Problem Set 1
EPS 528, Science of Complex Systems

Jonas Katona

November 16, 2022

Problem 1.

Solution. As discussed in class, we know that this insect outbreak model exhibits a saddle-node/fold bifurcation. This can be determined graphically by looking at when the curves $rx\left(1-\frac{x}{k}\right)$ and $\frac{x^2}{1+x^2}$ intersect for differing combinations of parameters r and k because this will happen iff x is a fixed point for the system. Doing this, we would see that every change in the stability of fixed points occurs when two fixed points form out of nowhere or annihilate and disappear as r or k are varied. From here, we can solve for the bifurcation curve, i.e., the combinations of values of r and k at which bifurcations occur, by noting the tangency, nondegeneracy, and transversality conditions: For a 1D continuous-time dynamical system $\dot{x} = f(\mathbf{r}, x)$, a saddle-node bifurcation occurs at some combination of parameter(s) $\mathbf{r}^* \in \mathbb{R}^m$ and point $x^* \in \mathbb{R}$ iff the following equations hold: $f(\mathbf{r}^*, x^*) = 0$, $\partial_x f(\mathbf{r}^*, x)|_{x=x^*} = 0$, $\partial_x^2 f(\mathbf{r}^*, x)|_{x=x^*} \neq 0$, and $\partial_{r_i} f(\mathbf{r}, x^*)|_{\mathbf{r}=\mathbf{r}^*} \neq 0$ for $i = 1, \ldots, m$.

In particular, for our system $\dot{x} = rx\left(1-\frac{x}{k}\right) - \frac{x^2}{1+x^2}$, this gives the following four conditions (dropping the $*$ superscripts):

$$rx\left(1-\frac{x}{k}\right) = \frac{x^2}{1+x^2} \Rightarrow (x=0) \vee \left(r\left(1-\frac{x}{k}\right) = \frac{x}{1+x^2} \overset{x\neq k}{\Longrightarrow} r = \frac{k}{k-x}\frac{x}{1+x^2}\right), \tag{1}$$

$$\partial_x\left[rx\left(1-\frac{x}{k}\right)\right] = \partial_x\left[\frac{x^2}{1+x^2}\right] \Rightarrow r\left(1-\frac{2x}{k}\right) = \frac{2x}{(1+x^2)^2}, \tag{2}$$

$$\partial_x^2\left[rx\left(1-\frac{x}{k}\right)\right] \neq \partial_x^2\left[\frac{x^2}{1+x^2}\right] \Rightarrow -\frac{2r}{k} \neq \frac{2-6x^2}{(x^2+1)^3} \Rightarrow \frac{r}{k} \neq \frac{3x^2-1}{(x^2+1)^3}, \tag{3}$$

$$\partial_r\left[rx\left(1-\frac{x}{k}\right)\right] \neq \partial_r\left[\frac{x^2}{1+x^2}\right] \Rightarrow x\left(1-\frac{x}{k}\right) \neq 0 \Rightarrow x \neq 0 \wedge x \neq k, \tag{4}$$

$$\partial_k\left[rx\left(1-\frac{x}{k}\right)\right] \neq \partial_k\left[\frac{x^2}{1+x^2}\right] \Rightarrow \frac{rx^2}{k^2} \neq 0 \Rightarrow x \neq 0 \wedge r \neq 0. \tag{5}$$

Within the physical constraints of our problem, we always know that $x \geq 0$ (our insect population cannot be negative), $k > 0$ (the carrying capacity cannot negative or zero), and $r > 0$ (the intrinsic population growth rate in absence of predation is positive). Using these constraints, I claim that (1) and (2) alone suffice in describing the bifurcation curve except when $x \neq \sqrt{3}$ (which we will check as an edge case). First off, $x = 0$ cannot be true at the bifurcation, because even if we go against (4) and (5) and assume that this is true, then equation (2) would imply that $r = 0$, a contradiction. Next, if we take $x = k$ contrary to the other condition in (4), then (1) would imply that $\frac{x}{1+x^2} = 0$, also a contradiction. Finally, if we assume that $\frac{r}{k} = \frac{3x^2-1}{(x^2+1)^3}$ contrary to (3), then we have two possible scenarios:

- $x = \frac{1}{\sqrt{3}} \Rightarrow \frac{r}{k} = 0 \Rightarrow r = 0$, a contradiction.

- $x \neq \frac{1}{\sqrt{3}}$, in which case we can let $k = \frac{r\left(x^2+1\right)^3}{3x^2-1}$. Then, this implies by (2) that

$$r\left(1 - 2x\frac{3x^2-1}{r\left(x^2+1\right)^3}\right) = \frac{2x}{\left(1+x^2\right)^2} \Rightarrow r = \frac{8x^3}{\left(x^2+1\right)^3},$$

which in turn implies by (1) that

$$\frac{8x^3}{\left(x^2+1\right)^3} = \frac{\frac{8x^3}{\left(x^2+1\right)^3}\frac{\left(x^2+1\right)^3}{3x^2-1}}{\frac{8x^3}{\left(x^2+1\right)^3}\frac{\left(x^2+1\right)^3}{3x^2-1} - x}\frac{x}{1+x^2} = \frac{\frac{8x^3}{3x^2-1}}{\frac{8x^3}{3x^2-1} - x}\frac{x}{1+x^2} = \frac{8x^3}{\left(1+5x^2\right)\left(1+x^2\right)}$$
$$\Rightarrow \left(x^2+1\right)^3 = \left(1+5x^2\right)\left(1+x^2\right) \Rightarrow x = 0, \pm\sqrt{3}. \tag{6}$$

Hence, by (6), since $x > 0$, we see that $\frac{r}{k} = \frac{3x^2-1}{\left(x^2+1\right)^3}$ leads to a contradiction only when $x \neq \sqrt{3}$. Otherwise, when $x = \sqrt{3}$, then for a special combination of k and r, our system is degenerate, and hence, we need to check this as an edge case. When $x = \sqrt{3}$, then (1) and (2) imply that $k = 3\sqrt{3}$ and $r = \frac{3\sqrt{3}}{8}$, and as we will see once we plot the bifurcation diagram, this is exactly where a cusp catastrophe takes place. Hence, depending on how we increase and decrease combinations of k and r near this critical combination of parameters, we will expect different changes in the dynamics of the system. This behavior will become more apparent and well-describable once we see the full bifurcation diagram in 3D space (k, r, x^*).

From here, we plot the set of curves described by equations (1) and (2) in $k - r$ parameter space. We can parameterize these using x by solving equations (1) and (2) simultaneously, and after some algebra, we arrive at the following parametric equations:

$$(k, r) = \left(\frac{2x^3}{x^2-1}, \frac{2x^3}{\left(x^2+1\right)^2}\right), \quad x > 0. \tag{7}$$

Using MATLAB, plotting equations (7) generates Figure 1. There are two fixed points in both the "outbreak" and "controlled" regions, four in the "bistable" region, and in all three regions, one of the fixed points is $x^* = 0$. As one leaves the "bistable" region enclosed by the two curves and into either the "outbreak" or "controlled" regions, saddle-node bifurcations occur as two fixed points combine and then disappear. Namely, if we order the fixed points in increasing order in x, then going into the "outbreak" region, the middle two fixed points annihilate each other, while going into the "controlled" region, the largest two fixed points annihilate each other.

Finally, we can generate the full 3D bifurcation diagram by using (1) and parameterizing the necessary surface plot in k and x, i.e., with the following parameteric equations:[1]

$$(k, r, x^*) = \left(k, \frac{k}{k-x^*}\frac{x^*}{1+\left(x^*\right)^2}, x^*\right), \quad x^*, k > 0. \tag{8}$$

Using (8), we can generate Figure 2 in MATLAB. The upper part of the loop in the lighter colors indicates the "outbreak" region while the lower part coming up from $x^* = 0$ corresponds to the "controlled" region. In between these two regions, if we draw a vertical line at fixed values of k and r and these puncture the surface at four locations, we are in the "bistable" region.

