Supplemental Material for Network Analysis of SFU Course Registrations

William Ruth[1] and Richard Lockhart[2]

[1]Corresponding Author - Department of Statistics and Actuarial Science , Simon Fraser University , Burnaby, BC Canada , wruth@sfu.ca
[2]Department of Statistics and Actuarial Science , Simon Fraser University , Burnaby, BC Canada

1 Simulation Output

In Ruth and Lockhart [2021], we present only a sample of the output generated by our simulations. Here we give a comprehensive presentation of our output. Section 1.1 gives the proportions of simulations where the infection develops into a full outbreak, where an outbreak is defined as an infection that affects at least 10 people. These proportions are given for all parameter settings and all terms. Section 1.2 gives sample infection trajectories for all parameter combinations and all terms. Trajectories are only given for simulations where the infection develops into a full outbreak, and at most 5 trajectories are given in each plot (for the sake of readability). Section 1.3 gives cumulative plots for the trajectories shown in Section 1.2. Section 1.4 gives average trajectories with ± 1 error bars for all parameter combinations and all terms. Note that averages are taken only over infections that develop into outbreaks.

	Threshold	50%	75%	90%	95%	100%
High Infectiousness	Spring 2019	3%	11%	29%	29%	48%
	Fall 2019	1%	4%	25%	27%	49%
	Spring 2020	2%	13%	15%	36%	57%
Medium Infectiousness	Spring 2019	0%	3%	12%	26%	36%
	Fall 2019	1%	5%	8%	17%	41%
	Spring 2020	0%	3%	10%	21%	44%
Low Infectiousness	Spring 2019	0%	0%	6%	2%	19%
	Fall 2019	0%	0%	0%	5%	28%
	Spring 2020	0%	0%	0%	6%	22%

Table 1: Proportions of 100 simulation runs where the infection develops into an active outbreak (i.e. infects at least 10 people) for various limits on the proportion of classes allowed to meet in person and various infectiousness regimes.

1.1 Develop Rates

See Table 1 for develop rates under all simulation regimes. Recall that 100 replicates were conducted at each parameter combination, and an infection is said to have developed if at least 10 total people are affected by the infection. See Ruth and Lockhart [2021] for details of the simulation setup.

1.2 Infection Trajectories

For each term, we present a sequence of figures. In each plot, we show a sample of trajectories for the number of infected students over the course of the term. For the sake of readability, each plot contains at most 5 such trajectories, with fewer if the number of developed infections is insufficient. The first figure in each term shows trajectories when all classes meet in-person under all of the three infectiousness regimes. Next we present a separate figure for each infectiousness level, with each figure containing trajectories for 50%, 75%, 90%, and 95% of classes meeting in-person. We then repeat this sequence of four figures for each term.

Figures 1-4 give the trajectories for Spring 2019. Figures 5-8 give the trajectories for Fall 2019. Figures 9-12 give the trajectories for Spring 2020.

1.3 Cumulative Trajectories

This section follows the same structure as Section 1.2, but here we give cumulative trajectories (i.e. the number of students who have been infected up to the specified day).

Figures 13-16 give the trajectories for Spring 2019. Figures 17-20 give the trajectories for Fall 2019. Figures 21-24 give the trajectories for Spring 2020.

1.4 Infection Trajectory Summaries

In this section, we give mean trajectories for each of the simulation regimes discussed in Sections 1.2 and 1.3. In each setting, we average only over developed infections. We also include ± 1 standard error bands. We follow the same structure as the previous sections.

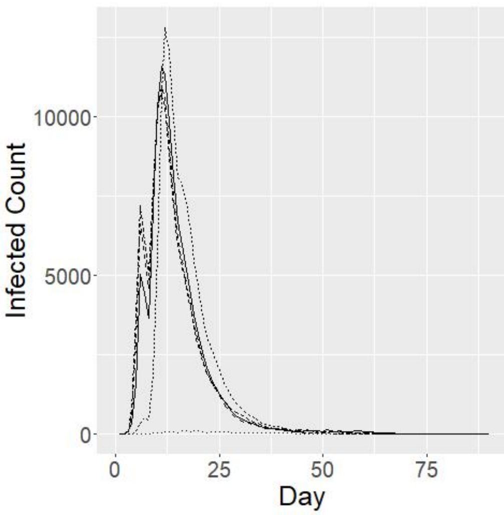
Figures 25-28 give the trajectories for Spring 2019. Figures 29-32 give the trajectories for Fall 2019. Figures 33-36 give the trajectories for Spring 2020.

2 Network Summaries

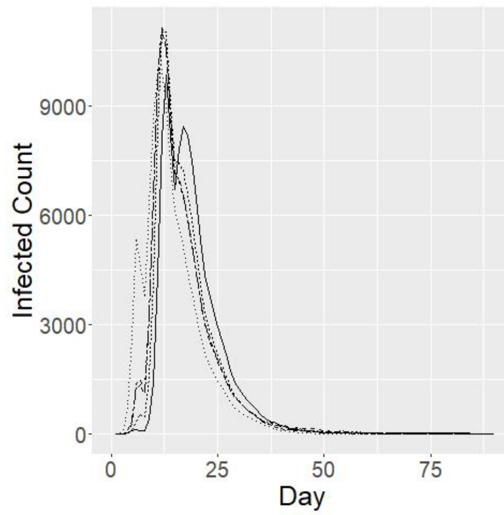
In Ruth and Lockhart [2021], we retain only those network summaries computed by Weeden and Cornwell [2020] which are most relevant to our simulation study. Here we present some other summary statistics which are helpful for understanding the network structure, but do not relate directly to how propagation occurs in our simulation.

2.1 The 2-Mode Network

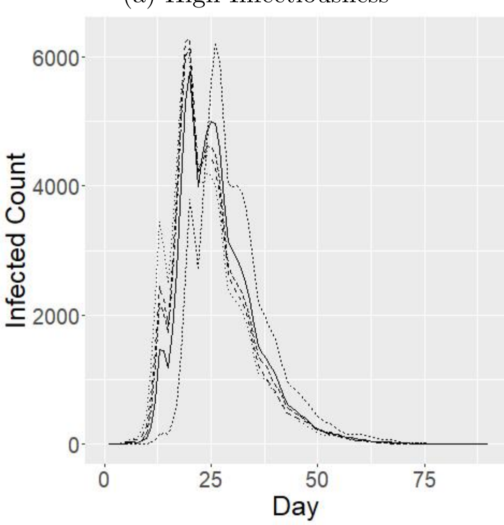
See Ruth and Lockhart [2021] for various definitions which are relevant to our network analysis. Here, we discuss only those quantities which are not defined therein.



(a) High Infectiousness

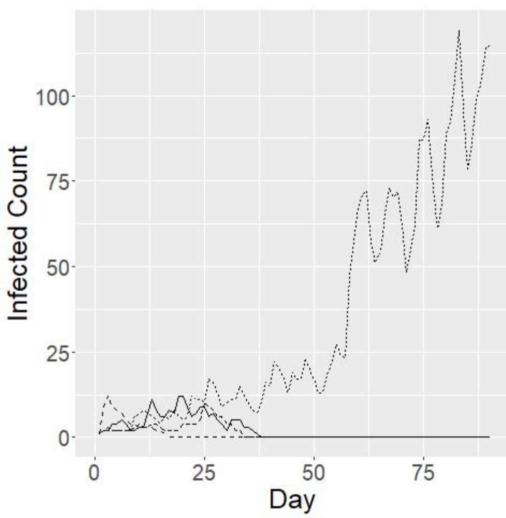


(b) Medium Infectiousness

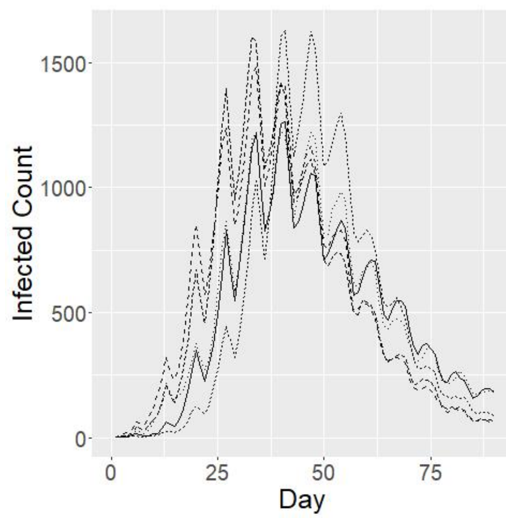


(c) Low Infectiousness

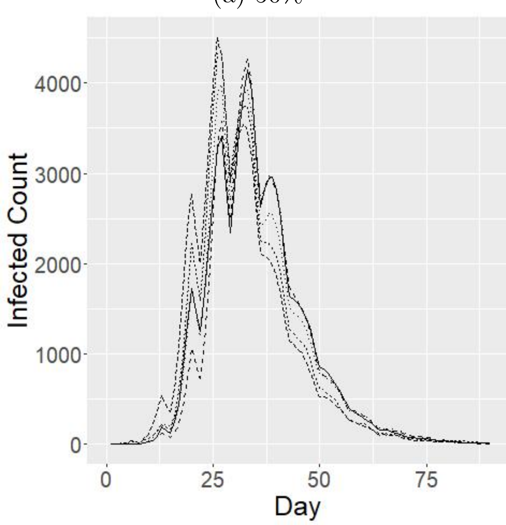
Figure 1: Sample trajectories when all classes are held in-person for Spring 2019.



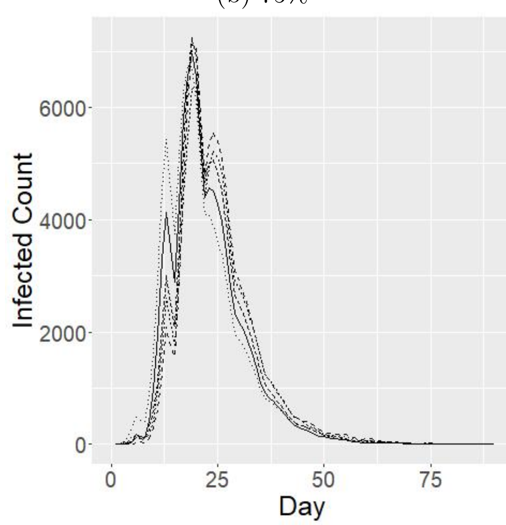
(a) 50%



(b) 75%

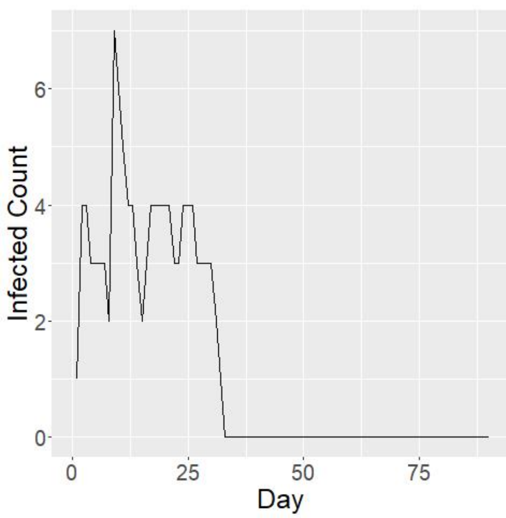


(c) 90%

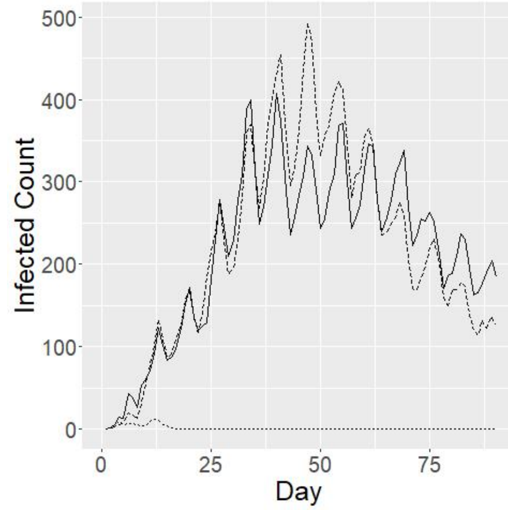


(d) 95%

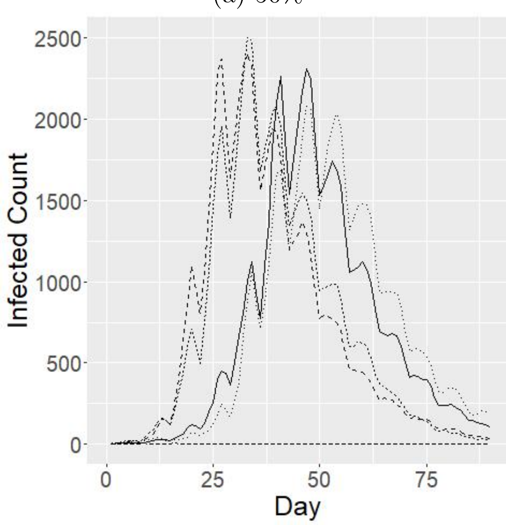
Figure 2: Sample trajectories from the high-infectiousness regime in Spring 2019.



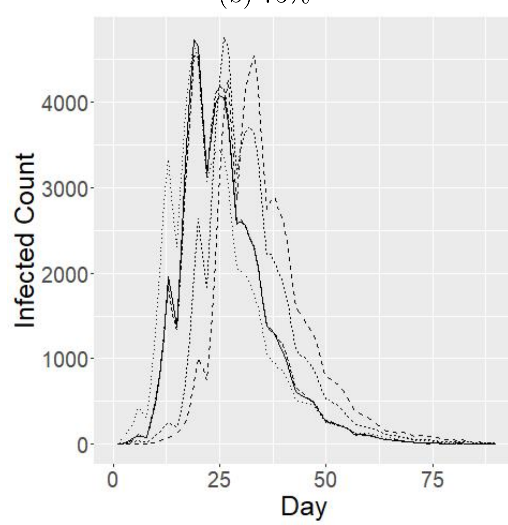
(a) 50%



(b) 75%

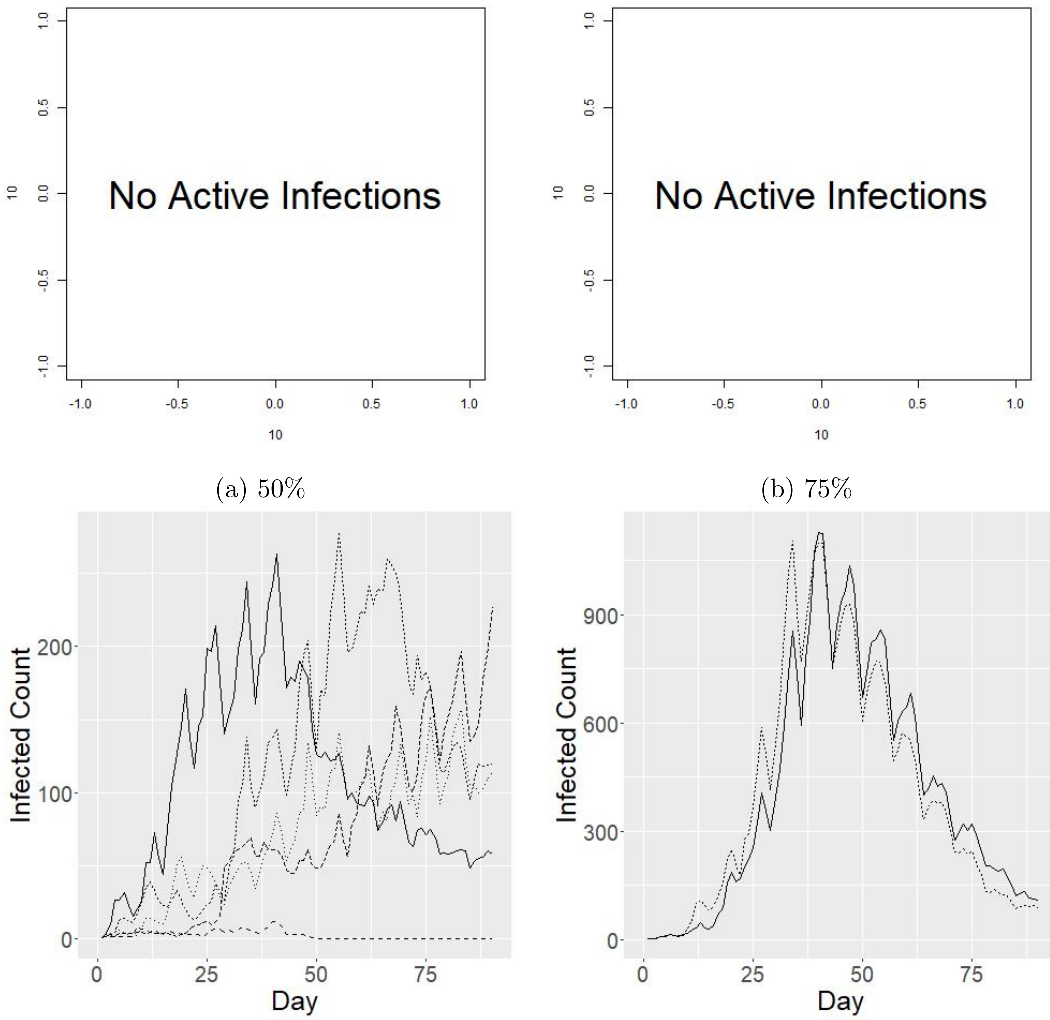


(c) 90%



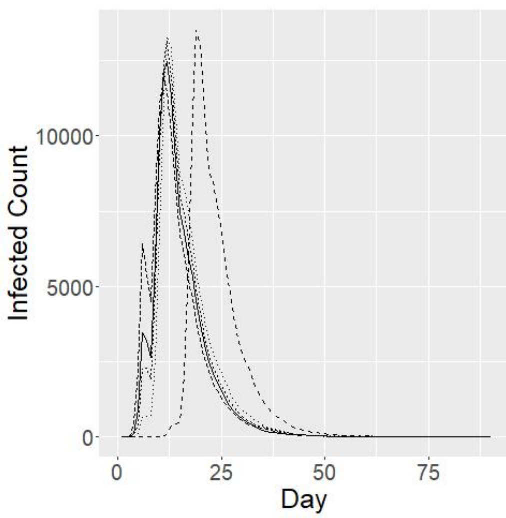
(d) 95%

Figure 3: Sample trajectories from the medium-infectiousness regime in Spring 2019.

