CSCI 4190 2021-22 Second Term Final project

I declare that the assignment here submitted is original except for source material explicitly acknowledged, and that the same or closely related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the following websites.

University Guideline on Academic Honesty:

<http://www.cuhk.edu.hk/policy/academichonesty/>

Faculty of Engineering Guidelines to Academic Honesty: <http://www.erg.cuhk.edu.hk/ergintra/upload/documents/ENGG_Discipline.pdf>

Student Name: Lam Wai To Keith

Student ID : 1155133260

**CSCI4190 Project Report**

**Abstract**

In social network analysis, it is very important for us to understand the diffusion of ideas, behaviors or even diseases. If we know how these spreading occur in a specific network, it may help us to solve real world problems, including how to stop the diffusion, or on the other hand spread the behavior across the network efficiently. Back in the early twentieth century, three models are developed to describe such a phenomenon, called the SIR, SIS and SIRS model, each for simulating a specific behavior spreading. To what degree do these models describe the real life situation accurately? We are going to answer this question in the following.

1. **Introduction**

In this project, I am going to investigate the spread of behavior and diffusion of ideas using three models, SIR, SIS and SIRS model. I am working on task set 4, which is to simulate epidemics. I will describe the process of the computer simulation of the models and present my findings through the dataset.

There are a few technologies that I have used to conduct my analysis. I used the SNAP library with Python to conduct analysis on the Facebook dataset with the following attributes:

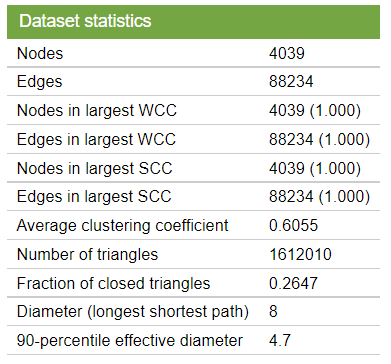


Fig. 1 Statistics of the dataset

Moreover, for the model construction, I used Python to access and manipulate the data, together with pandas and NodeXL to export and visualize my data in excel. Using these tools, I hope to model the spread of behavior in our social network, which gives us insights on social network analysis.

The main objective of this project is to visualize the effects of changing different parameters to the spread of disease. I am also interested in the deterministic factors that changes the speed of infection. There will be four main sessions to the report, where the first three will be on changing parameters on SIR, SIS and SIRS models respectively, and analyze how contagion probability and infectious period could affect the number of infections caused in a fixed time step. Finally in the last session I will demonstrate why is it important to choose the initial adopters of the behavior, to show that the network structure has a great impact on how the disease would spread.

1. **Results and Analysis**

*2.1 SIR Model*

The SIR model refers to three states in an epidemic, “susceptible”, “infectious”, and “removed”. Each node in the network will be initialized as “susceptible”, and after choosing a first group of nodes to be “infectious”, they can infect each susceptive neighbor with a fixed contagion probability. After a certain infectious period, the infectious node will turn into “removed” state, where they cannot be infected or infect others anymore. For this model, I will change the contagion probability and infectious period to see which parameter is more sensitive to cause more nodes to be infected. Also, to ensure the fairness of each experiment and make sure the results are significant, I will choose the node with the maximum out degree in the graph to start the infection. Unless specified, the parameters are fixed as follows:



Where the totalTimeStep variable denotes the total time step of the experiment. The normal result is as follow:

Fig. 2 Graph of nodes with different states in default settings

Later, we will compare the results with this graph.

*2.1.1 Changing the contagion probability*

In the first experiment we will change the contagion probability. This probability refers to how likely the node will infect its neighbor nodes. The following are results for

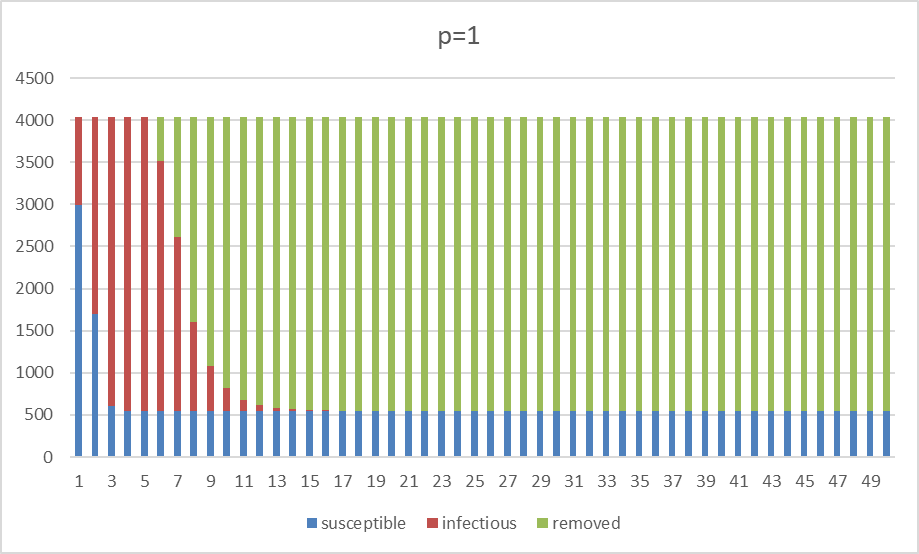
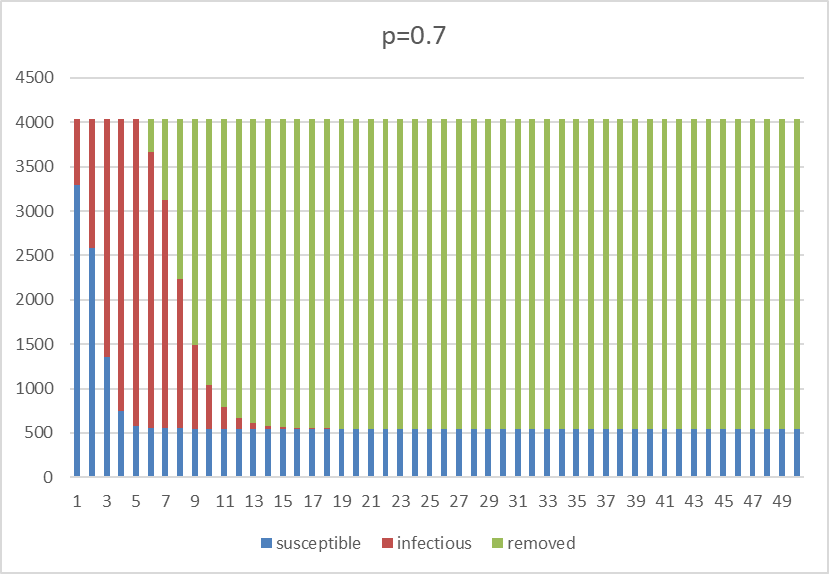
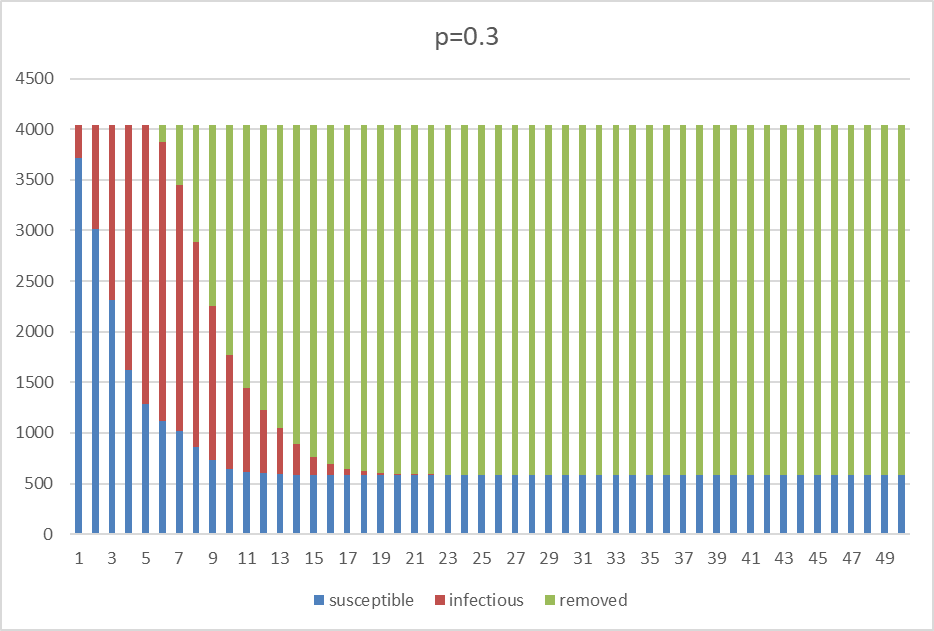
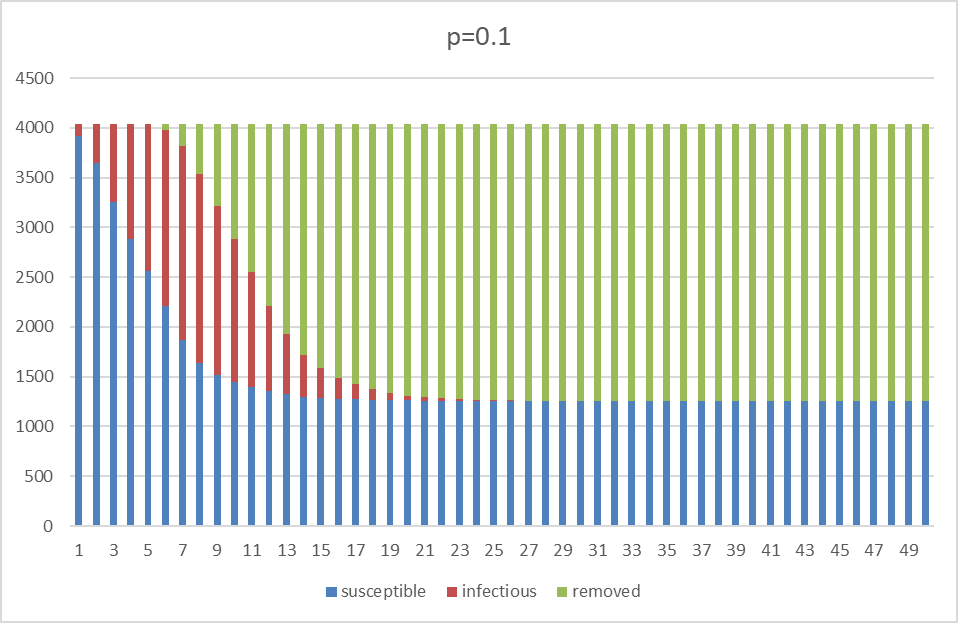