The code I used to generate these plots is attached below:

[1]Note that we must also plot $x^* = 0$, as that will always be an unstable fixed point for the system.

```
xq = 0 : 0.001 : 100;
top = 2 * xq .^ 3;
kq = top ./ (xq .^ 2 - 1);
rq = top ./ ((xq .^ 2 + 1) .^ 2);

figure(1)
plot(kq, rq)
xticks(0 : 5 : 50)
title('Bifurcation diagram of the insect outbreak model',...
    'FontSize', 16)
xlim([0 50])
ylim([0 0.8])
xlabel('k*(x)', 'FontSize', 16)
ylabel('r*(x)', 'FontSize', 16)
text(15, 0.7, 'outbreak', 'FontSize', 16, 'Color', 'r')
text(25, 0.35, 'bistable', 'FontSize', 16, 'Color', 'm')
text(10, 0.1, 'controlled', 'FontSize', 16, 'Color', 'b')

k = 0 : 0.02 : 15;
xq = (0 : 0.02 : 15).';
M = length(k);
N = length(xq);
kq = ones(N, 1) * k;
rq = (kq ./ (kq - xq)) .* (xq ./ (1 + (xq .^ 2)));
xq = xq * ones(1, M);

figure(2)
h1 = surf(kq, rq, xq);
set(h1,'LineStyle','none')
title('Full bifurcation diagram of the insect outbreak model',...
    'FontSize', 14)
xlim([0.01 15])
xticks(0 : 3 : 15)
ylim([0.01 0.8])
yticks(0 : 0.2 : 0.8)
zlim([0 15])
zticks(0 : 3 : 15)
xlabel('k', 'FontSize', 14)
ylabel('r', 'FontSize', 14)
zlabel('x*', 'FontSize', 14)
hold on
h2 = surf(ones(N, 1) * k, (0 : (10 / (N - 1)) : 10).' *...
    ones(1, M), zeros(N, M));
set(h2,'LineStyle','none')
hold off
```

□

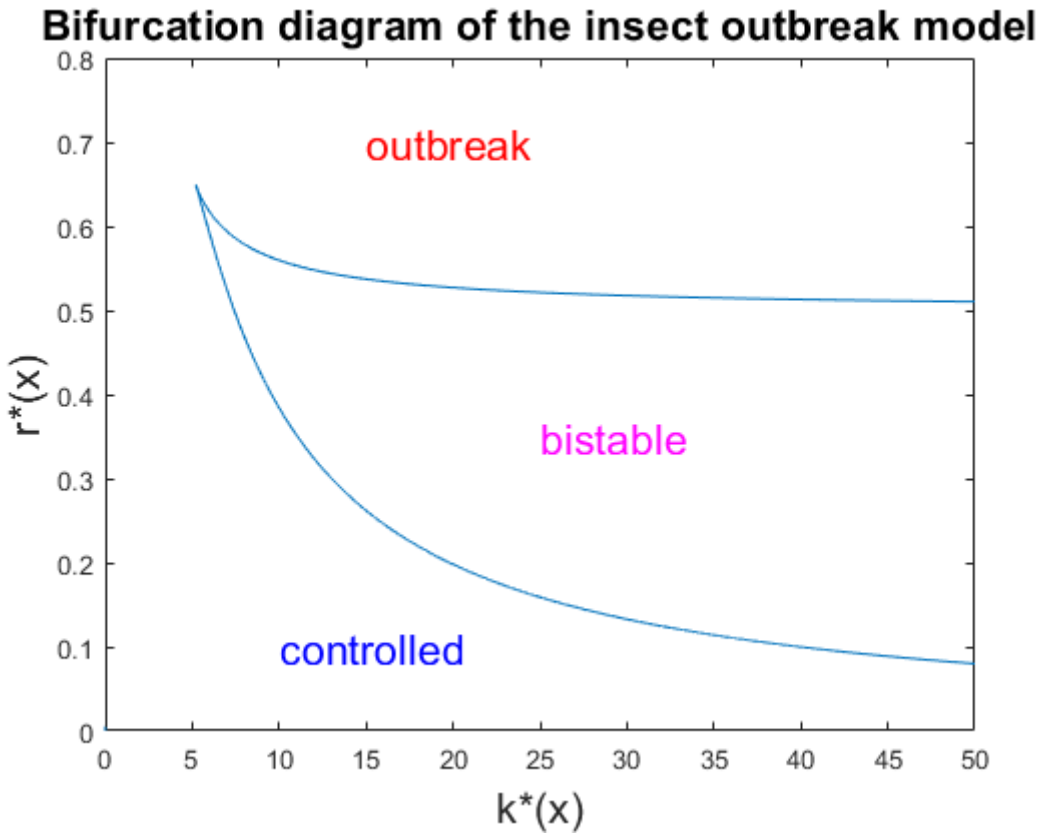

Figure 1: The 2D bifurcation diagram of the insect outbreak model described by $\dot{x} = rx\left(1-\frac{x}{k}\right) - \frac{x^2}{1+x^2}$ and projected onto $k-r$ space. The top curve asymptotes to $r = 0.5$ and the bottom one to $r = 0$ as $x \to 1$ and $x \to \infty$, respectively, and in both cases, this occurs as $k \to \infty$. Note the cusp catastrophe at $(k, r) = \left(3\sqrt{3}, \frac{3\sqrt{3}}{8}\right)$.

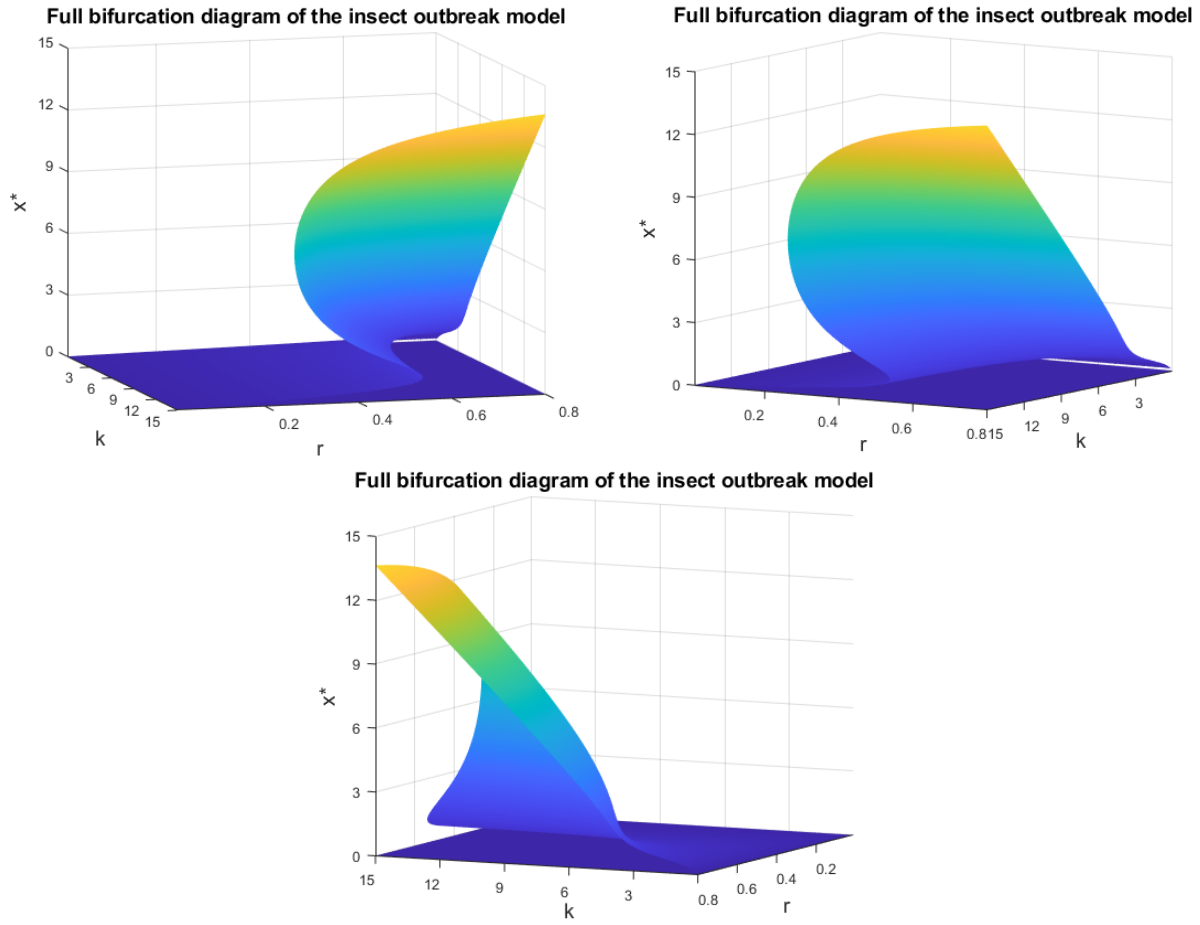

Figure 2: The full 3D bifurcation diagram of the insect outbreak model in three different viewing angles. Note the location of the cusp catastrophe in 3D space and the hysteresis loop. The hysteresis occurs as such: If we increase r too much from the darker portion of the upper surface (small or middle r), then the population jumps up to the region with the lightest colors (large r). From there, to get back to low x^*, we must ride the uppermost portion of the surface by decreasing r, perhaps even past where we started in the first place, until the population jumps back down to the second lowest fixed point, or the one right above $x^* = 0$.

Problem 2.