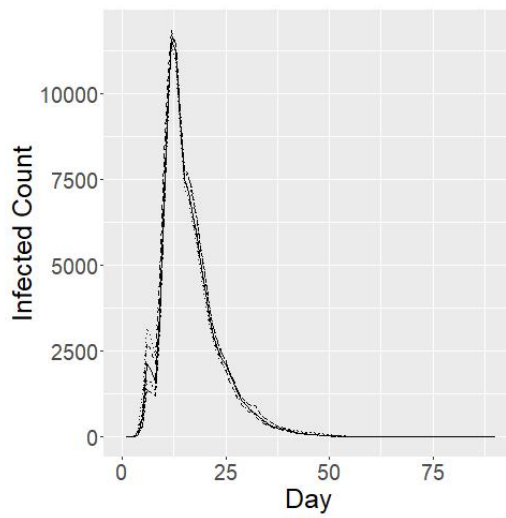


(a) 50%
(b) 75%
(c) 90%
(d) 95%

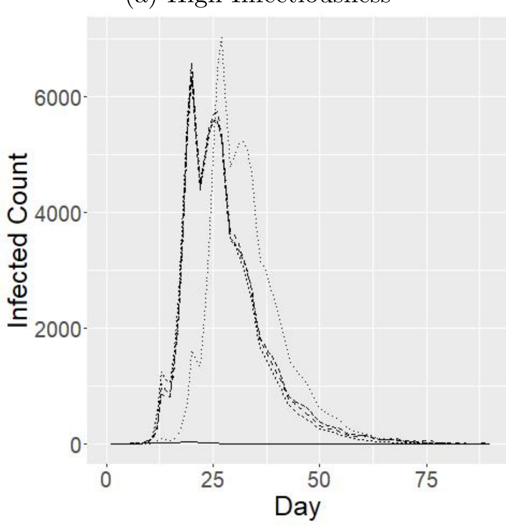
Figure 4: Sample trajectories from the low-infectiousness regime in Spring 2019.



(a) High Infectiousness

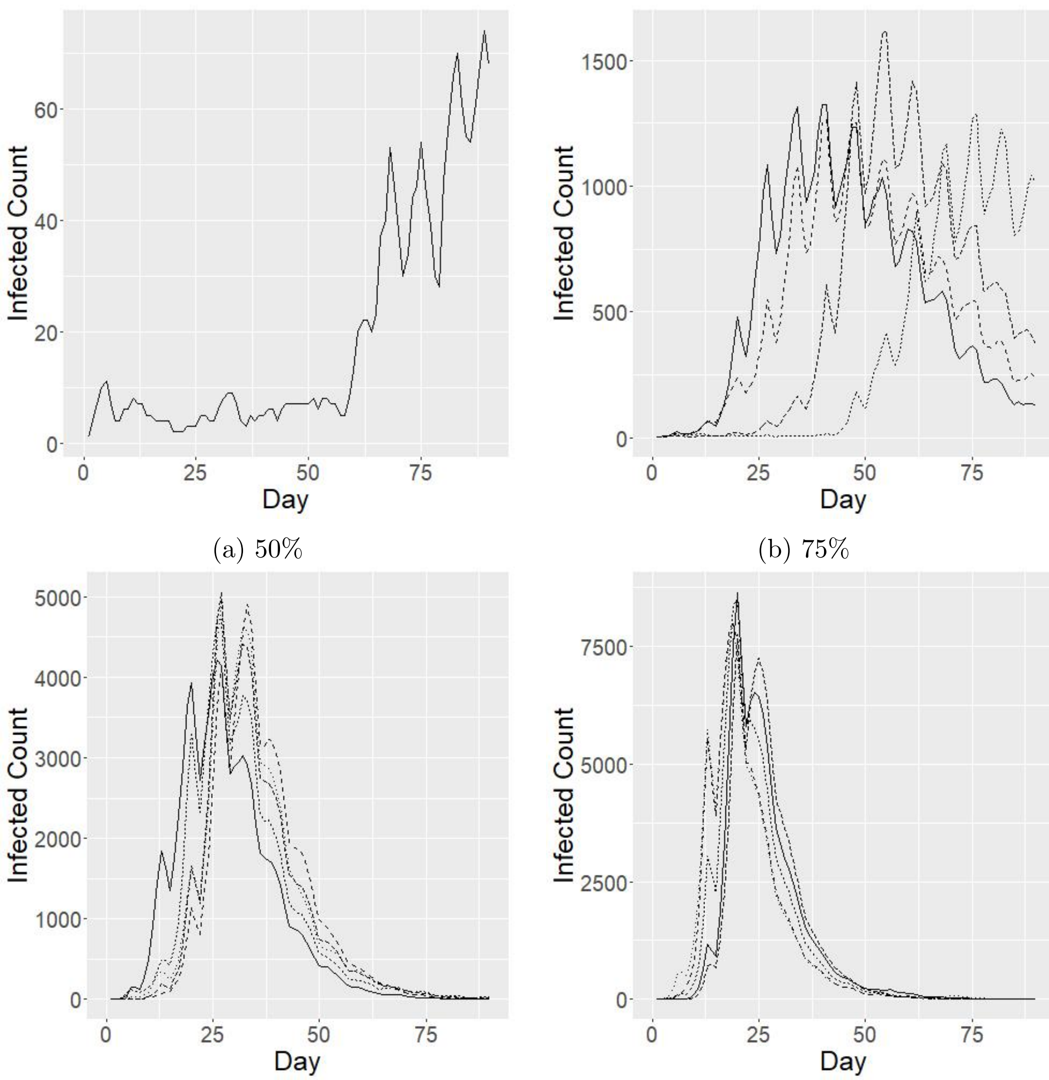


(b) Medium Infectiousness



(c) Low Infectiousness

Figure 5: Sample trajectories when all classes are held in-person for Fall 2019.



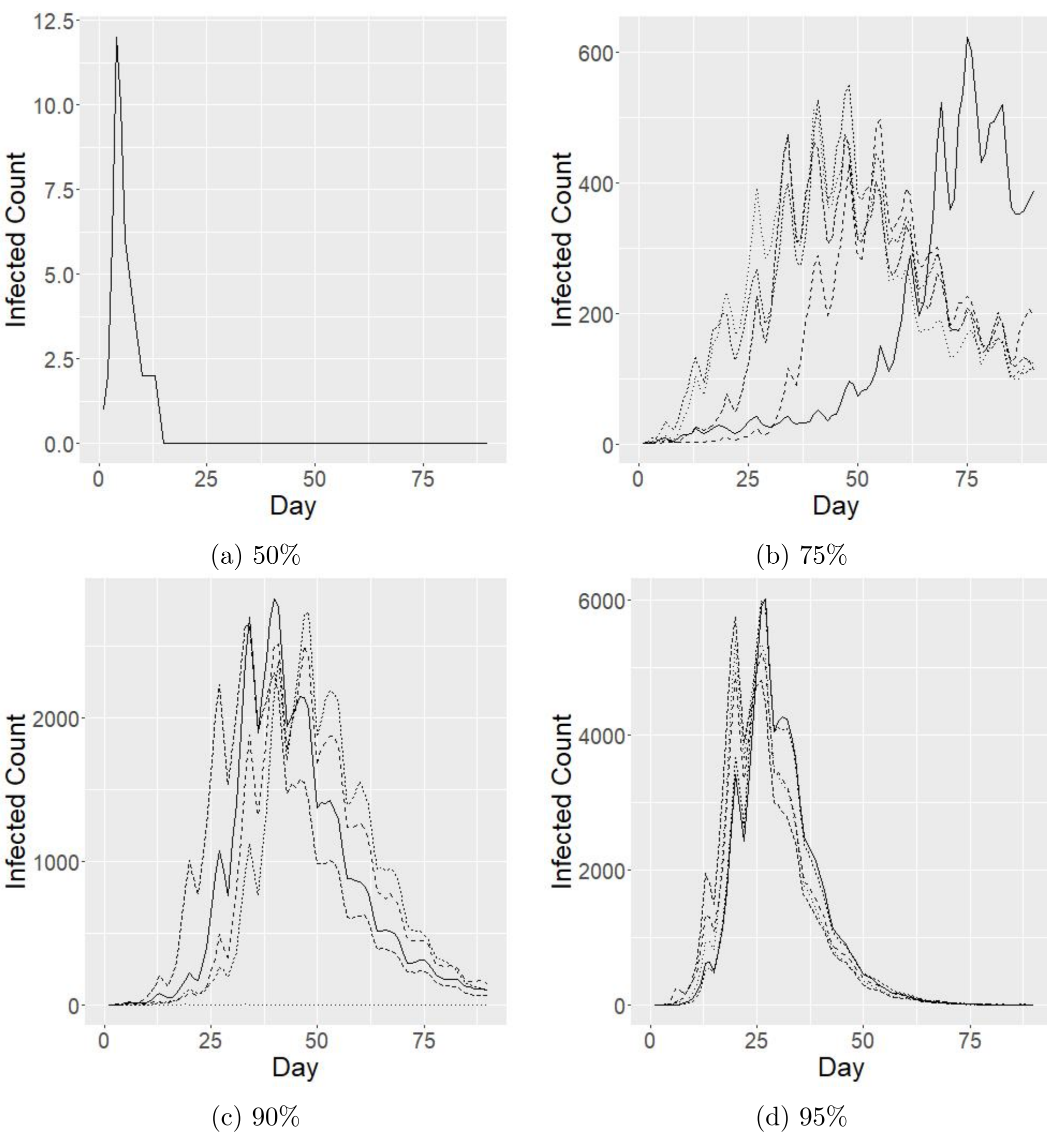
(a) 50%

(b) 75%

(c) 90%

(d) 95%

Figure 6: Sample trajectories from the high-infectiousness regime in Fall 2019.



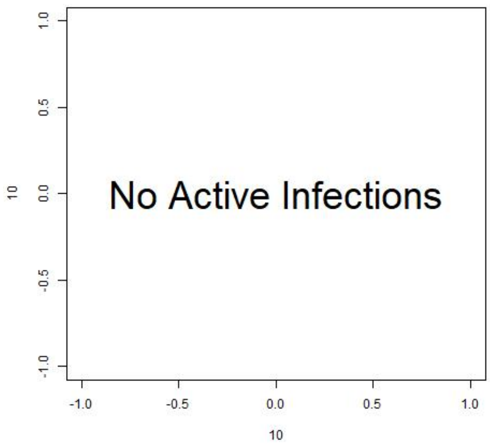
(a) 50%

(b) 75%

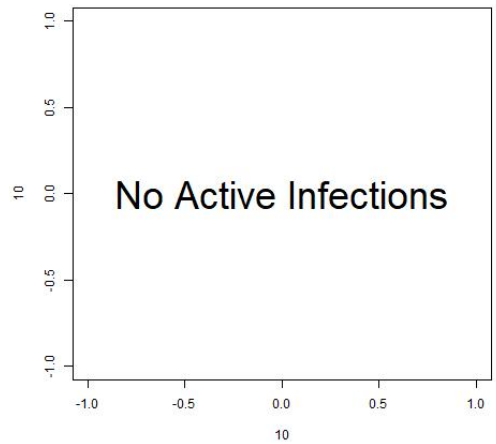
(c) 90%

(d) 95%

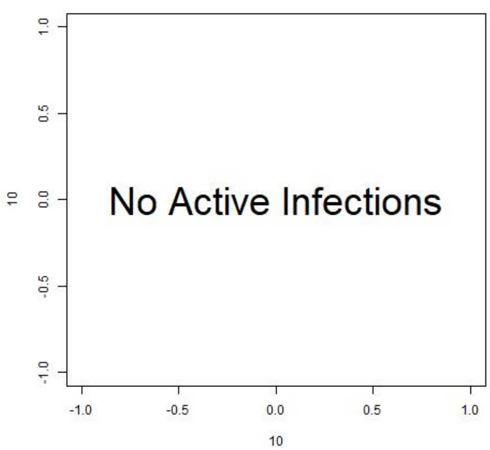
Figure 7: Sample trajectories from the medium-infectiousness regime in Fall 2019.



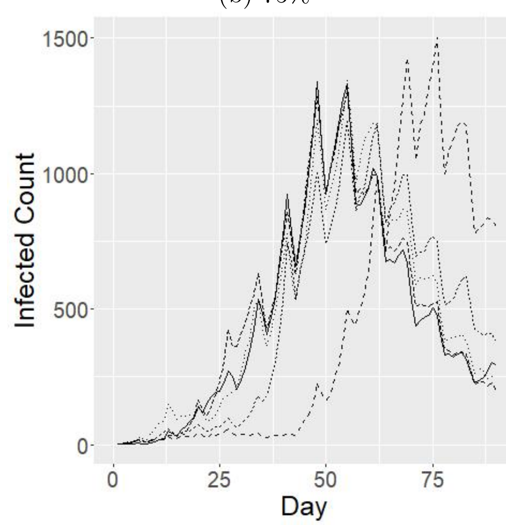
(a) 50%



(b) 75%

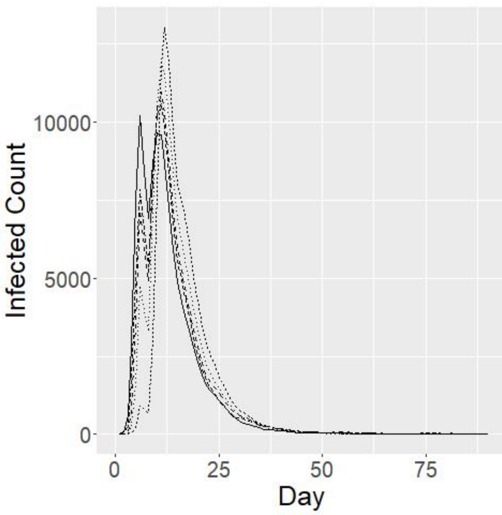


(c) 90%

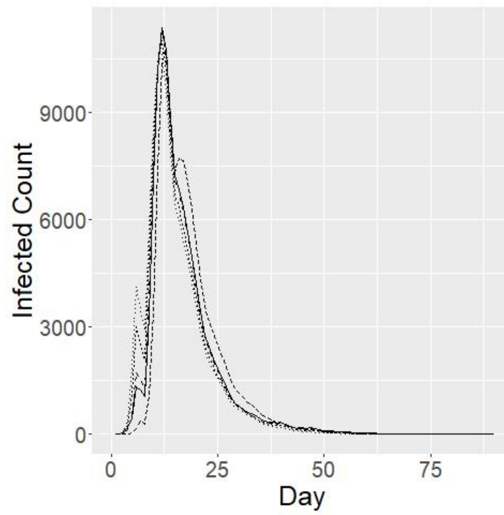


(d) 95%

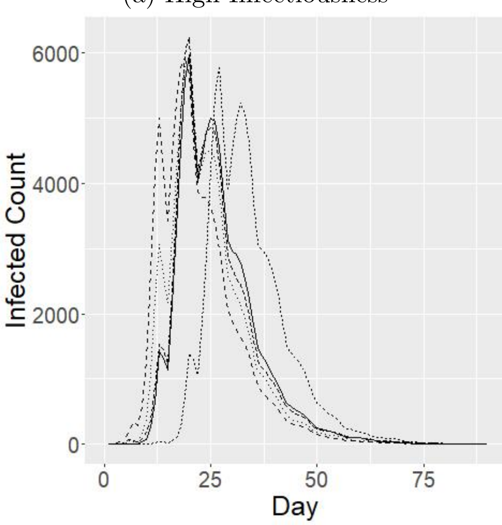
Figure 8: Sample trajectories from the low-infectiousness regime in Fall 2019.