Fig. 2 graph of nodes in each states with various contagion probability

In these graphs, we can make a few observations. First is that the red strip has variations in height and width. As the contagion probability goes up, the height of the strip increased, suggesting that at each round there are more people getting infected. This means the speed of which the disease spread will be faster. Also, as p increases, the strip decreases with width, indicating that the time it takes for the disease to die out is shorter. Finally, the total number of removed nodes varies over the contagion probability too. As shown in the graphs, when p=0.1, there are around 1300 nodes unaffected by the infection, but for p=0.1, only around 500 nodes remain untouched. This shows that the contagion probability also controls how many people are going to be in contact with the epidemic. However, even with p=1, we are not able to infect the whole network, this is because there are also nodes which has no connection to the giant connected component in the network, hence there is no way for the disease to infect them.

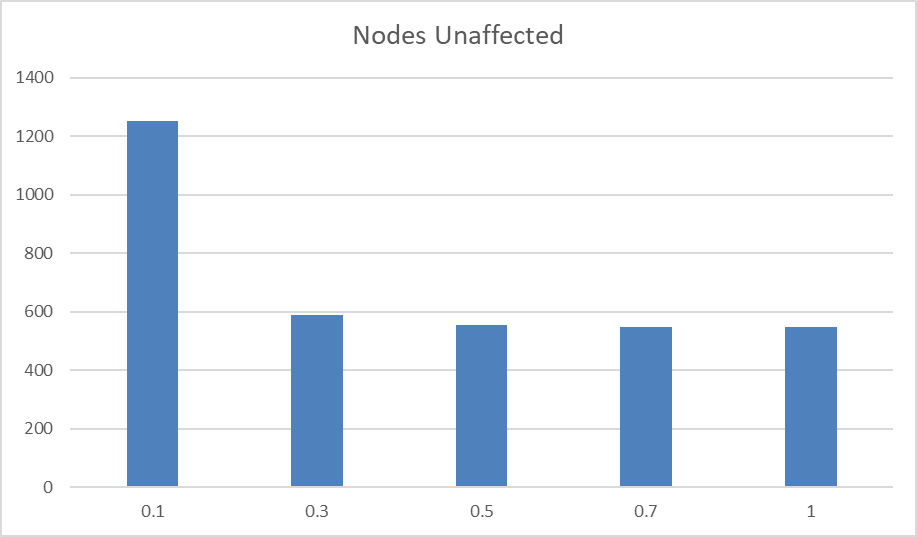
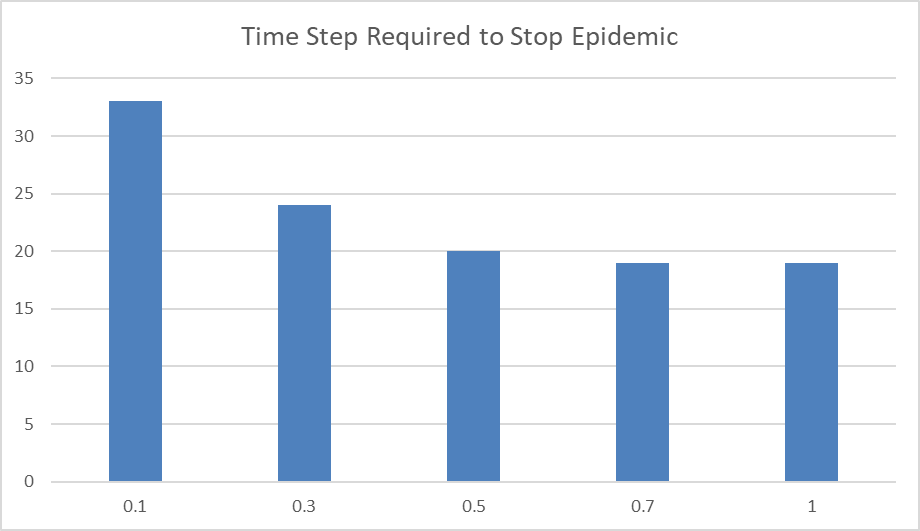


Fig. 3 Graphs to show correlation of probability with the time required to stop the epidemic and number of unaffected nodes

Moreover, we can see that if the contagion probability is low, then a small change in the number would cause a larger difference in the time required for the epidemic to stop, and the nodes affected. This is to say the parameter is more sensitive at lower values.

*2.1.2 Changing Infectious Period*

The next parameter we change will be the infectious period. This parameter determines how long the disease will stay on a specific node until it has been recovered. I have considered 5 values on the period: 1, 3, 5, 7, 9. We have the following results:

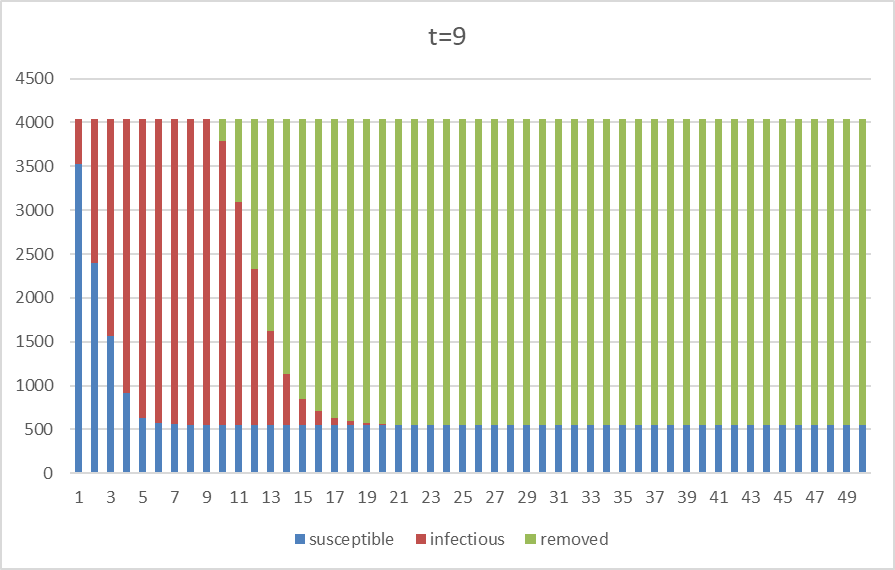
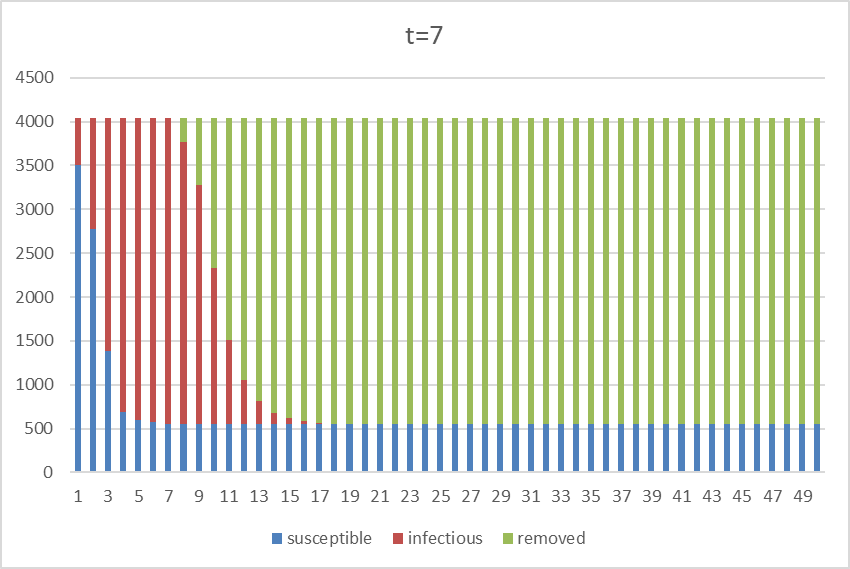
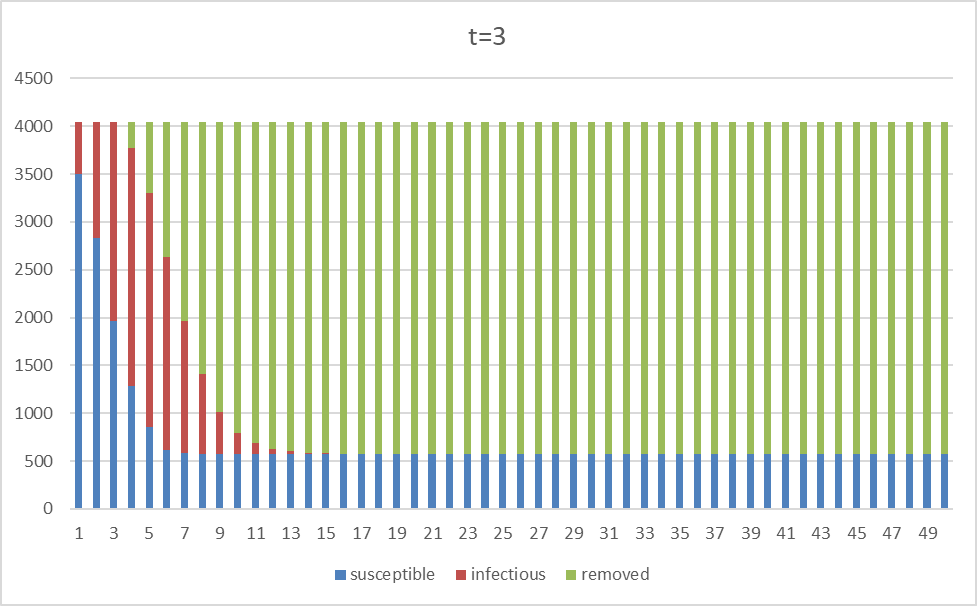
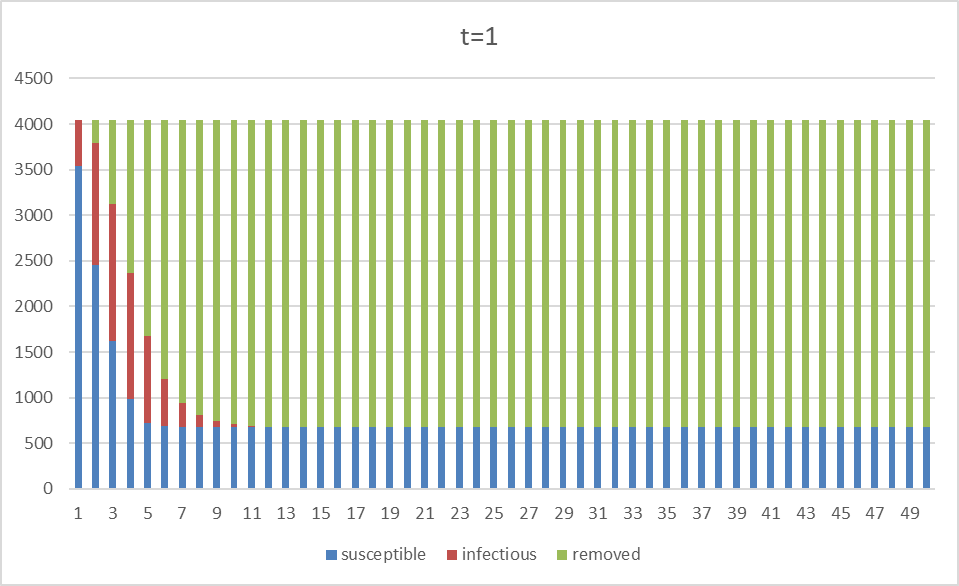