Solution. First off, here are the codes that I used to simulate the lattice-based random-walker SIS epidemic spreading model. I actually used two separate function scripts, one of which uses the other. The most foundational script is the first one:[2]

```
function [ numberinfected ] = randomwalk2DlatticeHWpart1(N, M, Q, L, n)
%N = linear dimension (lattice size is N^2)
%M = total population number
%Q = number of initial infected people
%L = days to recover
%n = number of total iterations

%start with agents in random locations
loc = randi(N, 2, M);
infections = zeros(1, M); %to track infections
%zero is healthy, nonzero is infected
infections(randperm(M, Q)) = L; %set Q random agent(s) to be infected
numberinfected = zeros(1, n); %to track number infected
numberinfected(1) = sum(infections > 0);

%sample directions (to add on at each iteration)
dirq = [0 0 0 1 -1; 0 1 -1 0 0];
for i = 2 : n
    infections = max(0, infections - 1); %allow infection to pass
    loc = mod(loc + dirq(:, randi(5, 1, M)), N);
    for j = find(infections > 0) % go through infected agents
        infloc = loc(:, j);
        % check where locations line up with infected
        equality = (loc == infloc);
        % check which were infected (but remove already infected)
        infections((sum(equality) == 2) & (infections == 0)) = L;
    end
    numberinfected(i) = sum(infections > 0);
end
```

The comments in the code describe how this code works pretty well, but I will still go through the code in better detail. At first, we start with all the M agents placed in random locations throughout the NxN lattice. This defines a 2xM vector, `loc`, of locations for each of the M agents, which are given as (x, y) coordinates on the lattice such that $1 \leq x, y \leq N$ and $x, y \in \mathbb{N}$. We also initialize a 1xM array tracking infections for each of the M agents, `infections`, which contains the days left until each agent is no longer infected. Consequently, all susceptible agents have a value of 0 in this vector, and to assign the initially infected people, we set Q random elements in this array to L. Finally, we want to track the number of infected patients at each iteration (of which there are n), which we do by defining an 1xn array `numberinfected`.

[2]**Note:** I have another version of this code which runs the same process but plots the lattice at each iteration, along with all of the infected and susceptible population members, and generates a movie at the end, as well as a plot of the number infected at each iteration. I also have versions of this latter code in 2D and 3D which use a continuous random walk, in which the directions and distances traveled are randomized

From here, the iteration procedure proceeds as follows. At start of each ith iteration, we subtract every element corresponding to an infected agent in `infections` by 1, indicating that one day has passed. We do this by taking `max(0, infections - 1)`, such that every element that was 0 (and hence corresponding to an uninfected agent) stays at 0 instead of going to -1. Then, for each of the M agents, we add a vector randomly chosen from the following: $(0,0)$, $(0,1)$, $(0,-1)$, $(1,0)$, and $(-1,0)$. This represents each of the five possible movements that each of the agents can take at each iteration, i.e., after each day. Of course, these addition operations will all be done modulo N, such that we satisfy the periodic boundary conditions. After doing this, we go through all of the currently infected agents, as indicated if they have a value greater than zero in `infections`. For each infected agent, we take its (x,y) coordinate and check which, if any, other agents lie on the same coordinate. For each of these agents that are currently uninfected/susceptible, if there are any, we set their corresponding value in `infections` to L+1, indicating that they have just been infected.[3] Finally, at the end of the iteration, we save the number of infected people in `numberinfected(i)`.

For this problem, we have to run multiple simulations, and of course, we could do that using the code written above. To automate the process, I wrote this second code below:

```
function [ ] = randomwalk2DlatticeHWpart2(N, M, Q, L, n)
numsimulations = 10;
data = zeros(numsimulations, n);

for i = 1 : numsimulations
    data(i, :) = randomwalk2DlatticeHWpart1(N, M, Q, L, n);
    disp(i)
end

numberinfected = mean(data);

figure
plot(1 : n, numberinfected)
title(['Infected vs. time, M=', num2str(M)])
xlabel('Time [days]')
ylabel('Number infected')
```

This code is pretty self-explanatory. Given values for N, M, Q, L, and n, the code runs `numsimulations` different simulations using `randomwalk2DlatticeHWpart1`, and then plots their average after completion. The plots are only differentiated for varying values of M, which is really all we need for this problem, since all of the other inputs are fixed.

Finally, the results of the plots above can be found below in Figure 3. Note that for large enough values of M, and in particular, $M > 1000$, the apparent fixed point at $I = 0$ becomes unstable and the system goes to some "fixed point" below M.[4]

□

for each population member at each iteration. If you would like some sample movies for each of these codes, please let me know.

[3]We set these to L+1 as opposed to L because to let the infection pass, we will subtract all nonzero elements in `infections` by 1 after this part of the loop. Hence, we want to make sure that all of these newly infected agents are still at L by the end of the iteration.

[4]I put fixed point in parentheses here because, technically, this is not a fixed point in the appropriate sense, since random fluctuations in the number of infected will always be present unless $I = 0$. However, it appears that the system approaches an *average* fixed point, i.e., as $n \to \infty$, the average number of infected

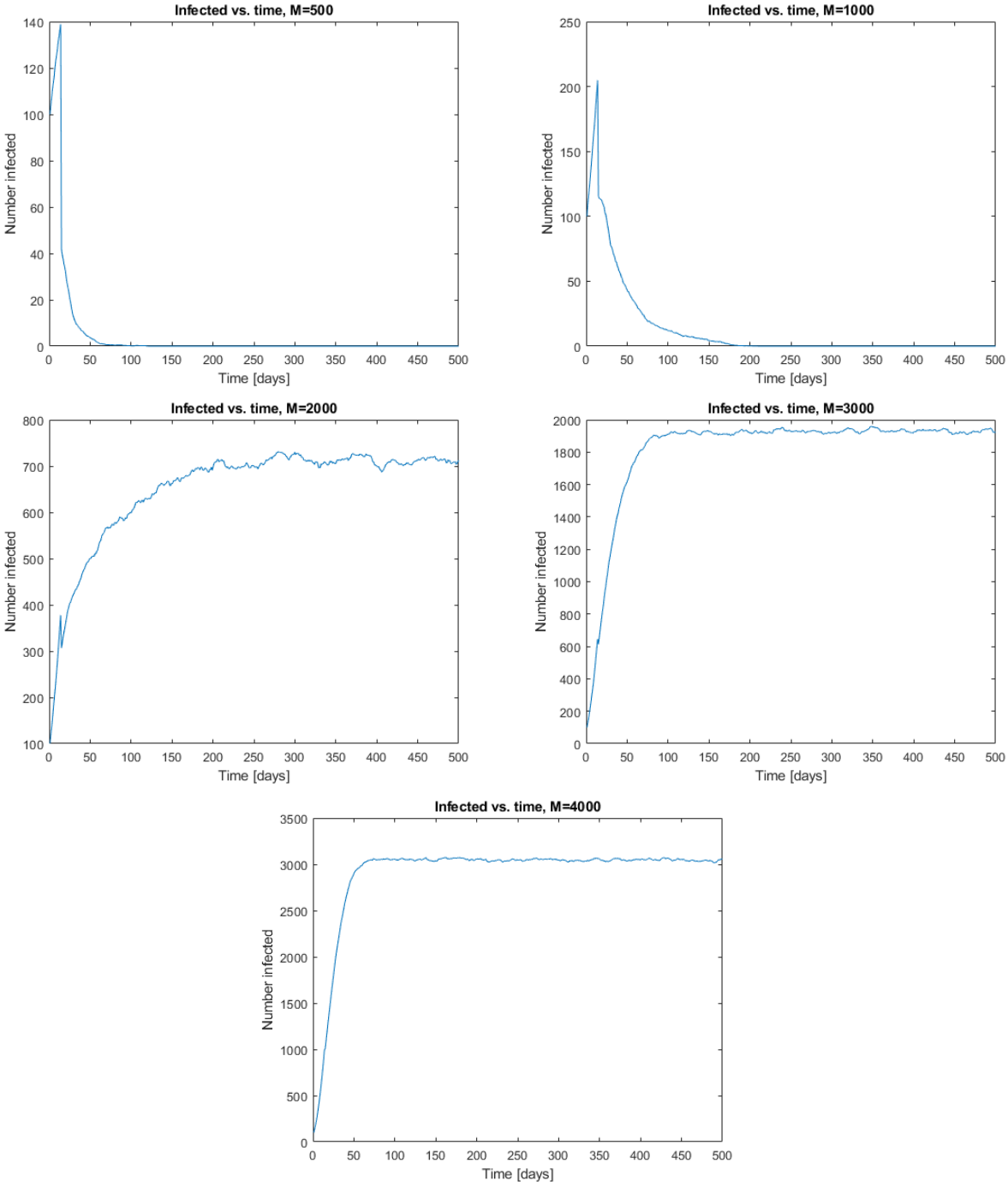

Figure 3: The five plots above summarize the average results over ten simulations each for our lattice SIS epidemic model ran with $N = 100$, $Q = 100$, $L = 14$, $n = 500$, and M as given in the title of each plot.

Problem 3. (Old version.) Below, you can find my work for when I took Professor Korenaga's words in part (b) literally when he said, " Calculate the probability of this person to spread infection at the next iteration," i.e., in rigorous terms, I calculated the probability and not the expected value. Expected value would make more sense logically within the context of the problem, and it is apparently what he intended, but I did not want to assume, and plus, in part (a), he actually uses "probability" in the way which I am familiar with. (I am an applied math student.) I do want to note that I had calculated the expected value in the first place because it seemed more natural, but redid it after second-guessing myself because that technically is not strictly the "probability," i.e., the chance of at least one agent being infected. I am keeping my work here because I did spend many hours trying to figure this out, and hence, you can see what I did. Cheers!