(a) High Infectiousness

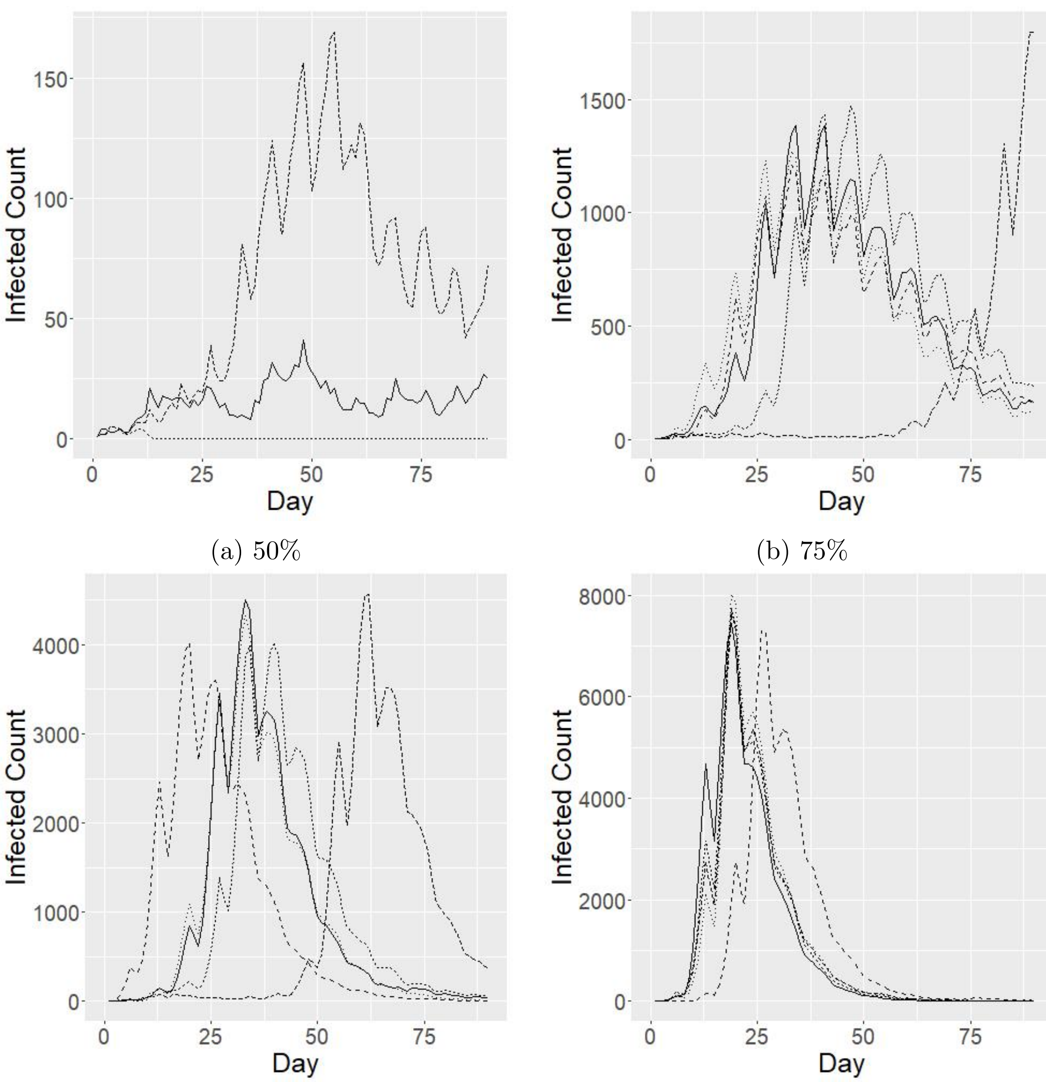


(b) Medium Infectiousness



(c) Low Infectiousness

Figure 9: Sample trajectories when all classes are held in-person for Spring 2020.



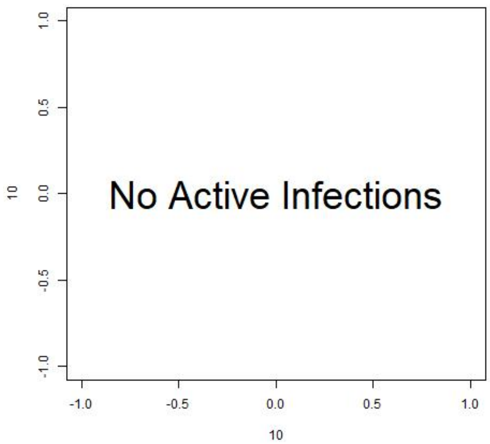
(a) 50%

(b) 75%

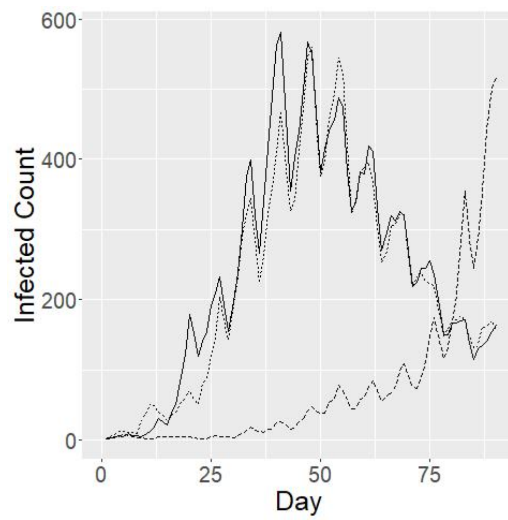
(c) 90%

(d) 95%

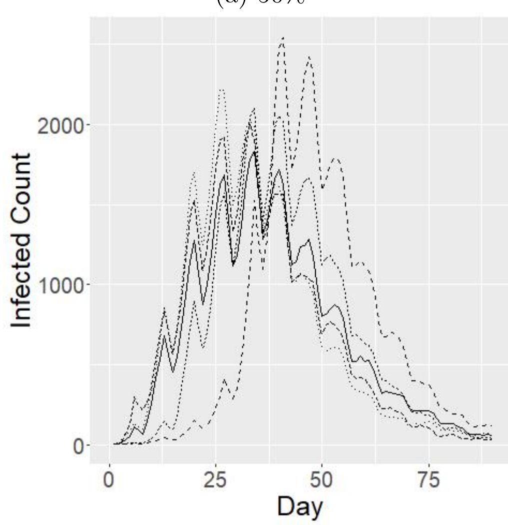
Figure 10: Sample trajectories from the high-infectiousness regime in Spring 2020.



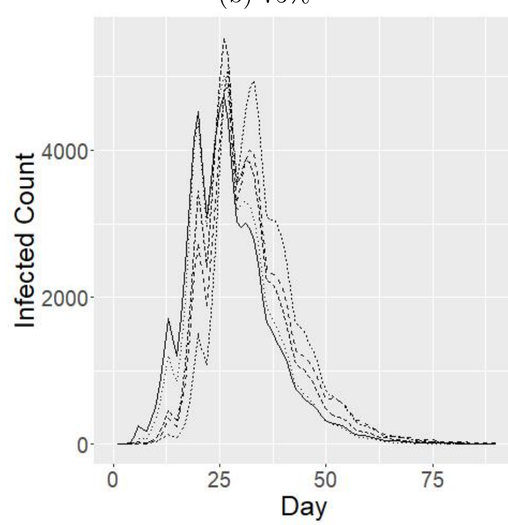
(a) 50%



(b) 75%

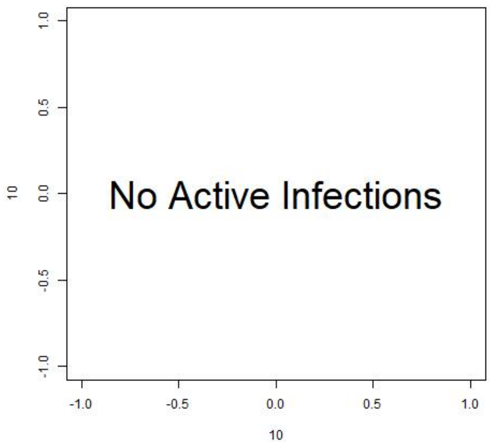


(c) 90%

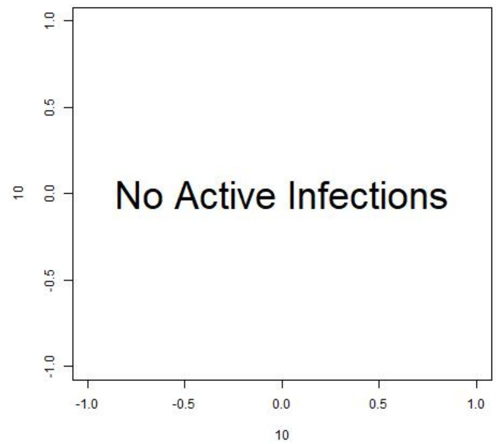


(d) 95%

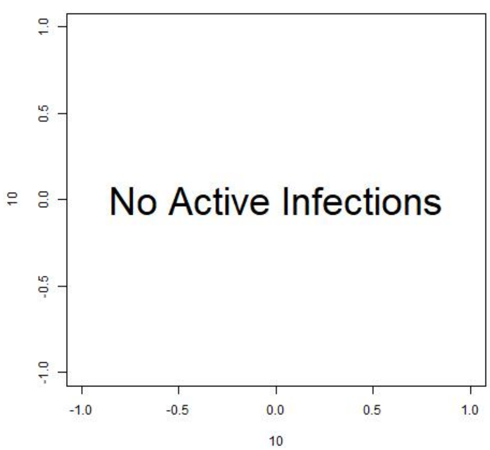
Figure 11: Sample trajectories from the medium-infectiousness regime in Spring 2020.



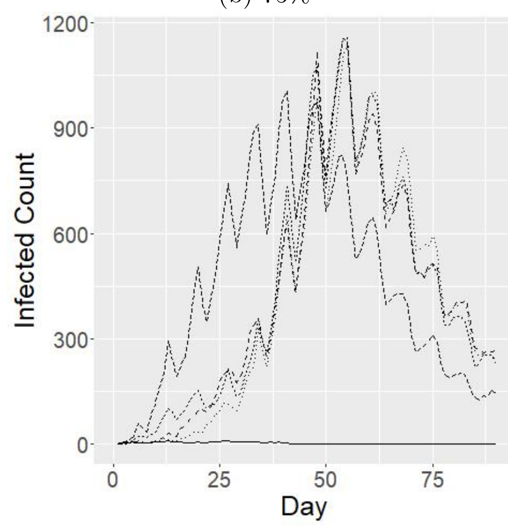
(a) 50%



(b) 75%

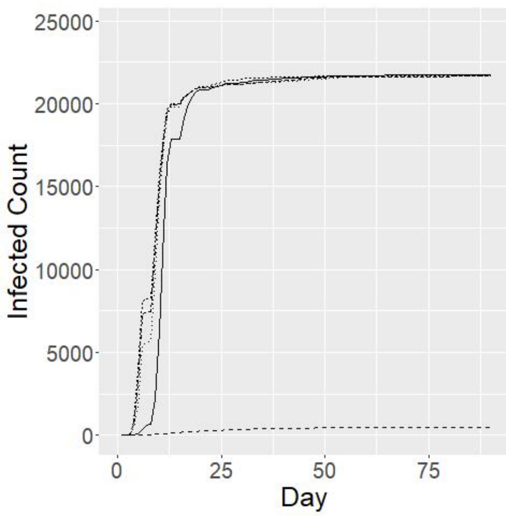


(c) 90%

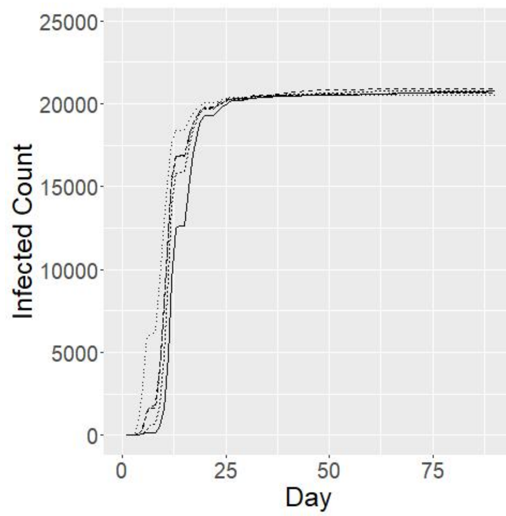


(d) 95%

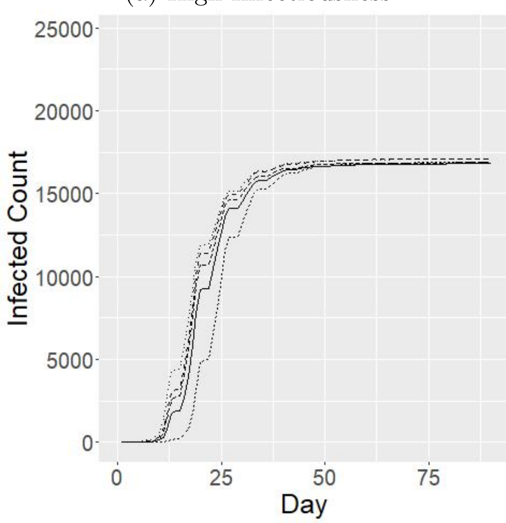
Figure 12: Sample trajectories from the low-infectiousness regime in Spring 2020.



(a) High Infectiousness

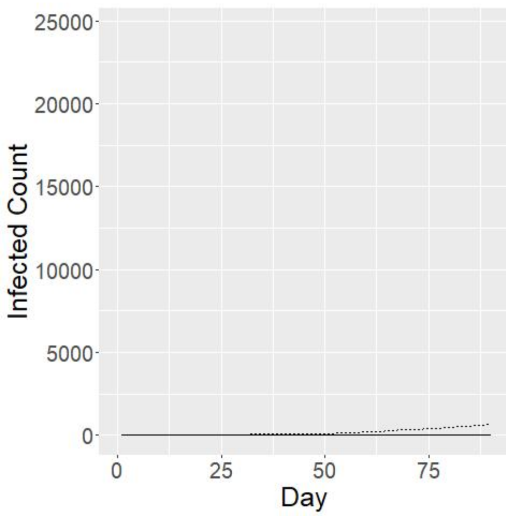


(b) Medium Infectiousness

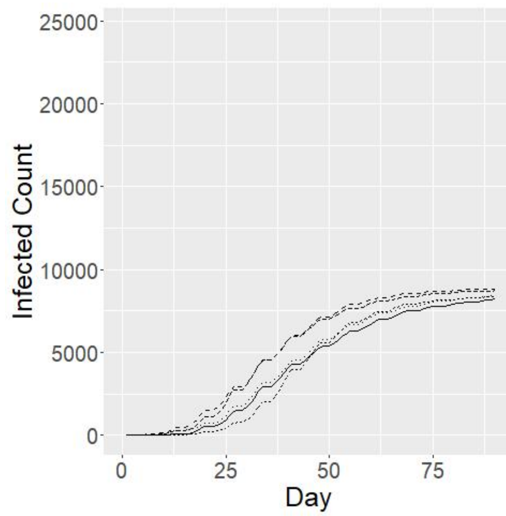


(c) Low Infectiousness

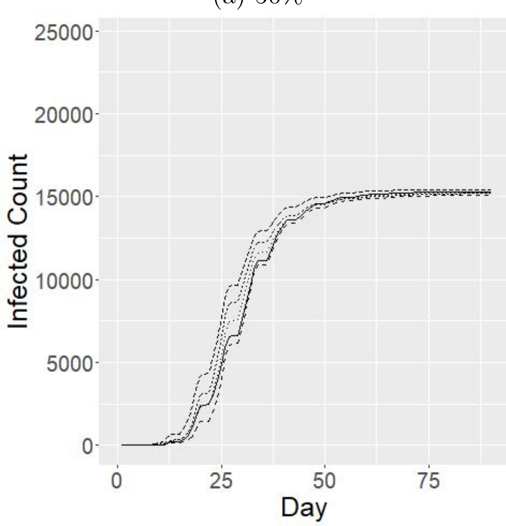
Figure 13: Sample cumulative trajectories when all classes are held in-person for Spring 2019.



