

# Section 0: Abstract

---

There are four major sections in this report. In section 1 we will give a brief introduction to the SIR model, the notation we will use, and the different variations we investigated. In section 2 we will introduce our implementations for the basic SIR model in details we got in midterm checkpoint, the spatial SIR model that we did for the final, and the variations we have proposed in the midterm report. We will show the updates details of both class and script implementations as well as the comparisons holding the same variables and our corresponding findings. Sections 3 will be a conclusion part, where we will summarize our results, discuss what we can do to improve our models, and point out interesting and meaningful directions for further investigation. Section 4 will be a bibliography.

## Section 1: Introduction of SIR model and used notation

---

SIR model is a classical model to predict the number of infected people by utilizing three major different states of a person:

1. Susceptible: the individual who has not caught the disease but they can be infected. Actually they are the only source in this model to get infected.
2. Infected: an individual has already been caught the disease and can spread the disease to susceptible individuals. They are the only source to spread the disease in the default settings.
3. Removed: the people who were previously infected but now either can't be infected or infect others.

We use the recommended model parameters  $b$  and  $k$  listed in the SIR.md in our script to simulate, where  $b$  represents the number of interactions each day that could spread the disease (per individual) and  $k$  represents the fraction of the infectious population which recovers each day. We will introduce the specific notations we used for simulation in part 3.

Note that to make the results easier to interpret, in the ODE-based model we further divide the group of removed into two groups of recovered and dead.

## Section 2: Basic SIR models (agent-based model and ODE model), our variations and the spatial SIR models

---

### Basic SIR models built in midterm project: some brief descriptions

---

#### Implementation of the agent-based SIR model

We implement the agent-based model following the instructions provided on class: first of all we constructed a class "Person" where we can update states of the person for the agent-based model. We specify three possible status for each agent: S (susceptible), I (infected) and R (Removed). When an agent has his/her status changed, we set the previous status to false and assign him/her the new status.

Then in the simulation, we also follow the similar implementations provide on class: we set fixed amount of contacts for each person and set the initial proportion of infected people to be 0.1%. Upon each contact, we update the status for each person and observe the trend of spreading.

#### Implementation of the ODE-based SIR model:

We implement the ODE-based model using the information of the basic SIR model in the ODE setting. The set of notations and equations are below:

```
S: the number of susceptible people
I: the number of infectious people (note the difference between the basic SIR model;
R: the number of removed people

To make the implementation easier, we also have:

s: the proportion of susceptible people
i: the proportion of infectious people (note the difference between the basic SIR model;
r: the proportion of removed people

Here, by our settings

 $s(t) = S(t)/N$ 
 $i(t) = I(t)/N$ 
 $r(t) = R(t)/N$ 

where N is the total number of people.

The set of ODE we are considering are:

 $ds/dt = -b*s(t)*i(t)$ , b here indicates infected individual has a fixed number b of contacts per day
 $di/dt = b*s(t)*i(t) - k*i(t)$ , where k is the fraction of infected turned into "removed" during a day
 $dr/dt = k*i(t)$ 
```

Then we use "solve\_ivp" to solve the system of equations. Still, we set initial proportion of infected people to 0.1%.

## Implementation of variations proposed in the midterm report

### Variation 1: We tried multiple choices of strength of the enforcement of masks and quarantine within the set-ups of both the agent-based model and ODE-based model:

For the agent-based model, we update the class of agents (called "Person"). Besides the given states S, I and R, we also consider the conditions which induce great effect on the infection trend that happen in our real life: whether people wear masks and whether people follow the advice to be quarantined. To initiate the class, we set self.susceptible be true and other attributes to be false since except for the given number of infected people, people are set default to be susceptible. Then we define some functions to represent transitions such as got\_infected and got\_removed by changing states of the booleans. We also include a transition\_to\_removed function to take the authorized 14 - days quarantine period into consideration. During the simulation, we set the total population to be 5000 and explore the influences brought by fraction of the infectious population which recovers each day and the number of interactions each day that could spread the disease.

Here, some values related to real-life scenarios (e.g: mask effectiveness and quarantine rate) are taken according to some related data and researches about Covid-19.

For the ODE-based model we add several new parameters. Here are the updated notations and set of equations:

```
I. About the updated parameters
p: mask effectiveness, default is 0
q: proportion of people who wear masks, default is 0
g: quarantine level, default is 1, which refers to no quarantine. 0 refers to perfect quarantine.
k: fixed fraction k of the infected group will recover during any given day
d: fixed fraction d of the infected group will die during any given day
b: infected individual has a fixed number b of contacts per day
```

Also, we change and add some notations. Here are the updated notations:

**N**: total population, default is 5000  
**r**: the proportion of recovered people and  
**d**: the proportion of deaths.

Here, by our settings:

$r(t) = R(t)/N$   
 $d(t) = D(t)/N$

The equations are also updated with the new parameters and implementations:

```
ds/dt = -b*g*(1-p*q)*s(t)*i(t)
di/dt = b*g*(1-p*q)*s(t)*i(t)-k*i(t)-d*i(t)
dr/dt = k*i(t), k here refers to the fraction of infected turned into "recovered" during a day
dd/dt = d*i(t), d here refers to the fraction of infected turned into "dead" during a day
```

We considered four cases:

Case A: no any kind of protections

Case B: only protection is mask, **while not** all people are wearing masks ( $p=0.8$ ,  $q=0.76$ )

