

CSCI4190 Introduction to Social Networks  
Project Report  
Task 4 Epidemics

Yu Chi To SID: 1155110447

April 28, 2020

# 1 Abstract

This social network project demonstrates a simple simulation of a few traditional but important epidemic models, such as SIR, SIS and SIRS models by using a Slashdot dataset as the underlying social network to be infected. The infectious behaviour will be simulated for observation in order to understand the characteristic of the three epidemic models and the interactions between the key parameters, infection probability, length of the period of infectious, length of the period of removed, number of initially infected and the network structure. Stanford Network Analysis Platform and Numpy is used for simulation of the epidemic models and social network. For epidemic models that require a long running time, Scikit-Learn is used to predict upon data simulated.

# Contents

<b>1</b>	<b>Abstract</b>	<b>i</b>
<b>2</b>	<b>Objective</b>	<b>1</b>
<b>3</b>	<b>Methodology</b>	<b>1</b>
3.1	Data Source . . . . .	1
3.2	Tool . . . . .	1
3.3	Experiment Design . . . . .	1
<b>4</b>	<b>Dataset Statistics</b>	<b>1</b>
4.1	Graph Statistics . . . . .	1
4.2	In-degree . . . . .	2
4.3	Out-degree . . . . .	3
4.4	Strongly Connected Components . . . . .	4
<b>5</b>	<b>Simulation Result</b>	<b>4</b>
5.1	Prediction on Termination on SIS and SIRS Epidemic Simulation Model . . . . .	4
5.2	Effect of Infection Probability . . . . .	6
5.2.1	SIR Simulation Result . . . . .	6
5.2.2	SIRS Simulation Result . . . . .	7
5.2.3	SIS Simulation Result . . . . .	7
5.3	Effect of Length of Infection Period . . . . .	7
5.3.1	SIR Simulation Result . . . . .	7
5.3.2	SIRS Simulation Result . . . . .	7
5.3.3	SIS Simulation Result . . . . .	8
5.4	Effect of Numbers of Initial infectious . . . . .	8
5.4.1	SIR Simulation Result . . . . .	8
5.4.2	SIRS Simulation Result . . . . .	8
5.4.3	SIS Simulation Result . . . . .	8
5.5	Effect of Length of Removed Period . . . . .	8
5.5.1	SIRS Simulation Result . . . . .	9
5.6	Effect of Network Structure on Epidemic Models . . . . .	9
5.6.1	SIRS $p = 0.7, r = 1, i = 3, s = 3$ Simulation Result . . . . .	9
<b>6</b>	<b>Analysis on SIR, SIRS, SIS Epidemic Models</b>	<b>9</b>
6.1	Difference and Similarities between SIR, SIRS and SIS Epidemic Models . . . . .	9
6.2	SIR Model . . . . .	10
6.3	SIRS Model . . . . .	11
6.4	SIS Model . . . . .	12
6.5	Effect of Network Structure on Epidemic . . . . .	13
6.6	Conclusion . . . . .	13
<b>7</b>	<b>Appendix</b>	<b>14</b>
<b>8</b>	<b>References</b>	<b>18</b>

## List of Figures

1	Miniature graphs of SIRS and SIS Epidemic Model Simulation Result . . . . .	5
2	Log transformed SIS simulation data of $s = 1$ . . . . .	12
3	SIRS $p=0.4, r=1, i=3, \text{init infected}=3$ . . . . .	14
4	SIRS $p=0.8, r=1, i=3, \text{init infected}=3$ . . . . .	15
5	SIS $p=0.3, r=1, i=3, \text{init infected}=3$ . . . . .	15
6	SIS $p=0.5, r=1, i=3, \text{init infected}=3$ . . . . .	16
7	SIR $p=0.7, r=1, i=20, \text{init infected}=3$ . . . . .	16
8	SIRS $p=0.7, r=1, i=20, \text{init infected}=3$ . . . . .	17
9	SIS $p=0.7, r=1, i=20, \text{init infected}=3$ . . . . .	17

## 2 Objective

Analysis on the 3 epidemic models, SIR, SIS, SIRS. And effect of network structure has upon the epidemic models.

## 3 Methodology

### 3.1 Data Source

A Slashdot dataset is chosen to be the underlying social network of this epidemic simulation. Slashdot is a science and technology related social news website which is famous for its specific user community. In 2002, Slashdot allowed users to tag each other as friends or foes. The network used contains friends/foes links between the users of the Slashdot in February[1].

### 3.2 Tool

Snap.py by the Stanford University was used to simulate the social network. Graph manipulation and social network simulation was done using Snap.py on Windows 10 and Windows Subsystem for Linux[2].

Numpy is used in the program for stating the state of each node in the social network and for after-processing the simulation data[3].

Scikit-learn is used to do linear regression on simulation data of SIS model for prediction in order to replace long computation of equilibrium points of SIS model[4].

A list of other libraries used for analysis of data is provided below[5].

- Matplotlib [6]

### 3.3 Experiment Design

In this project, the three SIR, SIRS, SIS epidemic models are simulated using SNAP.py on Python 3. Every node in the network is assumed to be a person and every directed edge in the network is assumed to be a contact. Initially, all nodes in the network is assumed susceptible to the epidemic(S). A random number of  $s$  people in the network would be infectious(I) by the epidemic. And for each contact, the disease will have  $p$  probability to transmit to the uninfected nodes in each wave( $t$ ). For each infected person, they will be infectious for  $i$  waves. And if they have a temporary immune to the epidemic after infection, the immunity will exist for  $r$  days. The epidemic ends at wave  $t_{end}$  if there is no infectious(I) in network.

A series of tests will be conducted by changing parameters  $p, r, i, s$  for observation on SIR, SIRS and SIS epidemic models. The observation will be based on average amount of infectious(I) and time( $t$ ) needed for the epidemic to end and no nodes in the network are infected.

As the time( $t$ ) needed for epidemic to end in SIRS and SIS models is large, the amount of iterations of infection waves will be limited to 100 iterations and the termination time of the epidemics will be predicted using regression on peak infectious values and the number of infectious of each wave.

The effect of network structure will also be investigated by randomly removing or adding a certain amount of edges to increase and decrease effective diameter of the network and the size of strongly connected components to investigate the effect of network structure in epidemics.

## 4 Dataset Statistics

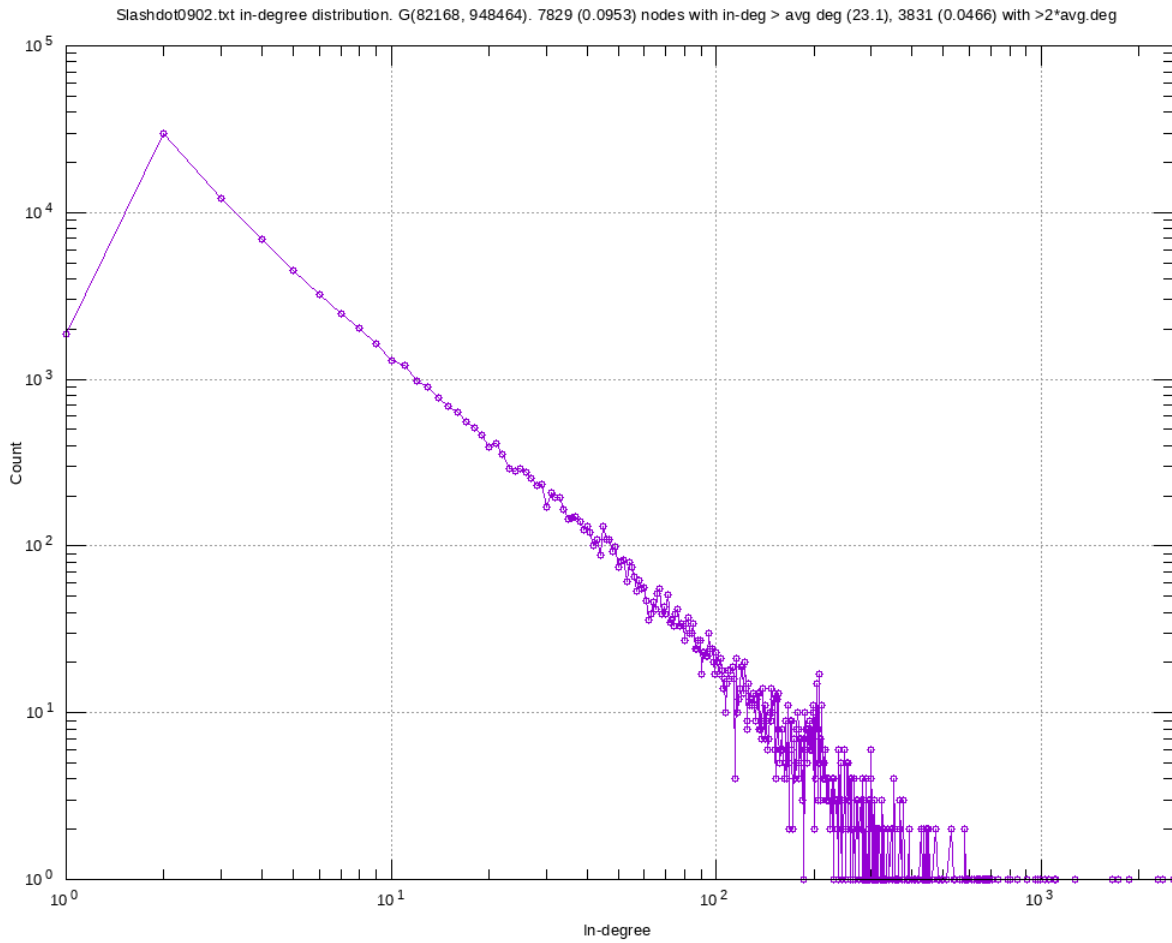
The table below shows the data generated using SNAP.py on the dataset Slashdot0922. 2009[1], [5]. The dataset is acquired from SNAP Stanford[5].