Fig. 4 Results of changing infectious period t

Directly from the graph, we have the following observations. The red strip this time is having increased in height and width when increasing the infectious period. This is expected as the longer the infectious period goes, the more nodes would stay in the infectious state across time. One interesting observation is that the number of nodes affected by the epidemic seemed to be similar across the graphs. This suggests that the infectious period does not really determine how many nodes will get infected across a long period.

*2.1.3 Short summary*

There are a few issues to the results before we proceed. There are a few reasons why the above results are not accurate enough. The main one is that our results are random according to the probability. We cannot recreate the above data, and sometimes the results fluctuate a lot in terms of when the epidemic would stop. Hence my choices of parameters are careful, so that we expect less of outliers to happen. For example, setting the max out degree node as our first infectious node guarantees that the epidemic can be spread without the effects of network structure (to be discussed at session 2.4), and setting the default infectious time period to be 5 ensures the disease would not die out easily in the first few time steps. This is to make sure even if we repeat the experiment, we can expect the similar outputs, so that the results are more convincing to interpret.

In general the SIR model has shown us that higher contagion probability with longer infectious period, yields a more dangerous and fast-spreading epidemic. This is as expected because every infectious node is more likely to infect others in a longer period of infection. This could be helpful in epidemic control, as if we can lower the contagion probability to a very low value, and shorten the infectious period by medical controls, we would expect to see less nodes being affected, and that the epidemic would stop earlier.

The SIR model is best at describing diseases or behavior such that no one could get infected for the second time. This includes diseases like the Spanish flu, where a patient develops antibodies to prevent another infection from others. However, that is not always the case, and hence we must look at the other models.

*2.2 SIS Model*

The SIS model is just a slight modification to the SIR model. Instead of having the “removed” state, after the infectious period, the node will return to the “susceptible” state. This means the node can be infected again and we would expect the epidemic to go on longer, or even goes on forever. Again, we use the default parameters stated in the SIR model.

*2.2.1 Changing Contagion Probability*

We again investigate the effects of contagion probability on the model. See the results below:

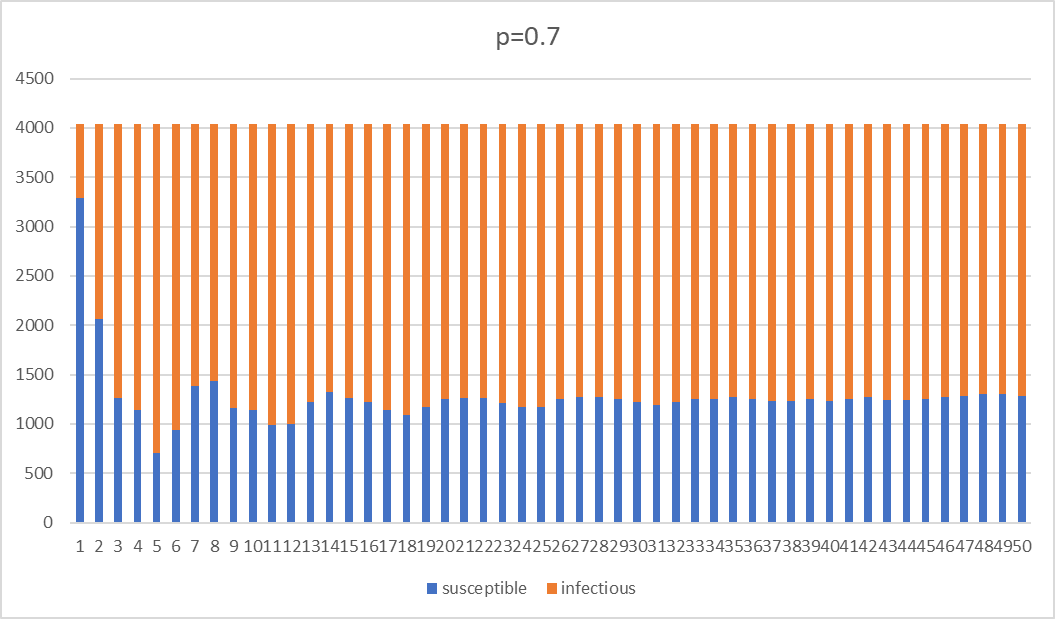
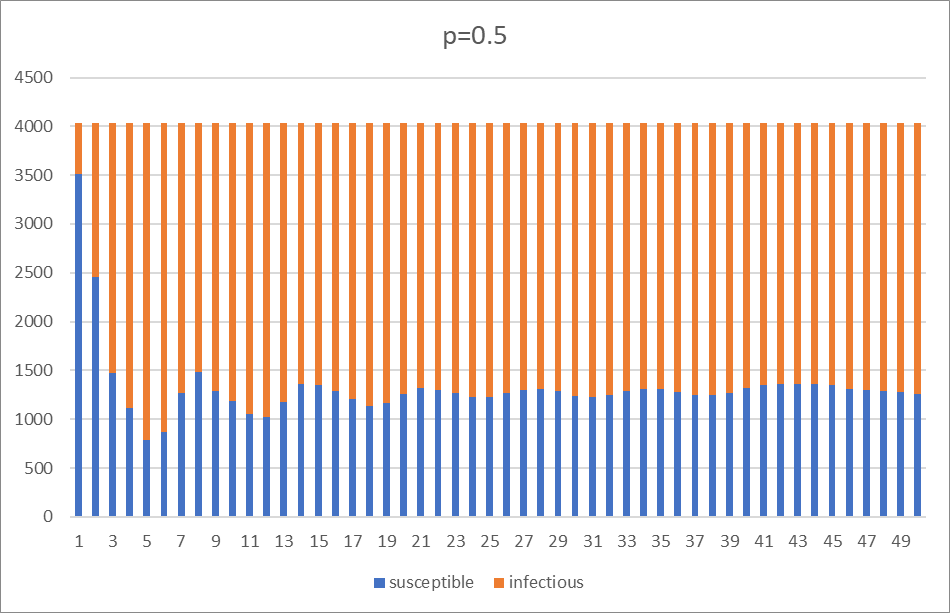
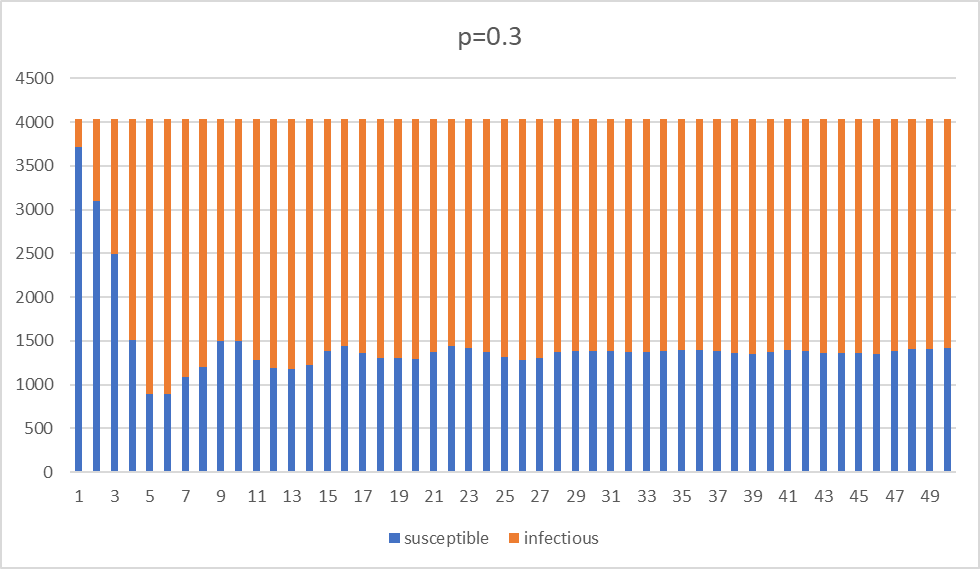
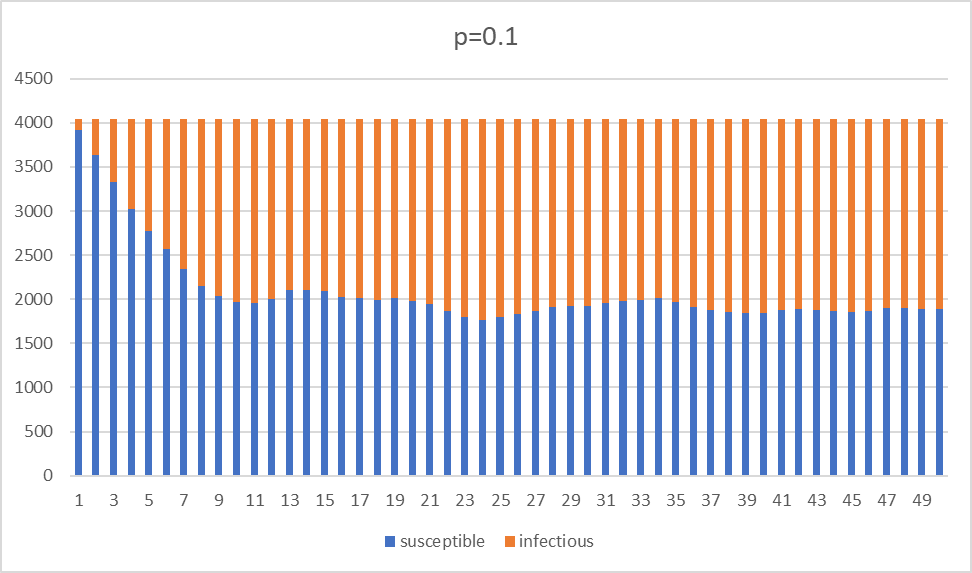