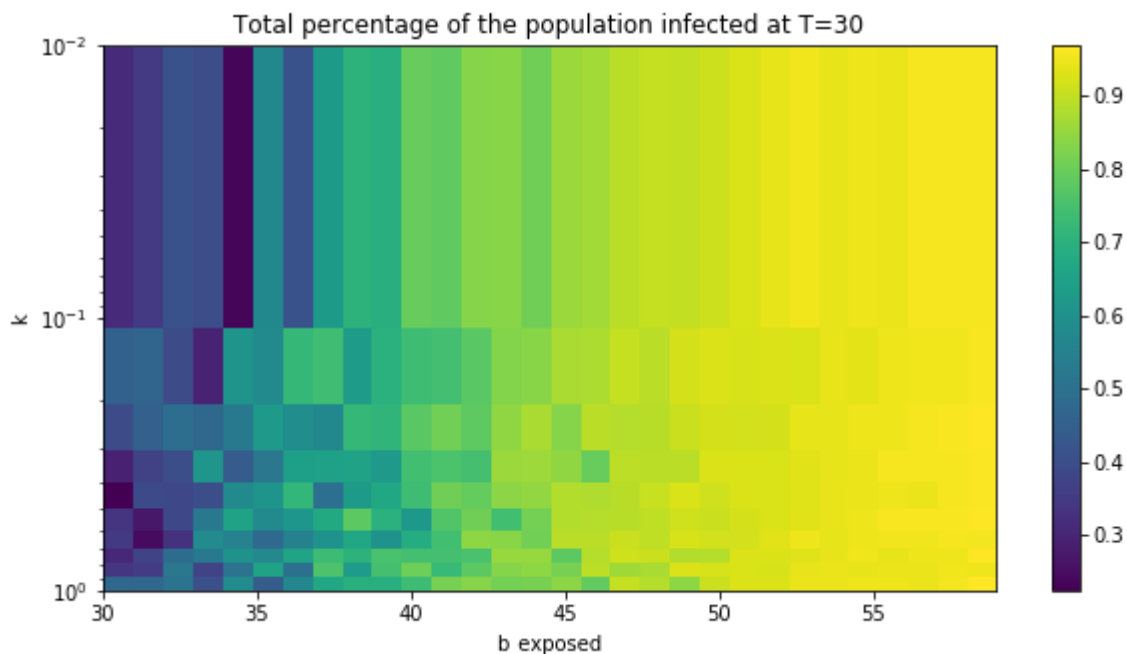
Case C: masks with patients quarantined ( $p=0.8$ ,  $q=0.76$ ,  $g=0.5$ )

Case D: High Pressure: lockdown (more strict quarantine) **and** mandatory masks (a much higher proportion of people wearing masks)

Here,  $p$ ,  $q$  and  $g$  are also taken according to some related data and researches about Covid-19.

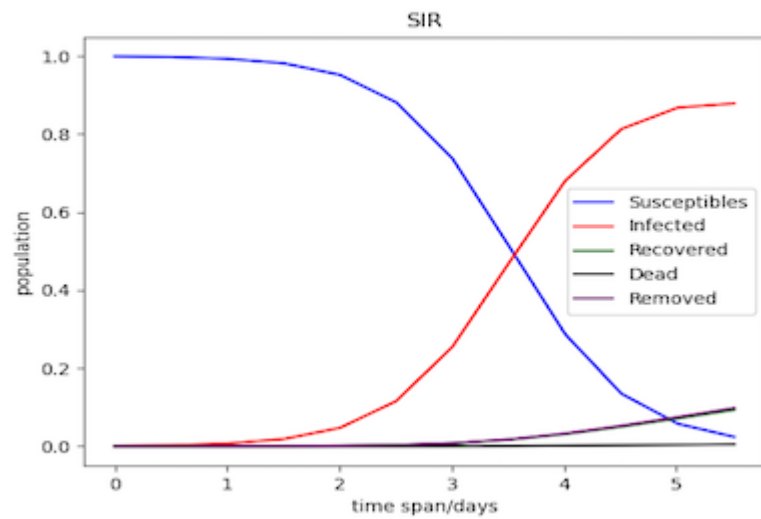
Here are some selected plots and observations:

## For the updated agent-based model

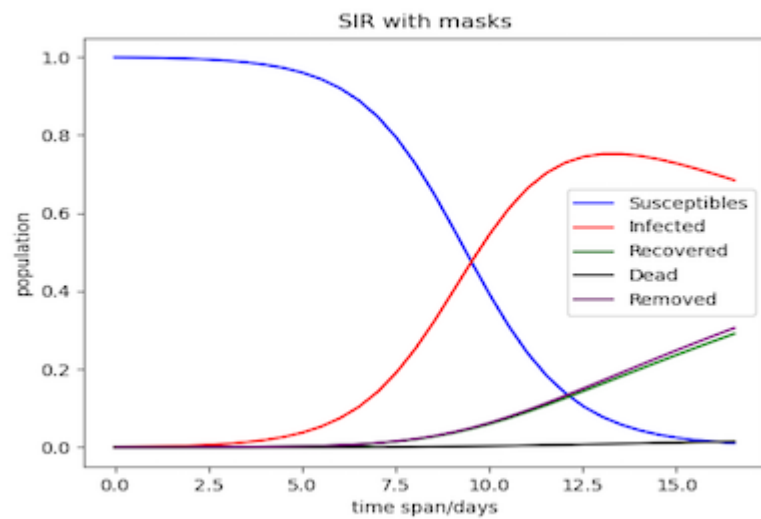


## For the updated ODE-based model

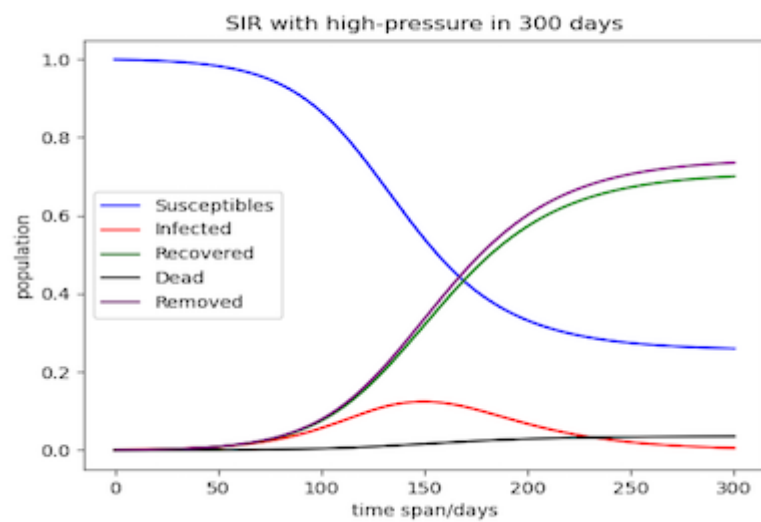
Plot1: no any kind of protections



Plot2: only with limited masks enforced



Plot3: High-pressure scenario



From the plots of the ODE-based model, we see that a fair amount of people wearing masks help slowing down the spread to some extent but it is still not effective enough. However without quarantine, all the population get infected in a short time period. However, with the help of quarantine and a more strict policy on wearing masks, the spread could be inhibited much more effectively. In the high-pressure settings, a fair amount of people will not get infected in a relatively long time period.