### 4.1 Graph Statistics

The table below shows statistics of the Slashdot0922 dataset[5] generated using SNAP.py[2].

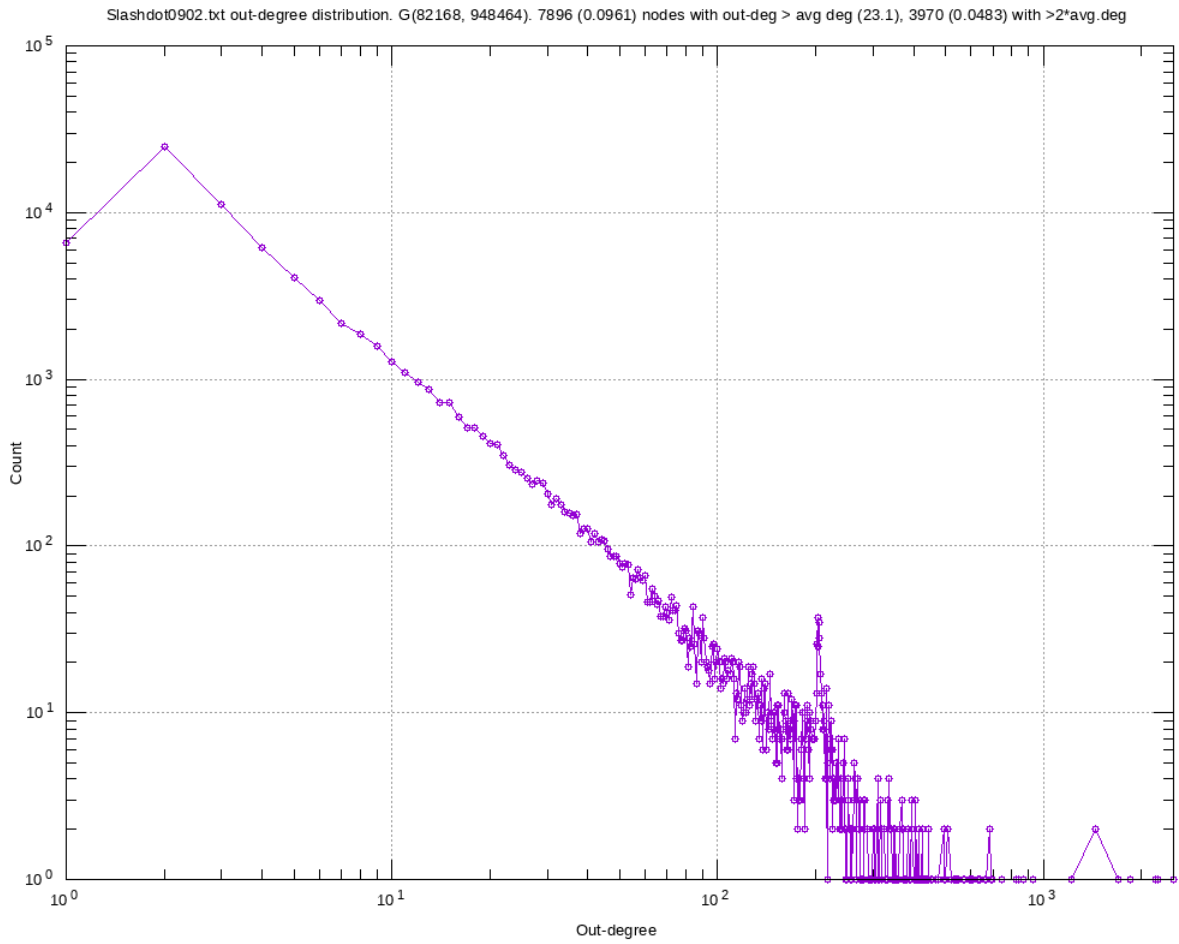
Slashdot0922:	Directed
Nodes:	82168
Edges:	948464
Zero Degree Nodes:	0
Zero In-degree Nodes:	0
Zero Out-degree Nodes:	3727
Non-zero In-out Degree Nodes:	78441
Unique Directed Edges:	870161 <sup>1</sup>
Unique Undirected Edges:	504230 <sup>2</sup>
Unique Bidirectional Edges:	365931
Self Edges:	78303
Zero Out-degree Nodes (Self Edges Removed):	10277
Non-zero In-out Degree Nodes (Self Edges Removed):	71891
Closed Triads/Closed Triangles:	602592
Open Triads/Open Triangles:	73175813
Fraction of Closed Triads:	0.008168
Average Clustering Coefficient:	0.063449
Size of Largest Strongly Connected Component:	0.86782
Weakly Connected:	True
Size of Largest Weakly Connected Component:	1.0
90% Effective Diameter:	4.688472 <sup>3</sup>
95% Effective Diameter:	4.9328
Approx. Full Diameter:	11

## 4.2 In-degree



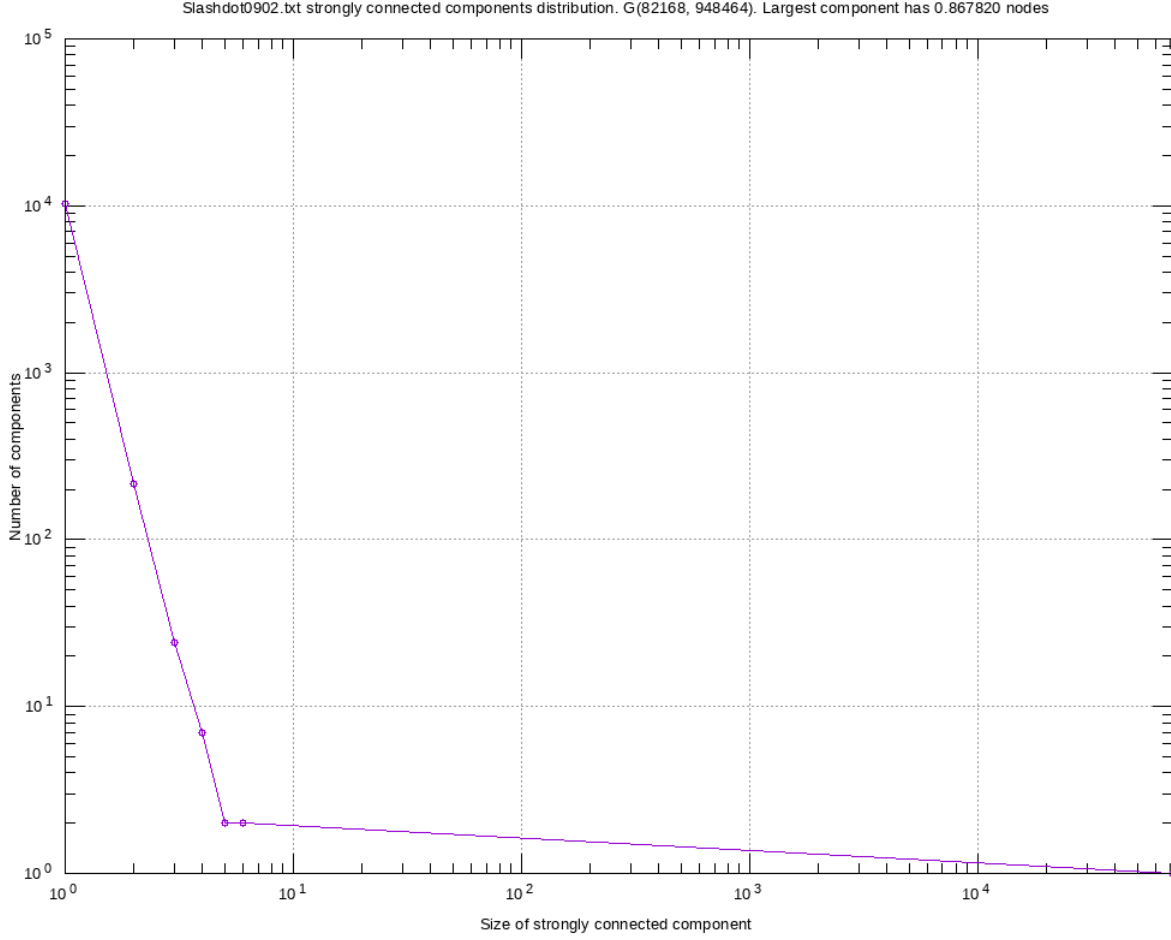
	Number of Edges
Max:	2553
Average:	1.4759
Min:	1