Fig. 5 Graphs of SIS model with various contagion probability

Surprisingly even if the contagion probability is low, the spread of the disease still goes on for a very long time. The SIS model seem to have less sensitivity on the contagion probability, because after a certain period, all graphs seem to have reach a relatively stable value. The reason behind could be that at every wave, a same number of nodes finished their infectious period comparing to the number of nodes that got infected again. A slight difference across the contagion probability, is that there are more fluctuations in the number of infectious nodes when the contagion probability is higher. This is because at every wave a node can infect more people faster, but at the same time must wait for more neighbor nodes to recover so to infect them again. Finally, we can also see that with less contagion probability, the plateau locates at a higher position, with more people being susceptible and less infectious nodes.

*2.2.2 Changing Infectious Period*

Again, we change the infectious period and see its effects:

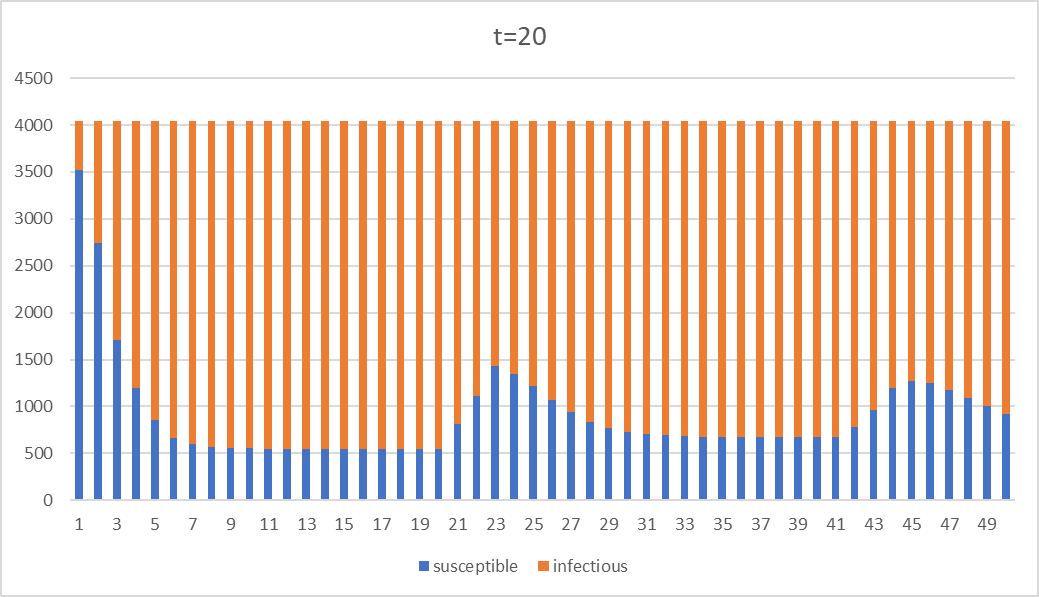
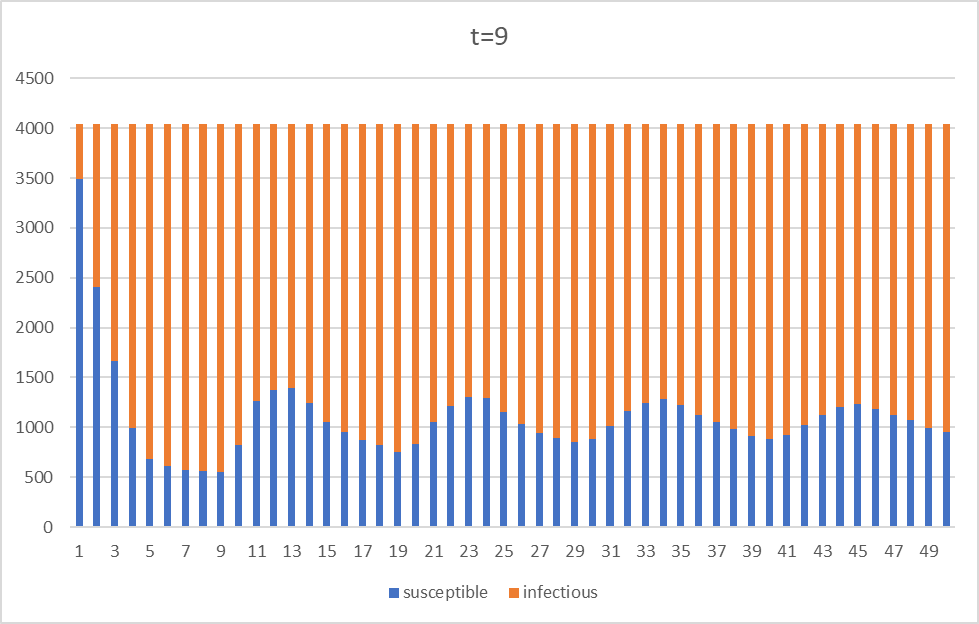
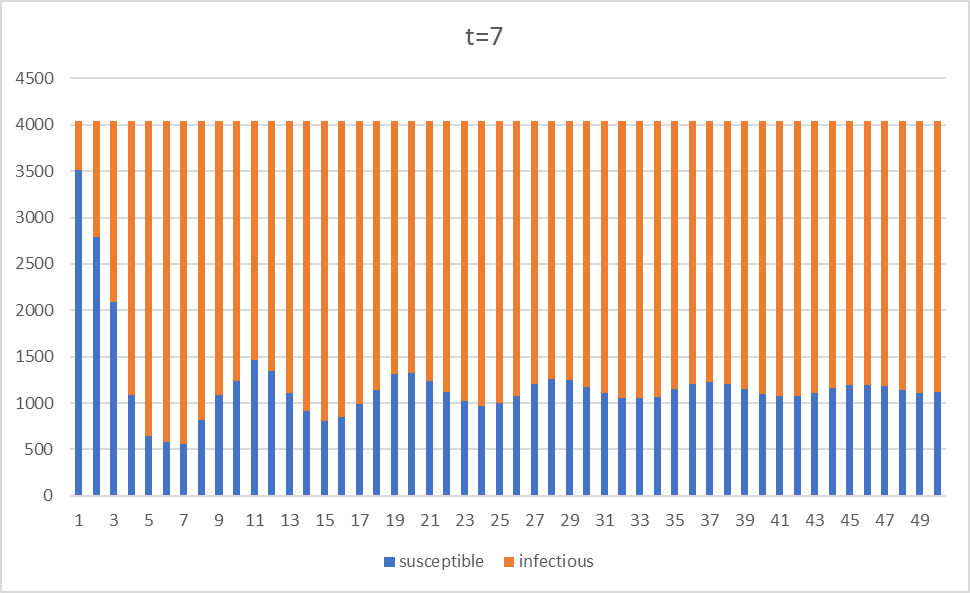
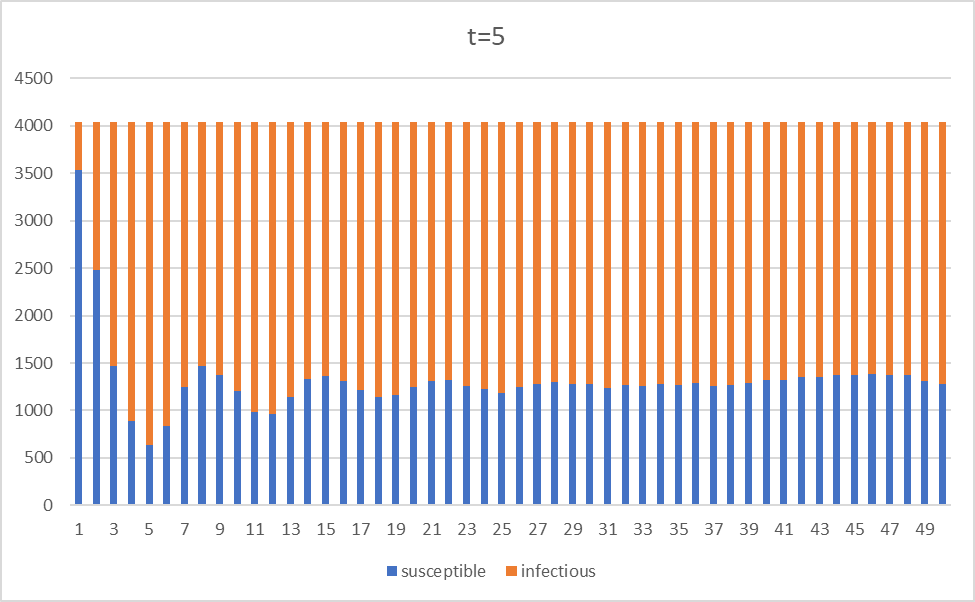
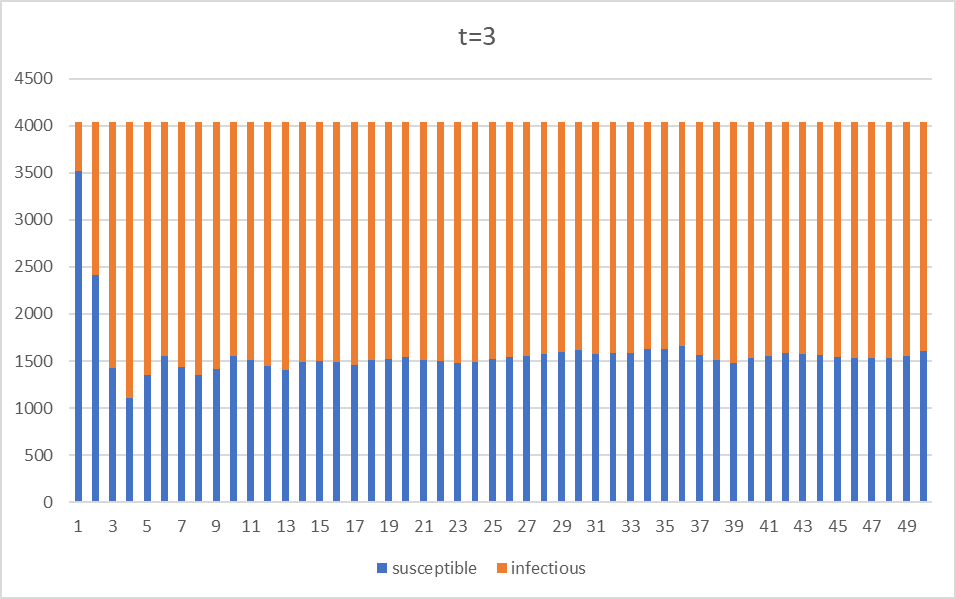
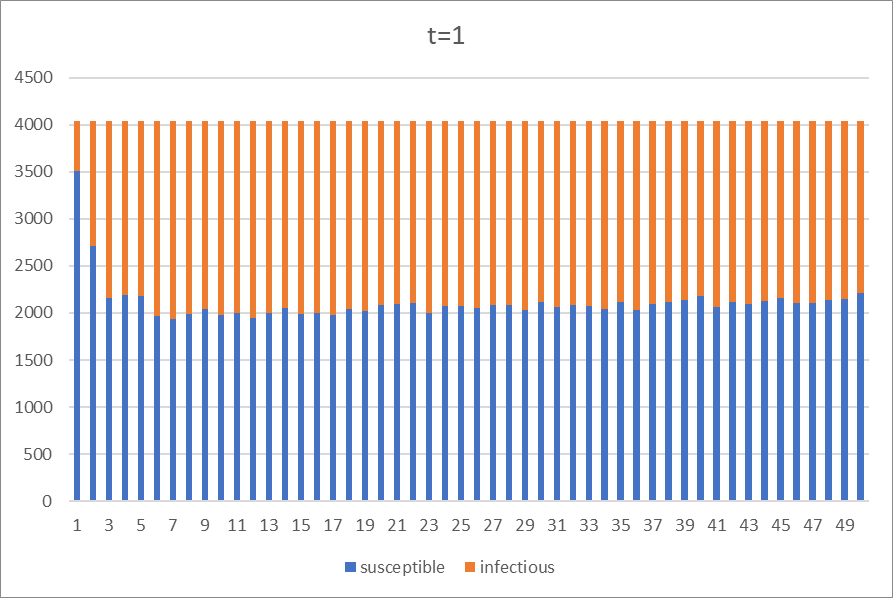


