SIR and SEIR Model for Disease Spreading COVID-19 Simulation

Mengchen Gong

mego6698@colorado.edu

Qinglu Sun

qisu4484@colorado.edu

Tianli Wu

tiwu5322@colorado.edu

1. INTRODUCTION

The outbreak of COVID-19 that started in Wuhan quickly swept the world in just three months. It is highly contagious, spreads quickly, has a long incubation period, is asymptomatic and remains contagious. Because of the lack of timely control and the high flow of people, the virus quickly spread across China, and then a blowout outbreak began globally. From the winter of 2019 to the spring of 2020, scientists analyzed and confirmed that the virus was a droplet infection. At the end of the summer of 2020, various countries analyzed the new subtypes of the new coronavirus and found that its infectivity increased while its lethality decreased. COVID-19 affects people around the world in every aspect, in the past months, we have seen the collapse of the medical system, we have become accustomed to this isolated lifestyle. The spread of covid-19 makes us realize the importance of studying the spread of infectious diseases in order to predict and control the spread of new crowns

1.1 Motivations

We hope to use disease spreading models to simulate the spread of infectious diseases in the population, perceive if the model fits the actual situation/or previous infection cases, thereby predicting the number of patients in the future. Determine the influence of policy in our model, gain insight of difference in models, why SEIR model fits COVID-19 best? and lastly what should we be aware of?

1.2 Related work

Parnell, D.[9] provided a rapid analysis of COVID-19 on mass sports gathering, he treated sports as a fundamentally network game, and showed sports gathering as a mirror to our network society and globalized operation. A few studies have been conducted on analyzing the spreading of COVID-19 using the SIR model since there are things in common between them. Chen et al.[7] proposed a time-dependent SIR model which tracks the probability of one susceptible becoming infected and the probability of one infected becoming susceptible to anticipate COVID-19 tendency. There are other studies focused on characterizing the spread of COVID-19, especially analyzing the impact on different genders, different ages groups and analyzing the average, mode and median incubation periods among the people based on the data they got [8]. Fanelli et al.[5] studied mainly on the susceptible-infected-recovered-deaths (SIRD) model to see the rate of infection, recovery and death in Italy, China and France and found that the recovery rate is about the same in different countries even though the infection rates are different because of the culture factors. Additionally, the study

found out the death rate is correlated to average age and health conditions the patients are in, which gave us an insight of the potential factors that we need to take into account before starting the simulations.

1.3 Challenges

Throughout the whole project, we've faced some challenges like limitations of the dataset; consider all different factors that are important for SIR and SEIR model, how to tilt each parameters in each modal to make sure it reflect back to the knowledge we've know from existing studies and papers, and so on.

1.3.1 Dataset

COVID-19 is still happening all around the world, for the dataset we got from Kaggle [2][3], the time range is limited that both of the dataset form Korean and the US are only for the beginning period. Therefore, analysis based on the 2 dataset we had was limited. We were only able to find the trend for the beginning 2 -3 months of the pandemic, which only corresponds to the beginning period of the pandemic.

Though there are companies like 1point3acress provided "get data" on their website that we were able to put in request for the access [12]. For all 3 requests we put in throughout 3 month, there was no reply from them so we were not able to get a better world wide dataset.

1.3.2 Math Model

There are certain assumptions that we are making that may not fit reality, we are making assumptions of 1. We are not considering the impact of population inflows and outflows on the model. 2. We are not considering asymptomatic populations. 3. Recovered population has antibodies and will not be infected again. And in the math model, we are assuming that the infection rate and recovery rate are fixed, which is not true.

1.3.3 SIR Model

We want to compare the simulation of the SIR model and the SEIR to see which model has better explaining of the outbreak.. S represents susceptible subjects. I represents infected subjects. And R represents recovered subjects. For the SIR simulation, we used an averaged rate of infection and recovery based on the U.S. data [2]. For the SIR model, we set the infected rate as the product of exposed rate and the infected rate in the exposed population. We applied three versions of the SIR model to explore the spreading

of the Covid-19. The first one is without isolation. The second one is isolation immediately. The third one is isolation delayed.

1.3.4   SEIR Model

There are countless parameters in the SEIR model that need to be taken into account. Based on the knowledge we knew from the COVID-19 transmission pattern right now, our simulation only used randomization to achieve partial parameters.