### 4.3 Out-degree



	Number of Edges
Max:	2511
Average:	1.3189
Min:	0

## 4.4 Strongly Connected Components



Given the network is weakly connected, and there is no nodes of zero in and out degrees, all nodes are connected to each other with at least one in or out edge. The small 90%, 95% effective diameter and the 0.86782 fraction of nodes in the largest strongly connected component demonstrate a large amount of nodes are interconnected and there is a probability of  $0.86782^2$  that there exists a path from any node in the node in the network to another node in the network. And after removing self edges, there are 0.010835 fraction of nodes have no outgoing edges in the network, implies they are only included in some other nodes' circles, forms sinks in the network but not connect themselves to each other.

## 5 Simulation Result

### 5.1 Prediction on Termination on SIS and SIRS Epidemic Simulation Model

Since SIS and SIRS epidemic model takes a long time to terminate, and is not able to terminate within 100 iterations, prediction of termination is done using regression. The regression is based on 1 observation.

**Observation 1:** The decrease of peak infectious value follows a concave up curve. In infectious graphs of SIRS and SIS epidemic models, the peak decreases and when linked up, forms a concave up curve.



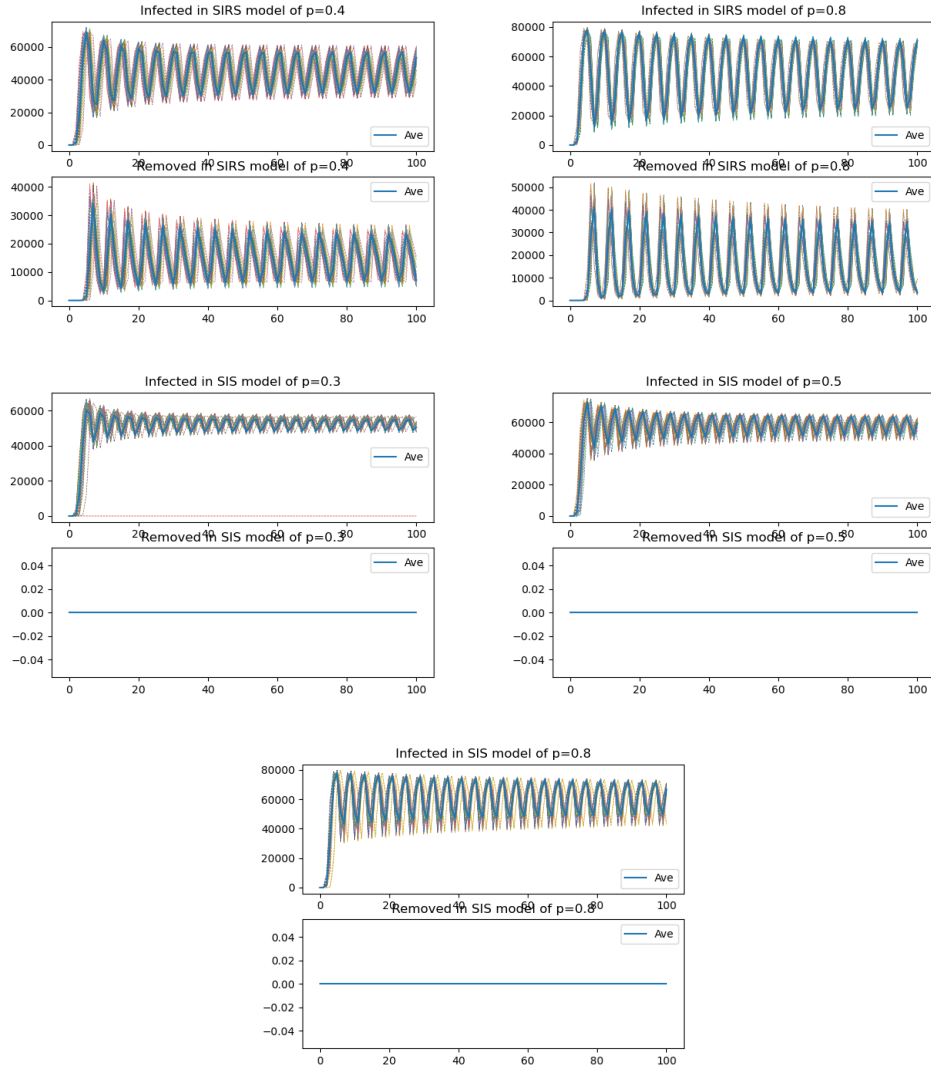
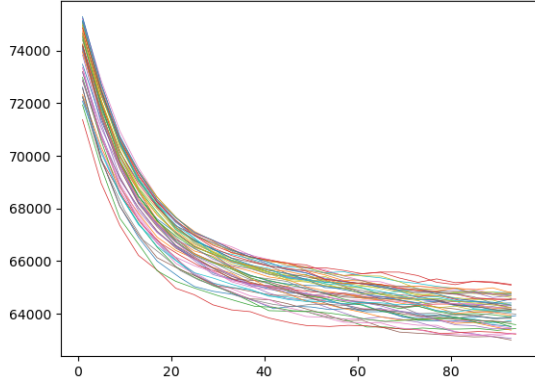
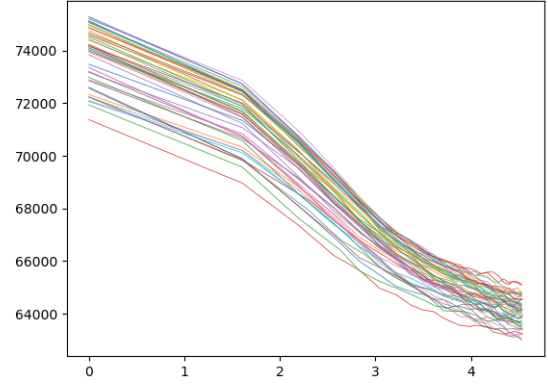


Figure 1: Miniature graphs of SIRS and SIS Epidemic Model Simulation Result

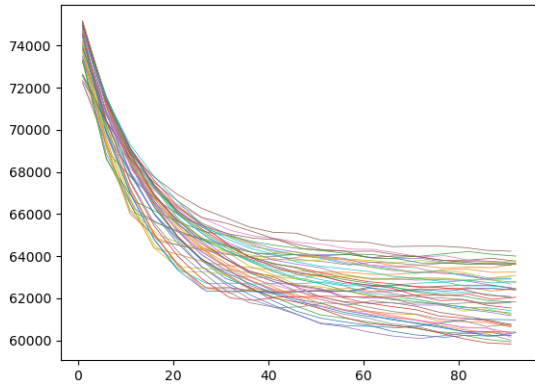
Therefore, an educated guess that the decay follows the logarithmic model  $y = w_1 \log(x) + w_0$  is made. SIS and SIRS data are preprocessed by logarithmic transformation for linear regression for epidemic termination prediction.



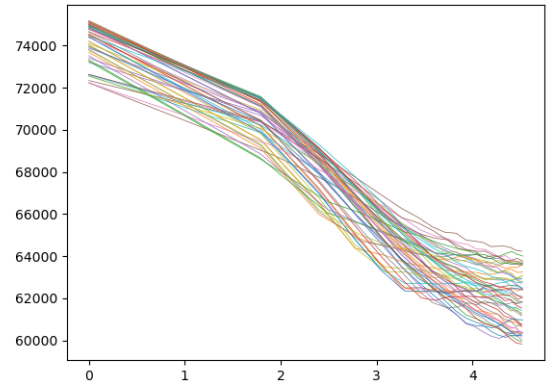
Raw SIS peak value diagram



Log transform processed SIS peak value diagram



Raw SIRS peak value diagram



Log transform processed SIRS peak value diagram

The result shows that the decay of peak values follows logarithmic model. Logarithmic transformation is applicable to some of the SIRS and SIS models.

## 5.2 Effect of Infection Probability

Model Simulation Parameters:  $i = 3, r = 1, s = 3$

The following are the result from 100 simulations of each  $p$  levels of each epidemic model respectively.