Fig. 6 Graphs of SIS model with various infectious periods

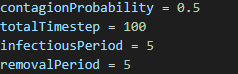
The graphs above demonstrated some interesting behavior. When t is small, we see that plateau is reached very quickly and steadily, with a higher number of nodes being susceptible. However, as t become larger, we start to see some wave-like patterns with different periods. The wavelength of the waves is approximately the infectious period and start to see a decrease in amplitude every time it oscillates. This is because with a very long infectious period t, when most of the nodes are infected, they ran out of nodes to infect and must wait until its neighbors to return to the susceptible state. The longer the infectious period, the more obvious is this effect, which is why I also attached an extreme case with t=20 for a reference. Note that in the graph of t=20, the infection quickly plateaus at around 500 nodes being susceptible, which is the minimum number of nodes that can stay susceptible throughout the time. During the plateau, as there are no more neighbors to infect, it has a flat slope and curves up after 20 time steps has passed where some nodes start to return to susceptible state.

*2.2.3 Summary*

The findings here are quite significant. We see that the contagion probability does not really hold back the epidemic once it started (obviously if the epidemic ends at the first few time steps, then it would be meaningless to discuss). We see that under the SIS model, an epidemic hardly goes to an end, but rather they reach a plateau where the number of susceptible and infectious nodes remains steady over time. However, if we change the infectious period, we can see that a wave-like pattern in the graph that is decreasing in amplitude. This is probably because when most neighbors are infected, the node must either wait until its neighbors to return susceptible, or the node returns susceptible by itself, hence creating such a pattern.

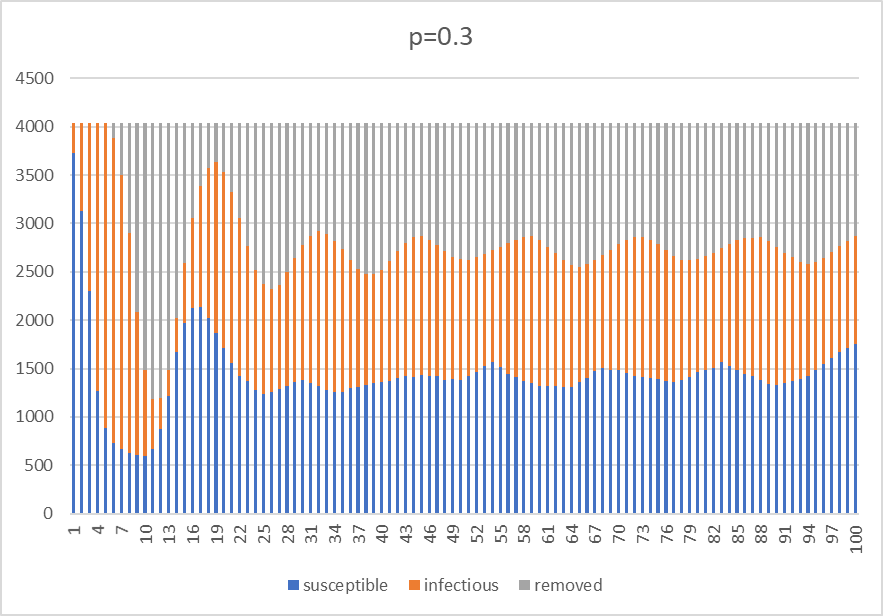
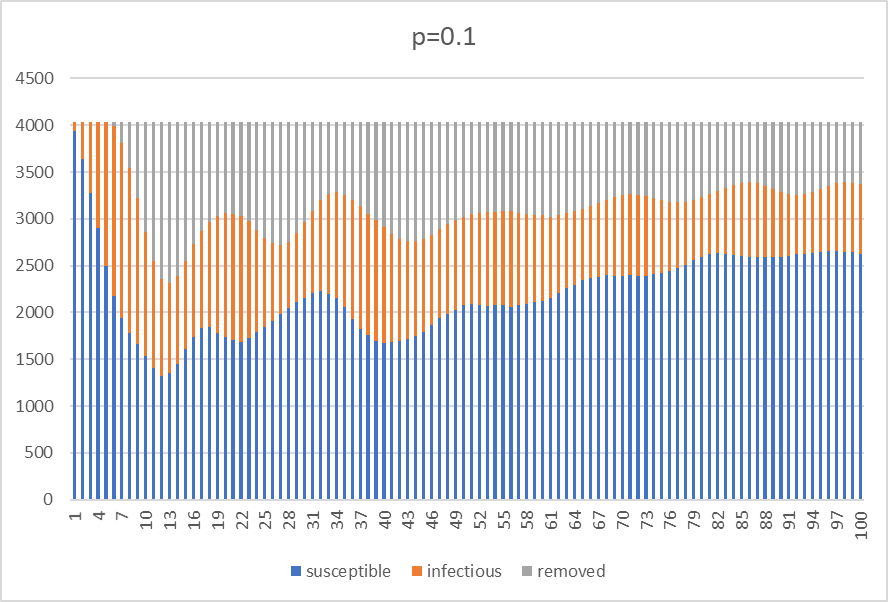
*2.3 SIRS Model*

Finally, we discuss the SIRS model, which is again just a modification to the previous SIR and SIS model. This time, after entering the removed state like the SIR model, we return the removed node to susceptible again after a certain “removed period”. Moreover, to present the data better, I set the total time step to 100 in order to see the effects of the parameters. So, this time we have:



*2.3.1 Changing Contagion Probability*

Using the same breakpoints as above, we yield the following:



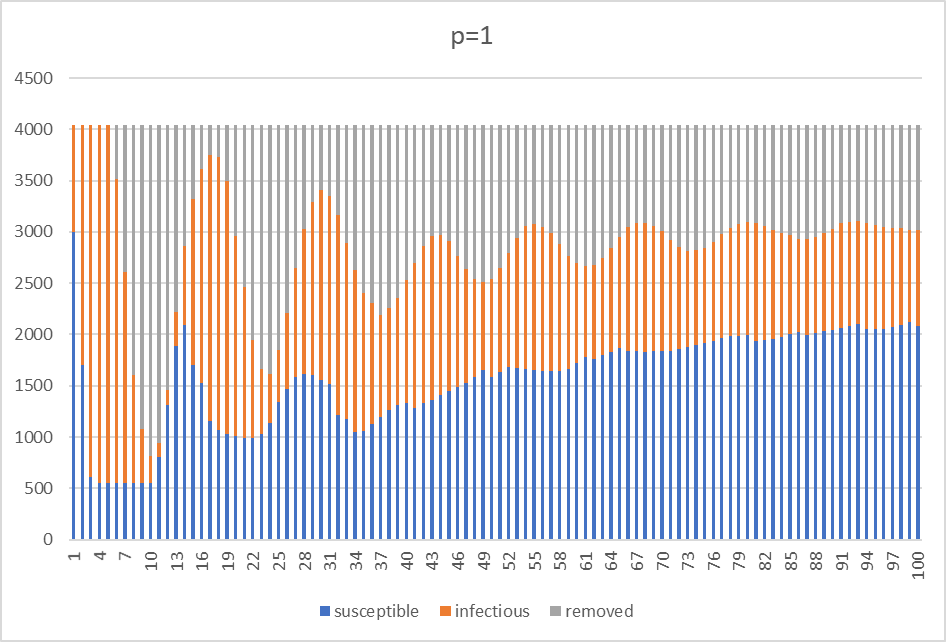
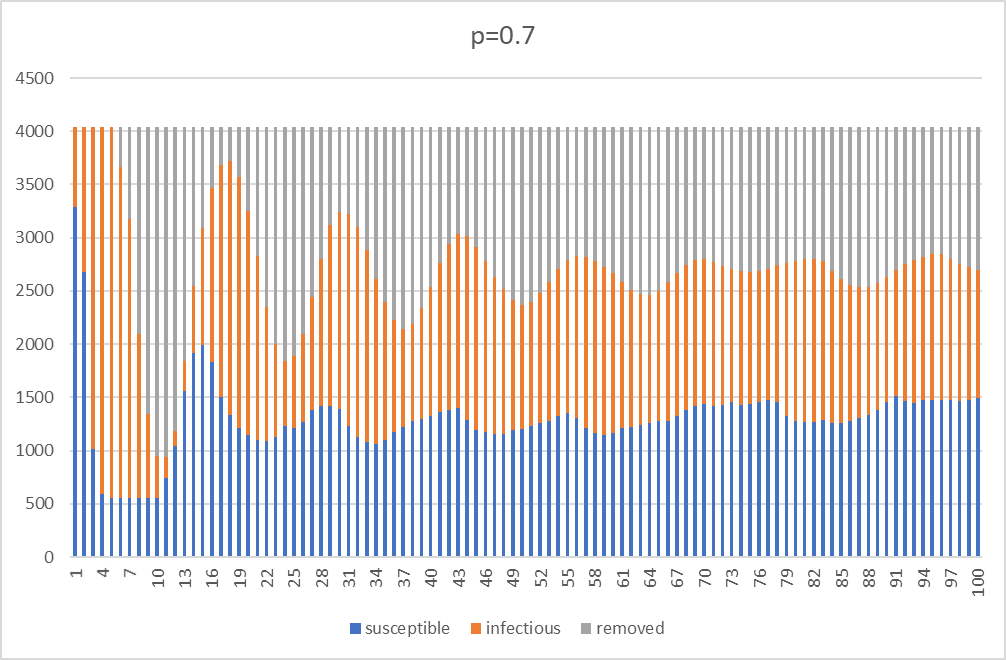
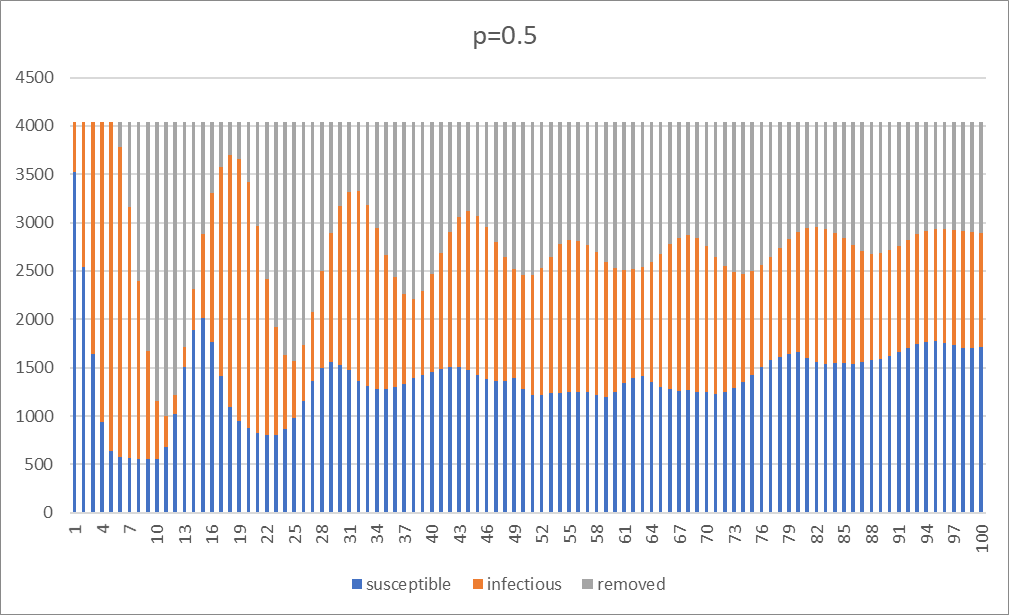


Fig. 7 Graphs of SIRS model with different contagion probability

From the graphs we can see that the SIRS model produces a more wave-like pattern regardless of the contagion probability. However, the lower the contagion probability, the higher the plateau ends up at. We also see that the fluctuation on the number of infectious and removed nodes is stronger when the contagion probability is higher. This is probably because with a higher probability, the more nodes will turn infected, hence more nodes enter the removed and susceptible state afterwards. But in general, for different contagion probability, we do not see an end to the epidemic even with a longer time step.