Superspreading events and normal spreading events for example. Based on Althoush's study, about 80% of transmissions are caused by 10%-20% of the infected population [1]. For our simulation, we only used an averaged rate of exposure, infection and recovery based on the U.S. data [2].

Tilting any parameter might affect the simulation results. For example, the number of days an exposed/infected person has before being put in isolation can change the result by a huge margin. In our SEIR model, we set the default delay for 2 days. However, in reality, it is not a single number that the number of delayed days can be changed by policies in the country, whether an exposed person knows he/she has been exposed. All those factors are hard to capture in the simulation.

1.3.5   Other Challenges

Disease Spreading in general has a lot of uncontrollable factors that will affect the result of the simulation. With the restricted scope for this project and the time we have, we would only be able to analyze a couple key factors, which are isolation at different stages, isolation length, delay before isolation, and probability a person gets exposed during isolation.

2.   METHODS

2.1   Math modal

The fraction of susceptible depends on the number of susceptible, contact between susceptible and fraction of infections, and the fraction of infected population, where b is the rate of transmission per S-I contact per unit time, k is the rate of recovery, and α is the rate at which an exposed person becomes infectious. For SIR, the Susceptible differential equation is $\frac{ds}{dt} = -bs(t)i(t)$. where s(t) is the susceptible fraction of the population, i(t) is the infected fraction of population. The removed differential equation is $\frac{dr}{dt} = ki(t)$, and the infected differential equation is $\frac{di}{dt} = bs(t)i(t) - k(i)$. For SEIR model, e(t) is the exposed fractional of the population, with adding the exposed statues, the susceptible differential equation is $\frac{ds}{dt} = -bs(t)i(t)$ the removed differential equation is $\frac{dr}{dt} = ki(t)$, the infected differential equation is $\frac{di}{dt} = \alpha e(t) - ki(t)$, and the exposed differential equation is $\frac{de}{dt} = bs(t)i(t) - \alpha e(t)$.

2.2   Kaggle Data

2.2.1   U.S. Data

The dataset includes numbers including cumulative positive, cumulative negative, cumulative recovered, death, total test results, death increase, positive increase, negative increase etc. We used data before the date of quarantine started in the US in order to determine the probability of infection and probability of recovered [2]. We also used this dataset to do visualizations of basic statistics of positive, recovered and death. We also used the data to compare with our prediction of fraction of positive and fraction of recovered.

2.2.2   Korean Data

The datasets include geological information of the patient, age, sex, infected by whom which is indicated by patient id, number of contact, policy, infecetion case which indicates that how the patient get infected and date of symptom, we used this dataset to perseve the infection pattern and analysis difference in policy and by using infection case to analysis what are major reason that the patient get infected, gain some insights of the communities activities [3].

2.3   SIR Model

The SIR model was built on the below assumptions:

- The recovered people will not get reinfected again.
- An infected patient will continuously spread COVID-19 until the patient is recovered.
- Patients will recover at night, so the patients can still spread COVID-19 on that day.
- Patients will stay at home when they need to be isolated.

In the network, simulation starts with a randomly selected node. A node represents a person. An edge represents that two individuals friend each other. If two individuals are friends and live in the same building, the weight of the edge is set to 2. If they live in different buildings, the weight is set to 1. The edge weight affects infection rate. The larger edge weight, the greater infected rate. In other words, people who live with infected people have a higher chance to get infected compared with the ones who live separately with infected people.

The algorithm of the SIR model without isolation is listed below. The breadth-first search is the cornerstone of building SIR simulation.

```
1. A starting node that was randomly picked was pushed to a
queue.
2. Start a while loop. The condition of the while loop is
that the queue is not empty.
   3. Get queue size. Loop through queue size.
      4. Pop one infected node from the queue.
       5. Check whether the node is recovered
          6. If the node is recovered, mark the node
recovered. Otherwise, add to the queue
    7. Get a list of susceptible neighbours of the infected
node.
       8. Loop through the list to check whether each
neighbour is infected
         9. If yes, add the node to the queue
```

The algorithm of the SIR model with isolation immediately is built based on the algorithm above. Removing 'otherwise, add to

the queue' (line 6) from the above algorithm becomes the algorithm of the SIR model with isolation immediately.