## Variation 2: We included the randomness of the number of interactions each day and the infectivity of diseases

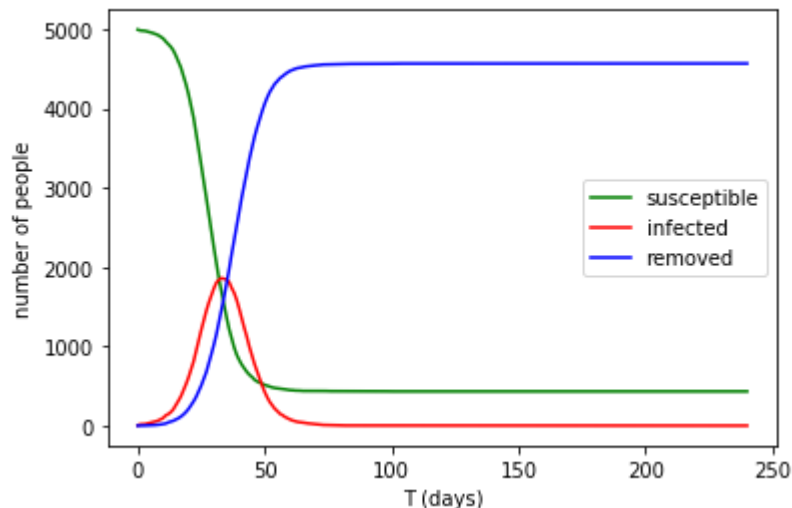
For this variation, we only need to update the simulation part. We used to fix the number of interactions per person/per day, and the probability that a close contact is infected, which will be used to represent the infectivity of a disease.

First, for the 'randomness of interactions' part, we set 'b' as the upper bound of interactions (i.e. an infected can interact with b people at most) and the following script helps us to take randomness into consideration:

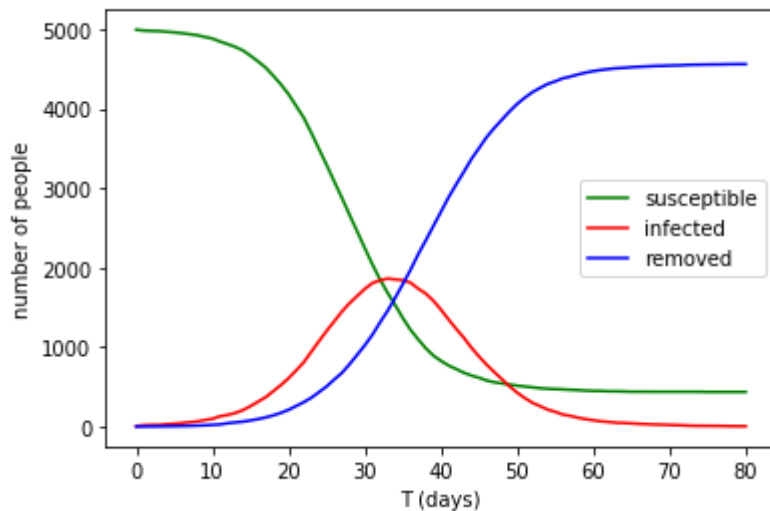
```
contacts = randint(N, size=b)
```

We consider both cases and obtain the plots:

Do not take the randomness of the number of interactions into consideration

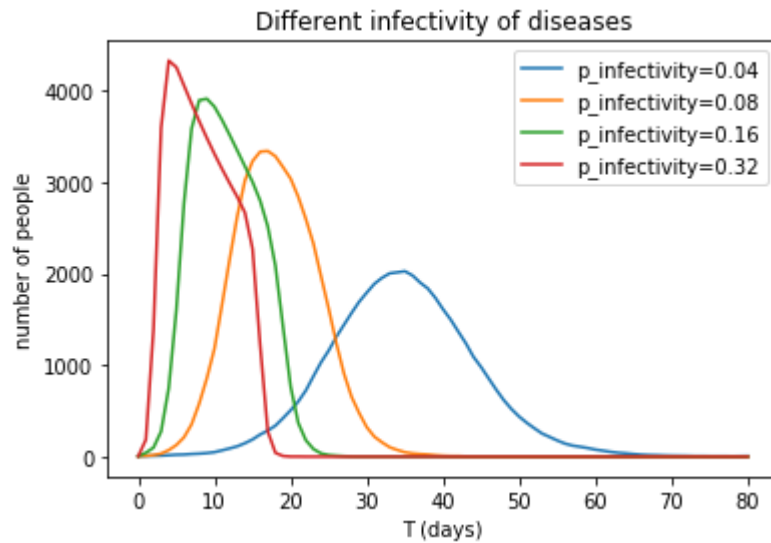


Take the randomness of the number of interactions into consideration



It is obvious that after taking the randomness into account, the disease spreads slower. But it takes longer time for the number of infected to get back to zero and the numbers of infected in total have no big difference.

As for the infectivity of diseases, instead of taking  $p = 0.037$ , we take  $p_{\text{infectivity}} = [0.04, 0.08, 0.16, 0.32]$  and plot for each value:



We can tell from this graph that infectivity matters much. The higher 'p' is, the rapidlier the disease will spread and the more people will get infected.