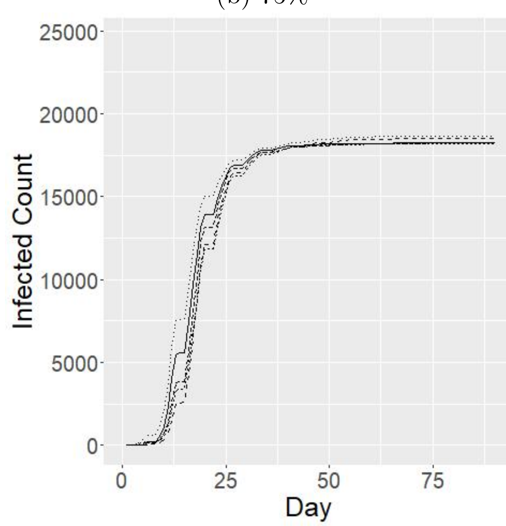
(a) 50%



(b) 75%

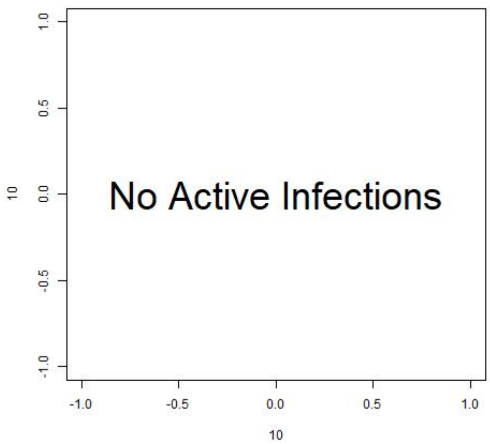


(c) 90%

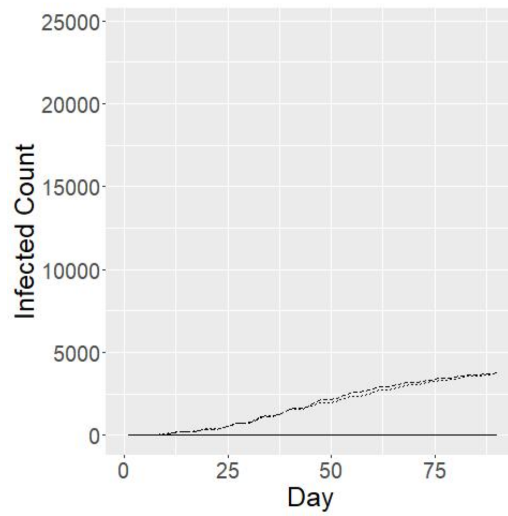


(d) 95%

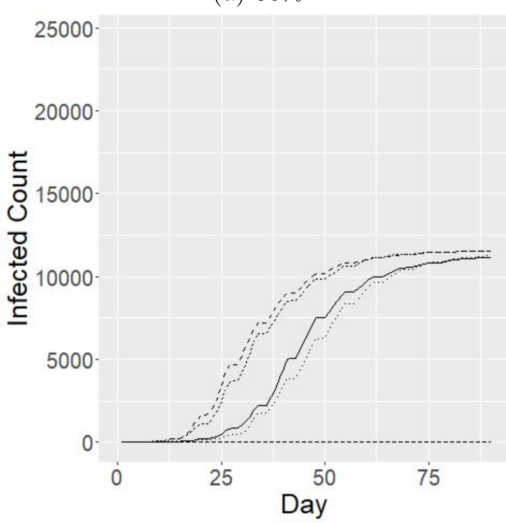
Figure 14: Sample cumulative trajectories from the high-infectiousness regime in Spring 2019.



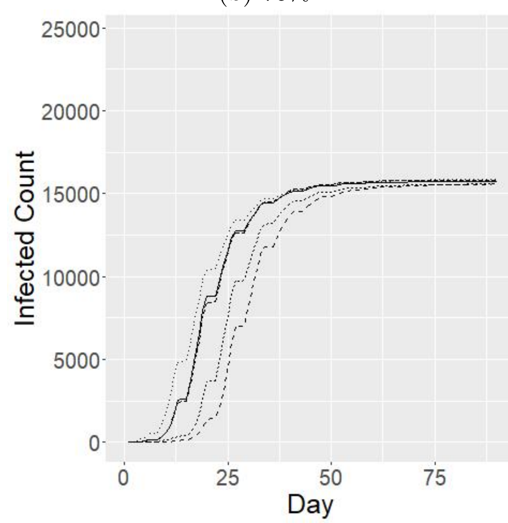
(a) 50%



(b) 75%

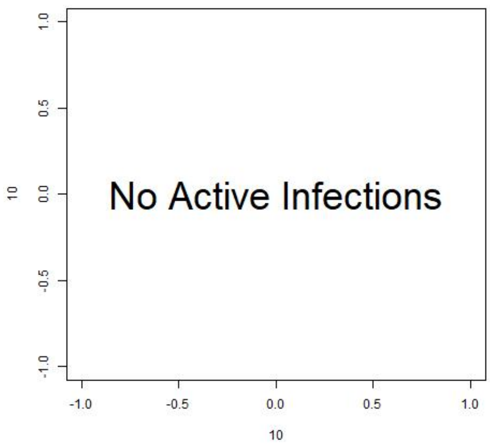


(c) 90%

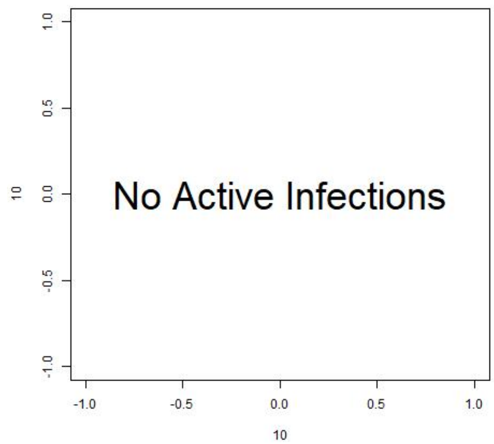


(d) 95%

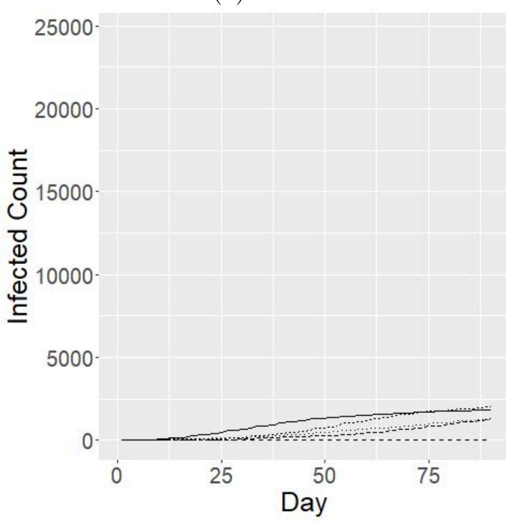
Figure 15: Sample cumulative trajectories from the medium-infectiousness regime in Spring 2019.



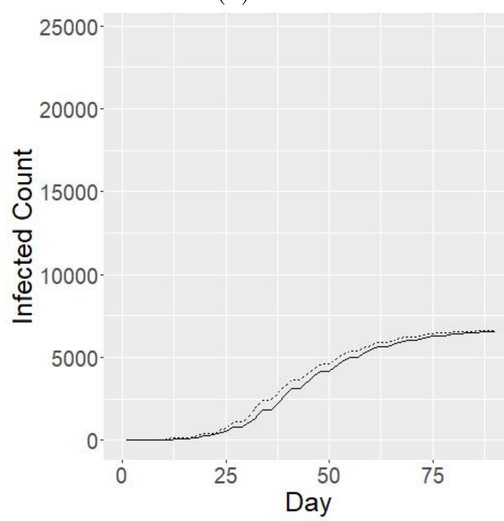
(a) 50%



(b) 75%

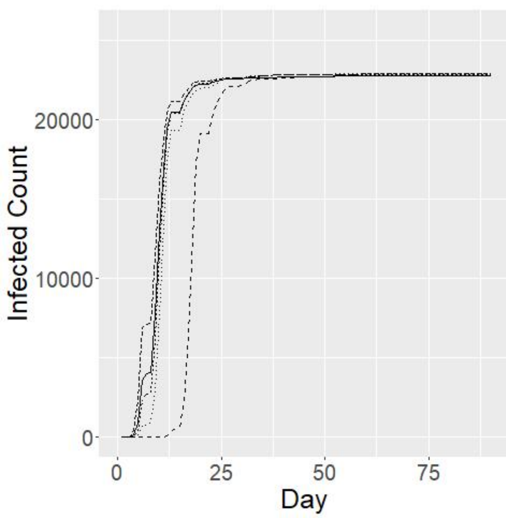


(c) 90%

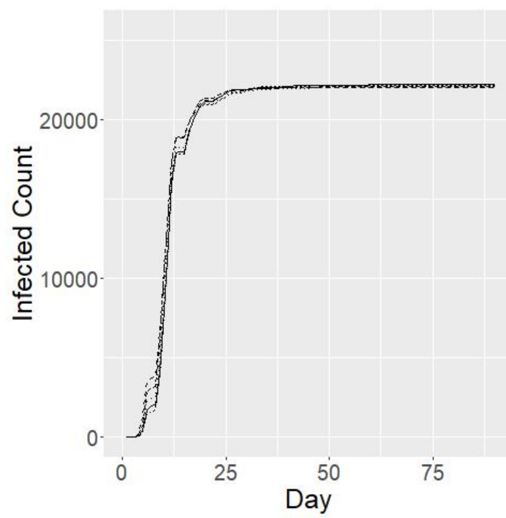


(d) 95%

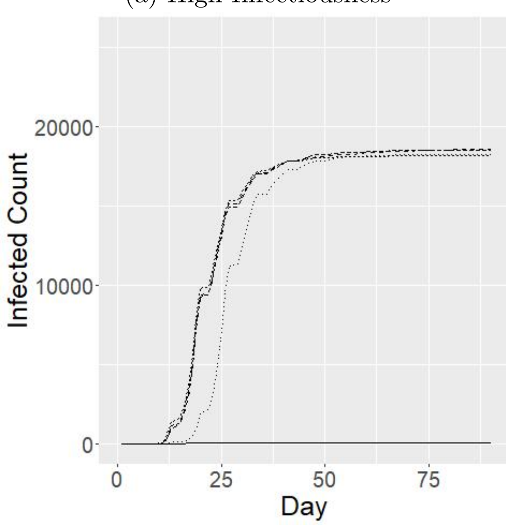
Figure 16: Sample cumulative trajectories from the low-infectiousness regime in Spring 2019.



(a) High Infectiousness

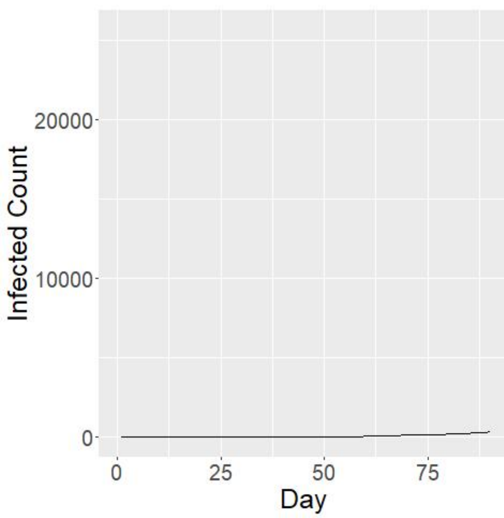


(b) Medium Infectiousness

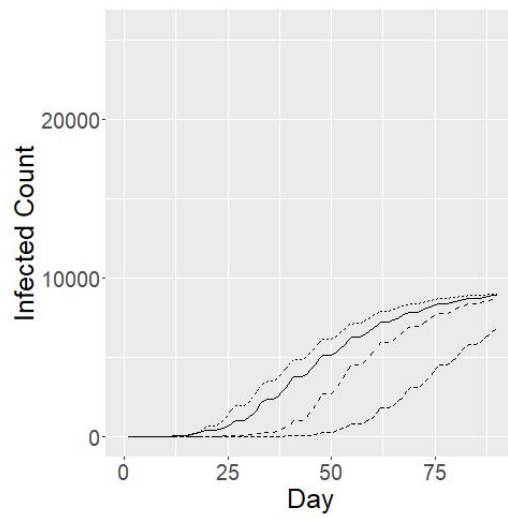


(c) Low Infectiousness

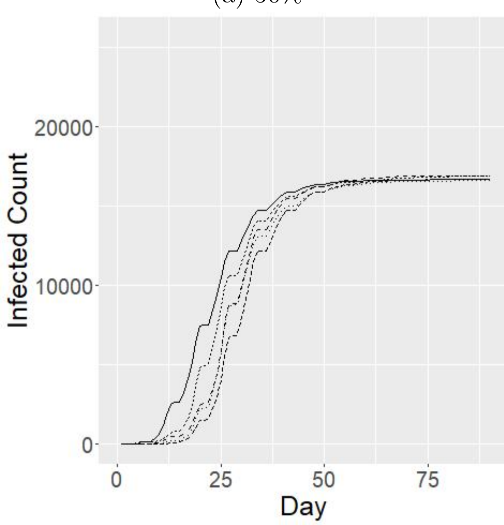
Figure 17: Sample cumulative trajectories when all classes are held in-person for Fall 2019.



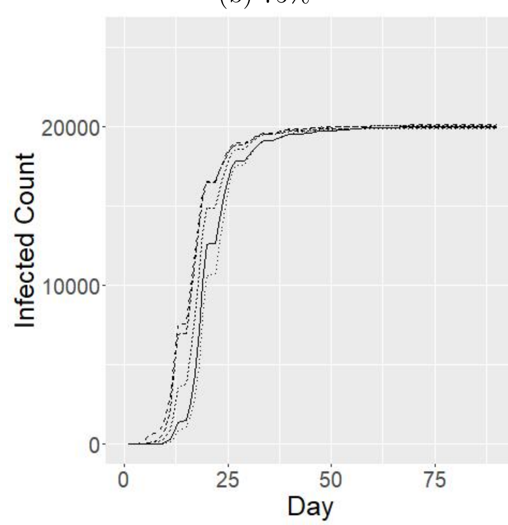
(a) 50%



(b) 75%

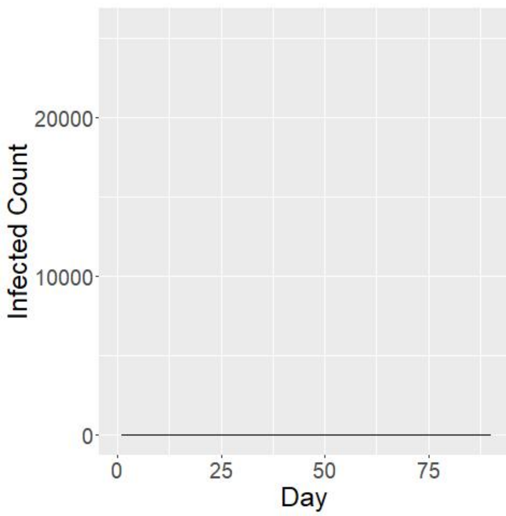


(c) 90%

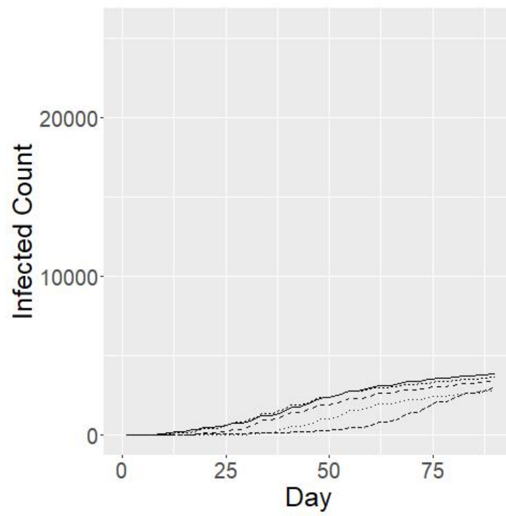


(d) 95%

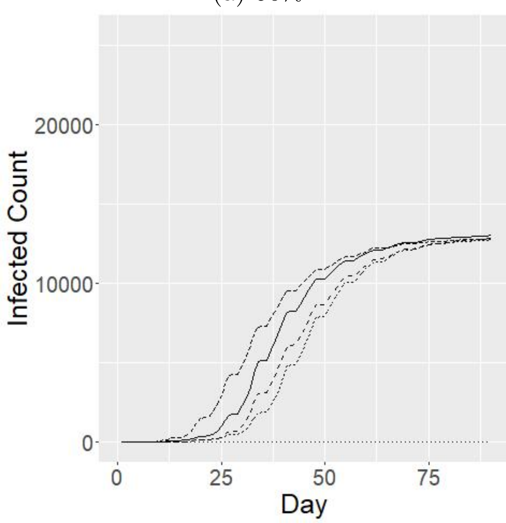
Figure 18: Sample cumulative trajectories from the high-infectiousness regime in Fall 2019.



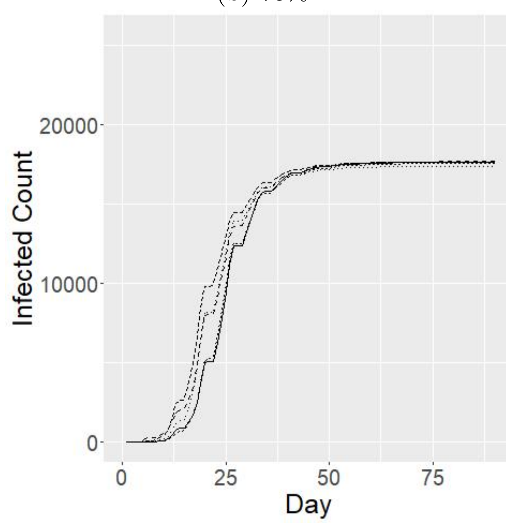
(a) 50%



(b) 75%

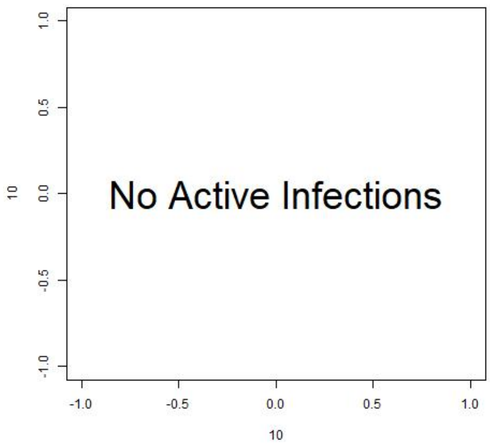


(c) 90%

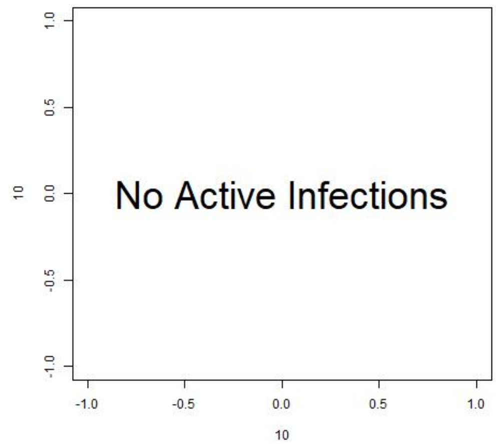


(d) 95%

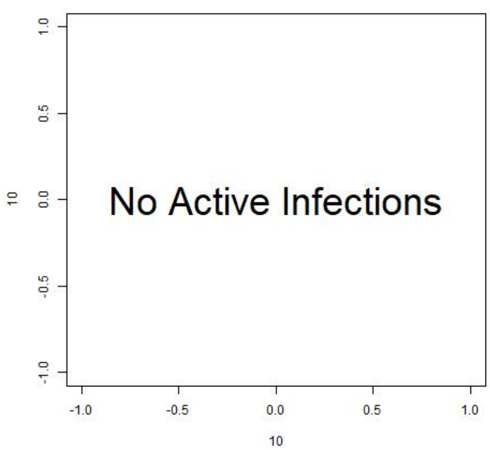
Figure 19: Sample cumulative trajectories from the medium-infectiousness regime in Fall 2019.