The SIR model with isolation immediately is extremely ideal. In the real world, people would not know they got infected immediately when they got infected. For example, test results will not be available immediately and people who don't have any symptoms as they get infected would get tests, and so on. Therefore, the SIR model needs to be improved. We considered that the infected node can spread the Covid-19 before isolation. Therefore, the SIR model with isolation delayed emerges.

The algorithm of the SIR model with isolation delayed is based on the SIR model with isolation. Adding a hashmap to track each infected node. Before the infected node is exposed several days in the public, keep adding the node to the queue if the node is not recovered.

2.4 SEIR Model

Built on top of the assumptions of SIR mode (3.3), we have a couple more assumptions below:

- Exposure duration: 6 days
- 0.6 exposure probability, 0.125 infected probability, 0.143 recovery probability
- An exposed person has ability to spread during its incubation time
- 0.5 chance an exposed person will be isolated with 2 days default delay
- All infected people will be isolated without delay
- Isolation simulation used a separated grid network with maximum 4 edges.

There are 3 main simulations we've done for the SEIR model with isolation length analysis, delay analysis, centrality analysis and lower exposed rate analysis during isolation time.

Set the whole spreading duration to 45 days, the first 2 modals are simalting in 100 nodes random network and 1000 nodes random network. For each simulation, we did 3 different simulations which are (1) no isolation, (2) isolation at infectious stage and (3) isolation 50% at exposure stage with 2 day delay, and 100% at infectious stage. For simulation (3), we plot the difference between 7-day isolation and 14-day isolation to see how the length of isolation will affect the trend of the COVID-19 spreading.

Before we apply the simulation on real-world data from Facebook100, we did the centrality analysis for the real-world data [4]. The goal is to see if the startnode is high or low (a) degree centrality, (b) harmonic centrality, (c) eigenvector centrality, (d) betweenness centrality will affect the spreading trend.

With knowledge of how centrality affects the spreading of the COVID-19, we then apply our 3 simulation model in the real-world data to see how the spreading trend looks, if it looks similar to the ones using a random network or if it is different. Same as above, isolate 7 days and 14 days has been done and analyzed.

After simulations for 3 different networks, we want to see how exposure probability during isolation and delay will affect the spreading trend. Exposure probability during isolation corresponds to how well the isolation period has been going. Unlike the SIR model, put in a grid graph for isolation in SEIR is trying to simulate the type isolation that puts each person in a single isolation room, the maximum contact they will have will be doctor, nurse and maximum direct neighbors, which is why we set the upper limit 4 edges during isolation. If the isolation method and process has been done very well, the exposure probability will be lower. In addition, we want to see how much delay will affect the spreading trend and if there is a delay immediately, will it be possible to stop the pandemic from getting worse.

3. RESULTS

We generated a bunch of plots to show our math modal, SIR model and SEIR model simulation spreading trend. We also used 3d plots to show how the probability of exposure during isolation will affect the spreading trend.

3.1 Math modal
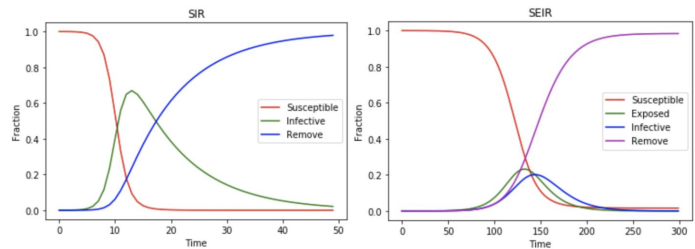
With fixed probability of infection, recovery and rate at which an exposed person becomes infectious, the fraction of population of removed status quickly converge to 1 within 30 days period which indicates our probabilities sets is too high and the result is not ideal, after adding the exposed stage which is SEIR, the fraction of population of the removed status converge to 1 in 180 days which seems to fit our observation for COVID-19 better. Since the COVID-19 has an incubation period, it is more appropriate and reasonable to use the seir model.
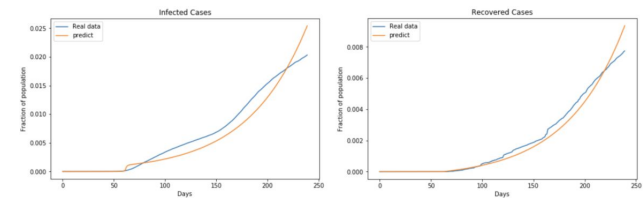
After implementing the SEIR_differential function, we used odeint and optimize curve fit for different values for fraction of susceptible, fraction of exposed, fraction of infected and fraction of recovery in order to find our best estimation for the probability of infection which is 0.021, probability of recovery which is 0.0058, to fit our real data from US. Overall our SEIR estimation fits the infected case and recovered and shows exponential growth. However we have also proved that the probabilities are not fixed and change overtime due to medical improvements

which will cause the probability of recovery increases and other factors such as quarantine policy which will cause the number of contacts to decrease.