*2.3.2 Changing Infectious and Removal Period*

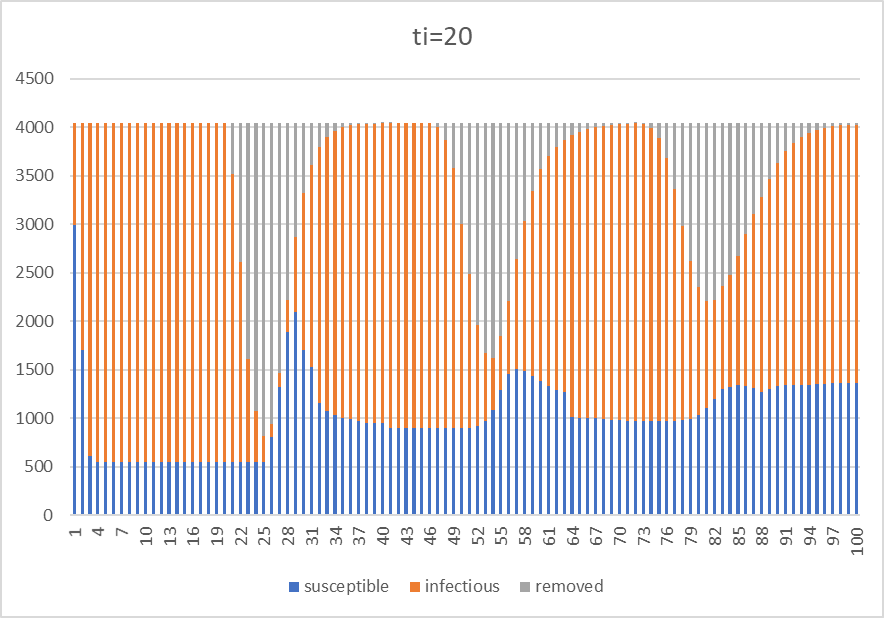
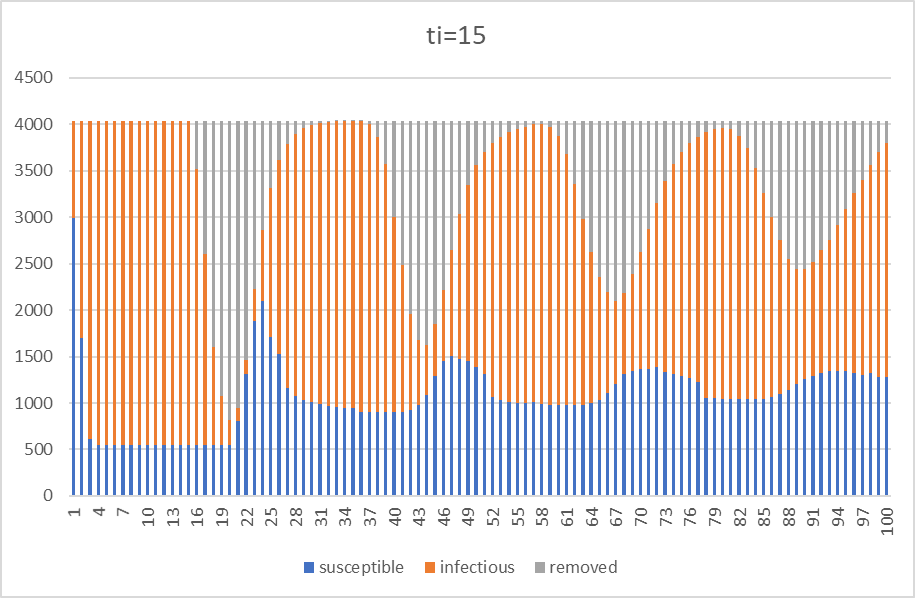
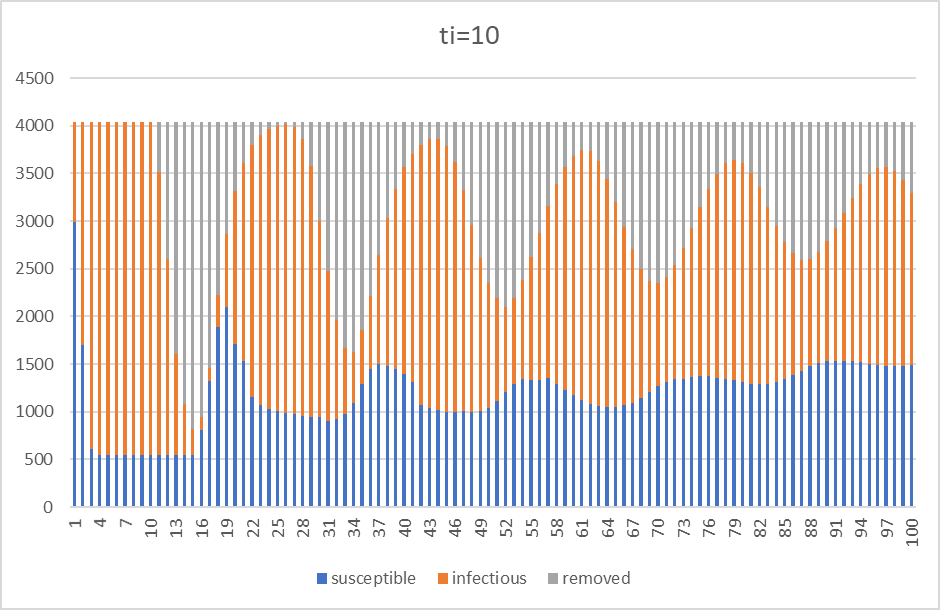
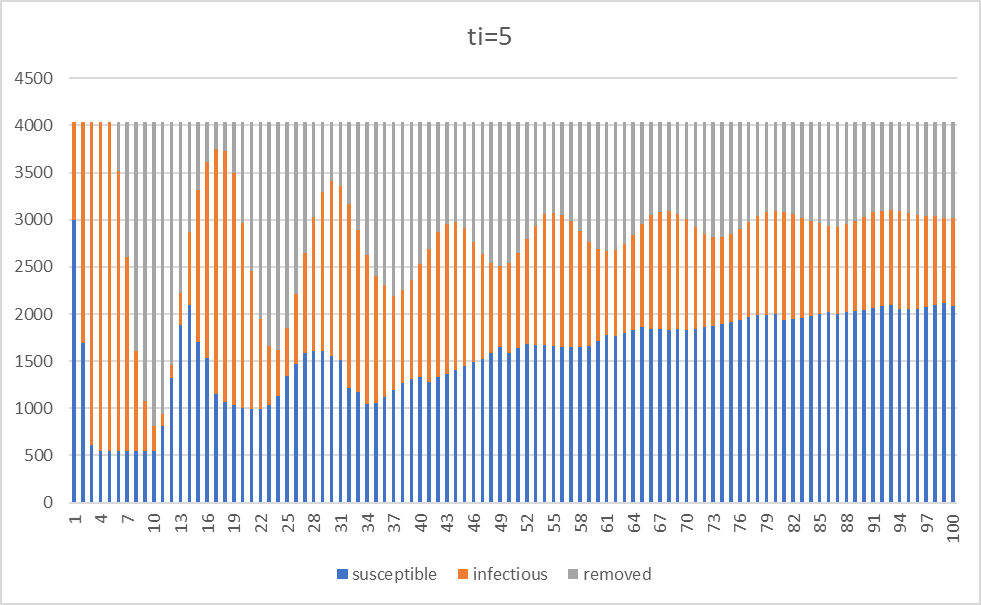
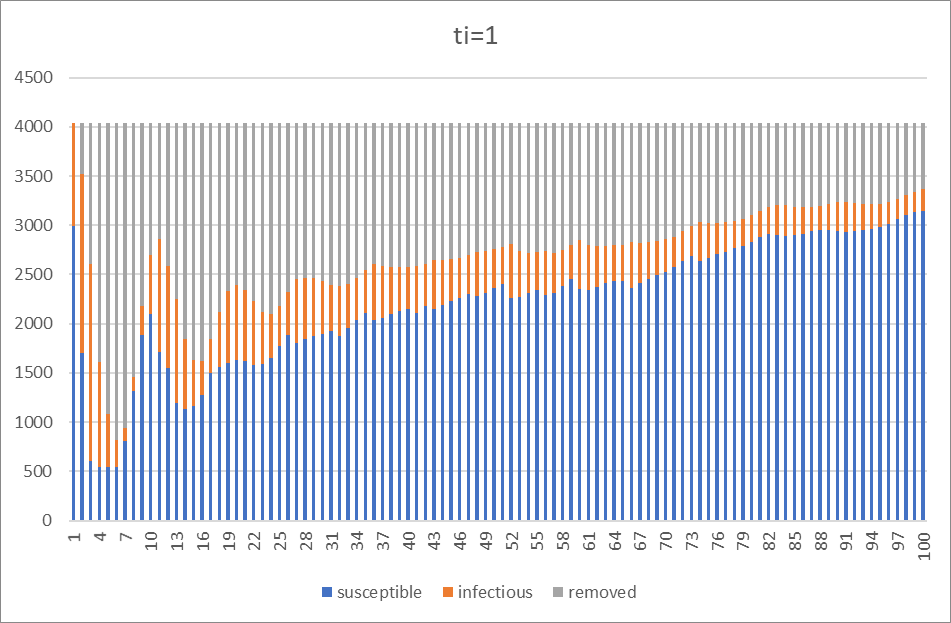


Fig. 8 SIRS model with various infectious period

As seen from the graphs, changing the infectious period has a large impact on the number of infected and removed nodes. When the infectious period is small, we see that the fluctuations in the number of infected nodes are small, so there are more nodes going into the susceptive state. We can also expect the epidemic to go to an end given a longer time step. However, as infectious period increase, there are more wave-like patterns in the graph. This is because the more nodes are infected, the slower is the nodes being able to infect their neighbors. The wave is again with the period equal to the infectious period, and each time decrease in amplitude. Finally, all graphs showed that the average number of infectious states is increasing, meaning that it is more likely that the epidemic is going to end. This trend is way more obvious when ti=1, where we see a steady increase in susceptive nodes. We then answer whether changing the removal period will give similar effects.