(a) *Solution.* Our square lattice, having linear dimension N, has $N \times N = N^2$ possible sites for the agents. If we choose one of these sites at random at any given iteration $k \in \{1, \ldots, n\}$, then the probability that there is some agent on that site is M/N^2, since there are M agents total on the lattice. Furthermore, since there are $S_k = M - I_k$ susceptible agents on the lattice at iteration k, the probability that this chosen agent is susceptible is $S_k/M = (M - I_k)/M$. Putting these two results together, since the events that there is an agent on a random lattice site and that agent is susceptible are independent,

$$p_s = \mathbb{P}\begin{bmatrix}\text{there is an agent} \\ \text{on a given spot}\end{bmatrix} \bigcap \begin{bmatrix}\text{a chosen agent} \\ \text{is susceptible}\end{bmatrix} =$$

$$\mathbb{P}\begin{bmatrix}\text{there is an agent} \\ \text{on a given spot}\end{bmatrix} \mathbb{P}\begin{bmatrix}\text{a chosen agent} \\ \text{is susceptible}\end{bmatrix} = \left(\frac{M}{N^2}\right)\left(\frac{S_k}{M}\right) = \frac{S_k}{N^2}$$

$$\Rightarrow \boxed{p_s = \frac{S_k}{N^2} = \frac{M - I_k}{N^2}}.$$

□

(b) *Solution.* Suppose that we have an infected agent, I, somewhere on the lattice. The ways that I can infect another agent depends on the position of that other agent in the prior iteration, i.e., we need to look at the number of ways this agent can come into contact with another susceptible agent, assuming that I does not recover fully and become susceptible again by the next iteration. By rotational symmetry, the possibilities of infection are the same if the infected agent moves anywhere in the next iteration. Thus, we can consider the infection possibilities for the following two scenarios, and for now, we will assume that I will still be infected at the next iteration.

Note: It is assumed that all of the arithmetic below is done mod N by the periodic boundary conditions. Hence, we omit mod N everywhere for sake of brevity.

(i) *I moves.* The chance of this happening is 4/5, as out of the five possible movements, four of these involve movement to the left, right, up, or down. By symmetry, it is enough to consider one of these scenarios, and hence, without loss of generality, suppose that I starts at some site $(x, y) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$ in

agents in the system will asymptote to some constant value. Furthermore, as we can see as M increases, these fluctuations appear not to scale with the size of M. Thus, these flucuations are in a sense "stable" and

the lattice and moves up to site $(x, y+1)$. The only way I can come into contact with another agent is if that other agent started at $(x-1, y+1)$ and moved right, started at $(x+1, y+1)$ and moved left, started at $(x, y+2)$ and moved down, or started at $(x, y+1)$ and stayed there. Also, there is an edge case: there could have been an agent which started at (x, y) and moved up alongside I. The probability of each of these cases happening is $1/5$, since each of these agents can move (or not move) in five possible ways, one of which will result in contact in each case. The only other requirement for this other agent to be infectable on the next iteration is that this agent becomes or is already susceptible once it moves to $(x, y+1)$, which makes the edge case quite marginal because any agent in that situation must have already been infected on the previous round with one day left to recover, such that it is susceptible on the next iteration.

From here, we calculate the probability that at least one infectable agent comes into contact with I on the next iteration, given that I moved. For agents that started out in $(x-1, y+1)$, $(x+1, y+1)$, $(x, y+2)$, or $(x, y+1)$, the chance of there being an infectable agent on those sites is $p_s + \frac{I_k}{LN^2}$, because they can either be susceptible or infected, but if they are infected, then they must have only one day left until recovery, which gives us the factor of $1/L$. For an agent which started out on (x, y), it could not have been susceptible at first. Thus, the chance of there being an infectable agent there is just $\frac{I_k}{LN^2}$. Furthermore, assuming that we have an infectable agent on either of these sites, the chance in each case that it lands on $(x, y+1)$ for the next iteration is $1/5$, as discussed earlier.

Hence, by the above considerations, we can reformulate this problem as the chance that we have at least one success, given that we have five trials. Out of these five trials, four will each succeed with probability $\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)$, while one will succeed with probability $\frac{I_k}{5LN^2}$. This probability is equal to $1 - \bar{p}_1$, where $\bar{p}_1$ is the probability that we have all failures. Thus, $\bar{p}_1 = \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)$, and in conclusion, the probability that I infects at least one person on the next iteration, given that I moved, is $1 - \bar{p}_1 = 1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)$.

(ii) *I stays.* The chance of this happening is $1/5$, as there are five possible movements any agent can take, but only one of these involves staying put. Then, I stays at (x, y), and the only way I will spread infection is if an agent in $(x-1, y)$ moved left, an agent in $(x, y+1)$ moved down, an agent in $(x+1, y)$ moved right, an agent in $(x, y-1)$ moved up, or an agent in (x, y) stayed there, and each of these happens with probability $1/5$. Furthermore, each of these agents would have to be infectable on the next iteration. The chance of there being an infectable agent on $(x-1, y)$, $(x, y+1)$, $(x+1, y)$, or $(x, y-1)$ is still $p_s + \frac{I_k}{LN^2}$ for reasons described in case (i). Similarly, the chance of there being an infectable agent on (x, y) is $\frac{I_k}{LN^2}$, since again, such an agent could not have started off susceptible. Hence, in the same manner as in part (i), we can think of our desired probability as the chance of at least one success over five trials, four of which succeed with probability $\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)$ and the last with probability $\frac{I_k}{5LN^2}$. Then, following the same steps, we deduce that the probability that I infects at least one person on the next iteration, given that I moved, is still $1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)$.

Putting parts (i) and (ii) together above, by the law of total probability, if we denote P the probability that I infects at least one person *given* that I will still be infected

seemingly bounded.

at the next iteration,

$$P = \mathbb{P}\left[I \text{ spreads infection } | I \text{ moves}\right] \mathbb{P}\left[I \text{ moves}\right] +$$
$$\mathbb{P}\left[I \text{ spreads infection } | I \text{ does not move}\right] \mathbb{P}\left[I \text{ does not move}\right] =$$
$$\left[1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right]\left(\frac{4}{5}\right) +$$
$$\left[1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right]\left(\frac{1}{5}\right) =$$
$$1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right).$$

Finally, since the probability that I infects at least one person given that I recovers by the next iteration is zero, we conclude that

$$p_i = \mathbb{P}\left[I \text{ does not recover}\right] P =$$
$$\left(\frac{L-1}{L}\right)\left[1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right] \Rightarrow$$
$$\boxed{p_i = \left(\frac{L-1}{L}\right)\left[1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right]}.$$

□

(c) *Solution.* Each infected person has some number of days left to recover, but we know that the number of days left ranges from 1 to L, giving us L choices. If an infected person recovers by the next iteration, they must have had only one day left to recover. Hence, this is one choice from the list $\{1, \ldots, L\}$ of possible days left to recover, such that $\boxed{p_r = 1/L}$. □

(d) *Solution.* Using (b) and (c), we write

$$dI = I\left(p_i - p_r\right) = I\left[\left(\frac{L-1}{L}\right)\left(1 - \left(1 - \frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right) - \frac{1}{L}\right] =$$
$$I\left[\left(\frac{L-1}{L}\right)\left(1 - \left(1 - \frac{M - \left(\frac{L-1}{L}\right) I_k}{5N^2}\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right) - \frac{1}{L}\right],$$

which already gives us a fixed point $I = 0$. If $I \neq 0$, then

$$\left(\frac{L-1}{L}\right)\left(1 - \left(1 - \frac{M - \left(\frac{L-1}{L}\right) I_k}{5N^2}\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right) - \frac{1}{L} = 0 \Rightarrow$$
$$(L-1)\left(1 - \left(1 - \frac{M - \left(\frac{L-1}{L}\right) I_k}{5N^2}\right)^4 \left(1 - \frac{I_k}{5LN^2}\right)\right) = 1. \tag{9}$$

(9) can be solved explicitly, but the expression is unbelievably complicated.

This is where I realized that Professor Korenaga actually meant expected value instead of the actual, true probability value. □

Problem 3. (New version.) In my final version of problem 3, I took the "probability" in Problem 3(b) to actually refer to the expected value in the more rigorous sense of the term, which makes better sense within the context of this problem. Parts (a) and (c) do not change.

(a) *Solution.* Our square lattice, having linear dimension N, has $N \times N = N^2$ possible sites for the agents. If we choose one of these sites at random at any given iteration $k \in \{1, \ldots, n\}$, then the probability that there is some agent on that site is M/N^2, since there are M agents total on the lattice. Furthermore, since there are $S_k = M - I_k$ susceptible agents on the lattice at iteration k, the probability that this chosen agent is susceptible is $S_k/M = (M - I_k)/M$. Putting these two results together, since the events that there is an agent on a random lattice site and that agent is susceptible are independent,

$$p_s = \mathbb{P}\begin{bmatrix}\text{there is an agent} \\ \text{on a given spot}\end{bmatrix} \bigcap \begin{bmatrix}\text{a chosen agent} \\ \text{is susceptible}\end{bmatrix} =$$

$$\mathbb{P}\begin{bmatrix}\text{there is an agent} \\ \text{on a given spot}\end{bmatrix} \mathbb{P}\begin{bmatrix}\text{a chosen agent} \\ \text{is susceptible}\end{bmatrix} = \left(\frac{M}{N^2}\right)\left(\frac{S_k}{M}\right) = \frac{S_k}{N^2}$$

$$\Rightarrow \boxed{p_s = \frac{S_k}{N^2} = \frac{M - I_k}{N^2}}.$$

□

(b) *Solution.* **Note:** For this problem, we will do all calculations for the lattice coordinates in modulo N arithmetic.