### Variation 3: We built a more robust and complete model: "SEIR" model with consideration of environmental factors.

In this model we add two new parameters:

v: environmental factor of virus transmission (for example, by transmission from close contact with wildlife)  
w: transmission rate from "Exposed" to infectious

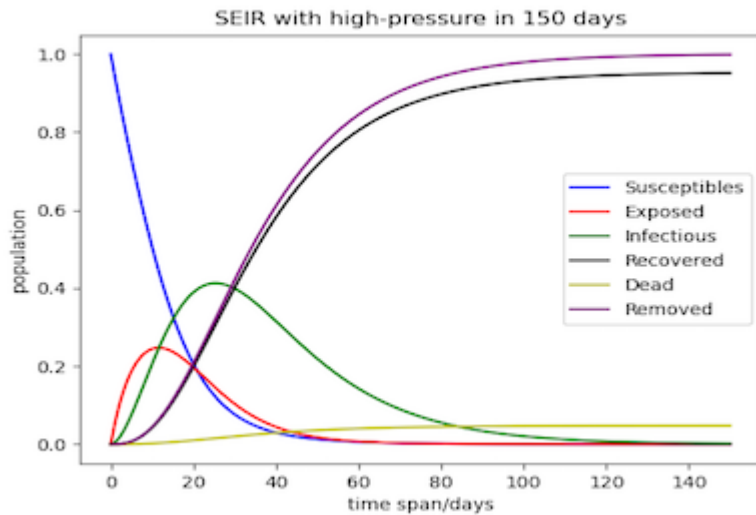
We add a new notation E and I:

E: the number of "exposed", which means infected but **not** yet infectious. **Note** that here we assume a person ne  
I: infectious, has the ability **to** spread the virus

Updates of equations for implementations:

```
y = transpose[s, e, i, r, d]
ds/dt = y0'(t) = -b*g*(1-p*q)*s(t)*i(t) - v*(1-p*q)*s(t) = -b*g*(1-p*q)*y0*y2 - v*(1-p*q)*y0
de/dt = y1'(t) = b*g*(1-p*q)*s(t)*i(t) + v*(1-p*q)*s(t) - w*e(t) = b*g*(1-p*q)*y0*y2 + v*(1-p*q)*y0 - w * y1
di/dt = y2'(t) = w*e(t)-k*i(t)-d*i(t) = w*y1 - k*y2 - d*y2
dr/dt = y3'(t) = k*i(t) = k*y2
dd/dt = y4'(t) = d*i(t) = d*y2
```

We still considered the four same cases as the updated SIR model. Here is the result we get from the similar high-pressure setting as the updated SIR ODE-based model in variation1:



We observe that even we set the environmental transmission factor to be relatively small, since the quarantine for infectious people will not stop this transmission process, all people get infected really soon. Also, under the settings of a relatively short "incubation period" of the virus, the spread does not seem to be slowed significantly.

Comparing to real-life data for the Covid-19 and our model in variation1, we could probably propose that the environmental transmission might not be really significant in the actual world, or we should expect a much more rapid increasing of patients.

## Spatial agent-based model implemented for final submission

### Updated class "Person\_spatial"

Based on the basic SIR model we built in the midterm checkpoint, We add a (2-dimensional) spatial component to our simulations to confine the people in the square  $[0,1] \times [0,1]$ . The difference that showed in our coding part is that we created a new class named "Person\_spatial" and takes in a new input "orig\_b\_exposed" to get us the corresponding interaction radius in the prompt,  $q$ . The relationship between  $p$  and  $q$ , given by the prompt, is  $b = N * \pi * (q^2)$ .

Also, we add the new feature, "position" (called pos in the class, which is formed by pos\_x and pos\_y), to each of the agent. We used random.uniform to initialize people's position uniformly within the domain at random. Then, at each time step, each individual takes a step of length  $p$  in a random direction, we achieve this by using  $*self.dpos = p * self.dpos / np.linalg.norm(self.dpos)*$ , which means each person can move with a step length of  $p$  at any direction once they do not travel out of bounds. If the person would go out the unit square, the person just stays at the same location. This has been implemented correctly in our functions.

### Updated simulations:

Compared to the script we had in the midterm checkpoint, we include a new parameter " $q$ ". People will interact with all the other people within radius  $q$  after they move according to a step length of  $p$ . It means that we do not need the parameter  $b$  anymore since  $q$  is the parameter that decides the interaction now.

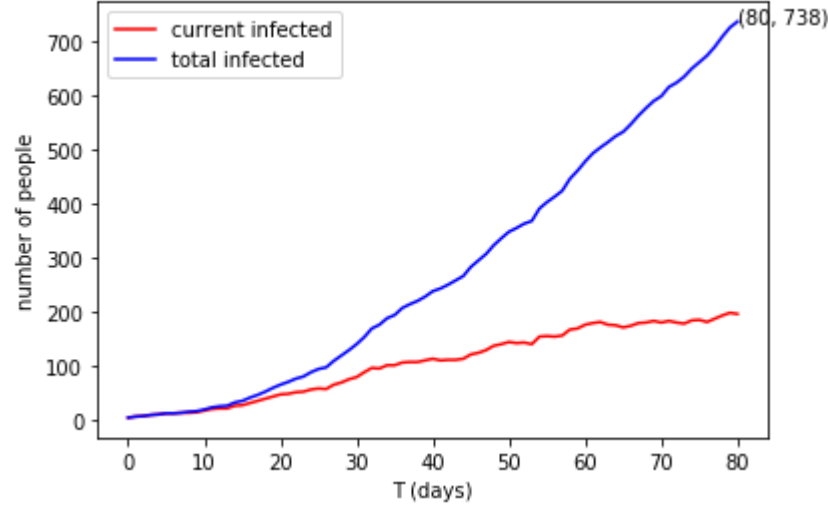
Answer to the question **If infected individuals start at the center of the square, how does the total number of individuals infected throughout the simulation change with the parameter  $p$ ?**

**This question requires the comparison of choosing different step length  $p$  given the initial position is the center of the square:**