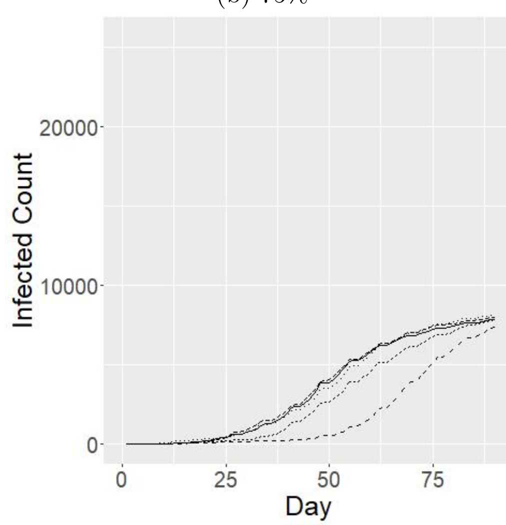
(a) 50%



(b) 75%

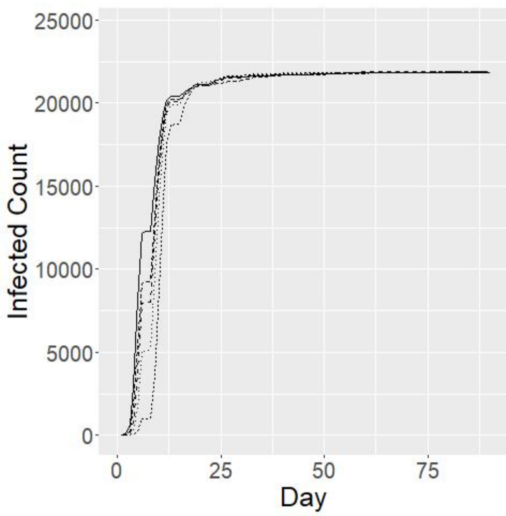


(c) 90%

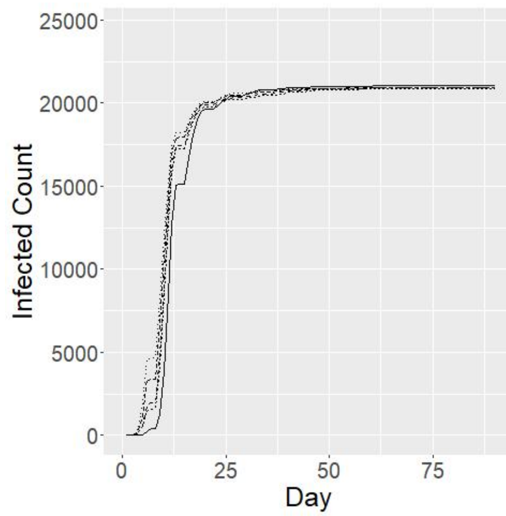


(d) 95%

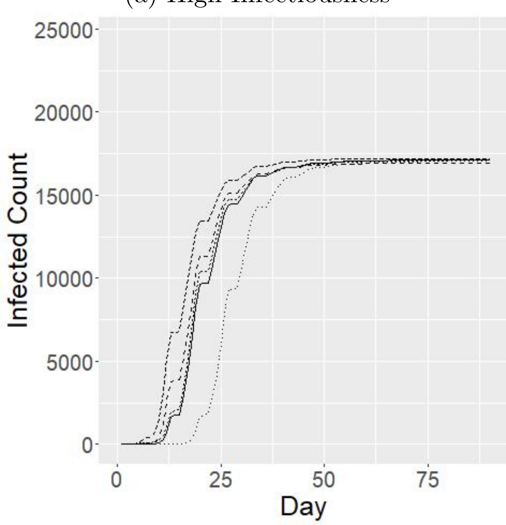
Figure 20: Sample cumulative trajectories from the low-infectiousness regime in Fall 2019.



(a) High Infectiousness

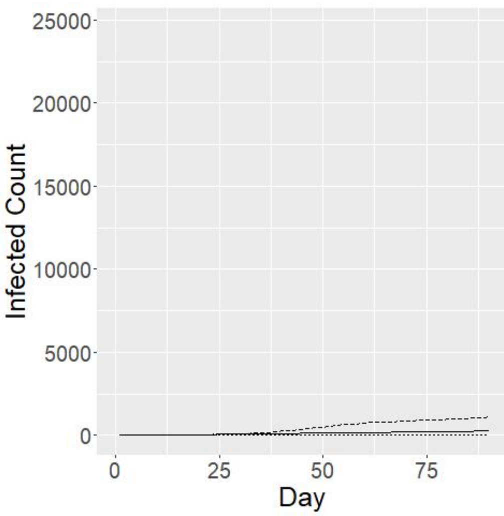


(b) Medium Infectiousness

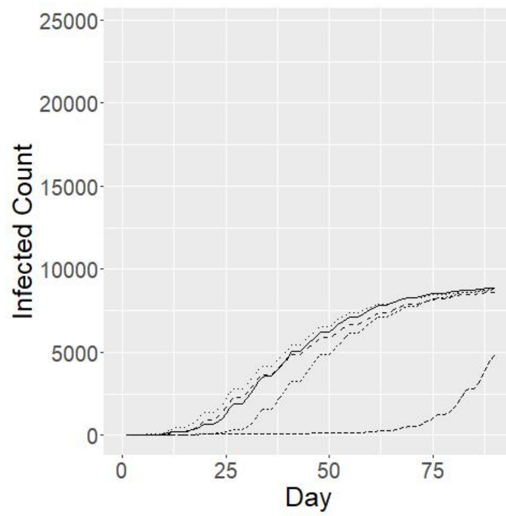


(c) Low Infectiousness

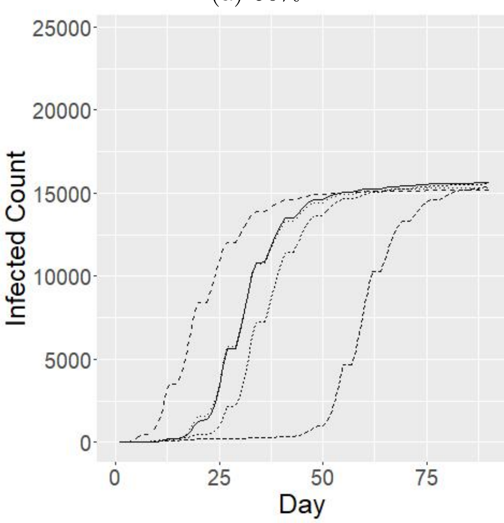
Figure 21: Sample cumulative trajectories when all classes are held in-person for Spring 2020.



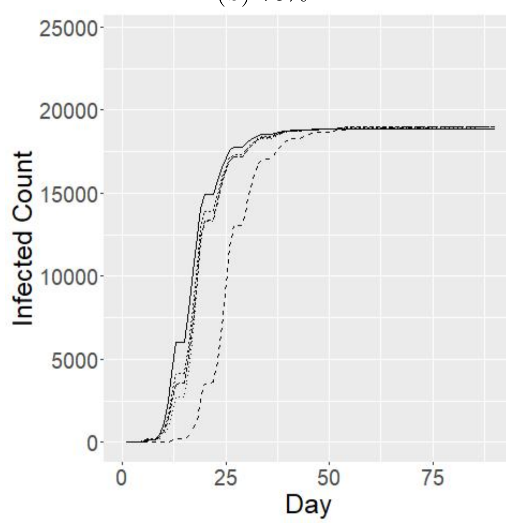
(a) 50%



(b) 75%

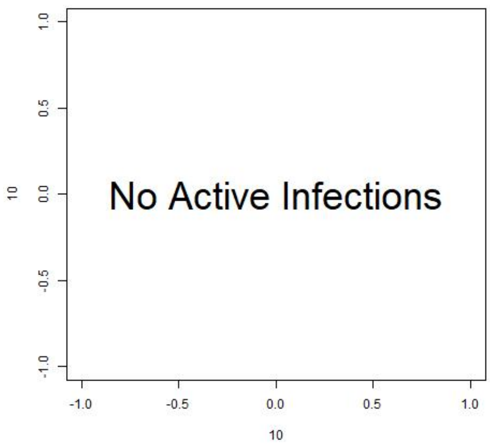


(c) 90%

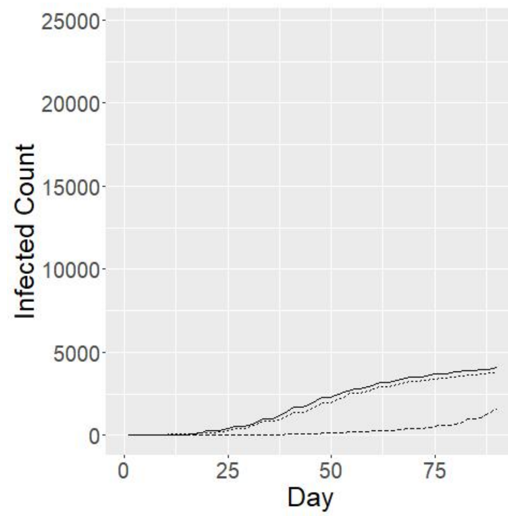


(d) 95%

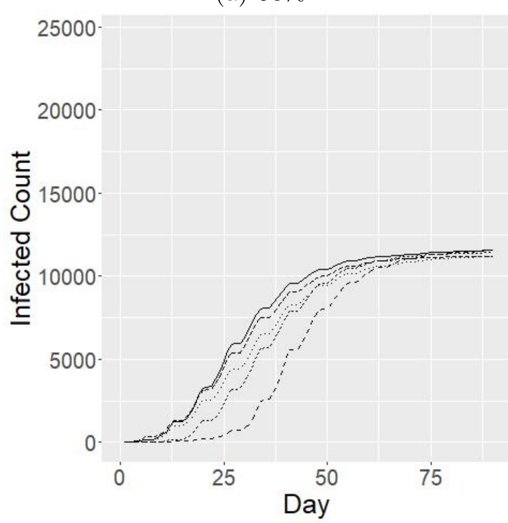
Figure 22: Sample cumulative trajectories from the high-infectiousness regime in Spring 2020.



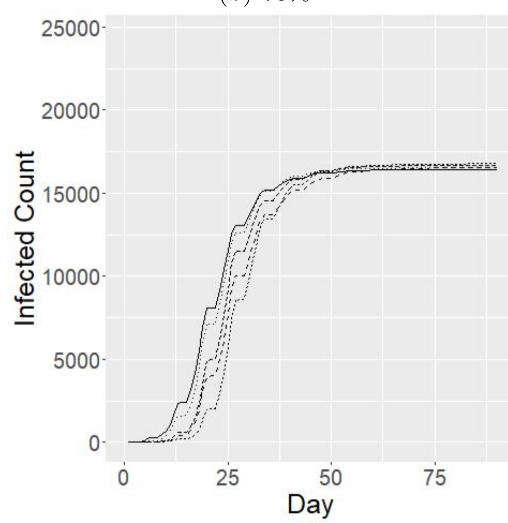
(a) 50%



(b) 75%

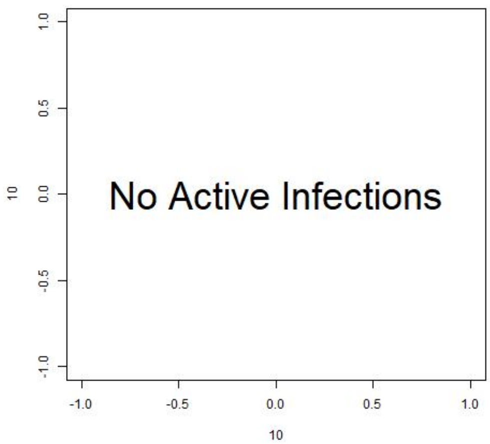


(c) 90%

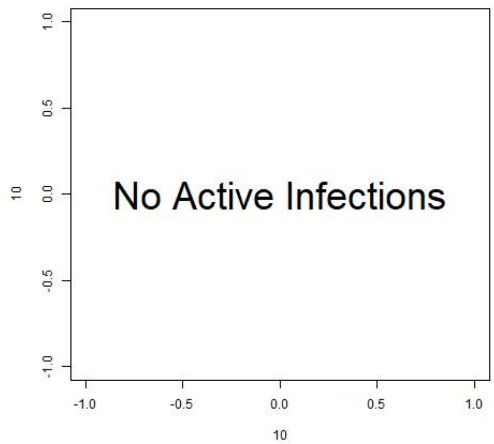


(d) 95%

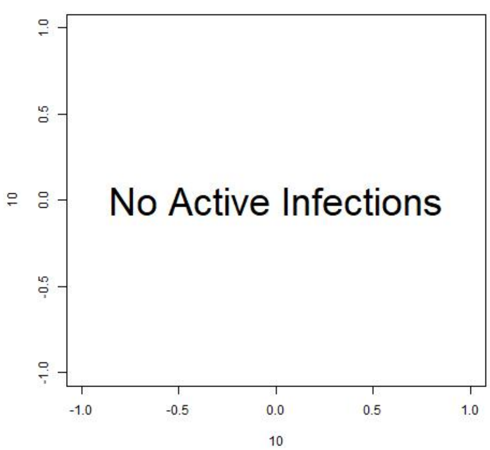
Figure 23: Sample cumulative trajectories from the medium-infectiousness regime in Spring 2020.



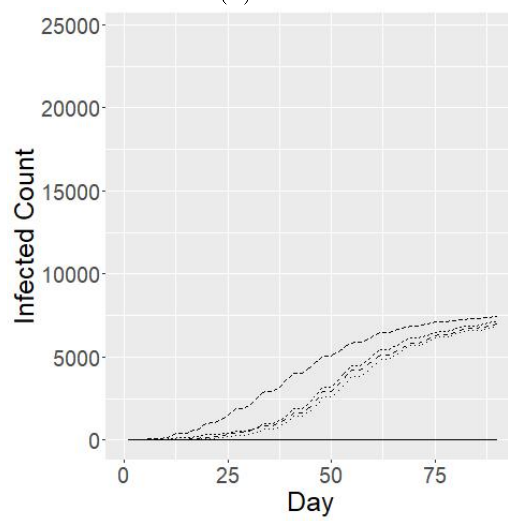
(a) 50%



(b) 75%

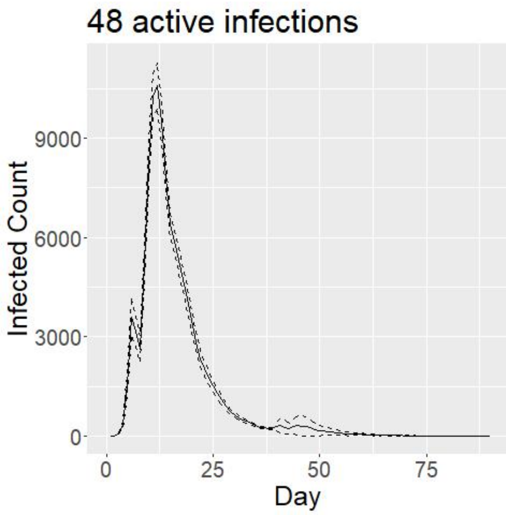


(c) 90%

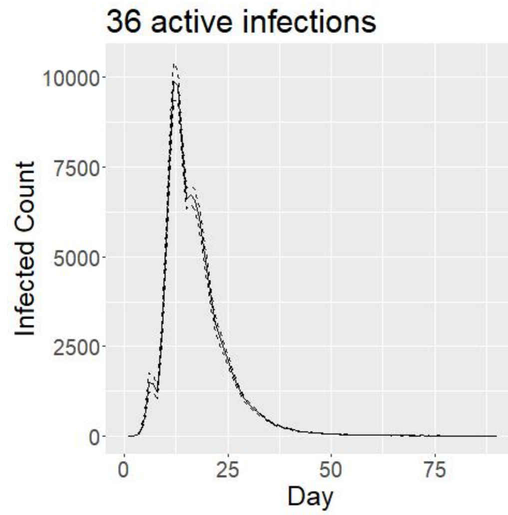


(d) 95%

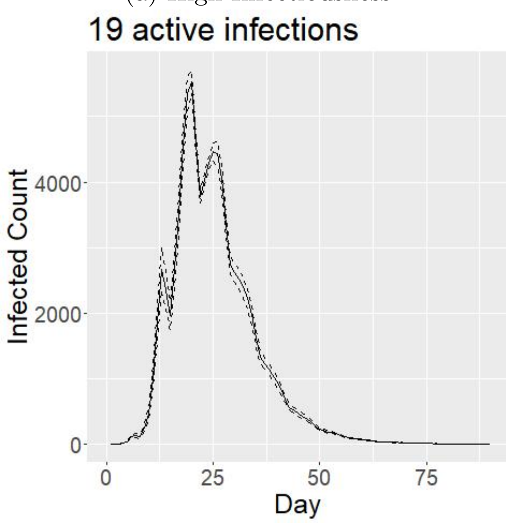
Figure 24: Sample cumulative trajectories from the low-infectiousness regime in Spring 2020.