### 5.2.1 SIR Simulation Result

SIR	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	p=0.9	p=1.0
Avg. Termination ( $t$ ):	19.12	17.58	16.41	15.37	14.28	13.7	13.0	12.08	11.6	11.37
Avg. Fraction of infected:	0.66472	0.85094	0.92685	0.96445	0.98348	0.99281	0.99729	0.99925	0.99991	1.0

### 5.2.2 SIRS Simulation Result

SIRS	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	p=0.9	p=1.0
Predicted Avg. Termination (t):	$5.9 + e^{12.378}$	$5.76 + e^{16.010}$	$5.26 + e^{23.423}$	$5.14 + e^{23.639}$	$5.08 + e^{23.153}$	$4.9 + e^{23.494}$	$4.78 + e^{26.516}$	$4.82 + e^{31.904}$	$4.74e^{40.203}$	$12.0^4$
Avg. First Infection Peak:	5.9	5.76	5.26	5.14	5.08	4.9	4.78	4.82	4.74	4.74
Avg. Wave Between Peak(t):	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	$_{-5}$
Avg. Max Fraction of infectious:	0.50035	0.70016	0.79625	0.85564	0.90212	0.92747	0.94059	0.96067	0.96716	0.97088

### 5.2.3 SIS Simulation Result

SIS	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	p=0.9	p=1.0
Predicted Avg. Termination (t):	$6.0 + e^{6.5067}$	$5.66 + e^{35.1872}$	$5.32 + e^{31.7277}$	$5.18 + e^{30.3105}$	$5.06 + e^{29.5604}$	$4.96 + e^{29.5536}$	$4.76 + e^{30.7959}$	$4.92 + e^{37.5641}$	$4.68 + e^{49.4019}$	$_{-6}$
Avg. First Infection Peak:	6.0	5.66	5.32	5.18	5.06	4.96	4.76	4.92	4.68	4.68
Avg. Wave Between Peak(t):	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Avg. Max Fraction of infectious:	0.52313	0.68936	0.78111	0.85573	0.89919	0.92837	0.94482	0.95975	0.96489	0.97377

## 5.3 Effect of Length of Infection Period

Model Simulation Parameters:  $p = 0.7, r = 1, s = 3$

The following are the result from 50 simulations of each  $i$  levels of each epidemic model respectively.

### 5.3.1 SIR Simulation Result

SIR	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=20
Avg. Termination (t):	9.18	11.44	12.84	14.14	15.44	16.74	18.16	19.34	20.78	21.88	31.88
Avg. Fraction of infected:	0.94104	0.98917	0.99729	0.99925	0.99978	0.99993	0.99998	0.99999	1.0	1.0	1.0

### 5.3.2 SIRS Simulation Result

SIRS	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=20
Predicted Avg. Termination (t): <sup>7</sup>	$3.68 + e^{4.6743}$	$4.16 + e^{19.4561}$	$4.87 + e^{26.8814}$	901.4842	1357.87	1913.31	3033.82	4475.72	7813.95	16386.95	$_{-8}$
Avg. First Infection Peak:	3.68	4.16	4.87	5.48	6.28	7.06	7.94	8.96	9.8	12.54	$11.0^9$
Avg. Wave Between Peak(t):	3.01	4.01	5.02	6.03	7.05	8.07	9.07	10.07	11.14	12.21	$\approx 23^{10}$
Avg. Max Fraction of infectious:	0.54819175	0.83670018	0.94382144	0.98115787	0.99390809	0.99800506	0.99927344	0.9996624	0.99984203	0.9999243	1.

<sup>4</sup>actual average termination is taken as when p=1.0, the SIRS model simulations succeeded in terminating

<sup>5</sup>SIRS p=1.0 does not persists in this social network

<sup>6</sup>An increasing slope is observed in linear regression

<sup>7</sup>Predicted Average Termination for  $i \geq 4$  used linear regression due to a positive  $R^2$  score in linear regression, not log regression and linear data plotted are not skewed and diverged

<sup>8</sup>Non-negative slope detected in linear regression and fraction of infectious reach 1.0 periodically

<sup>9</sup>infectious fraction reach 1.0 at  $t = 11$  in average

<sup>10</sup>Observation from figure of SIRS  $p = 0.7, i = 20, r = 1, init\ infect = 3$

### 5.3.3 SIS Simulation Result

SIS	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=20
Predicted Avg. Termination <sup>11</sup> : ( $t$ )	$3.62 + e^{4.0091}$	$4.26 + e^{24.959}$	$4.8 + e^{30.138}$	$5.54 + e^{39.511}$	1602.9	2497.2	3664.3	5512.0	16532.27	16386.95	<sup>12</sup>
Avg. First Infection Peak:	3.62	4.26	4.8	5.54	6.18	7.0	7.94	8.66	9.54	12.18	$11.0^{13}$
Avg. Wave Between Peak( $t$ ):	2.0055	3.0026	4.0026	5.0111	6.0280	7.0338	8.0364	9.1039	10.090	11.163	$\approx 21^{14}$
Avg. Max Fraction of infectious:	0.56421	0.83312	0.94710	0.98251	0.99392	0.99800	0.99928	0.99971	0.99985	0.99992	1.0000

## 5.4 Effect of Numbers of Initial infectious

Model Simulation Parameters:  $p = 0.7, r = 1, i = 3$

The following are the result from 50 simulations of each  $s$  levels of each epidemic model respectively.

### 5.4.1 SIR Simulation Result

SIR	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
Avg. Termination ( $t$ ):	13.48	13.06	12.84	12.58	12.64	12.64	12.42	12.5	12.5	12.32
Avg. Fraction of infected:	0.99727	0.99731	0.99729	0.99730	0.99732	0.99729	0.99733	0.99726	0.99735	0.99731

### 5.4.2 SIRS Simulation Result

SIRS	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
Predicted Avg. Termination : ( $t$ )	$5.3 + e^{25.788}$	$5.04 + e^{24.616}$	$4.86 + e^{26.881}$	$4.76 + e^{26.255}$	$4.7 + e^{26.806}$	$4.7 + e^{27.585}$	$4.52 + e^{29.612}$	$4.36 + e^{28.599}$	$4.44 + e^{27.806}$	$4.34 + e^{28.121}$
Avg. First Infection Peak:	5.3	5.04	4.86	4.76	4.7	4.7	4.52	4.36	4.44	4.34
Avg. Wave Between Peak( $t$ ):	5.0267	5.0189	5.0211	5.0178	5.0211	5.0178	5.0267	5.0356	5.03	5.0355
Avg. Max Fraction of infectious:	0.946	0.94776	0.94382	0.94591	0.94373	0.94148	0.93761	0.93893	0.94044	0.93847

### 5.4.3 SIS Simulation Result

SIS	s=1	s=2	s=3	s=4	s=5	s=6	s=7	s=8	s=9	s=10
Predicted Avg. Termination : ( $t$ )	$5.12 + e^{18.581}$	$5.0 + e^{30.855}$	$4.8 + e^{30.138}$	$4.84 + e^{31.349}$	$4.66 + e^{30.956}$	$4.54 + e^{30.303}$	$4.42 + e^{30.479}$	$4.48 + e^{31.245}$	$4.52 + e^{31.340}$	$4.4 + e^{31.886}$
Avg. First Infection Peak:	5.12	5.0	4.8	4.84	4.66	4.54	4.42	4.48	4.52	4.4
Avg. Wave Between Peak( $t$ ):	4.0061	4.0052	4.0026	4.0026	4.007	4.0052	4.0061	4.007	4.0052	4.007
Avg. Max Fraction of infectious:	0.94478	0.94469	0.9471	0.94224	0.94001	0.94015	0.94142	0.94017	0.93989	0.93943

## 5.5 Effect of Length of Removed Period

Model Simulation Parameters:  $p = 0.7, i = 3, s = 3$

The following are the result from 50 simulations of each  $r$  levels of each epidemic model respectively.

Only SIRS model will be tested as SIS model have no removed state and SIR model will not convert back to susceptible state after removed.