Suppose that we have an infected person I at some site $(x, y) \in \{1, \ldots, N\}^2$ in the lattice. First off, we will consider all the possible places that a susceptible (or soon to be susceptible) agent will be to become infected upon the next iteration. We will call such agents "infectable." All infectable agents lie within two steps from (x, y) on the initial iteration. They must, as the agents will only come into contact on the next iteration if they both move towards each other or if one stays while the other moves towards it. With this introduction to our situation, we consider the following two possibilities for infection, assuming that I does not recover by the next iteration:

(i) *Case 1: I moves.* The chance of this happening is 4/5, as out of the five possible movements, four of these involve movement to the left, right, up, or down. By symmetry, it is enough to consider one of these scenarios, and hence, without loss of generality, suppose that I starts at some site $(x, y) \in \{1, \ldots, N\}^2$ in the lattice and moves up to site $(x, y + 1)$. The only way I can come into contact with another agent is if that other agent started at $(x - 1, y + 1)$ and moved right, started at $(x + 1, y + 1)$ and moved left, started at $(x, y + 2)$ and moved down, or started at $(x, y + 1)$ and stayed there. Also, there is an edge case: there could have been an agent which started at (x, y) and moved up alongside I. The probability of each of these cases happening is 1/5, since each of these agents can move (or not move) in five possible ways, one of which will result in contact in each case. The only other requirement for this other agent to be infectable on the next iteration is that this agent becomes or is already susceptible once it moves to $(x, y + 1)$, which makes the edge case quite marginal because any agent in that situation must have already been infected on the previous round with one day left to recover, such that it is susceptible on the next iteration.

From here, we calculate the probability that each infectable agent comes into contact with I on the next iteration, given that I moved. For agents that started out in $(x-1, y+1)$, $(x+1, y+1)$, $(x, y+2)$, or $(x, y+1)$, the chance of there being an infectable agent on those sites is $p_s + \frac{I_k}{LN^2}$, because they can either be susceptible or infected, but if they are infected, then they must have only one day left until recovery, which gives us the factor of $1/L$. For an agent which started out on (x, y), it could not have been susceptible at first. Thus, the chance of there being an infectable agent there is just $\frac{I_k}{LN^2}$. Furthermore, assuming that we have an infectable agent on either of these sites, the chance in each case that it lands on $(x, y+1)$ for the next iteration is $1/5$, as discussed earlier.

Hence, the probability that agents begin in $(x-1, y+1)$, $(x+1, y+1)$, $(x, y+2)$, or $(x, y+1)$ and get infected is $\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)$ each, and the probability that an agent starts out in (x, y) and gets infected is $\frac{I_k}{5LN^2}$. Any agent which starts anywhere else is not infectable by I in the next iteration. This allows us to calculate the expected value of infections. Let X_M be the random variable that represents the number of agents that I infects on the next iteration when I moves. Then, we can write $X_M = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{1}_{E_{(a,b)}}$, where $E_{(a,b)}$ is the event that moving I infected an agent that started at point $(a, b) \in \{1, \ldots, N\}^2$ and $\mathbb{1}_S$ is the indicator function of any set S. Thus, by linearity of expectation and the fact that $\mathbb{E}\left[\mathbb{1}_{E_{(a,b)}}\right] = \mathbb{P}\left[E_{(a,b)}\right]$,

$$\mathbb{E}\left[X_M\right] = \mathbb{E}\left[\sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{1}_{E_{(a,b)}}\right] = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{E}\left[\mathbb{1}_{E_{(a,b)}}\right] = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{P}\left[E_{(a,b)}\right] =$$
$$\mathbb{P}\left[E_{(x-1,y+1)}\right] + \mathbb{P}\left[E_{(x+1,y+1)}\right] + \mathbb{P}\left[E_{(x,y+2)}\right] + \mathbb{P}\left[E_{(x,y+1)}\right] + \mathbb{P}\left[E_{(x,y)}\right] =$$
$$4\left(\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right) + \frac{I_k}{5LN^2} = \frac{4}{5}p_s + \frac{I_k}{LN^2}.$$

Therefore, if I moves, the expected number of infected agents is $\frac{4}{5}p_s + \frac{I_k}{LN^2}$.

(ii) *Case 2: I does not move.* The chance of this happening is $1/5$, as there are five possible movements any agent can take, but only one of these involves staying put. Then, I stays at (x, y), and the only way I will spread infection is if an agent in $(x-1, y)$ moved left, an agent in $(x, y+1)$ moved down, an agent in $(x+1, y)$ moved right, an agent in $(x, y-1)$ moved up, or an agent in (x, y) stayed there, and each of these happens with probability $1/5$. Furthermore, each of these agents would have to be infectable on the next iteration. The chance of there being an infectable agent on $(x-1, y)$, $(x, y+1)$, $(x+1, y)$, or $(x, y-1)$ is still $p_s + \frac{I_k}{LN^2}$ for reasons described in case (i). Similarly, the chance of there being an infectable agent on (x, y) is $\frac{I_k}{LN^2}$, because again, such an agent could not have started off susceptible. Thus, since the probability of each of these necessary movements occurring is $1/5$ each, the probability that agents begin in $(x-1, y)$, $(x, y+1)$, $(x+1, y)$, or $(x, y-1)$ and get infected is $\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)$ each, and the probability that an agent starts out in (x, y) and gets infected is $\frac{I_k}{5LN^2}$. If we define X_S to be the random variable that represents the number of agents that I infects on the next iteration when I stays put, then we can similarly write $X_S = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{1}_{F_{(a,b)}}$, where $F_{(a,b)}$ is the event that non-moving I infected an agent that started at point

$(a,b) \in \{1,\ldots,N\}^2$, giving us the following computation:

$$\mathbb{E}[X_S] = \mathbb{E}\left[\sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{1}_{F_{(a,b)}}\right] = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{E}\left[\mathbb{1}_{F_{(a,b)}}\right] = \sum_{b=1}^{N}\sum_{a=1}^{N} \mathbb{P}\left[F_{(a,b)}\right] =$$
$$\mathbb{P}\left[F_{(x-1,y)}\right] + \mathbb{P}\left[F_{(x,y+1)}\right] + \mathbb{P}\left[F_{(x+1,y)}\right] + \mathbb{P}\left[F_{(x,y-1)}\right] + \mathbb{P}\left[F_{(x,y)}\right] =$$
$$4\left(\frac{1}{5}\left(p_s + \frac{I_k}{LN^2}\right)\right) + \frac{I_k}{5LN^2} = \frac{4}{5}p_s + \frac{I_k}{LN^2}.$$

Therefore, if I does not move, the expected number of infected agents is still $\frac{4}{5}p_s + \frac{I_k}{LN^2}$.

Finally, we can put cases (i) and (ii). As stated earlier, if we let M be the event that I moves and S the event that I does not move, then $\mathbb{P}[M] = 4/5$ and $\mathbb{P}[S] = 1/5$. Hence, via the law of total expectation, if we denote X as the random variable which represents the number of agents that I infects on the next iteration, then $X|M = X_M$ and $X|S = X_S$, such that

$$\mathbb{E}[X] = \mathbb{E}[X|M]\,\mathbb{P}[M] + \mathbb{E}[X|S]\,\mathbb{P}[S] =$$
$$\left(\frac{4}{5}p_s + \frac{I_k}{LN^2}\right)\left(\frac{4}{5}\right) + \left(\frac{4}{5}p_s + \frac{I_k}{LN^2}\right)\left(\frac{1}{5}\right) = \frac{4}{5}p_s + \frac{I_k}{LN^2}.$$

However, note that this analysis assumes that I does not recover by the next iteration. If I recovers by the next iteration, which we will see happens with probability $1/L$ in part (c), then all bets are off — no infections will take place. On the other hand, the chance that I is still infected by the next iteration is $1 - 1/L = (L-1)/L$. Therefore, our final answer is actually $\boxed{p_i = \frac{L-1}{L}\left(\frac{4}{5}p_s + \frac{I_k}{LN^2}\right)}$, as desired.[5] □

(c) *Solution.* Each infected person has some number of days left to recover, but we know that the number of days left ranges from 1 to L, giving us L choices. If an infected person recovers by the next iteration, they must have had only one day left to recover. Hence, this is one choice from the list $\{1,\ldots,L\}$ of possible days left to recover, such that $\boxed{p_r = 1/L}$. □

(d) *Solution.* Using (b) and (c), we write

$$dI = I(p_i - p_r) = I\left(\frac{L-1}{L}\left(\frac{4}{5}p_s + \frac{I_k}{LN^2}\right) - \frac{1}{L}\right) =$$
$$I\left(\frac{L-1}{L}\left(\frac{4}{5}\frac{M - I_k}{N^2} + \frac{I_k}{LN^2}\right) - \frac{1}{L}\right),$$

which already gives us a fixed point $I^* = 0$. If $I^* \neq 0$, then

$$\frac{L-1}{L}\left(\frac{4}{5}\frac{M - I^*}{N^2} + \frac{I^*}{LN^2}\right) - \frac{1}{L} = 0 \Rightarrow I^* = \frac{L(4LM - 5N^2 - 4M)}{(L-1)(4L-5)},$$

such that our two fixed points for the system are $\boxed{I^* = 0, \frac{L(4LM - 5N^2 - 4M)}{(L-1)(4L-5)}}$. □

[5] **Note:** If we had reformulated our SIS model such that almost-recovered agents, i.e., those who were

(e) *Solution.* Before doing anything, we should classify the stability of these two fixed points, which will be useful in part (e). Consider that as M increases, the stability of $I^* = 0$ appears to change in the simulations, and over some critical value of M, $I^* = 0$ goes from stable to unstable, and another stable fixed point appears above 0. This is the hallmark of a transcritical bifurcation, and hence, we should expect one in our analysis below.