(a) High Infectiousness

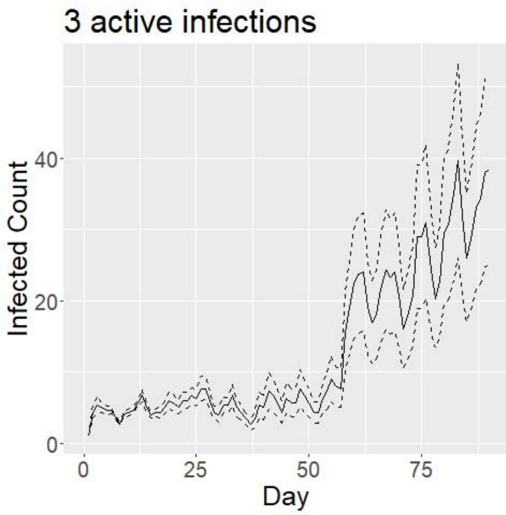


(b) Medium Infectiousness

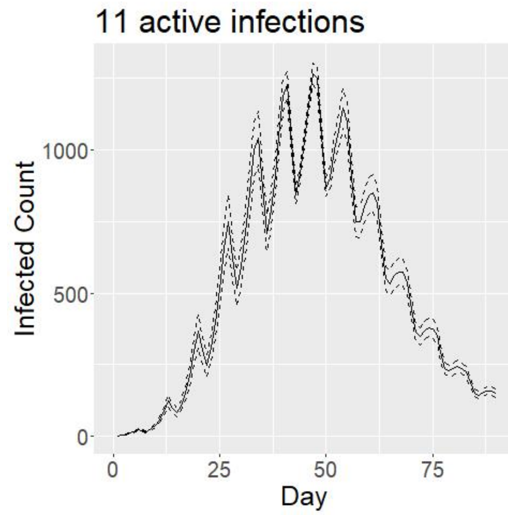


(c) Low Infectiousness

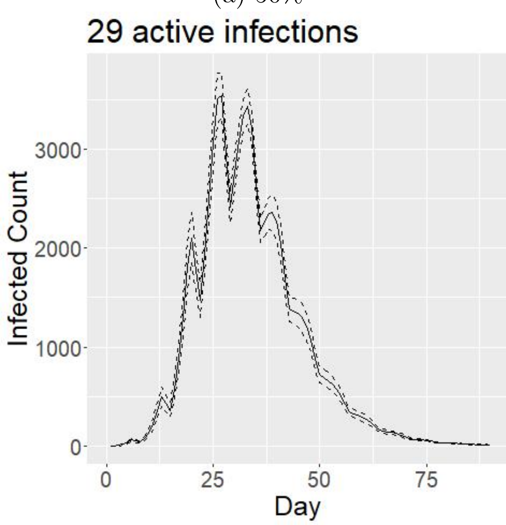
Figure 25: Trajectory summaries when all classes are held in-person for Spring 2019.



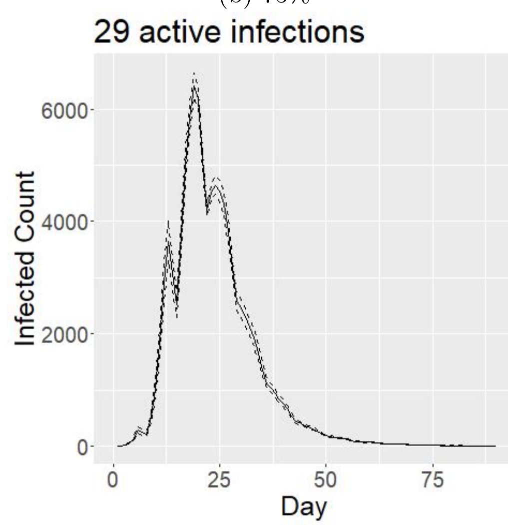
(a) 50%



(b) 75%

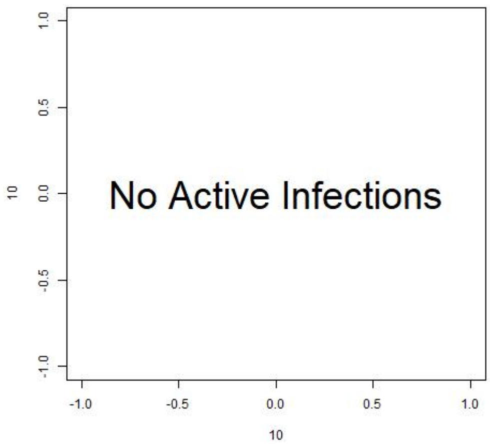


(c) 90%

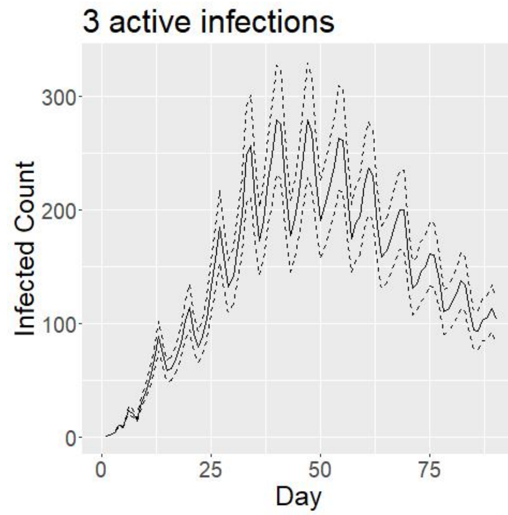


(d) 95%

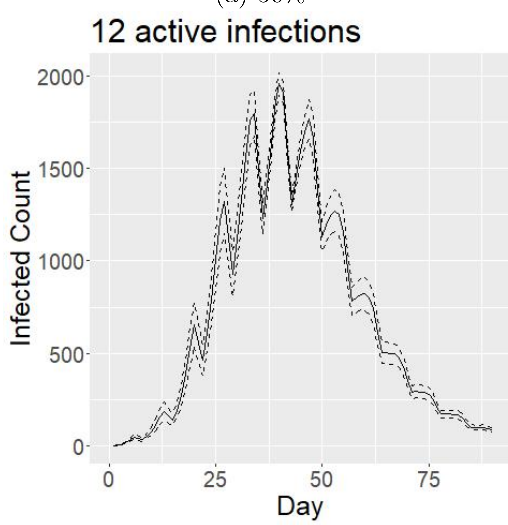
Figure 26: Trajectory summaries from the high-infectiousness regime in Spring 2019.



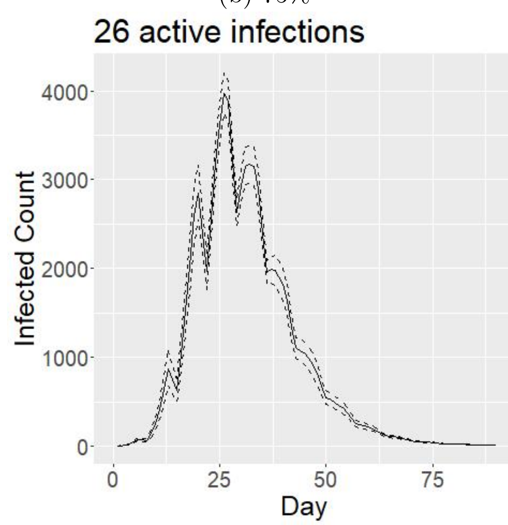
(a) 50%



(b) 75%

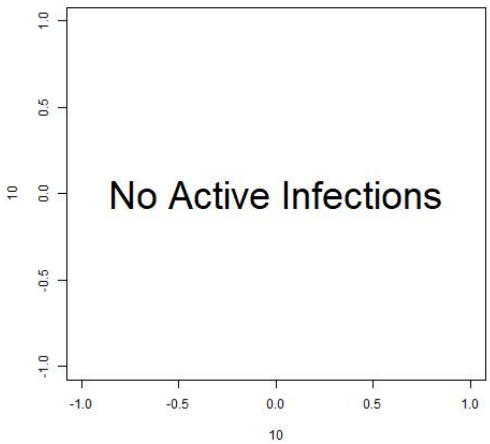


(c) 90%

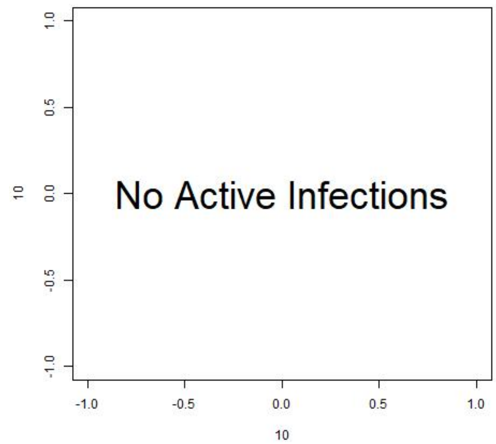


(d) 95%

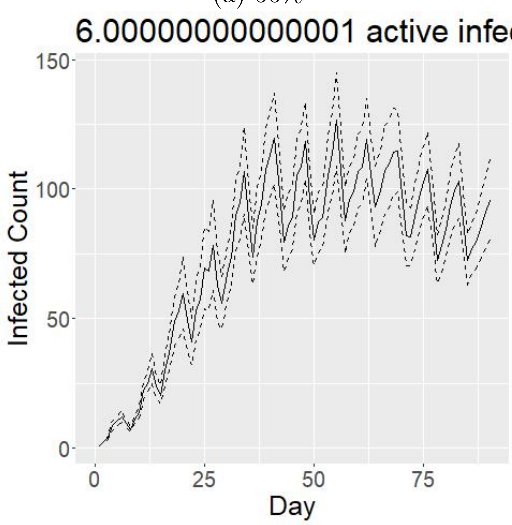
Figure 27: Trajectory summaries from the medium-infectiousness regime in Spring 2019.



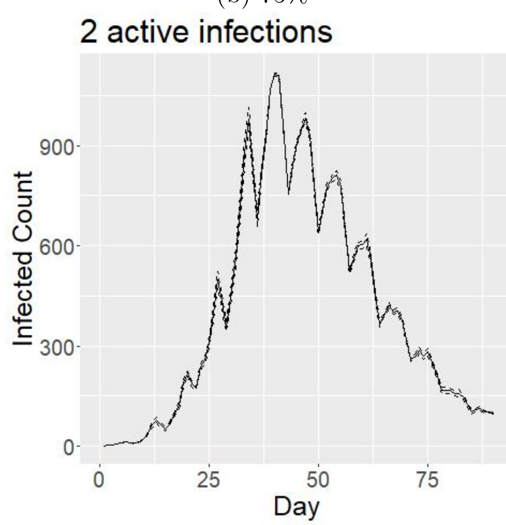
(a) 50%



(b) 75%

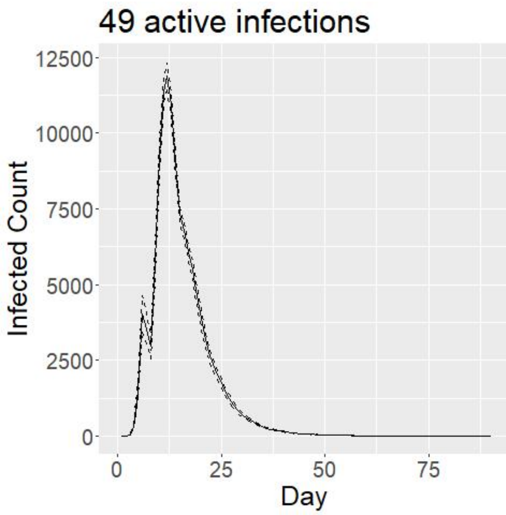


(c) 90%

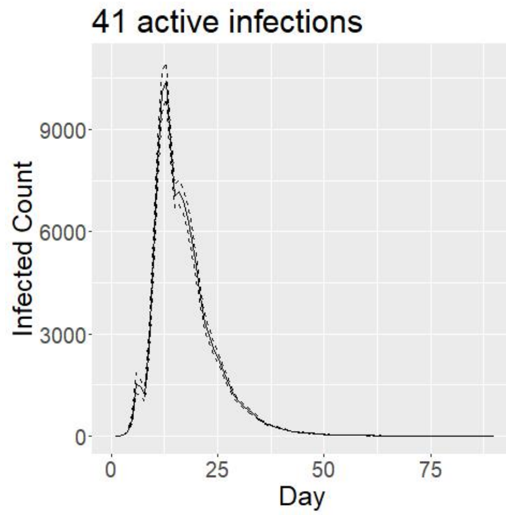


(d) 95%

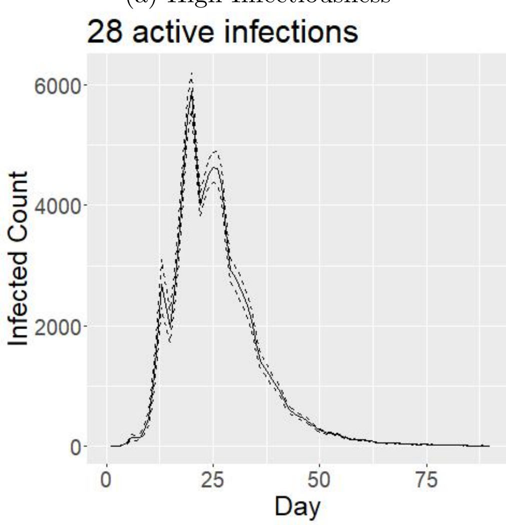
Figure 28: Trajectory summaries from the low-infectiousness regime in Spring 2019.



(a) High Infectiousness

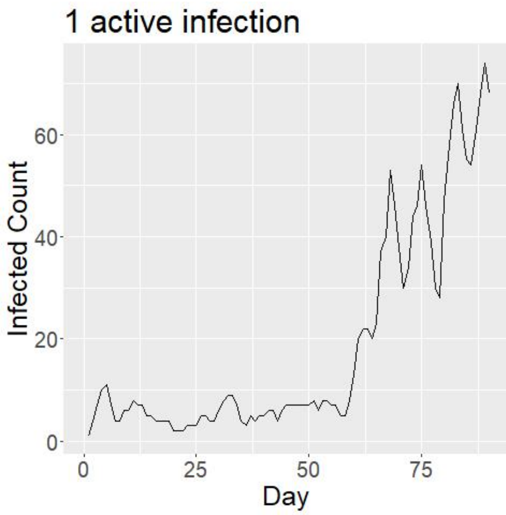


(b) Medium Infectiousness

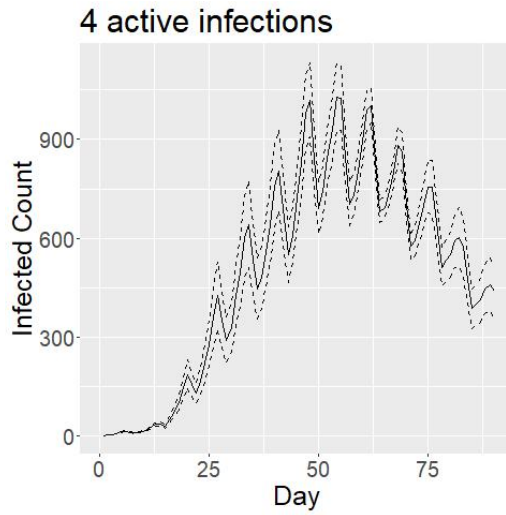


(c) Low Infectiousness

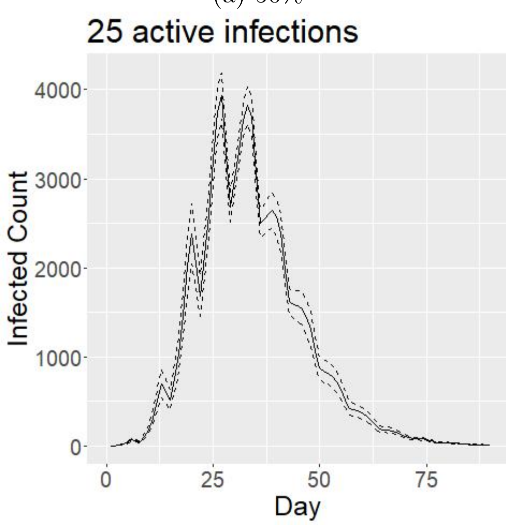
Figure 29: Trajectory summaries when all classes are held in-person for Fall 2019.