<sup>11</sup>For  $i \geq 5$ , the peak:time graph approaches linear

<sup>12</sup>Non-negative slope detected in linear regression of data and fraction of infectious reach 1.0 periodically

<sup>13</sup>infectious fraction reach 1.0 at  $t = 11$  in average

<sup>14</sup>Observation from figure of SIS  $p = 0.7, i = 20, r = 1, init\ infect = 3$

### 5.5.1 SIRS Simulation Result

SIRS	r=1	r=2	r=3	r=4	r=5	r=6	r=7	r=8	r=9	r=10
Predicted Avg. Termination (t):	$4.86 + e^{26.881}$	$4.8 + e^{25.254}$	$690.55^{15}$	$12.72^{16}$	13.46	12.9	12.74	12.82	12.96	12.98
Avg. First Infection Peak:	4.86	4.8	4.9	4.88	4.88	4.82	4.86	4.88	4.88	4.9
Avg. Wave Between Peak(t): <sup>17</sup>	5.0211	6.0467	7.2965	8.7042	10.3333	-	-	-	-	-
Avg. Max Fraction of infectious:	0.94382	0.94588	0.95088	0.94825	0.94518	0.94684	0.94659	0.94637	0.94475	0.94523

## 5.6 Effect of Network Structure on Epidemic Models

New Network Statistics after removing 50% edges from original graph:

Slashdot0922:	Directed
Nodes:	82168
Edges:	474232
Zero Degree Nodes:	1791
Zero In-degree Nodes:	1817
Zero Out-degree Nodes:	13463
Non-zero In-Out Degree Nodes:	68679
Unique Directed Edges:	474232
Unique Undirected Edges:	394657
Self Edges:	68568
BiDir Edges:	227718
Closed Triangles:	41812
Open Triangles:	23656244
Fraction of Closed Triads:	0.001764
Connected Component Size:	0.945039
Strongly Connected Component Size:	0.505501
Approximate Full Diameter:	11
90% Effective Diameter:	5.024003

We can see that the 90% effective diameter increased from 4.688472 to 5.024003, and the size of largest strongly connected component decreased from 0.86782 to 0.505501. This shows a sparser network, and with 0.0218 fraction of nodes are isolated.

### 5.6.1 SIRS $p = 0.7, r = 1, i = 3, s = 3$ Simulation Result

The following result is based on 25 simulations of SIRS model with  $p = 0.7, r = 1, i = 3, s = 3$ .

Predicted Avg. Termination(t):	$5.36 + e^{24.509}$
Avg. Max Fraction of infectious:	0.85682
Avg. First Infection Peak:	5.36
Avg. Max Fraction of infectious:	5.0333

## 6 Analysis on SIR, SIRS, SIS Epidemic Models

### 6.1 Difference and Similarities between SIR, SIRS and SIS Epidemic Models

In formulation, the SIR model and the SIS model are highly related to the SIRS epidemic model. The SIRS model can be seen as a generalization of the SIR and SIS model with  $r = \infty$  or  $r = 0$  respectively. Therefore, there shall be some pattern between the models in common. For example, the pattern of changing of fraction of

<sup>15</sup>Linear regression used due to the large s.d. of time and magnitude of peaks

<sup>16</sup>0.88 fraction of simulations terminates

<sup>17</sup>r=6 to r=10 have all simulations terminates without second peak arising

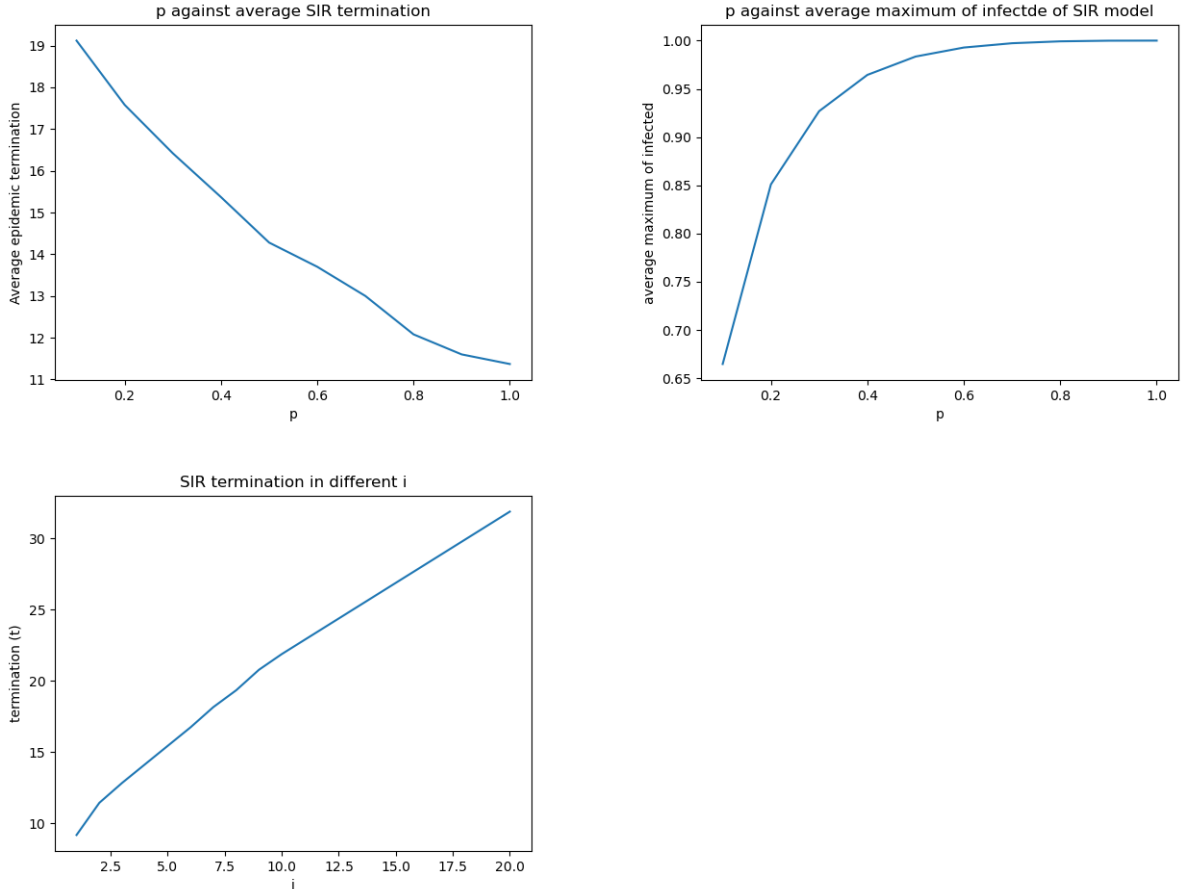
infected/infectious, and the pattern of changing in termination. And this can be observed from the simulation results also.

However, the difference in immunity of the three models lead to certain different characteristics of the models, such as a promised termination of epidemic in SIR model in finite graph as there is no replenish of susceptible nodes. Or a predicted direct backward transmission of SIS model compared to SIRS model which does not allow the direct backward transmission due to the existence of removed state forming a barrier.

But in general, the reasons leading to come of the characteristics of the models should be similar due to their high similarity.

## 6.2 SIR Model

From the simulation result, there exists an inversely proportional relationship between  $p$  and termination of the epidemic. While the number of infected is proportional  $p$ . The relationship between  $p$  and termination of epidemic or the number of infected is also non-linear by observation from simulation result. The decrease of termination with respect to increasing  $p$  shows an exponential decay, and the increase of maximum infected show a logarithmic increase.



For  $i$ , we can observe that the average termination of the epidemic increases linearly when the length of infectious period increased. The average fraction of infected shows a logarithmic increase also.

To explain the exponential decay of SIR epidemic termination when  $p$  increased, at each wave  $a$ , there will be  $k^a$  contacts. And the expected number of infected will be  $p^a k^a$ , which indicates an exponential growth of the number of infected. However, there is finite person in social network, and the infected will not become susceptible again. Without replenish in susceptible, the epidemic is unable to continue to transmit, thus terminates.

$$\sum_{i=1}^a p^i k^i \leq |\text{population in social network}|$$

For the logarithmic increase observed when  $p$  or  $i$  increased, it is related to the increase of probability of infection. Firstly, the probability distribution of nodes having  $\geq \alpha$  edges follows an exponential decrease function according to the power law. And a node with more edges connected to has a higher probability of infection of  $(1 - (1 - p)^{|\text{neighbors}|})$ . Therefore we can make an assumption that the nodes with more neighbors will be more likely to be infected before the nodes with less neighbors due to a higher probability of infection of the person. This matches the real world scenario of epidemic and the formulation of basic reproduction number.

And both changing  $p$  or  $i$  will change the actual probability of infection of the susceptible node in the graph, this can be explained using the function of probability of being not infected  $(1 - p)^{i|\text{infected neighbors}|}$ . Therefore, in later stage of the epidemic, the susceptible nodes is likely to have few edges, so their actual probability of infection is lower. Thus lead to the logarithmic growth of number of infected.

For  $s$ , we can observe a decreasing trend of termination of epidemic when  $s$  increases, while the number of infected is not affected. Changing the number of initially infected affects the termination of the epidemic only is because it changes the number of infected at the beginning. Moreover, with more nodes infected, we can expect to have more nodes have in-links from the infected nodes. Thus increasing the speed of infection at the beginning phase of the epidemic and therefore the epidemic enters the later phase and terminates faster.