**Fix  $q = 0.0437$  (referring to  $b = 30$ ) and the time period equals to 80:**

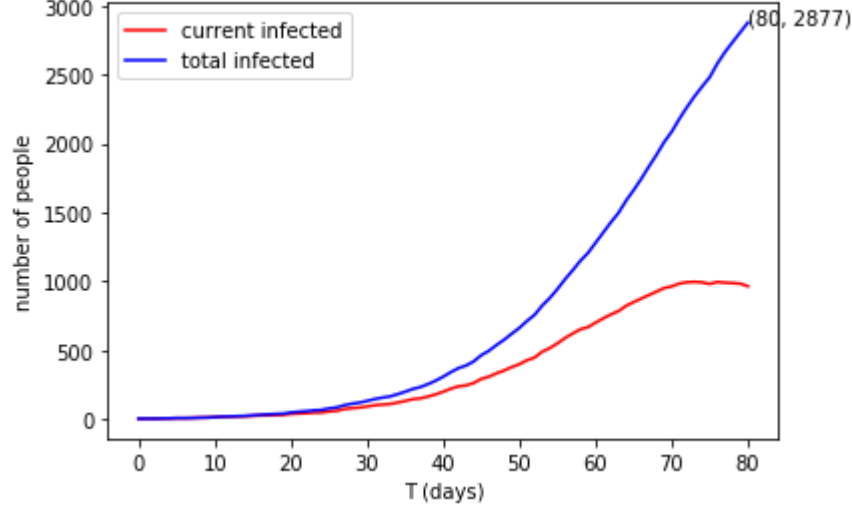
$p = 0.01$

Spatial Component Introduced with  $p=0.01$   $q=0.043701937223683165$



$p = 0.4$

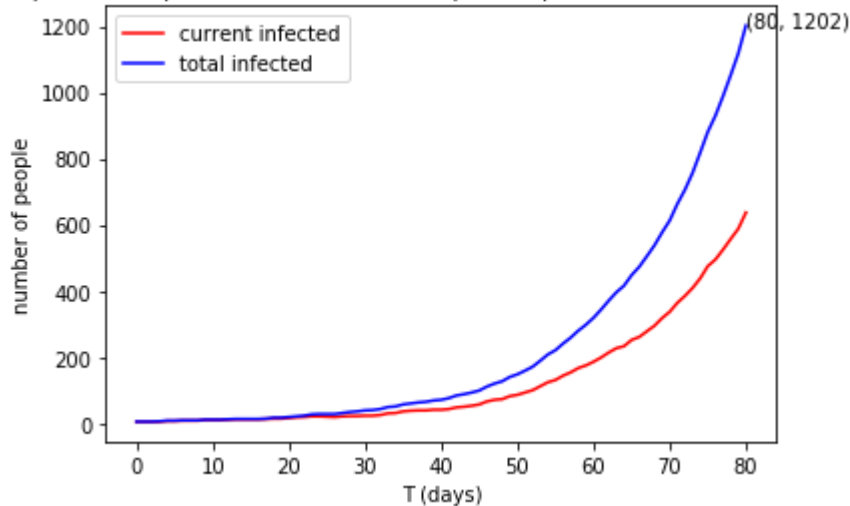
Spatial Component Introduced with  $p=0.4$   $q=0.043701937223683165$



$p = 0.7$

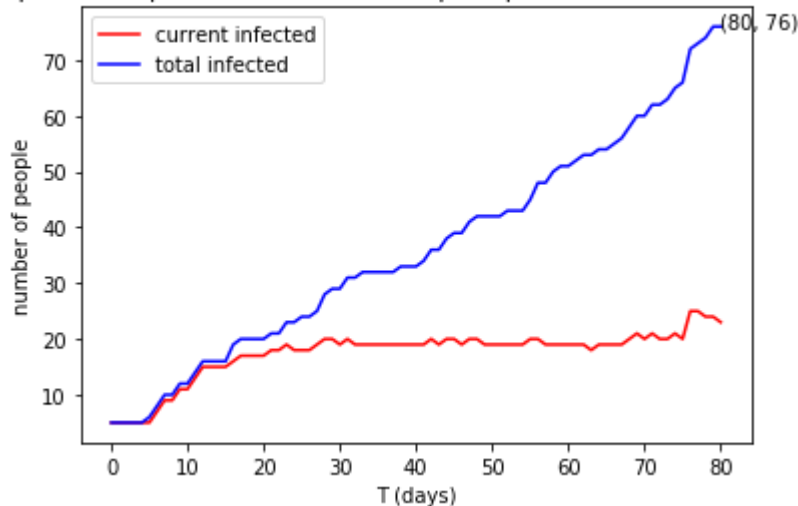


Spatial Component Introduced with  $p=0.7$   $q=0.043701937223683165$



$p = 1$

Spatial Component Introduced with  $p=1$   $q=0.043701937223683165$



## Analysis

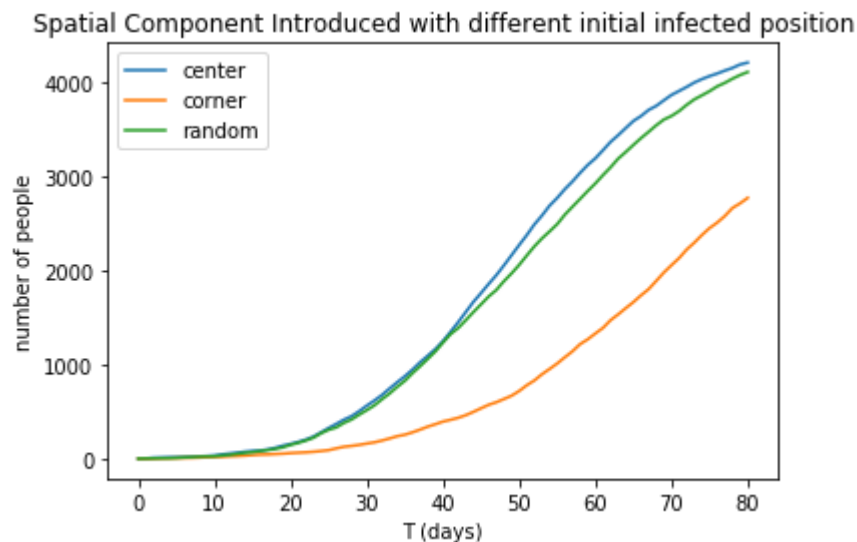
From the graphs we can see that when  $p$  is smaller than or equal to 0.01, the number of total infected people grows relatively slow and the current infected won't reach it's maximum before  $T = 80$ . As  $p$  gets larger, the number of total infected people grows faster and the current infected can reach it's maximum before  $T = 80$ . The reason is that when  $p$  is small, the step size of people is small and thus people can only effect a small group of new surrounding people. As the the step size gets larger, people can get into a brand new neighborhood and thus get more uninfected people infected. However, we can also find that when  $p$  grows larger than 0.7, the speed of total and current infected people growth goes down drastically compared to the same time period of different  $p$ , the reason is that we set the people to stay at the original place when they step out of the designated area, so in this situation they are more likely to just stay at the original position and thus get fewer people infected.