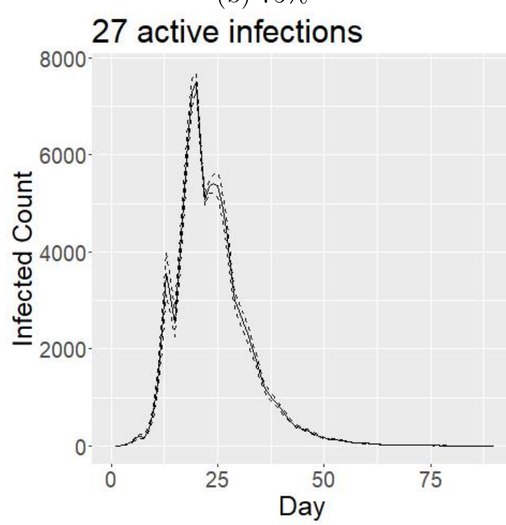
(a) 50%



(b) 75%

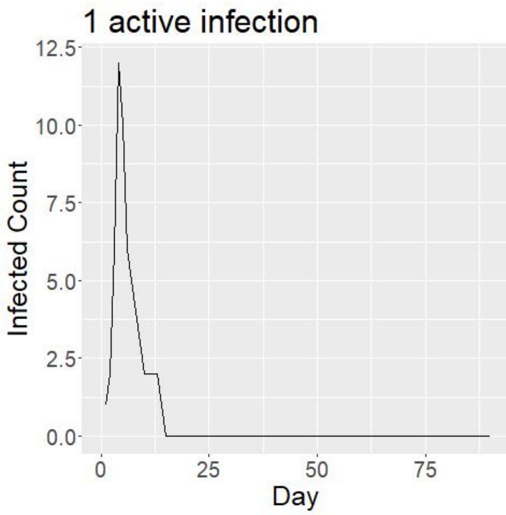


(c) 90%

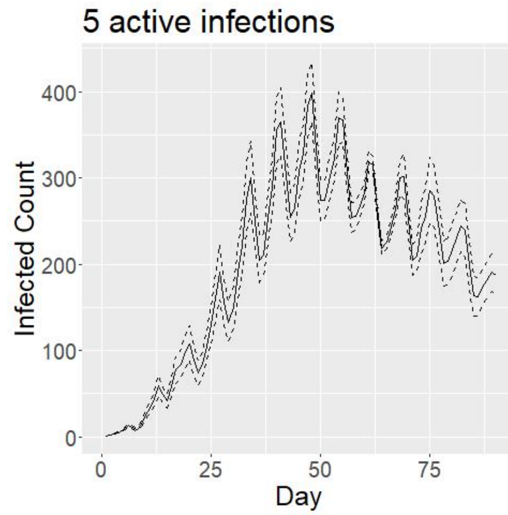


(d) 95%

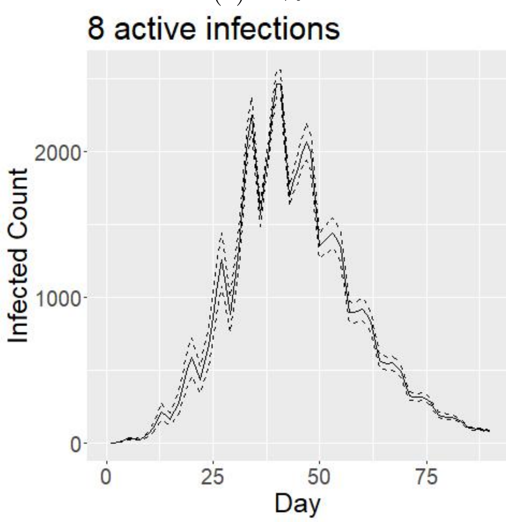
Figure 30: Trajectory summaries from the high-infectiousness regime in Fall 2019.



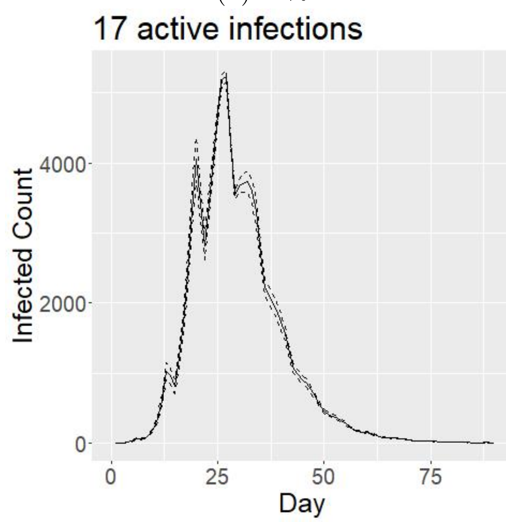
(a) 50%



(b) 75%

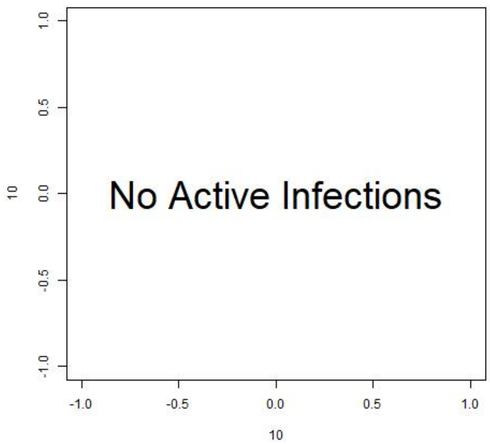


(c) 90%

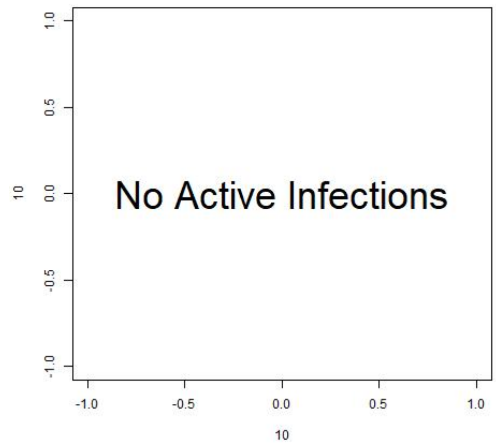


(d) 95%

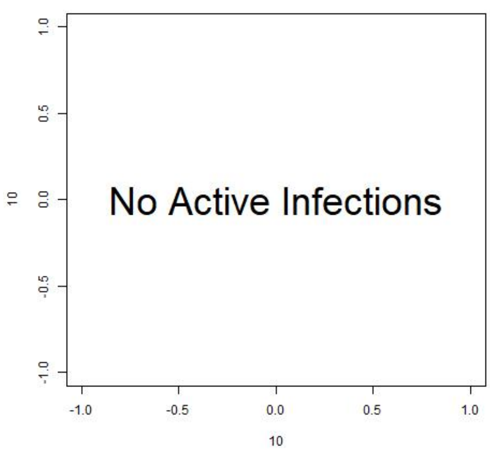
Figure 31: Trajectory summaries from the medium-infectiousness regime in Fall 2019.



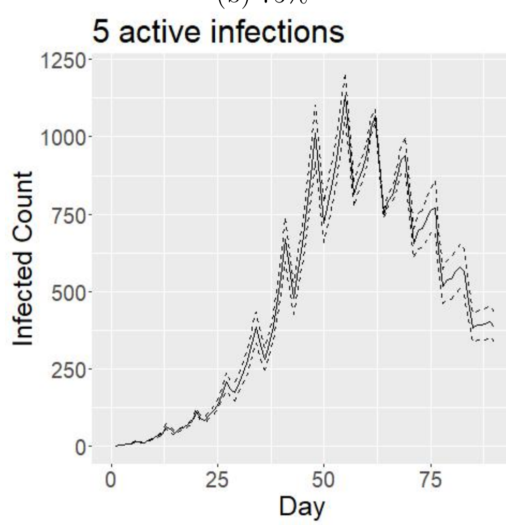
(a) 50%



(b) 75%

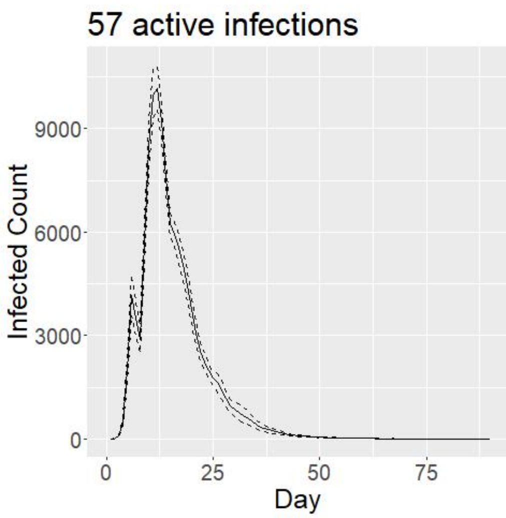


(c) 90%

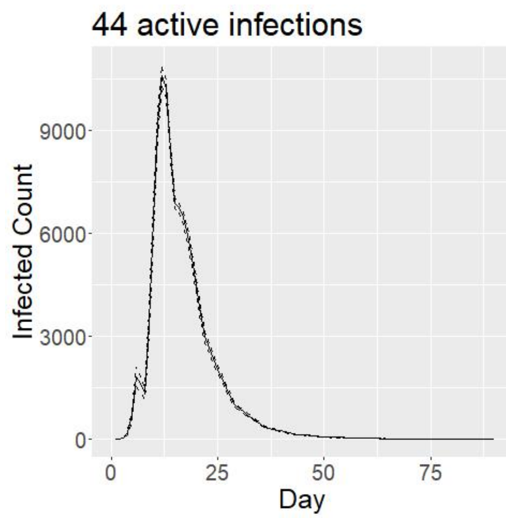


(d) 95%

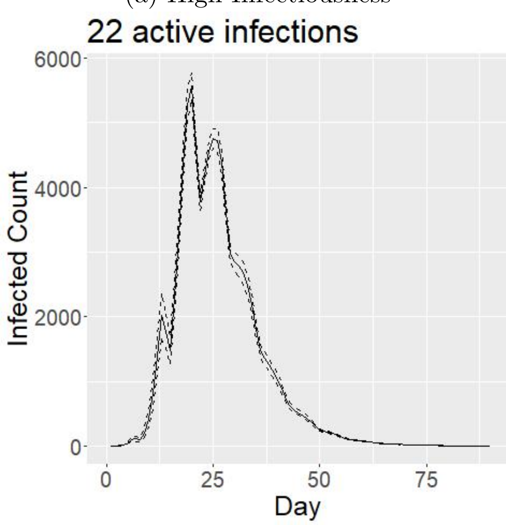
Figure 32: Trajectory summaries from the low-infectiousness regime in Fall 2019.



(a) High Infectiousness

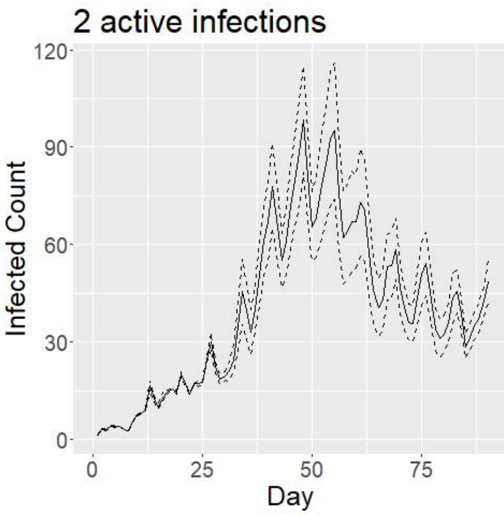


(b) Medium Infectiousness

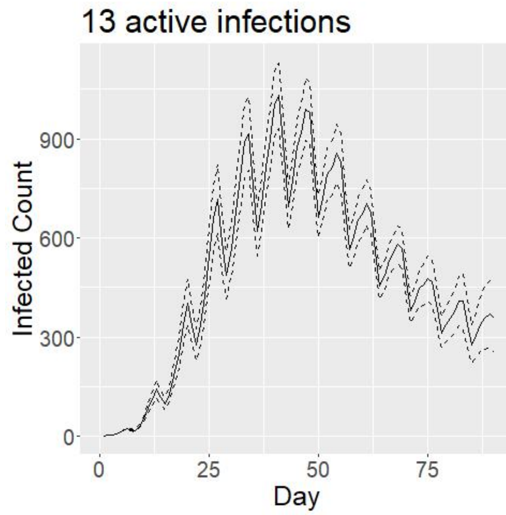


(c) Low Infectiousness

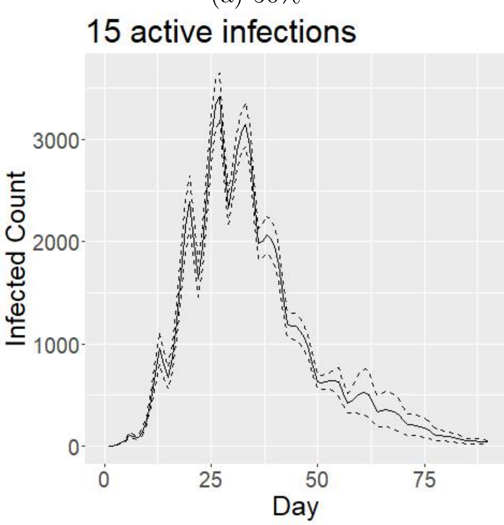
Figure 33: Trajectory summaries when all classes are held in-person for Spring 2020.