### 6.3 SIRS Model

The SIRS model consists of 4 changable parameters  $p, r, i, s$  in the experiment. And the simulations of SIRS epidemic model mostly failed to terminate within 100 waves. Therefore, regression is used to estimate the termination.

First, on the post-processing of the result, we can observe from the figures of SIRS model simulation (fig. 3, 4) that there shows a non-linear decreasing trend on the number of maximum infectious nodes in the graph over time. This indicates the drop follows an exponential decreasing function. Thus a logarithm transformation is needed to before fitting the regression line. However, in experiments on the effect of infectious period on the SIRS epidemic model, a linear regression is directly used on the data as the peaks shows a near linear decreasing trend and fitting with  $y = w_1 \log(x) + w_0$  exhibit a high error. This is likely to be caused by the increase of  $i$ . From the results of SIR model, we can already observe that termination of SIR-modeled epidemics will terminate later when  $i$  increased. The SIR model is similar to a single peak in SIRS infection model. Therefore, with the same magnitude of decrease in number of infected of each peak, but the separation of each peak increased, we can expect the curvature of the function to decrease, thus tends to be a straight line. The increase in the average number of waves between peaks supports the argument. Moreover, when  $i$  increased, the probability of a node not to infect decreases. Therefore, the peak should increase compared to the peaks of smaller  $i$ . Therefore, the estimated decreasing function tends to be a straight line when  $i$  is large and the number of iterations is small.

For SIRS epidemic model, when  $p$  or  $i$  increases, the average maximum fraction of infectious increases following logarithmic increase. The reason is the same as SIR epidemic model, the increase in the probability of infection within a period increases the expected number of nodes of infectious, and the increase in probability affects the nodes with more in-links heavier than those with less in-links. Thus causing the logarithmic increase.

When  $p$  increases, the predicted termination increases exponentially until  $p = 1$ . For  $p = 1$ , the epidemic must transmit to any other susceptible node that the infectious node has outgoing edges to. This causes every node on the graph to be infected stage by stage. And forms a barrier by the currently infectious and removed nodes for the nodes recovered from removed state back to susceptible state. This can be perceived as a concentric circle, with the center at the initial infectious. The infectious transmit to any nodes it has outgoing edge to, and forms a circular ring of  $i$  width of band of infectious which expands outwards with time. And there is another ring of width  $r$  inside the ring of infectious for the removed nodes. And There is a circle within the ring of removed which are the infected nodes converted back to susceptible according to the SIRS model. Therefore the removed forms a natural barrier or protection to the infected, but now susceptible again nodes. Thus SIRS terminates when  $p = 1.0$  in a static social network like in the simulation, but not persists indefinitely like what the basic reproduction number suggests. For  $p < 1$ , it follows the basic reproduction number, a greater  $p$  may lead to  $R_0 > 1$ , indicating the epidemic may persists with a probability greater than 0. With infection probability less than 1, there may exists susceptible nodes providing as a bridge between the infectious ring and the susceptible circle in the model mentioned previously, allowing the epidemic to persists with oscillation in the number of infectious. And the termination increases exponentially when  $p$  increases, is because with the infection probability increased, it would be harder for a node to not get infected, thus increasing the probability the epidemic to start another wave of infection within the strongly connected social network when many of the nodes have get infected, but converted back to susceptible after  $i + r$  steps.

When  $i$  increases, average maximum fraction of infectious increases. And the increase follows logarithmic growth. When  $i$  increases, the width of the ring of infectious increases, but as  $i$  increases in an finite graph, the probability of overlap of these rings from different initial infectious increases. Thus the increase of average maximum fraction of infectious decreases when  $i$  increases. And from the data of  $1 \leq i \leq 3$ , we can observe an exponential growth in the termination of the epidemic. As  $i$  increases, the probability for a node to get infected increases. Therefore we can expect there is a higher chance for the epidemic to persists longer.

When  $s$  increases, the predicted termination of the epidemic fluctuates. Combining this observation with the average maximum fraction of infectious, I believe change in number of initially infected will not affect the SIRS epidemic in terms of termination.  $s$  increases may speed up the transmission of the epidemic in the beginning stage, but it will not affect the epidemic other than decreasing the time where the first peak of infectious comes.

For  $r$ , from the observation, when  $r$  is large enough, the SIRS epidemic terminates. This shows that  $r$  is inversely proportional to the termination of the epidemic. And increase in  $r$  can effectively increase the number of steps between each peak of infectious in the social network. This can be explained as increasing the width of the barrier between infectious nodes and the susceptible nodes that are recovered from the epidemic using the model mentioned in the explanation of  $p = 1$ .

And from observation, the average number of waves between the peaks of infectious in SIRS model is  $i+r+1$ . The  $i+r$  part is because of the time needed for the infectious nodes to recover and convert back to susceptible state. While the 1 may be a constant of the time needed for re-transmission in this social network graph.

## 6.4 SIS Model

The SIS model is another commonly discussed model like the SIR model. The nodes recovered from the epidemic will be in susceptible state again without immunity to the epidemic. Thus it is more likely for the SIS epidemic model to persist indefinitely. And except that SIS model has no Removed state, the mechanism of SIS model and SIRS model are highly similar. Therefore except for special reasons, the explanations will be referred to the suitable explanation in SIRS.

From the simulation results, when  $p$  increases, the predicted termination of the epidemic generally increases. This is because when the infection probability increases, it is less likely for the epidemic to cease as the probability of a node to not get transmitted is low. And the average maximum fraction of infectious increases exponentially with a ratio similar to that of the SIRS epidemic simulation. The reason of that is the same as explained in the part explaining the interaction of  $p$  and SIRS except that there is no removed state nodes forming a barrier from the susceptible nodes. And the epidemic persists when  $p = 1$  in SIS model as there is no removed nodes blocking the re-transmission of the epidemic.

For  $i$ , as stated in SIRS model analysis, when  $i$  increases, the curvature of the decreasing function of the peaks in SIS model decreases, causing the curve tends to be a linear line when  $i$  is large and the number of iterations is small. And the logarithmic increase of average maximum fraction of infectious is explained in SIRS.

For  $s$ , in  $2 \leq i \leq 10$ , it shows a similar predicted termination. However, when  $i = 1$ , the predicted value of termination of the SIS simulation is typically smaller, but the processed simulation data are highly uniform except for 1 outlier.

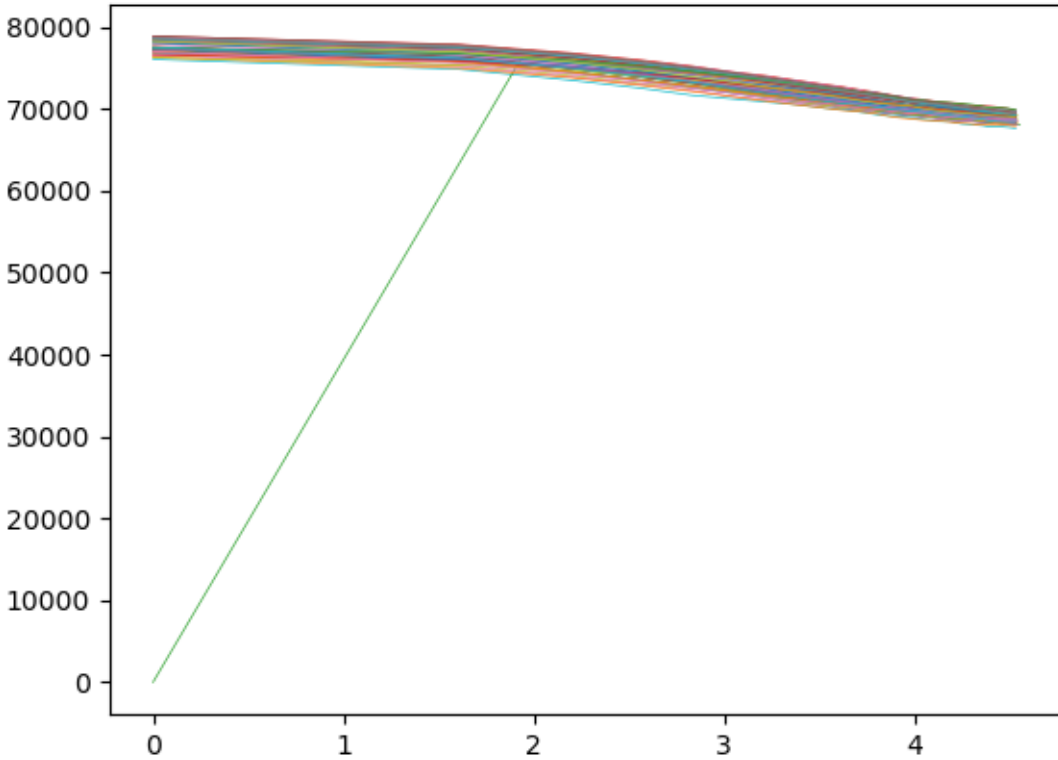


Figure 2: Log transformed SIS simulation data of  $s = 1$



Therefore, the difference on termination between  $i = 1$  and  $i = 2$  might be exhibiting a positive relationship between the number of initially infected and the termination of the epidemic. And the positive relationship might be due to the property of having no removed state in SIS model to block the transmission and allowing the node removed from infectious state to be converted back to infectious in the next step. Thus the increase of initially infected increases the stability and efficiency of the transmission in the graph, and helps in increasing the termination of the epidemic.