Answer to the question **Choose an interesting parameter of  $p$  using question 1. How does the simulation qualitatively differ when the initial infected individuals start in a single corner of the square vs. the center of the square vs. being randomly spread out?**

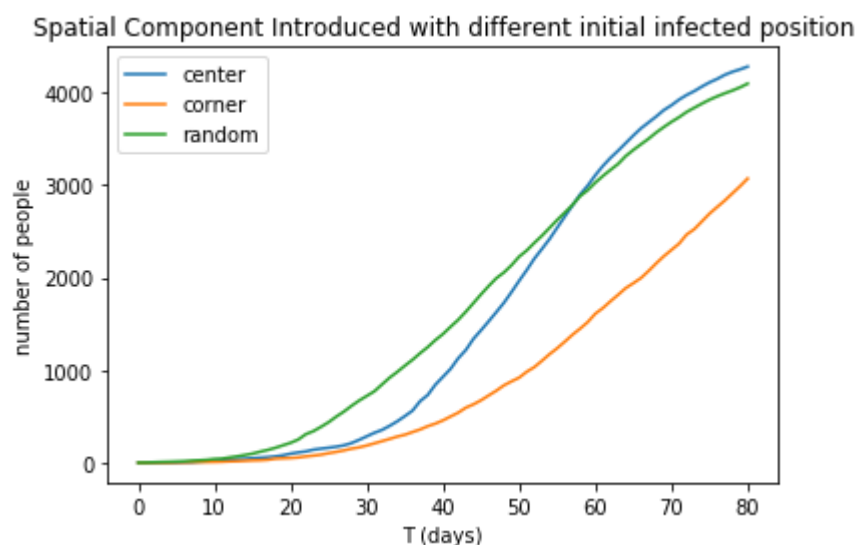
**This question requires the comparison of choosing different initial places of patients:**

In this simulation we will fix the step size  $p = 0.05$  since at the step size the number of total infected people and current infected people appear to grow fastest so we may observe the most apparant pattern in this simulation. We set the initial position as the variable this time:

Set the corner position to be  $[0,0]$ :



Set the corner position to be  $[1,1]$ :



## Analysis

From the graphs we can see that given the fixed  $p$ , the situations are apparently different if peoples' initial positions are corners, centers or randomly set respectively. We can see that the number of infected people at each time in these three different situations have similar tendency even if we change the positions of corners, that is  $\text{center} \geq \text{random} > \text{corner}$ . It's reasonable to get this result since the step size we choose is relatively small, so even if the random initial position is not center, it's very likely people won't stay at the postion,i.e., people don't have the tendency to step out the boundary. Therefore, the number of infected people would be slightly smaller when their initial position is randomly choosen than their initial position is set at center. However, the situation is totally different if the peoples' initial position is at the corners since they are very likely to go out the boundary and thus they are at the same large probability that they will stay at the boundary then the percentage of population they would interact is much smaller than the counterpart of the above two intial positions.

# Spatial PDE SIR model implemented for final submission

## Updated ODE\_SIR.py to transform our ODE model to be a PDE.

We discretize the square into an  $M \times M$  grid and set  $M = 200$ . We represent the population at each time with three  $M \times M$  arrays,  $s(x, t)$ ,  $i(x, t)$ , and  $r(x, t)$ , where  $x = (i, j)$  is the index on the  $M \times M$  grid. The random walk for the agent based model becomes diffusion in the continuous limit, so we have a system of PDEs:

1.  $\partial s(x, t)/\partial t = -b * s(x, t) * i(x, t) + p * \Delta s(x, t)$
2.  $\partial r(x, t)/\partial t = k * i(x, t) + p * \Delta r(x, t)$
3.  $\partial i(x, t)/\partial t = b * s(x, t) * i(x, t) - k * i(x, t) + p * \Delta i(x, t)$

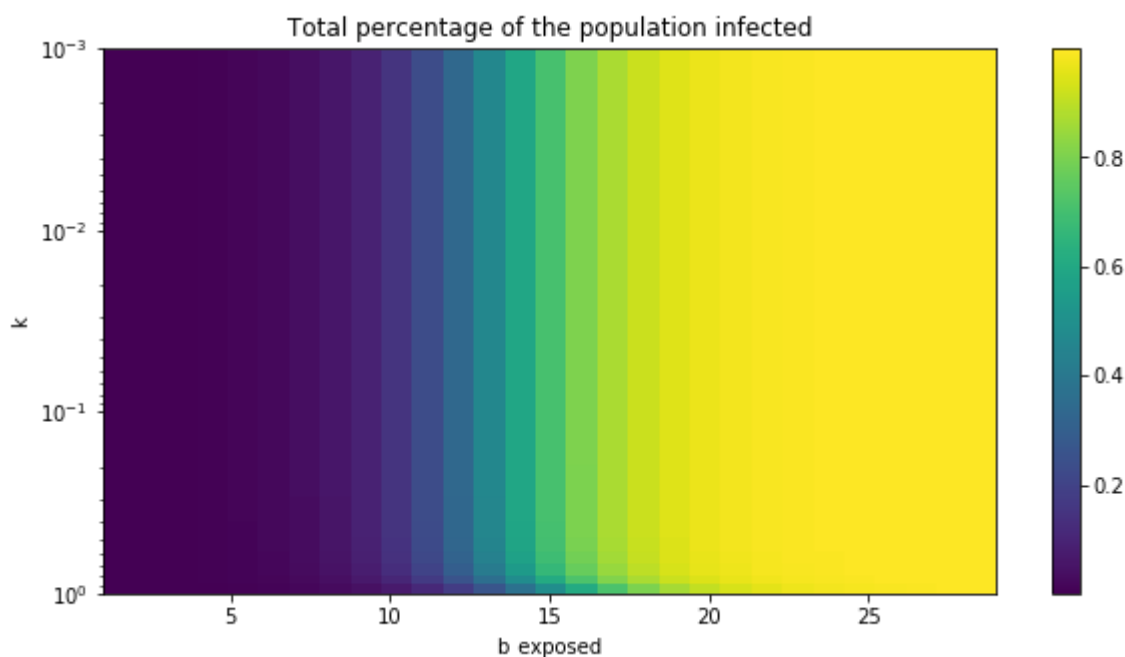
We provide the parameters as follows and notice that we used the parameter  $b$  again and used  $p$  in another way:

```
M: grid size
I: initial infected M*M array,
S: initial susceptible M*M array
k: recovery rate
b: same as before, infected individual has a fixed number b of contacts per day
p: use the parameter p to weight the diffusion term
```