For discrete dynamical systems of the form $x_{n+1} = f(x_n)$, one usually looks at the value of $f'(x)$ at the fixed point x^* in question. If $|f'(x^*)| < 1$, then x^* is stable, and if $|f'(x^*)| > 1$, then x^* is unstable. Else, when $|f'(x^*)| = 1$, we must look at higher-order terms to determine the stability. In particular, if the third condition occurs in combination with a parameter value, then a bifurcation occurs there.

For our system defined above,

$$f(I_n) = I_n + dI_n = I_n + I_n\left(\frac{L-1}{L}\left(\frac{4}{5}\frac{M-I_n}{N^2} + \frac{I_n}{LN^2}\right) - \frac{1}{L}\right) =$$
$$\frac{L-1}{L}I_n\left(1 + \frac{4}{5}\frac{M-I_n}{N^2} + \frac{I_n}{LN^2}\right) \Rightarrow f'(I^*) =$$
$$\frac{L-1}{L}\left(1 + \frac{4}{5}\frac{M-I^*}{N^2} + \frac{I^*}{LN^2}\right) + \frac{L-1}{L}I^*\left(-\frac{4}{5N^2} + \frac{1}{LN^2}\right) =$$
$$\frac{L-1}{5L^2N^2}\left(4LM + 5LN^2 - 8LI^* + 10I^*\right). \tag{10}$$

Hence, using (10), we can test the stability for $I^* = 0, \frac{L\left(4LM-5N^2-4M\right)}{(L-1)(4L-5)}$ as $L, M, N \in \mathbb{N}$ each change.

(i) *Case 1:* $I^* = 0$. Then,

$$f'(I^*) = \frac{L-1}{5LN^2}\left(4M + 5N^2\right),$$

such that I^* is stable when

$$-1 < \frac{L-1}{5LN^2}\left(4M + 5N^2\right) < 1 \iff$$
$$-\frac{5LN^2}{L-1} < 4M + 5N^2 < \frac{5LN^2}{L-1}. \tag{11}$$

Since $L \in \mathbb{N}$, we know that $L \geq 1 \Rightarrow -\frac{5LN^2}{L-1} < 0$. However, $4M + 5N^2 > 0$, such that the lower inequality in (11) will always hold. Thus, we only need

$$4M + 5N^2 < \frac{5LN^2}{L-1} \Rightarrow 4M < \frac{5LN^2}{L-1} - 5N^2 \Rightarrow 4M(L-1) < 5N^2 \tag{12}$$

as the necessary condition for stability of $I^* = 0$. Similarly, when $4M(L-1) > 5N^2$, I^* will be unstable, and when $4M(L-1) = 5N^2$, some kind of bifurcation occurs, to be determined.

infected on the previous iteration and become susceptible on the next, could not be immediately reinfected afterwards, then our answer would only include the p_s term, i.e., $p_i = 4p_s/5$.

(ii) *Case 2:* $I^* = \frac{L\left(4LM-5N^2-4M\right)}{(L-1)(4L-5)}$. Then,

$$f'\left(I^*\right) = \frac{L-1}{5L^2N^2}\left(4LM + 5LN^2 + (10-8L)\,\frac{L\left(4LM-5N^2-4M\right)}{(L-1)\,(4L-5)}\right) = \frac{L-1}{5L^2N^2}\left(\frac{L\left(4M+5N^2+5LN^2-4LM\right)}{L-1}\right) = \frac{1}{5LN^2}\left(4M+5N^2+5LN^2-4LM\right),$$

such that I^* is stable when

$$-1 < \frac{1}{5LN^2}\left(4M+5N^2+5LN^2-4LM\right) < 1 \iff$$
$$-5LN^2 < 4M+5N^2+5LN^2-4LM < 5LN^2 \iff$$
$$4LM - 10LN^2 < 4M + 5LN^2 < 4LM. \tag{13}$$

$M \leq N^2$ will always hold due to physical constraints, which implies that $4LM \leq 4LN^2 < 10LN^2 \Rightarrow 4LM - 10LN^2 < 0 < 4M + 5LN^2$, such that the lower bound in (13) will always hold. Thus, we just need $4M + 5LN^2 < 4LM \Rightarrow 5LN^2 < 4M\,(L-1)$ for stability of I^*, and similarly, I^* is unstable when $5LN^2 > 4M\,(L-1)$.

Comparing the results of cases (i) and (ii), we find that the conditions for the stabilities of $I^* = 0$ and $I^* = \frac{L\left(4LM-5N^2-4M\right)}{(L-1)(4L-5)}$ are precisely swapped; one is stable when the other is unstable, and vice-versa. Furthermore, when $4M\,(L-1) = 5N^2$, then these fixed points coincide at $I^* = 0$. Hence, we conclude that at $4M\,(L-1) = 5N^2$, a transcritical bifurcation occurs, as we guessed earlier.

Let us compare our stability analysis above with the results from Problem 2. For our parameters, L and N are fixed, such that $I^* = 0$ is stable when $4M\,(14-1) < 5(100)^2 \Rightarrow M < 12500/13 \approx 962$. Unfortunately, according to our simulations, we should still have a stable fixed point at $I^* = 0$ when M is as large as $M = 1000$, which already indicates a flaw in our model.

Before testing our analysis directly from parts (a)-(d) with the actual agent-based modeling results, as a sanity check, Professor Korenaga stated in class that, in a sense, the model from parts (a)-(d) has a "natural" connection with the continuous-time ODE SIS epidemic model from class. To show that this is true (albeit, rather heuristically), we want to look at the limit when the time interval represented by each iteration gets arbitrarily small, say, within a size Δt. Hopefully, we can take Δt at some time to approach a time derivative. We know that

$$dI_k = I_{k+1} - I_k \Rightarrow \frac{dI_k}{\Delta t} = \frac{I_{k+1}-I_k}{\Delta t} = \frac{\Delta I_k}{\Delta t}. \tag{14}$$

Thus, since $\lim_{\Delta t \to 0} \frac{\Delta I_k}{\Delta t} = \frac{dI(t_k)}{dt}$, we should define necessary new parameters such that the limit $\Delta t \to 0$ makes sense on the left side of (14).

$$\frac{dI_k}{\Delta t} = \frac{I_k}{\Delta t}\left(\frac{L-1}{L}\left(\frac{4}{5}\frac{S_k}{N^2} + \frac{I_k}{LN^2}\right) - \frac{1}{L}\right) = I_k\left(\frac{4}{5}\frac{L-1}{L}\frac{S_k}{N^2\Delta t} + \frac{L-1}{L}\frac{I_k}{LN^2\Delta t} - \frac{1}{L\Delta t}\right) = \frac{4}{5}\frac{L-1}{LN^2\Delta t}I_kS_k - \frac{1}{L\Delta t}I_k + \frac{L-1}{L^2N^2\Delta t}I_k^2. \tag{15}$$

Clearly, the last term in (15) should disappear to get our model in the form desired in lecture. This term corresponds only to cases when an infected agent recovers and then is immediately reinfected *on the same iteration*, and heuristically, we should expect this term to vanish as $\Delta t \to 0$ in the continuous time limit since the time interval on which any agent could do something like this would go to zero. Furthermore, note that in the continuous limit, L should scale with Δt somehow; after all, as Δt goes to zero, the terms with L should still make sense. This leads us to define $\alpha := (L\Delta t)^{-1}$, where α is the recovery rate, and in terms of L, $L = (\alpha \Delta t)^{-1}$.