## 6.5 Effect of Network Structure on Epidemic

With half of the edges randomly removed from the graph, the 90% effective diameter increased from 4.688472 to 5.024003 while the size of the strongly connected component decreased from 0.86782 to 0.505501. This made the network sparser than the original network and with the distance between nodes increased.

We can observe a smaller predicted termination and a smaller average maximum fraction of infectious in the SIRS simulation with  $p = 0.7, i = 3, r = 1, s = 3$ . This demonstrates the effect of connectedness and average distance of the graph on the epidemic. With a higher average distance between the nodes and a smaller strongly connected component, the epidemic is harder to transmit, causing a shorter period will be needed for the epidemic to terminate and less people will be infected. This demonstrates the effect of quarantine and the current “social distancing” where people stays at home and connect to each other via the internet instead of having physical social contacts, which with these measures, the edge between people decreases and the size of strongly connected component shrinks. And when these properties of the network decreases, the epidemic terminates faster as the number of contacts decreases, so as the basic reproduction number. Thus showing the effectiveness of these measures and the relationship of distance and connectedness of the graph and the epidemic.

## 6.6 Conclusion

In conclusion, both infection probability and length period of infectious have huge effect on the three epidemic models. And the number of initial infected have less significant effect to the epidemic models. And change in length of period of removed helps to terminate the epidemic earlier, and bring a termination to the epidemic through quarantine. And the connectedness and the distance between the nodes can also affect the transmission of the epidemic. Thus showing the importance of “social distancing” and staying at home in the time when an epidemic occurs epidemics.

## 7 Appendix

Key figures used for observations or miniature version used will be provided in larger scale.

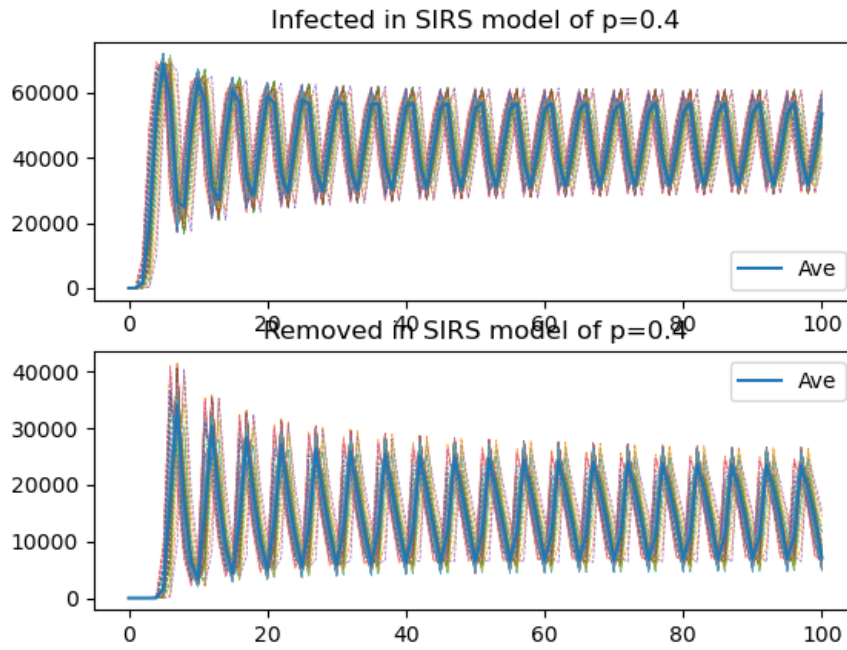


Figure 3: SIRS Model with  $p=0.4$   $r=1$   $i=3$  initial infected=3 simulation using Snap.py

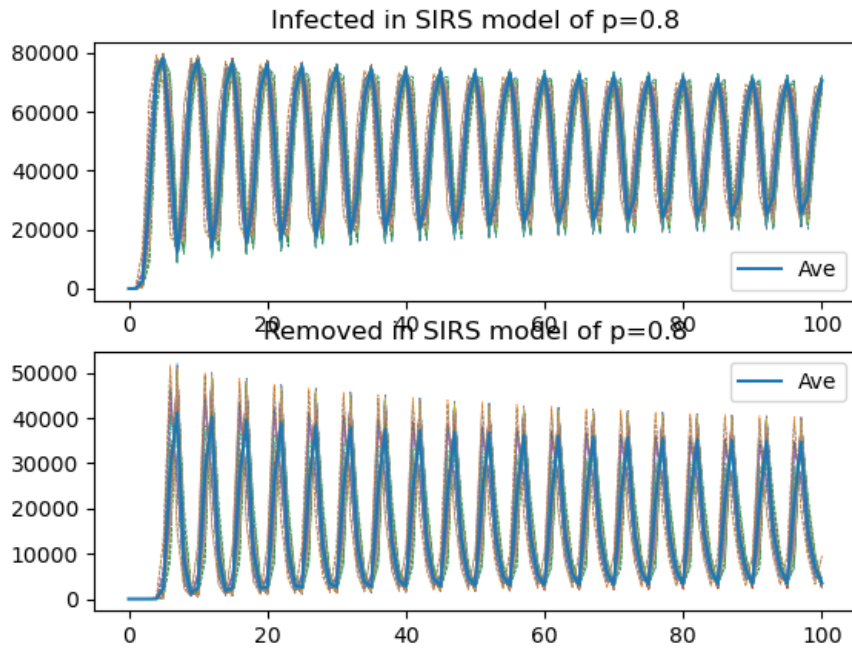


Figure 4: SIRS Model with  $p=0.8$   $r=1$   $i=3$  initial infected=3 simulation using Snap.py

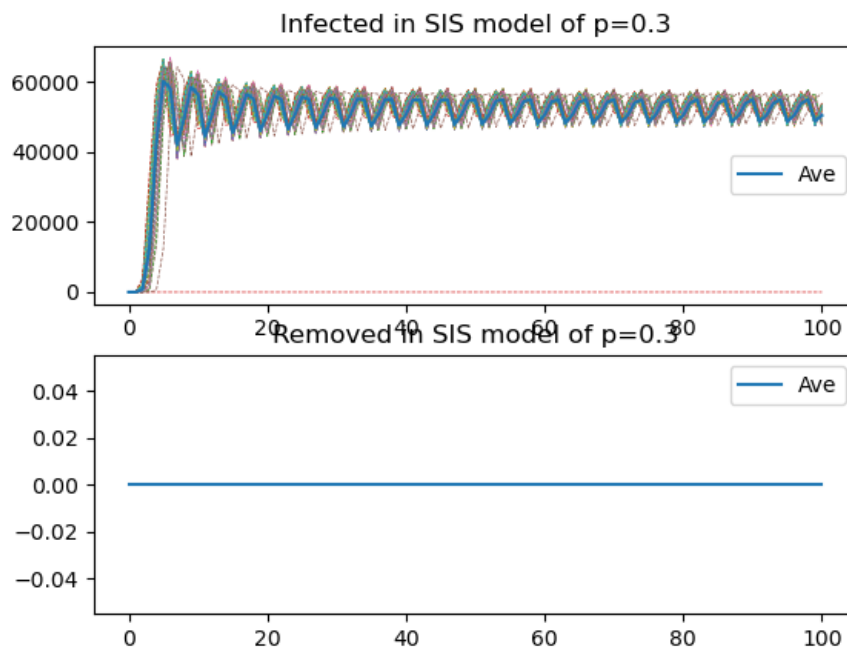


Figure 5: SIS Model with  $p=0.3$   $r=1$   $i=3$  initial infected=3 simulation using Snap.py