3.2 Data exploration
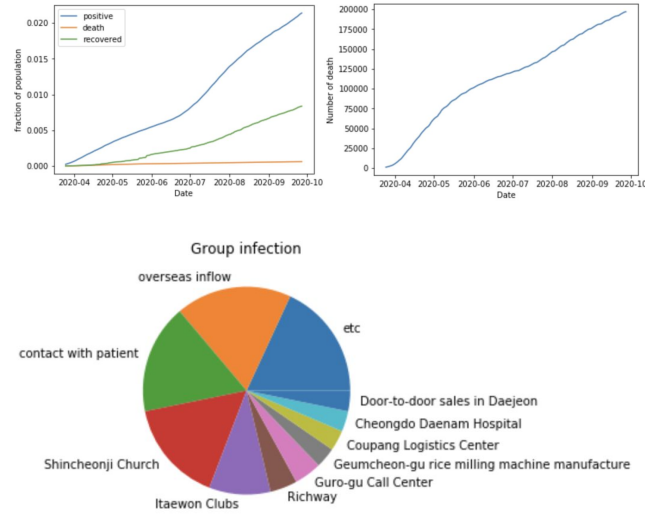
For the first two plots which shows the fraction of positive, fraction of death population and fraction of recovered population for US, in the plot we can see that the fraction of positive population and fraction of recovered population is increasing as time goes, and in the first plot, even though the fraction of death population seems small, we plotted the number of cumulative death on the other plot, as we can see, the cumulative death increases as well, and reaches 200000 on 2020-10.

According to the Korean dataset, 71% of infection cases are group infection, where most group infections are overseas inflow, contact with patients and large gatherings which in Korean data, the large gatherings are church, clubs, business centers and hospitals. There are 5213 confirmed for gathering at Shincheonji Church, 1304 confirmed for contact with patients and 949 confirmed for overseas inflow. Where we have theories about different communities such as residential communities where the flow of people is small, people in residential communities typically will not have contact with each other, but friend and family members will walk around in these communities, the types of activities residential areas can have are random walks or small gatherings between friend or family. Large crowded communities, such as business centers, grocery stores. people's walking trajectory is not fixed, but there will be crowds of people in some areas/stores. The crowd eventually gathers to a certain area such as the cashier counter in stores. Isolated communities, such as hospitals, cruise ships and churches which have fixed prayer. In hospitals Each patient has its own small room, not only there are patients with COVID-19, there are also patients with other kinds of illness these people are included in the susceptible people in the procedure, and probably would have higher infection probabilities. We can certainly observe that those large crowded communities and isolated communities have more confirmed cases than residential communities.

Lastly, we looked at the policies of the government, from 2020-02-04, Korea has implemented the policy of special immigration procedure for travelers from China, and policy of emergency use authorization of diagnostic kit and on 2020-03-19, Korea has implemented the policy of special immigration procedure for travelers from all the countries. On 2020-04-01, Korea has implemented the policy of mandatory 14-day self-quarantine from all the countries. Korea also implemented policy of drive-though screening center by local government, mask distribution, social distancing campaign, school closure, electronic wristbands, local government administrative orders, close bars and clubs on difference timestamp, we would like to explore influence for policy in our model as well such as self-quarantine.