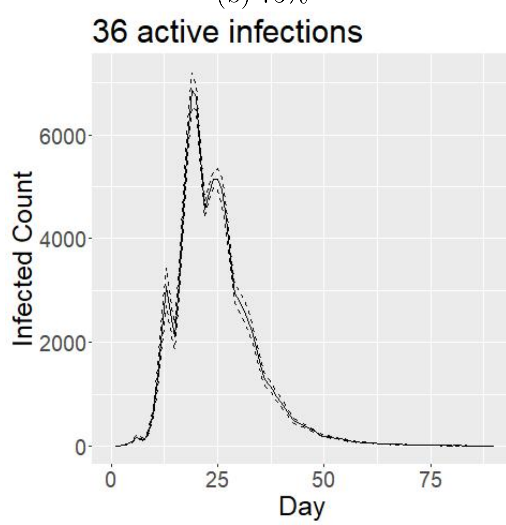
(a) 50%



(b) 75%

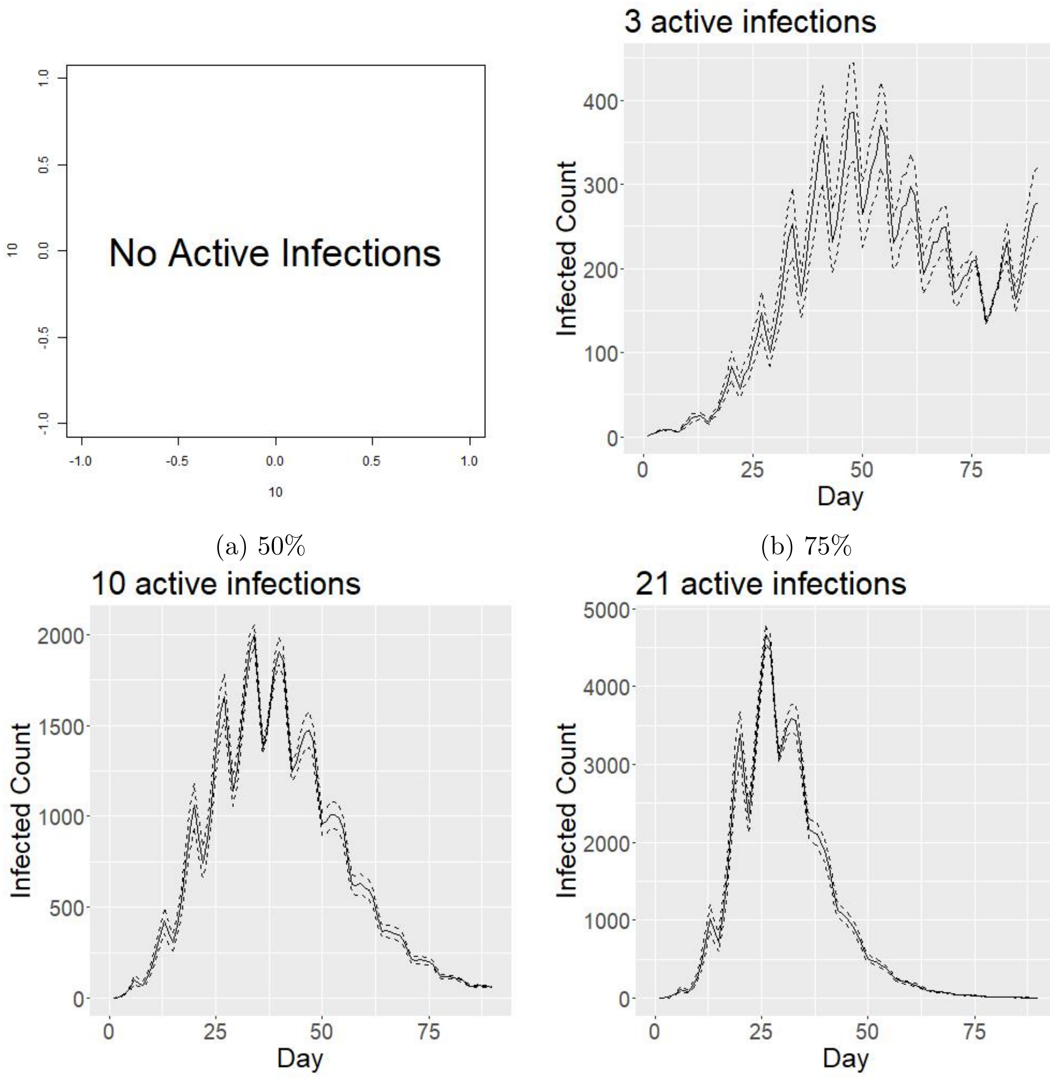


(c) 90%



(d) 95%

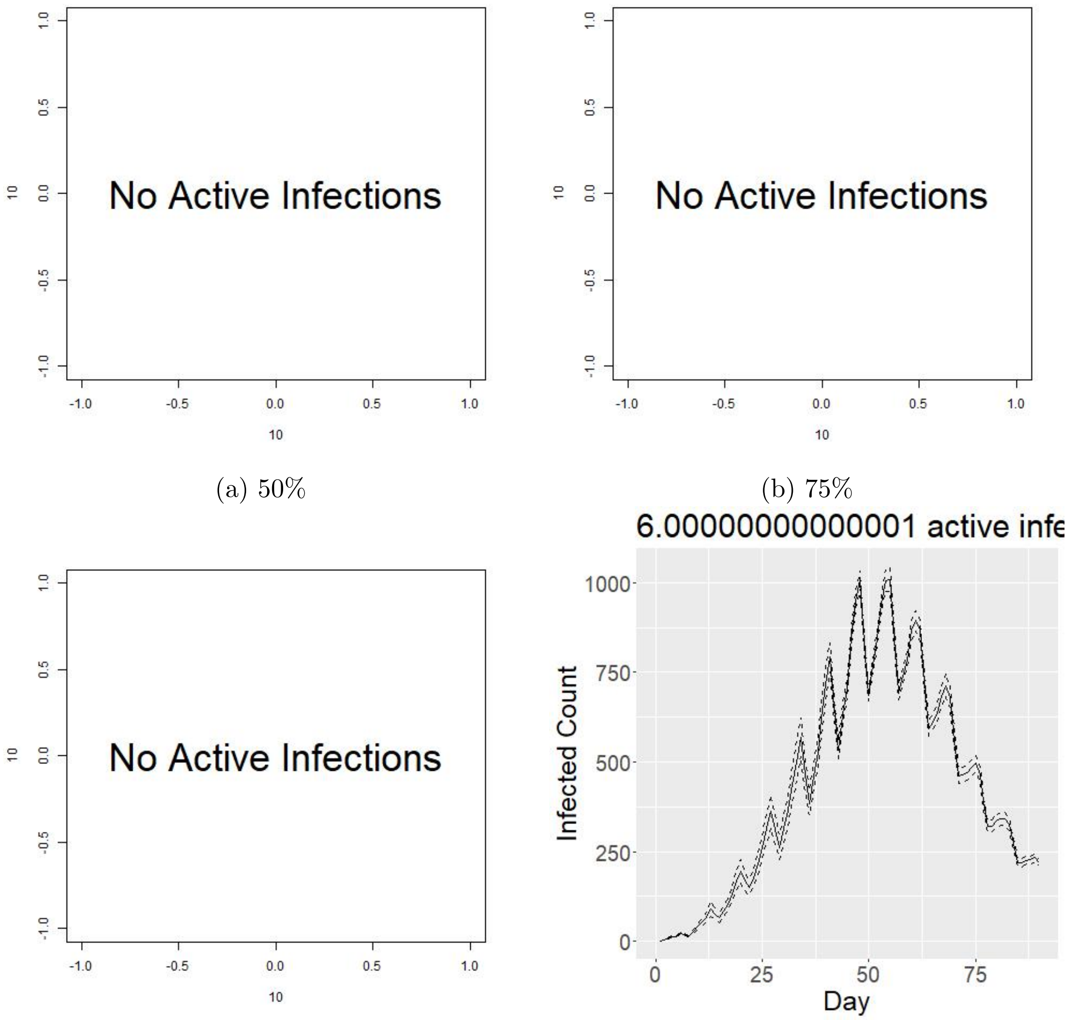
Figure 34: Trajectory summaries from the high-infectiousness regime in Spring 2020.



(a) 50% (b) 75%

(c) 90% (d) 95%

Figure 35: Trajectory summaries from the medium-infectiousness regime in Spring 2020.



(a) 50%

(b) 75%

(c) 90%

(d) 95%

Figure 36: Trajectory summaries from the low-infectiousness regime in Spring 2020.

First is "betweenness centralization". The general idea is to count, for each student i, the number of pairs of students j, k for which there is a geodesic connecting j to k which passes through i and divide by the number of geodesics between j and k. These numbers then need to be summarized. We used the package `igraph` in `R` and the particular routine called `centr_betw` to compute the summary called in that package "betweenness centralization"; see Csárdi and Nepusz [2006]. The routine weused computes, for a given student i, the number of geodesics from some j to some k which pass through vertex i. These numbers are divided by the number of geodesics from j to k. The result is averaged then rescaled by dividing through by the largest possible value to get a number between 0 and 1. We report the results with some hesitation; we are not persuaded that our computation is scaled in the same way as those reported by Weeden and Cornwell [2020]. If we are computing the same thing then the SFU value is *much* lower (a factor of nearly 10) than those reported for Cornell.

Next, Table 2 records some features of the largest bi-component of the 2-mode network. A group of nodes (a set of students and courses) is a 'bi-component' if you cannot remove a single node in such a way as to disconnect the graph. (This is a typical mathematician's strategy – define a term by a negation.) In all four datasets most nodes are in this largest bi-component. In particular over 90% of students are in the largest bi-component in every case. The figures for courses are rather lower because there are a number of small courses with very few registrants; wehave not looked at these courses in detail.

2.2 The 1-Mode Network

The 'clustering coefficient' deserves some explanation. Consider three students i, j, and k. If i and j are in a class together and j and k are in a class together it might or might not be the case that i and k are in a class together. We count the number of triples i, j, k where

Table 2: Social Network Measures for the 2-mode (student-to-course) graph. Data from part of Table 2 in Weeden and Cornwell (2020, Table 2) with SFU added. Network density is given $l/(nm)$ which compares the number of edges to the maximum possible for a bipartite graph. The figure reported in WC is 0.0000 which compares the number of edges to the number of edges on the (not bipartite) complete graph on the same set of vertices.

	Cornell		SFU	
	Fall 2019	Spring 2019	Fall 2019	Spring 2020
# students (n)	22,051	24,071	25,089	23,836
# courses (m)	6,072	3,546	3,789	3,552
# edges (l)	118,314	108,554	117,588	107,902
Network density	0.0009	0.0013	0.0012	0.0013
Number of Components	—	149	150	140
Largest component				
Proportion: students	0.991	0.971	0.966	0.969
Proportion: courses	0.976	0.970	0.969	0.968
Betweenness centralization	0.098	0.036	0.028	0.030
Largest bi-component				
Proportion: students	0.945	0.902	0.903	0.910
Proportion: courses	0.730	0.943	0.948	0.946

all three connections exist and divide by the number of triples i, j, k where i and j share a class and j and k share a class. This number is large in networks where the existence of the two links makes the third link likely to exist. The figures here – in the 0.45 to 0.5 range – are fairly high; students with common interests tend to show up as a fully connected group of 3. This point is explored in some depth in Weeden and Cornwell [2020]. Notice that there are two effects which tend to push these numbers up. First, if students i and j are in a class together it is relatively likely that this class is large.[1] Second, if students are in the same major and at similar stages in their programs then they are likely to share classes with other students in their major. Turned around by Bayes theorem we see that the fact that i and j are in a class together makes it more likely that they are in the same major. The same applies to j and k. Consequently the chance that i and k are in the same major is elevated and thus the chance they share a class is elevated.

3 Relationship Between Network Summaries and Simulation Results

In Ruth and Lockhart [2021], we present results for a logistic regression model fit to predict develop rate using log-number of k-step paths in the enrollment network for spring 2019 with various class size thresholds. We also used AIC to select a subset of path lengths, retaining only $k = 1, 2, 3$.

Here, we present the corresponding results for all terms. Using stepwise selection with AIC, we chose to retain $k = 1, 2, 3, 4$ in fall 2019, and $k = 1$ in spring 2020. Plots of fitted

[1] Suppose we have a class with C students registered in it. If we pick a registration at random from the l records in our data set then we have chance C/l of selecting a registration in this class; this number is proportional to C. If we pick two different registrations at random then the chance they are both in this class is $C(C-1)/(l(l-1))$ which is proportional to C^2. As a consequence of these scalings the conditional probability that two records are in a class of size C given they are both in some class is roughly proportional to C.

Table 3: Social Network Measures for the projected 1-mode (student-to-student) graph. Data from Table 2 in Weeden and Cornwell (2020, Table 2) with SFU added. SFU data reflects only the largest component of the network.

	Cornell Fall 2019	Spring 2019	SFU Fall 2019	Spring 2020
# unique edges (l)	5,832,358	3,706,679	4,064,905	3,500,917
Network density	0.024	0.013	0.013	0.013
Clustering coefficient (transitivity/closure)	0.480	0.444	0.445	0.432
Average geodesic	2.466	2.779	2.730	2.722
Network diameter	10	16	16	15
Average k-step reach centrality				
$k = 1$	0.024	0.014	0.014	0.013
$k = 2$	0.594	0.449	0.460	0.445
$k = 3$	0.921	0.894	0.903	0.906
$k = 4$	0.966	0.938	0.947	0.951

and observed develop rates are given in Figures 37-39.

See Ruth and Lockhart [2021] for more details.

References

Gábor Csárdi and Tamás Nepusz. The igraph software package for complex network research. *InterJournal, complex systems*, 1695(5):1–9, 2006.

William Ruth and Richard Lockhart. Network analysis of SFU course registrations. 2021.

Kim A. Weeden and Benjamin Cornwell. The small-world network of college classes: Implications for epidemic spread on a university campus. *Sociological Science*, 7:222–241, 2020.

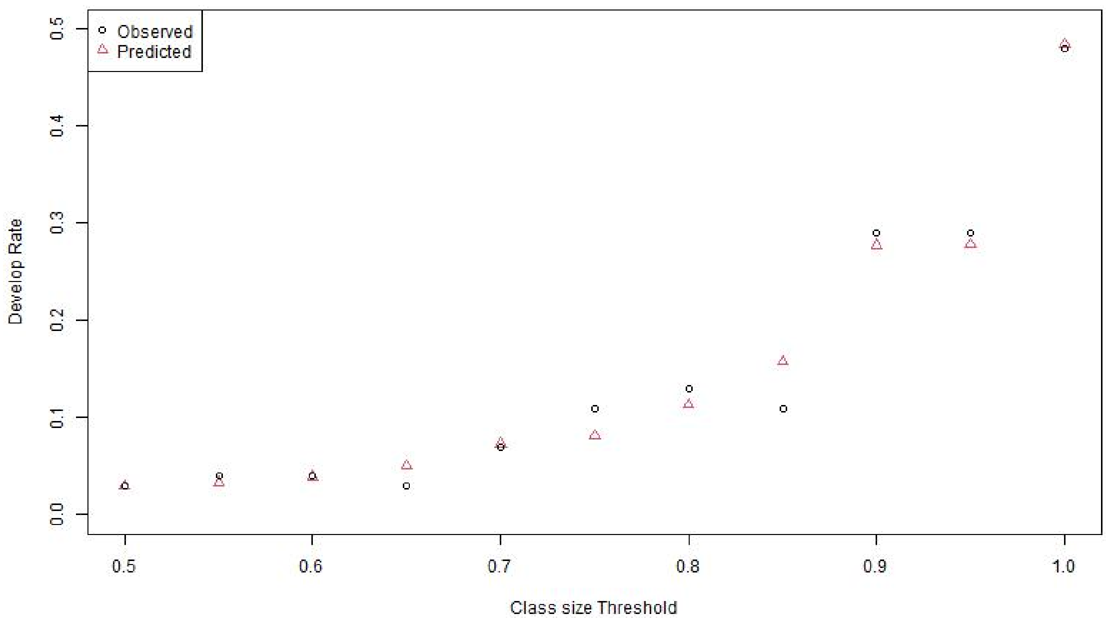


Figure 37: Observed and fitted develop rates in spring 2019 using numbers of k-step paths for $k = 1, 2, 3$.

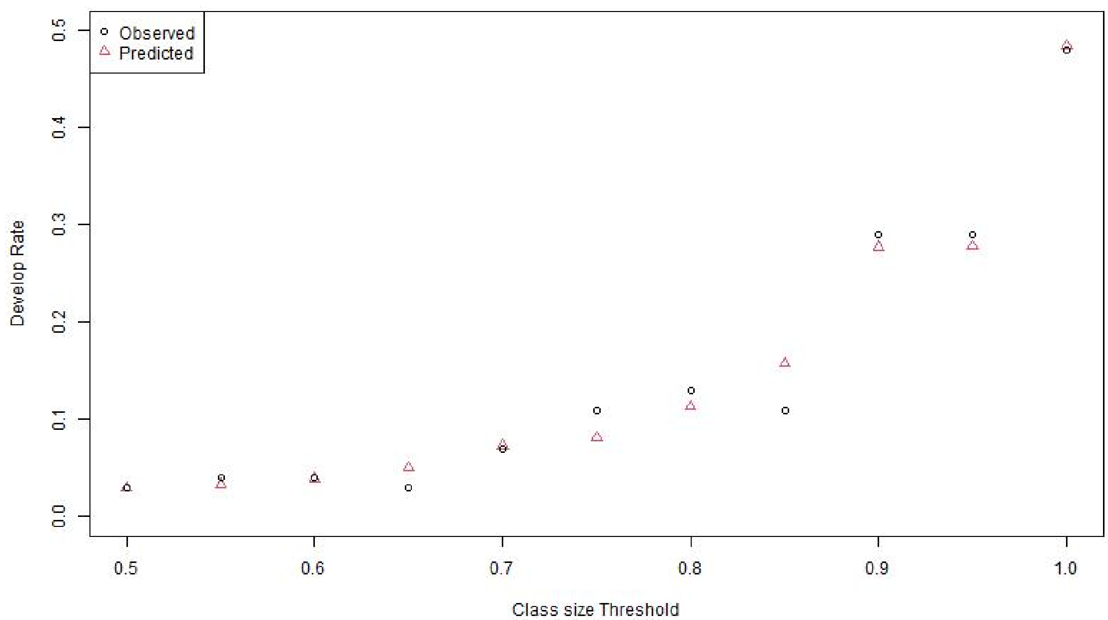


Figure 38: Observed and fitted develop rates in fall 2019 using numbers of k-step paths for $k = 1, 2, 3, 4$.

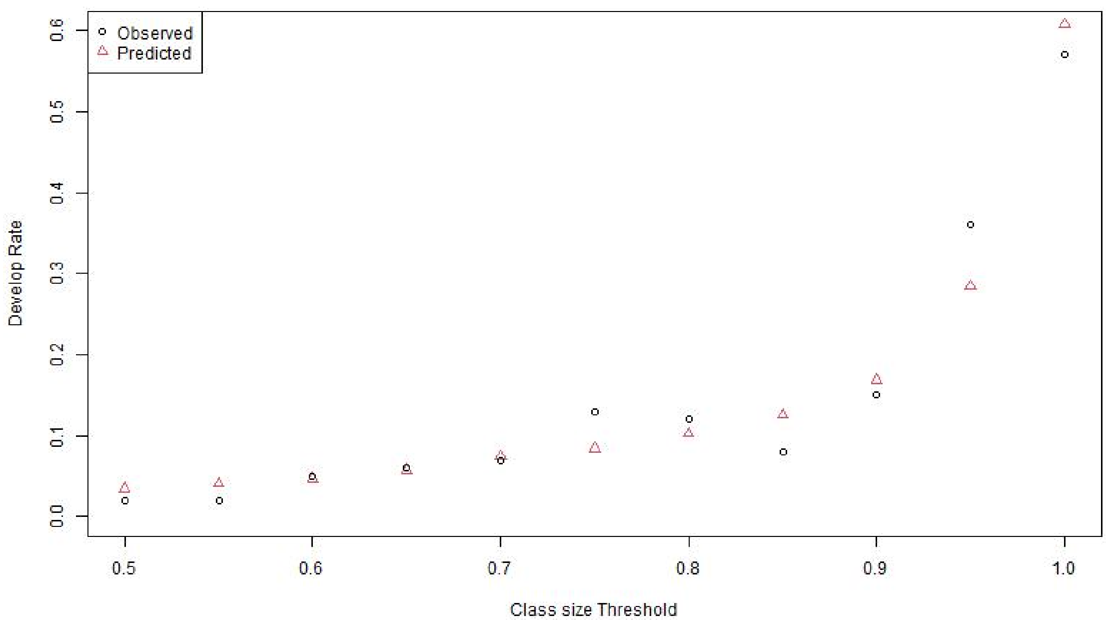


Figure 39: Observed and fitted develop rates in spring 2020 using numbers of k-step paths for $k = 1$.