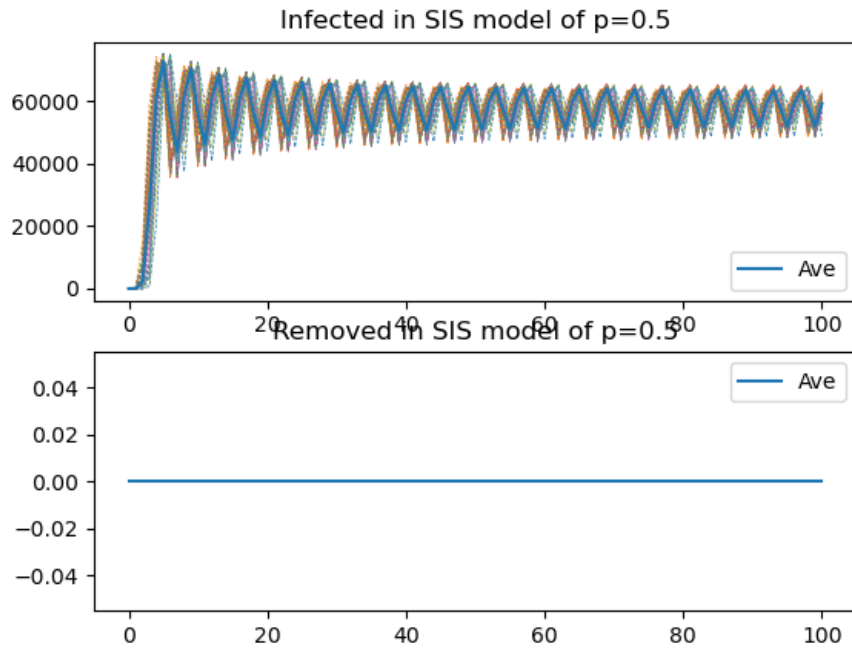


Figure 6: SIS Model with  $p=0.5$   $r=1$   $i=3$  initial infected=3 simulation using Snap.py

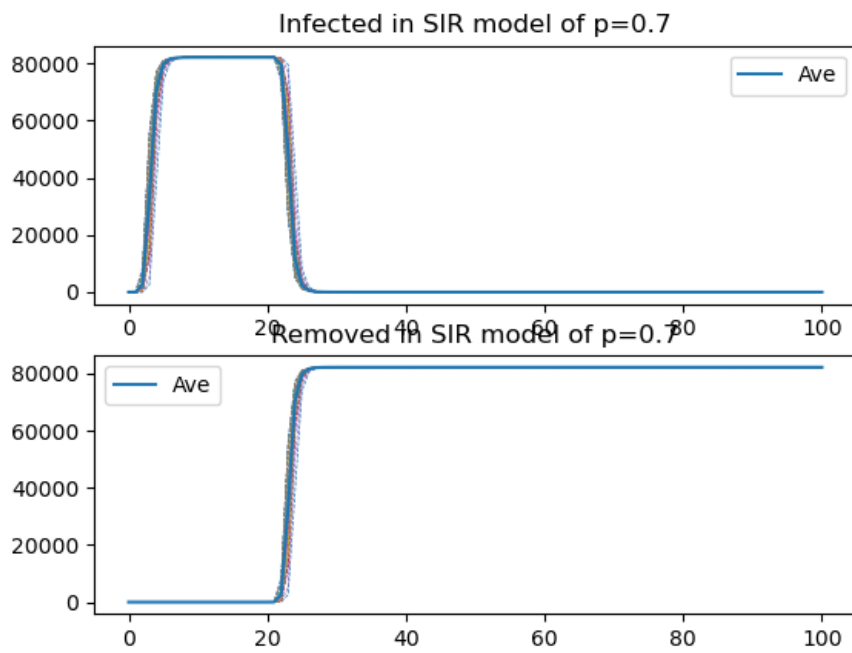


Figure 7: SIR Model with  $p=0.7$   $r=1$   $i=20$  initial infected=3 simulation using Snap.py

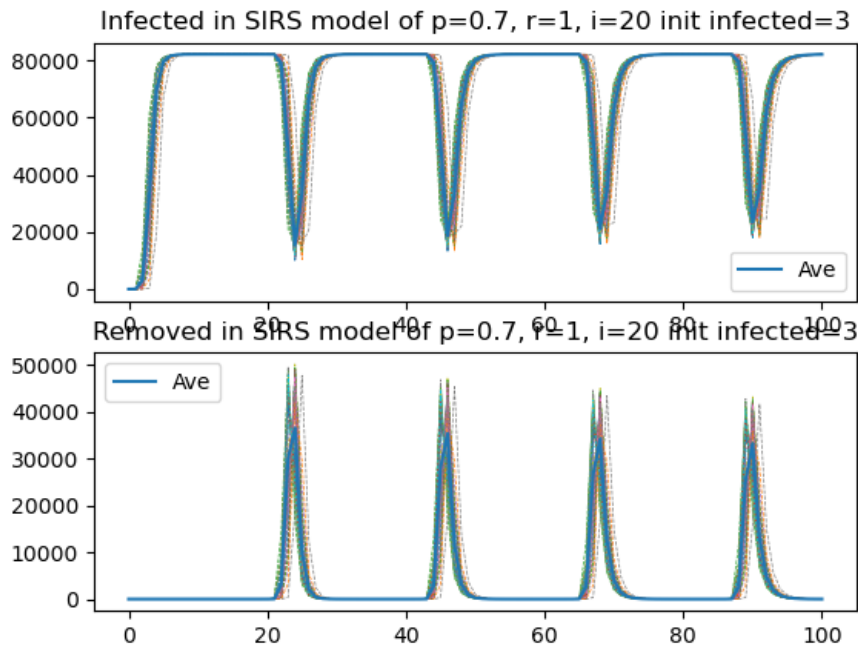


Figure 8: SIRS Model with  $p=0.7$   $r=1$   $i=20$  initial infected=3 simulation using Snap.py

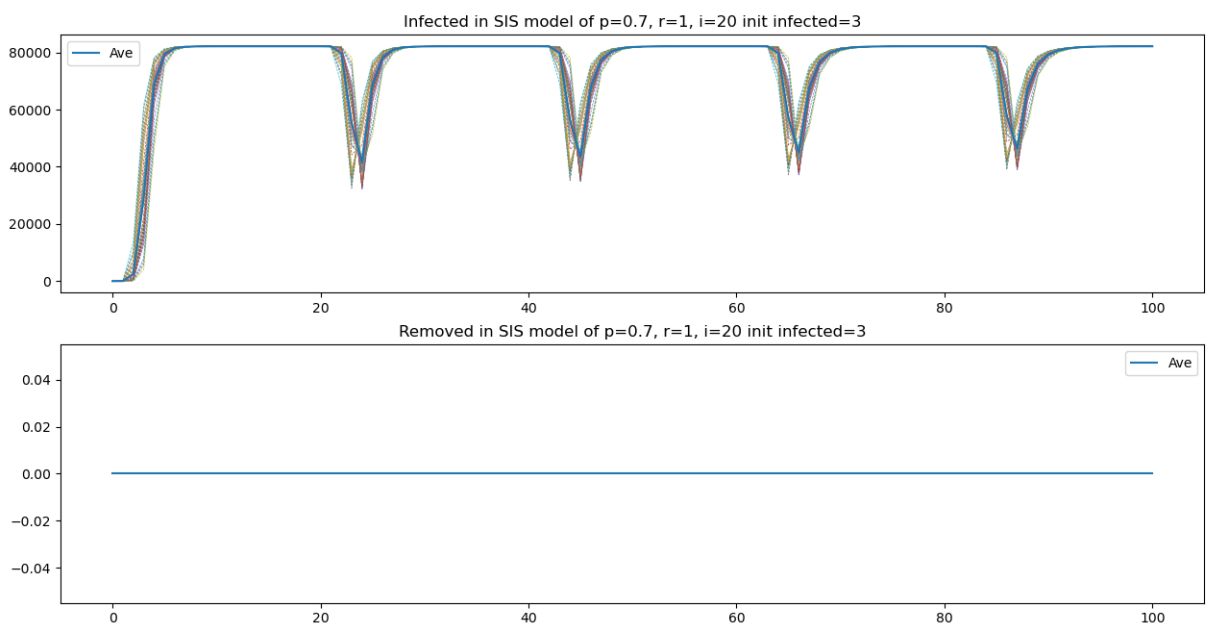


Figure 9: SIS Model with  $p=0.7$   $r=1$   $i=20$  initial infected=3 simulation using Snap.py

## 8 References

- [1] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters*, 2008. arXiv: 0810.1355 [cs.DS].
- [2] J. Leskovec and R. Sosič, “Snap: A general-purpose network analysis and graph-mining library,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 1, 2016.
- [3] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, bibinitperiodI. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: <https://doi.org/10.1038/s41592-019-0686-2>.
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [5] J. Leskovec and A. Krevl, *SNAP Datasets: Stanford large network dataset collection*, <http://snap.stanford.edu/data>, Jun. 2014.
- [6] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007. DOI: 10.1109/MCSE.2007.55.
- [7] D. Easley and J. Kleinberg, *Networks, crowds and markets: reasoning about a highly connected world*. Cambridge University Press, 2019.