## Updated simulations:

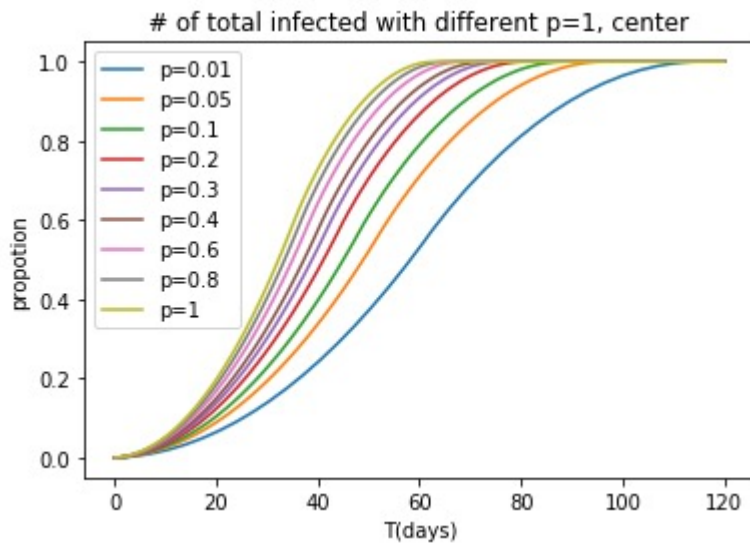
Answer to question 1 If infected individuals start at the center of the square, how does the total number of individuals infected in the simulation change with the parameter  $p$ ?

As before, this question requires the comparison of choosing different step length  $p$  given the initial position is the center of the square. Also, we pick  $b=15$  and  $k=0.05$  according to the phase transition.

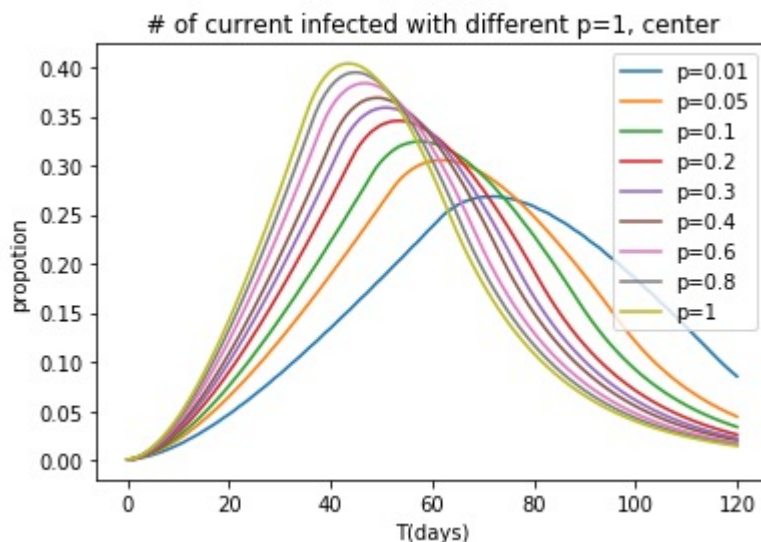


Fix  $b = 15$  and the time period equals to 80:

## Total Infected:



## Current Infected:



## Analysis

From the graphs we can see that when  $p$  is smaller than or equal to 0.1, the number of total infected people grows relatively slow and the current infected won't reach its maximum before  $T = 80$ . As  $p$  gets larger, the number of total infected people and days of the current infected reach their maximum get earlier and earlier. Also, the maximum of the current infected comes earlier than the total infected since the total infected is dependent of the variation of the number of current infected. The reason is that when  $p$  is small, the step size of people is small and thus people can only effect a small group of new surrounding people. As the step size gets larger, people can get into a brand new neighborhood and thus get more uninfected people infected.

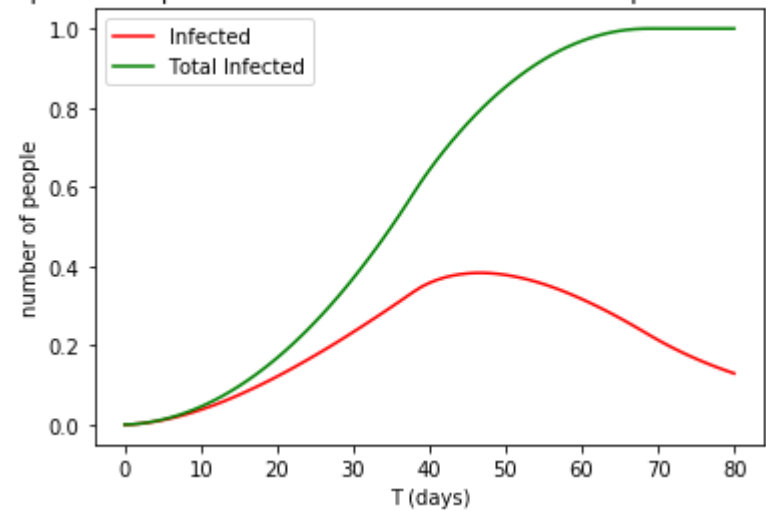
Answer to question 2 **Choose an interesting parameter of  $p$  using question 1. How does the simulation qualitatively differ when the initial infected individuals start in a single corner of the square vs. the center of the square vs. being randomly spread out? You can investigate this by choosing initial conditions for  $i(x,0)$  appropriately.**