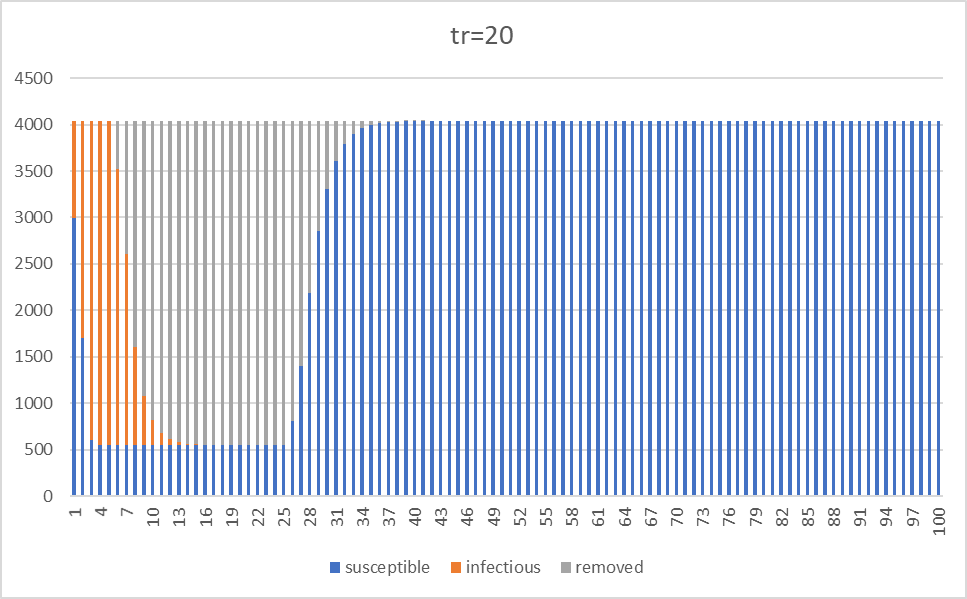
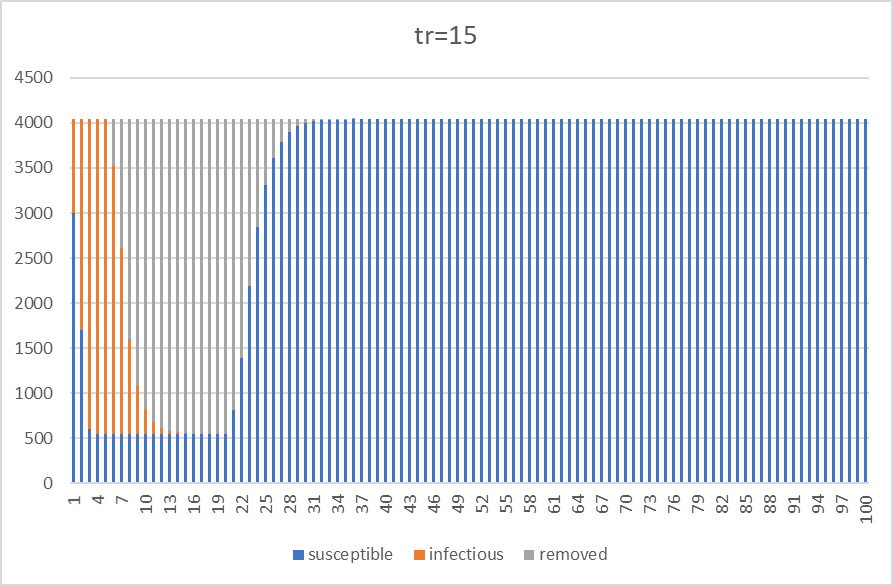
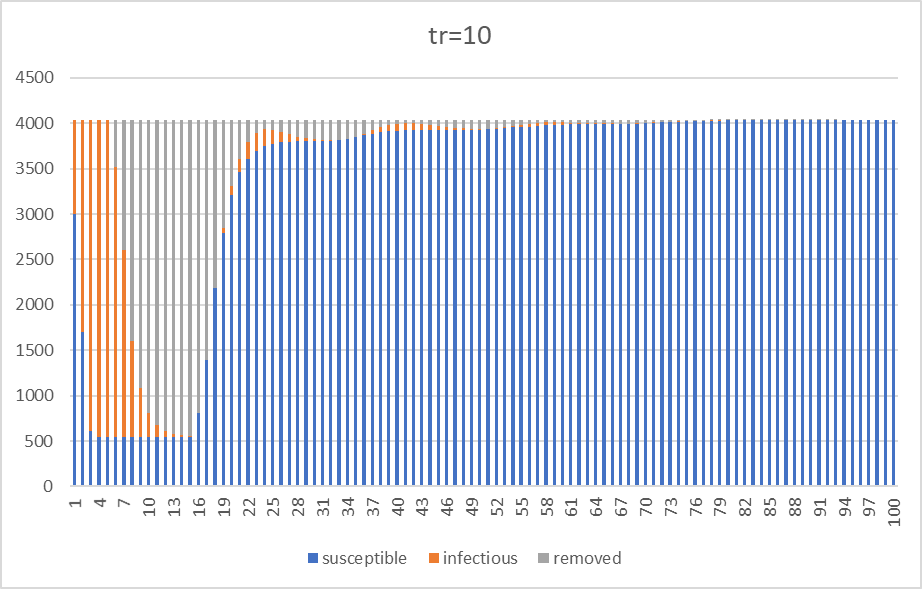
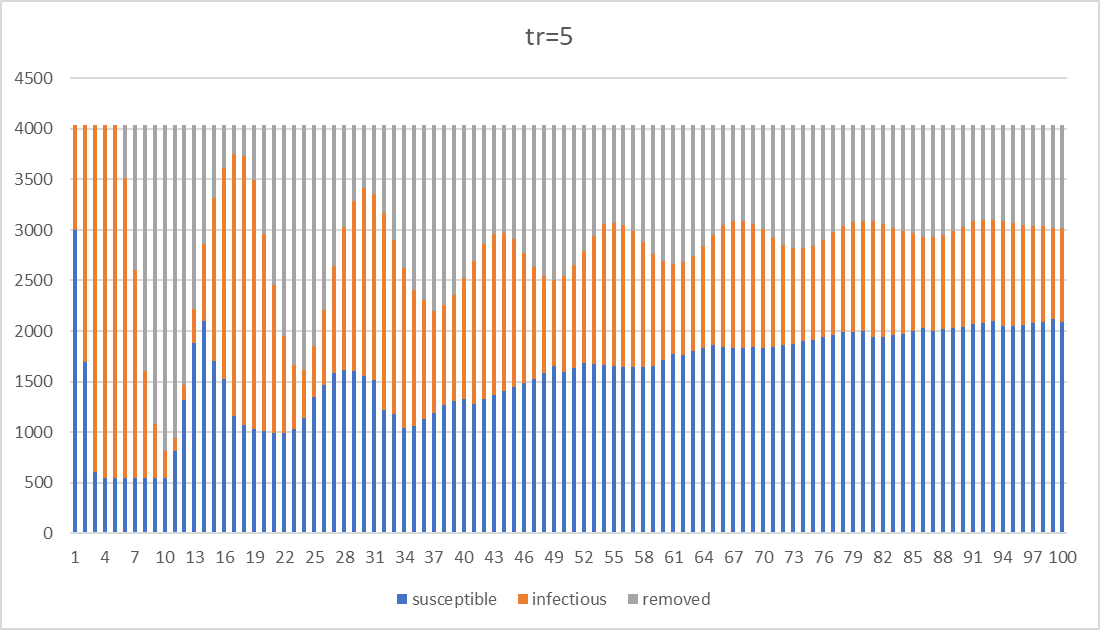
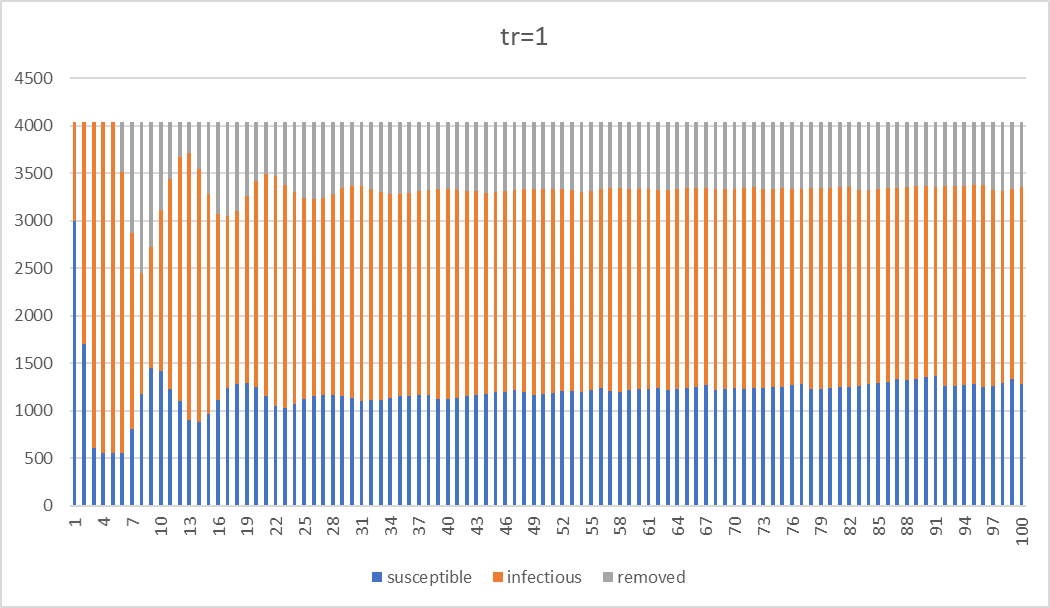


Fig. 9 SIRS model with various removal period

From the above, we can see that increasing the removal period can effectively stop the epidemic. When the removal period is short, the epidemic will stay for a very long time without any big fluctuations. However, when the removal period is longer, we see that most nodes are infected quickly and enter the removed state, decreasing the number of infectious nodes drastically. The epidemic also come to an end faster when we increase the removal period. While changing infectious period might not help a lot in stopping the epidemic, increasing the removal period could be a very good solution to stop the epidemic.

*2.3.3 Summary*

In the SIRS model we find more relationships among the spread of behavior with its parameters. While changing contagion probability and infectious period might bring some change to the network, changing the removal period of the nodes has the most impact to the graph. Taking this concept to ground, when dealing with an epidemic, if the disease allows our body to produce an antibody for a long time, the epidemic is likely to stop soon. Therefore, some disease can be more infectious than others, despite its transmission.

*2.4 Epidemics and Network Structure*

In the following session we discuss the effects of the network structure on epidemics. In the previous sessions we fixed the first infectious node to be the node with the most out links. This is to ensure our experiment will not be affected by how the network is structured. But how does a node’s out degree affect our investigation? We present the following data:

The following graphs are obtained from the SIR model, by changing the first infectious node. To prevent too many calculations, I have randomly picked 200 nodes for my assessment. The nodes are sorted according to their out degree. For every node, I will run the SIR simulation on it and measure the time stamp where the epidemic has stopped. The results are as follows:

Fig. 10 Graph of Out Degree Versus Stop Time of SIR model

In this graph, the x-axis represents the out degree of the nodes, where the y-axis represents the time where the epidemic stops. From the graph we see a slightly positive relationship between the two variables, with a lot of fluctuations in between. This is because the stopping time does not only depend on the first node’s out degree. It is more dependent on the cluster density around the first node. Although the out degree contributes some to the density, that does not mean the neighbors are all having similar out degrees, hence having a lot of fluctuations. But in general, this graph shows us that choosing the first node to infect the disease also affects the outcome of the epidemic. Especially when we choose a node with a higher out degree, we would expect the epidemic to go on longer, and affect more nodes.

1. **Conclusion**

Throughout this project we have established a few results. By applying different models to the same data set, we see that the contagion probability has different effects to the eventual outcome of the epidemic. In the SIR model, the contagion probability affects the total people who might got the disease in the epidemic. But in the SIS model, this probability only controls where the equilibrium is located. In SIRS we also see that changing the contagion probability does not bring a very significant effect to the outcome. As for infectious period, all three models show that the longer the period, the more people will get infected. The infectious period also produces a wave-like pattern to all models, and it is more significant when the period is longer. Finally, the choice of the first node to be infected is also important as it determines the rate of the first wave of infection. The higher the out degree of the first infected node, the more likely that the epidemic will go on longer.

There are several weaknesses to my research. First, the probability in this model leads to non-replicable results, meaning there could be some inaccuracies in the experiment when the sample size is small. More data should be collected and normalized to prevent such an error. Moreover, the models I applied here is very basic and somewhat inaccurate of describing real life situations. We could add attributes like different infectious period and contagion probability for different nodes, or even adopt a changing network to deploy our models to achieve more accurate solutions.

I hope that this project can demonstrate how the parameters of the epidemic models can affect the outcomes of an epidemic, which is helpful in social network analysis. Although the data presented might not be perfect, it somewhat represents certain graphical behavior in our society. The term “epidemic” here might not just mean the spread of disease, but also including the spread of a certain idea, ideal, or a behavior, which makes this whole experiment more meaningful than just applying in studying disease transmission.