Finally, note from earlier that the factors involving p_i also involved the probability that each infected agent came into contact with a susceptible agent on the next iteration. As the time interval Δt goes to zero, we would expect this probability also to go to zero, and hence, we are missing a factor of Δt somewhere. To resolve this, note how for our lattice construction, we assume that each agent travels with speed 1, such that after 1 unit of time, the agent finally reaches the point it would have reached after one iteration in the original, non-coarse-grained model. If we want to break up our iterations into time intervals $\Delta t \ll 1$, we have to extend our lattice accordingly and insert some "dummy" sites on the lattice that correspond to intermediate locations. But since each agent can take a step up, down, left, right, or neither at each iteration, this is just a 2D random walk. Hence, in order to travel a distance d, m timesteps are required, where $m \approx (d/l)^2$ and l is the stepsize. For our case, $m = 1/\Delta t$ and $d = 1$. Accordingly, solving for l, $1/\Delta t \approx (1/l)^2 \Rightarrow l = \sqrt{\Delta t}$. This means that we need to replace $N \to N/\sqrt{\Delta t}$, since that represents how many more spots we need to consider *on average* when calculating the probability of two or more agents coming into contact at a given time.

Putting our work in the above two paragraphs together with (15), we find that

$$\begin{aligned}
\frac{dI_k}{\Delta t} = \frac{4}{5}\frac{(\alpha\Delta t)^{-1} - 1}{(\alpha\Delta t)^{-1}\left(N/\sqrt{\Delta t}\right)^2 \Delta t} I_k S_k - \frac{1}{(\alpha\Delta t)^{-1}\Delta t} I_k + \frac{(\alpha\Delta t)^{-1} - 1}{(\alpha\Delta t)^{-2}\left(N/\sqrt{\Delta t}\right)^2 \Delta t} I_k^2 = \\
\frac{4}{5}\frac{1 - \alpha\Delta t}{(N^2/\Delta t)\,\Delta t} I_k S_k - \alpha I_k + \frac{\alpha\Delta t - \alpha^2 (\Delta t)^2}{(N^2/\Delta t)\,\Delta t} I_k^2 \Rightarrow \\
\frac{\Delta I_k}{\Delta t} = \frac{dI_k}{\Delta t} = \frac{4}{5}\frac{1 - \alpha\Delta t}{N^2} I_k S_k - \alpha I_k + \frac{\alpha\Delta t - \alpha^2 (\Delta t)^2}{N^2} I_k^2. \qquad (16)
\end{aligned}$$

Hence, using (14) and taking the limit of both sides of (16) as $\Delta t \to 0$, we find that

$$\frac{dI_k}{dt} = \frac{4}{5N^2} I_k S_k - \alpha I_k, \qquad (17)$$

and defining $\mu := 4/(5N^2)$, (17) reproduces the same coarse-grained ODE model given in lecture.

Finally, we should compare our difference equation predictions with those from the actual agent-based model. The comparisons between these models can be found in Figure 4. Clearly, there is a rather noticeable discrepancy between the results that the agent-based model gave us and the ones we derived via probability theory, but two general trends we may notice are that (1) when M is near the bifurcation value, the discrepancy between the plots is the largest, and (2) as M increases, the difference equation results get closer to the agent-based simulation results.

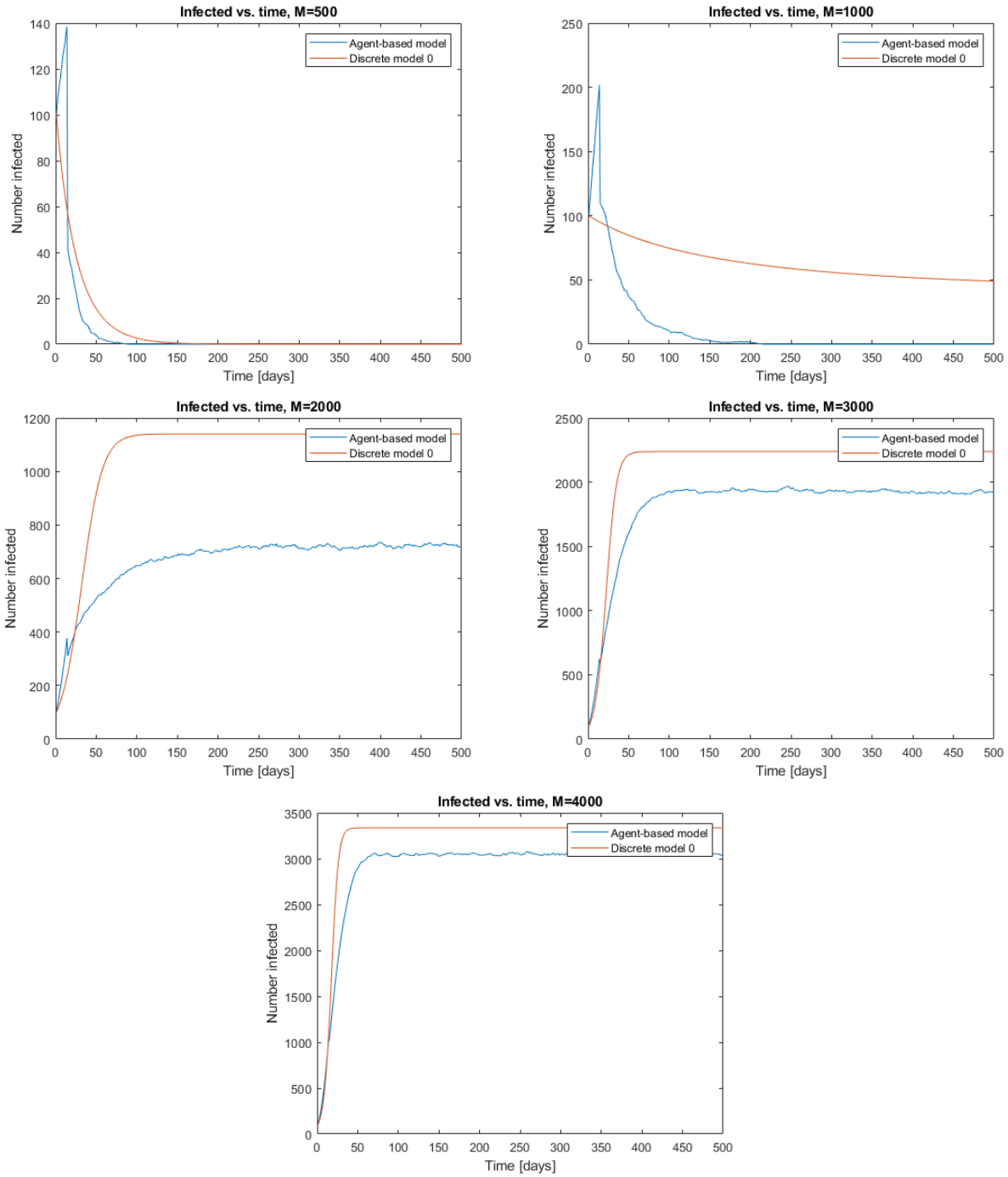

Figure 4: The five plots above summarize the average results over ten simulations each for our lattice SIS epidemic model ran with $N = 100$, $Q = 100$, $L = 14$, $n = 500$, and M as given in the title of each plot in comparison with the uncorrected difference equation we derived in parts (a)-(d).

Below, I will list and discuss some possible causes for the discrepancies evident in Figure 4:

- *We are undercounting, but only towards the start of the simulation.* The prior spatial distribution of infected people will certainly affect the probability that a given infected person will recover. In particular, near the beginning of the simulation, the number of infected people will climb much quicker, since no infected people can recover at that point, but there will come a point at around $n = L$ when infected people will start recovering. This phenomenon would explain the small spike we see in the number of infected for the agent-based model which becomes more pronounced as M gets smaller. However, this does not explain the discrepancies as n gets large.
- *We are overcounting.* We are double-, triple-, or even quadruple-counting cases where two or more infected move to the same lattice site and infect the same agents there. For instance, in our current model, if two separate infected people moved to the same spot and infected two other people there on the next iteration, we would effectively count those as four newly infected people, even though it is the same pair of people being infected.
- *We are overcounting, again.* The probability of finding susceptible and infected people across the domain is not uniform, even though we assumed so by how we used p_s uniformly regardless of which infected agent we were considering at each iteration, as well as when considering the probability of finding soon-to-recover infected. Here is a counterexample demonstrating the implications of this phenomenon: Suppose that you had two scenarios, each with the same number of infected and susceptible people on lattices 1 and 2 of the same sizes. However, say that on lattice 1, you had all the infected people in a large clump, or several clumps, with the susceptible people scattered elsewhere. On the other hand, say that on lattice 2, the infected people were independently and identically distributed across the whole domain. Which case would you expect to have a greater increase in infections at the next iteration? Obviously, the case of lattice 2.

  Intuitively, we would expect that this clumping occurs in cases where L is large; since the infected stay infected for a while when L is large, it is easy for regions to become dominated by infected people. In the limit as $L \to \infty$, you should actually get a random walk where the infected represent the walkers and susceptible are sort of just ambient space, and the amount of susceptible agents and their distribution controls the diffusion speed and direction of the random walkers.