**We choose  $p = 0.6$  for question 2. The reason is that we assume (according to agent-based model) the number of infected people will grow much slower if it**

starts at corner. So in order to better present three lines in one graph, we have to take a 'p' which should not be small. Also, 'p' can not be very large in case the number of infected grows too fast when start at random.:

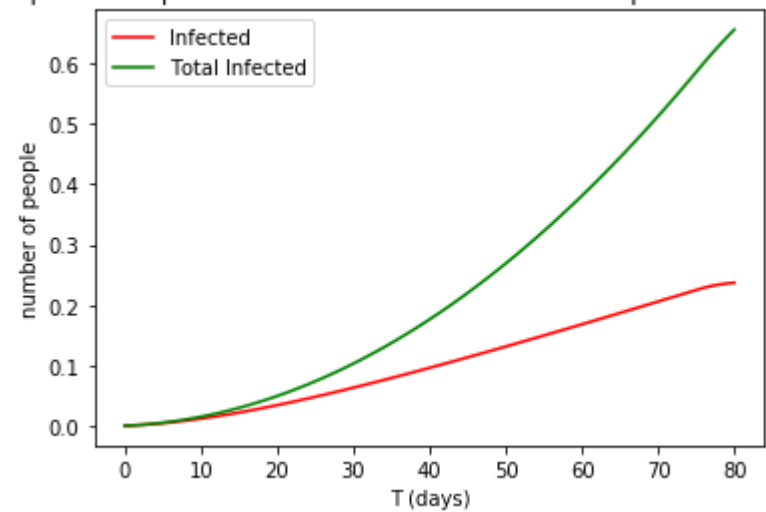
Initial position is the center:

Spatial Component Introduced with initial infected position at center



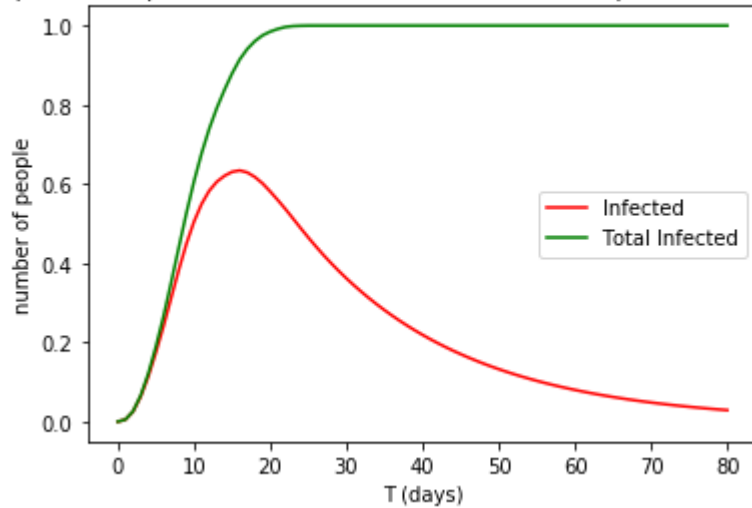
Initial position is the corner:

Spatial Component Introduced with initial infected position at corner



Initial position is chosen randomly:

Spatial Component Introduced with initial infected position at random



## Result:

From the graph we can see that the order of speed of number of total infected people is random > center > corner. This is an interesting observation since we expect center has the top speed.

## Section 3: Conclusion

### Summarization of our results

We proposed three variations in our midterm checkpoints and we implemented all of them for the final submission.

For the first variation discussing the influences brought by different strengths of the enforcement of masks and quarantine, we can draw a conclusion that

For the second variation, we observe that the speed would be slower of both the disease spread and number of infected to go back to zero.

For the third variation about SEIR model, we find that

The result we got for spatial agent-based model is that when we fix  $q = 0.0437$  and the time period equals to 80, we get that when  $p$  is smaller than or equal to 0.01, the number of total infected people grows relatively slow and the current infected won't reach it's maximum before  $T = 80$ . As  $p$  gets larger, the number of total infected people grows faster and the current infected can reach it's maximum before  $T = 80$ . Choose  $p = 0.05$ , we find that the number of infected people at each time in these three different situations have similar tendency even if we change the positions of corners, that is center  $\geq$  random  $>$  corner. We then implemented the PDE model, in this model we capture a stable pattern such that the growth of both current infected and total infected is fastest when the initial position is chosen randomly, then slower when the initial position is the center. The speed is slowest when the initial position is the corner.

### Limitations of our models

In our model, the probability is the same for infected people to get others infected, but in the reality, the situations are very different between a infected person to interact with other people in different locations. For example, a infected person can easily get others infected in hotels and processions but less likely to infect others in a forest. Also, people's activity area is far from being regular circle or rectangular set in our model, so this factor can seriously infect the accuracy of the simulation result. Finally, the recover probability  $k$  should be different for people at different age since people's immune system generally get weaker after a certain age.

## Interesting directions for further investigation

---

We found there is an idea to implement a network SIR model on the Internet. This model adopts the network approach to represent people as nodes and contacts that potentially spread virus as links in a network. Then the limitations we mentioned above can be partly solved since now we can link any two people in the designated network according to multiple factors rather than simply confine an infected person in a regular-sized shape. Therefore every node can have different number of links to others nodes, which represents the fact that everyone can have different number of interactions with other people. For example, people who may live in rural area may have fewer contacts with others through their life than individuals live in urban centers with high population density. We can also differentiate interactions that with different probability to spread the disease by weighting the links. In terms of the variance of strength of each person's immune system, we can also weight the nodes to implement this idea.

## Section 4: Bibliography

---

- Craig, B. & Phelan, T. & Siedlarek, J. & Steinberg, J. (2020). Improving Epidemic Modeling with Networks. *ECONOMIC COMMUNITY*
- Tolles.J & Luong.T (2020). Modeling Epidemics With Compartmental Models. *JAMA Guide to Statistics and Methods*