by → **Number of nodes**

num/frac: 1

Then create initial states by first pressing **Reset**, followed by **Set**.

Now you need to pay careful attention to the structure of the graph. If the index case happens to be the root, then the number of secondary infections is still equal to the number of gray leaves of the tree. If the index case is a leaf, then this number is 1 if the root turned gray and 0 otherwise.

- (b) In our simulations the number of secondary infections ranged from 0 to 7, with a mean of 0.95. □

Exercise 9.76.

- (a) Before looking at the output, ask yourself what you would expect in terms of the mean number of removed hosts at the end and the variability between runs.

- (b) Now look at your output. Sort by the column that represents your output variable. Find the minimum, maximum, and mean of the column.
- (c) Do your data match the expectations that you articulated in point (a)?

Sample Solution:

- (a) One would expect that on average about 51 hosts (the index case plus 50 other hosts) experience infection. Because the degree of node 0 is rather large, the distribution should be close to normal, with a standard deviation of about 5.
- (b) We found a minimum of 37, a maximum of 60, and a mean of 50.67.
- (c) This conforms rather well to our expectations. □

Exercise 9.77.

- (a) Set up and run a batch processing experiment as for the previous exercise. What would you expect in terms of the output column?
- (b) Sort the output file by the column that represents your output variable. Find the minimum, maximum, and mean.
- (c) Do your data match the expectations that you articulated in point (a)? How do they compare with the data you collected for the next-generation model?

Sample Solution: We found a minimum of 1, a maximum of 101, and a mean of 49.77. The results seem to be evenly distributed in the interval from 1 to 101. The mean is very similar to the one we found in the previous exercise, but the variability is much larger than for the outcomes of the simulations for the next-generation model. □

Exercise 9.78. Derive Eq. (9.16).

Sample Solution: The probability that a successful transmission from host j^* to a given adjacent host i occurs before host j^* gets removed is

$$p = \int_0^{\infty} \beta e^{-\beta t} \int_t^{\infty} \alpha e^{-\alpha \tau} d\tau dt = \int_0^{\infty} \beta e^{-\beta t} e^{-\alpha t} dt = \frac{\beta}{\alpha + \beta}.$$

Exercise 9.79. Derive Eq. (9.18). □

Sample Solution: Consider hosts j^* and i in a network with $N = 2$ nodes; assume that j^* is infectious at time $t = 0$ and host i is susceptible. The probability that j^* infects host i by time $t = 1$ is b . The probability that none of these hosts has changed the state by time 1 is $(1 - a)(1 - b)$, and the conditional probability of an effective contact by time 2 given that the state is still the same by time 1 is again b . And so on. Thus for $a < 1$ and $b < 1$ we get

$$p = \sum_{t=0}^{\infty} (1 - a)^t (1 - b)^t b = b \sum_{t=0}^{\infty} (1 - a)^t (1 - b)^t = \frac{b}{1 - (1 - a)(1 - b)} = \frac{b}{a + b - ab}.$$

Note that this formula also applies to the special cases when $a = 1$ or $b = 1$ where we get $p = b$, although $(1 - a)^0(1 - b)^0$ is undefined. □

Exercise 9.80.

- (a) Set up and run a batch processing experiment as for Exercise 9.76. What would you expect in terms of the output column?
- (b) Sort the output file by the column that represents your output variable. Find the minimum, maximum, mean, and standard deviation.
- (c) Do your data match the expectations that you articulated in point (a)? How do your data compare with those that you found for Exercises 9.76 and 9.77?

Sample Solution: We found a minimum of 26, a maximum of 98, and a mean of 51.93. The mean was similar to the means in the previous exercises. We observed as much variability on the high end as for the continuous-time model, but not as much variability on the low end. \square

Exercise 9.81. Try to find a plausible explanation for the differences in variability that you observed between Exercises 9.76, 9.77 and 9.80. Be sure to look up our sample solution and compare notes.

Sample Solution: In all three models, the occurrence of at least one effective contact between the index case 0 and adjacent hosts $1, \dots, 100$ over a *given time interval* are independent events. The relevant time interval is the period of infectiousness of the index case.

In next-generation models, the length of this interval is fixed; it corresponds to exactly one time step. Thus the Central Limit Theorem (CLT) applies and we should see relatively little variability in the number of secondary infections.