Out of all of these above, I would hypothesize that the largest sources of discrepancy are related to the spatial complexity in the agent-based SIS model. If I were able to show you the movies I made of the simulations for middle values of M, i.e., those around $M = 1000$ and $M = 2000$, then you would easily be able to see how much spatial organization there is and the sheer cohesion of the clumps of infected people that form around the initial places where we first placed the infected at the start of the simulation. Although, all of the above three bullets indicate that we are in overcounting the number of newly infected at each iteration, which among other things easily explains what we observe in Figure 4.

Below, I will try my best to address each of these three above discrepancies in order. (After all, that roughly follows the complexity of each fix.) For the first case, it is

simple enough to rewrite our difference equation as follows:

$$I_{n+1} = \begin{cases} I_n\left(1 + \frac{4}{5}\frac{M-I_n}{N^2}\right), & 1 < n \leq L-1 \\ \frac{L-1}{L} I_n\left(1 + \frac{4}{5}\frac{M-I_n}{N^2} + \frac{I_n}{LN^2}\right), & n \geq L. \end{cases} \tag{18}$$

The correction in (18) is simple. Up until $n = L$, there are no infected people who will recover on the next iteration. Thus, $p_i = \frac{4}{5}p_s$ and $p_r = 0$ when $n < L$, but once $n \geq L$, we will use our usual difference equation. This gives us (18) above.

Taking care of the overcounting will be trickier, but it is doable explicitly. We first need to calculate the expected number of infected people in the next round who will be the only infected in their new site, the expected number of pairs of infected (for double-counting), the expected number of triplets of infected (for triple-counting), etc.[6] We will denote these expected values as E_1, E_2, etc. respectively, and in the end, we need to replace the Ip_i term with $\sum_{k=1}^{\infty} E_k N_k$, where N_k is the expected number of infections at a lattice site given that there are k infected agents who are at that site, because of course, this number will depend on how many infected agents have crammed themselves around or in it.

We start off by handling the E_k's. For a given lattice site, if any agent is going to reach that site by the next iteration, that agent must have started at that site or at a site which is at most one step away. Hence, there are five possible sites where an infected agent could have came from. The probability that any of these sites has an infected agent is $\frac{L-1}{L}\frac{I_k}{N^2}$, which follows because there are I_k infected agents in total, N^2 sites in total, and every infected agent has a $\frac{L-1}{L}$ chance of staying infected at the next iteration. Furthermore, the probability that such an infected agent moves correctly, i.e., to the given lattice site, is $1/5$. Hence, we can reformulate our problem as a Bernoulli experiment with five trials, each with success probability $\frac{L-1}{5L}\frac{I_k}{N^2}$.

Since the above formulation holds at any coordinate in our lattice, we can use the same indicator formulation of expectations that we used in part (b). Using that, because there are N^2 total sites in our lattice, we deduce that E_k is equal to N^2 multiplied by the probability that our Bernoulli experiment above has *exactly* k successes and $5-k$ failures. Therefore, $E_k = N^2 \binom{5}{k} \left(\frac{L-1}{5L}\frac{I_k}{N^2}\right)^k \left(1 - \frac{L-1}{5L}\frac{I_k}{N^2}\right)^{5-k}$.

Finally, we move on to calculate the N_k's. We start by referring back to part (b) and our indicator formulation to solve for p_i. Our probabilities for the placement of the infectable agents do not change in either case; we just have to restrict the number of possible prior places, because we now know that some number of these were occupied by infected agents. From there, we can just take the average over all possible configurations for the infected agents in each case.

Actually, I just realized that an issue with this formulation is that we have not accounted for the possibility that multiple infected agents could be occupying one of those five possible sites outlined earlier. But unfortunately, since I am unfortunately out of time, I shall attach some plots which show the results from my first correction in (18). Indeed, my hypothesis was correct: The small spike near the beginning of the agent-based model was captured by this correction to the difference equation model, as we can see in Figure 5.

[6] We do not need to consider the expected number of slots with no infected, because of course, these will contribute nothing to the calculation of dI.

I should also close off by saying that, to correct the final, third bullet on the list regarding spatial organization, we would have to track and calculate the local infection probabilities for each agent with days to recover greater than 1.[7] This would involve considering all nearest neighbors and next-nearest neighbors for each given infected agent that will stay infected next round, and keeping track of their infected vs. susceptible statuses. Then, we could just sum over all of the directly-calculated expected number of infections for each infected agent, and we would have to do this for each iteration.

However, at this point, the computational benefits of using a difference equation approximation would be severely limited, as you would have to keep track of the spatial data at each iteration; at that point, you might be better off just using the agent-based model as is. Nevertheless, I had an interesting question: Since it appears that the degree of spatial organization depends off of some combination of the given parameters, is there some uniform way to scale p_i such that it would "adjust" as a measure of spatial localization imposed by certain parameters? Suppose that we call this adjusted probability $\tilde{p}_i$. Perhaps $\tilde{p}_i$ would scale with the entropy in the system; the greater the entropy, the more that $\frac{\tilde{p}_i}{p_i} \to 1$, but otherwise, we should always expect $\tilde{p}_i < p_i$. Maybe we can even impose that $\tilde{p}_i = C(L, M, N)\, p_i$, in which case we could determine C empirically via more simulations as in Problem 2. Of course, we would expect $0 \leq C(L, M, N) \leq 1$ for any $L, M, N \in \mathbb{N}$.

□

[7] If they have only one day to recover, then they cannot infect another agent on the next iteration, since they will have recovered by then.

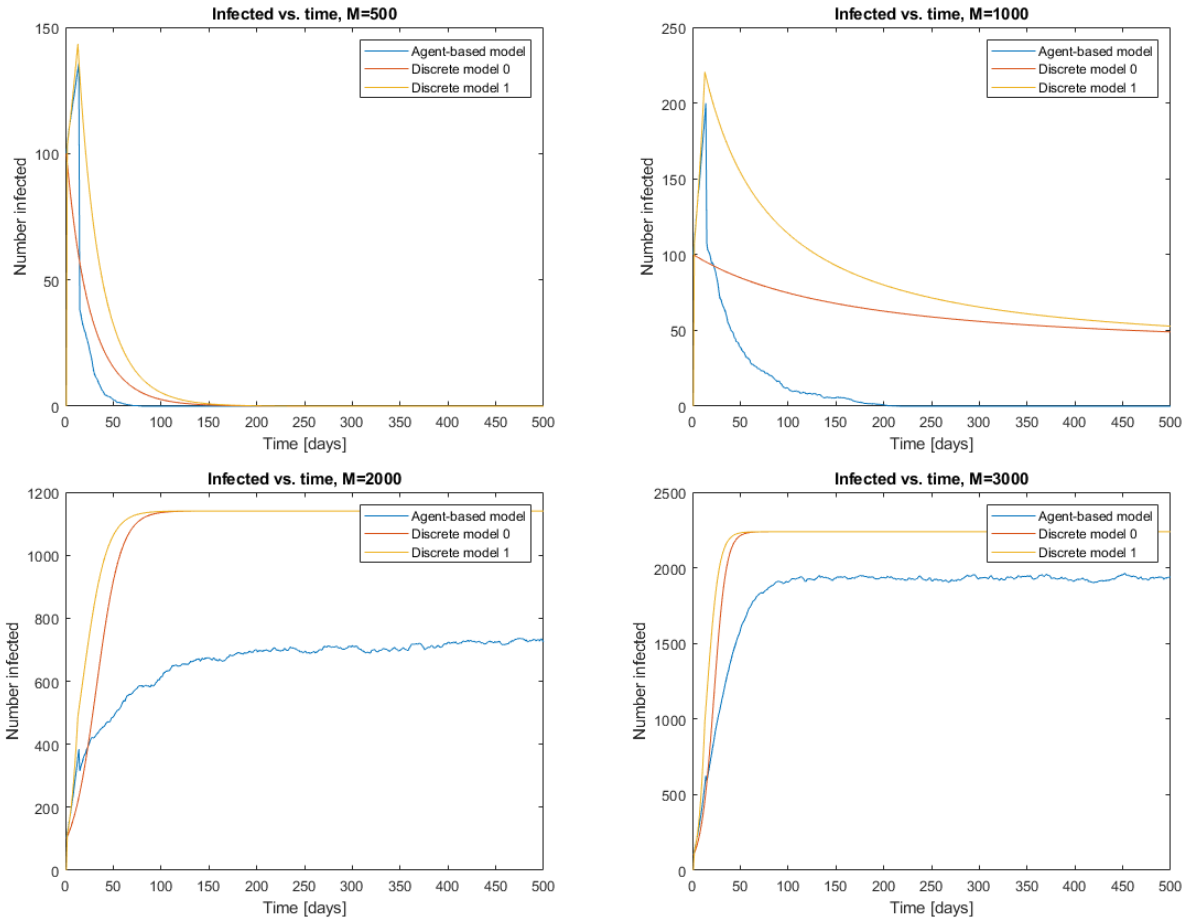

Figure 5: The five plots above summarize the average results over ten simulations each for our lattice SIS epidemic model ran with $N = 100$, $Q = 100$, $L = 14$, $n = 500$, and M as given in the title of each plot in comparison with the uncorrected and first correction to the difference equation we derived in parts (a)-(d).