In continuous-time models, in most cases this interval is rather short, so that the index case will cause only very few secondary infections. Recall what you saw when looking at the prevalence plots that illustrated the variability in removal times: Very short times to removal are the most common ones, but there will occasionally be a host for whom this interval is very long. Such a host will cause many more secondary infections than an “average” host. In the literature such hosts are sometimes called *superspreaders*. This explains the huge variability that we observed for the continuous-time model.

General discrete-time models fall somewhere in between these extremes. Occasionally, the index case may have an unusually long period of infectiousness and become a superspreader. Thus the variability at the high end will be similar to the one we observe in continuous-time models. But now very short durations of infectiousness are ruled out; the length of the time step Δt is a lower bound in discrete-time models. Thus we should not expect to observe as much variability at the low end as in continuous-time models. \square

Exercise 9.82. Did you get a higher or a lower value than the one you found using **Metrics**? If you noticed a discrepancy, how would you explain it?

Sample Solution: Most likely, you got an overestimate by a large margin. The reason is that by the time the index case gets removed, most likely some other hosts will already have caused secondary infections. \square

Exercise 9.83.

- (a) Explain why, in general, the right-hand sides of Eq. (9.17) and the second line of Eq. (9.18) are only approximations. Would they give overestimates or underestimates?
- (b) Show that the estimate (9.18) gives the correct value of R_0 in next-generation models.
- (c) Show that the estimates give the correct value for R_0 if the contact network is a tree or forest, regardless of how we model time.
- (d) Show that these estimates become arbitrarily good when the contact network is K_N and $N \rightarrow \infty$.

Sample Solution:

- (a) In our derivation of the formulas for p (and hence R_0), we had tacitly assumed that the host i at the other end of the edge $\{j^*, i\}$ will stay susceptible until either an effective contact with host j^* or removal of host j^* occurs. This will not always be the case though, as in general host i can be infected earlier from another host in generation 1 or higher. Our derivation of p is correct, but it was defined as the probability that an *effective* contact between host j^* and an adjacent host i occurs, while the definition of R_0 requires a *successful* contact

Let p^- be the probability of a successful contact between the index case and a given host i who is initially susceptible. Then $p^- \leq p$, but the inequality may be strict if host i can become infectious via a different path of transmissions.

In Eq. (9.15) we should really have used p^- instead of p . Our approximations of R_0 will be overestimates if the inequality $p^- \leq p$ is strict. However, as part (c) of this exercise shows, in the context of a star tree where we first encountered (9.15) we must have $p = p^-$ so that Eq. (9.15) gives the correct value of R_0 in this case.

- (b) In next-generation models, secondary infections can be caused by hosts other than the index case only after the index case has already been removed at time step 1. Thus the problem outlined in part (a) cannot occur in these models.
- (c) Host i can become infected only if the infection travels along a path in the contact network from host j^* to host i . In trees and forests there are no cycles, so that every such path must contain the edge $\{j^*, i\}$.
- (d) In large populations with uniform mixing, the probability that the index case will infect any given host i is very small. In particular, the probability that j^* will make contact, prior to time $T_{j^*}^R$, with a host who has already been infected via a different path, will be very small. In the limit when $N \rightarrow \infty$ this probability becomes negligible. \square

Exercise 9.84. Do your observations confirm the predictions about the degree distribution that we made in Section 9.3.5?

Sample Solution: The values of the mean degree that you see in the **Command Center** should be close to λ , and the histogram of the degree distribution in the **Network Metrics** plot should show a distinct peak near λ . For large values of N and $\lambda = \frac{N}{2}$ you should not see any nodes with degrees near 0 or N . \square

Exercise 9.85. What do you observe? Do you get connected components with a range of different sizes? If the component is large, is it always the same one? How can you tell from the plot?

Sample Solution: About 42% of the time, the connected component of the index case will be very small. About 58% of the time, the connected component will contain about 58% of all nodes. Your number will most likely be different from 58%, but almost certainly be larger than 50%. Because connected components are pairwise disjoint, there can only be one connected component of this size. \square

Exercise 9.86. In what respect are the results similar to the ones of the previous exercise? In what respect are they different?

Sample Solution: The general pattern will be the same, but now the sizes of the largest component differ slightly for different networks. In all cases they should be close to 58% of all nodes. This makes sense, because now you are looking at components in different graphs, while previously you looked at the same component in the same network. \square

Exercise 9.87.

- (a) Analyze the data that you collected in the above experiments. Sort the output columns from smallest to largest values. Classify observed outbreaks into minor and major ones. Record the numbers of minor and major outbreaks and also the number of runs where no secondary infections whatsoever occurred. Record the maximum size of minor outbreaks and both the maximum and minimum sizes of major ones. Compute mean numbers of hosts who experienced infection for minor and major outbreaks separately.
- (b) Are the results for the 4 models similar enough that we could attribute all differences to random fluctuations?

Sample Solution:

- (a) In the continuous-time simulations for the Erdős-Rényi networks, we observed 38 runs with 10 or fewer hosts who had experienced infection, and classified those runs as minor. We also observed 61 runs with at least 61 removed hosts; these definitely indicated major outbreaks. In one run we observed 26 removed hosts, and it was not immediately clear whether to classify this as a major or minor outbreak; we went with “major.”

There were 26 runs with only one removed host, which translates into no secondary infections whatsoever. The mean number of removed hosts in all outbreaks that we classified as minor was 2.34.

The largest outbreak that we observed affected 90 hosts. The mean number of removed hosts in all outbreaks that we classified as major was 78.47.

In the continuous-time simulations for complete graphs, we observed 47 runs with 11 or fewer hosts who had experienced infection, and classified those runs as minor. We also observed 53 runs with at least 43 removed hosts; these were classified as major outbreaks.

There were 27 runs with only one removed host, which translates into no secondary infections whatsoever. The mean number of removed hosts in all outbreaks that we classified as minor was 2.74.

The largest outbreak that we observed affected 94 hosts. The mean number of removed hosts in all outbreaks that we classified as major was 79.62.

In the simulations of next-generation models with a complete graph, we observed 22 runs with 5 or fewer hosts who had experienced infection, and classified those runs as minor. We also observed 78 runs with at least 60 removed hosts; these were classified as major.

There were 16 runs with only one removed host, which translates into no secondary infections whatsoever. The mean number of removed hosts in all outbreaks that we classified as minor was 1.5.

The largest outbreak that we observed affected 94 hosts. The mean number of removed hosts in all outbreaks that we classified as major was 80.72.

In the simulations of next-generation models with Erdős-Rényi networks we observed 17 runs with 6 or fewer hosts who had experienced infection and classified those runs as minor. We also observed 83 runs with at least 61 removed hosts; these were classified as major.

There were 11 runs with only one removed host, which translates into no secondary infectious whatsoever. The mean number of removed hosts in all outbreaks that we classified as minor was 1.76.

The largest outbreak that we observed affected 94 hosts. The mean number of removed hosts in all outbreaks that we classified as major was 79.86.

- (b) There was no significant difference in the mean final sizes of the major outbreak. Between the two next-generation models there were also no significant differences in the observed proportions of outbreaks that were classified as minor.

The differences between the next-generation and continuous-time models with regard to the proportions of observed minor outbreaks are too large to be entirely attributed to statistical fluctuations alone.

For the two continuous-time batches, more minor outbreaks were observed in the simulations with complete graphs. It is not immediately clear whether the latter might be entirely due to statistical fluctuations. \square

REFERENCES

- [1] Just W, Callender H, LaMar MD. Online Appendix: Disease transmission dynamics on networks: Network structure vs. disease dynamics. In: Robeva, R, editor. Algebraic and discrete mathematical methods for modern biology. New York: Academic Press; 2015. Available from http://textbooks.elsevier.com/web/product_details.aspx?isbn=9780128012130.
- [2] Centers for Disease Control and Prevention. 10 ways STDs impact women differently from men <http://www.cdc.gov/nchhstp/newsroom/docs/STDs-Women-042011.pdf>. Accessed: 2014-09-18.
- [3] Just W, Callender H, LaMar MD. Exploring transmission of infectious diseases on networks with NetLogo Available from <https://qubeshub.org/iontw> and also <http://www.ohio.edu/people/just/IONTW/>.
- [4] Just W, Callender H, LaMar MD, Toporikova N. Online Appendix: Transmission of infectious diseases: Data, models, and simulations. In: Robeva, R, editor. Algebraic and discrete mathematical methods for modern biology. New York: Academic Press; 2015. Available from http://textbooks.elsevier.com/web/product_details.aspx?isbn=9780128012130.