3.3 Centreality

We computed harmonic centrality, eigenvector centrality, betweenness centrality and degree centrality to observe the impact of highest centrality nodes and lowest centrality nodes in our network simulation using Caltech 36 from facebook100.

harmonic	eigenvector	betweenness	degree
Highest 1:709	Highest 1:709	Highest 1:709	Highest 1:709
Highest 2:90	Highest 2:90	Highest 2:223	Highest 2:90
Highest 3:223	Highest 3:623	Highest 3:90	Highest 3:223
Lowest 1:437	Lowest 1:437	Lowest 1:751	Lowest 1:713
Lowest 2:169	Lowest 2:169	Lowest 2:744	Lowest 2: 536
Lowest 3:147	Lowest 3:147	Lowest 3:743	Lowest 3:533

3.4 SIR Model

We started to simulate our SIR Model with a fixed number of nodes. We got exactly a similar figure with the one created from the math. In addition, we explore the difference between the model of immediately isolated people who are infected and the model of doing nothing. In addition, we started to simulate our SIR Model using real world data. Besides experiment between the model of immediate isolation and the model of doing nothing, we also explore the SIR model with isolation of specific days delay. Last but not least, we explore how centrality affects the simulation.

### 3.4.1 Simulation with Fixed Number of Nodes

The graph is generated by 1000 nodes with random edges assigned from 1 to 9. The results of the SIR Model without isolation and the SIR Model with immediate isolation are listed below.
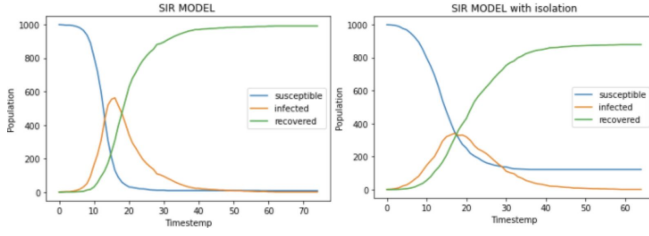
Compared with the model without isolation (left), the SIR model with immediate isolation (right) has less total number of infected people, lower peak level of infection, and slower increase in the number of daily infected cases. From above, we can get that immediate isolation is a valid way to control the spreading of COVID-19.

3.4.2 Simulation Using Real World Data

In addition, we simulated the SIR Model on a network with real data[4]. The results of the SIR Model without isolation and the SIR model with immediate isolation are listed below. The blue line represents the number of susceptible people. The orange line represents the number of infected people. The green line represents the number of recovered people.
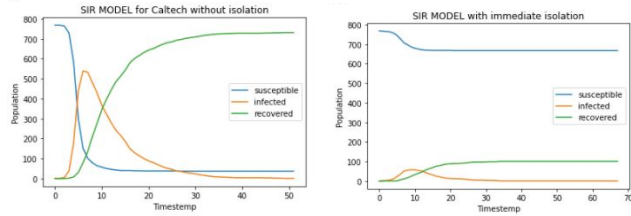

The left figure shows the result of the model without isolation. The right figure shows the results of the model with immediate isolation. Obviously, if the infected people isolate themselves immediately once they get infected, the pandemic can be controlled. It is an ideal mode. But in reality, people could not know they are infected when they are infected.

We simulated one day delay model and two day delay model. The results are listed below.
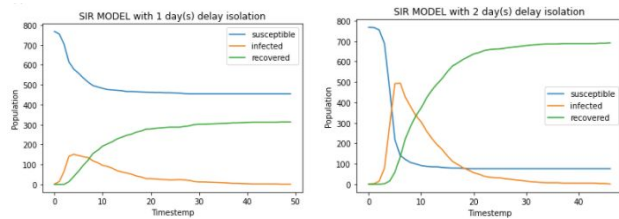
For the one day delay model (left) the pandemic can still be controlled, but the result of the two day delay model (right) is the same as the result of the model without isolation. From above, it seems to us that the less time between getting infected and isolation, the better the pandemic is to be controlled.

3.4.3 Centreality analysis
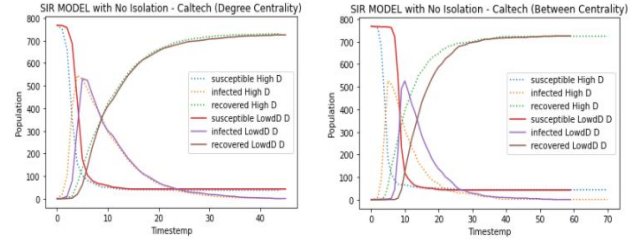
We start our model with a random node. Therefore, we want to check whether centreality affects the spreading. Next, we tested our model by two simulations for each centrality. One simulation is started with the node with highest centreality. And the other one is started with the node with lowest centrality in the giant component. The dotted line represents the result of the node with high centrality. The full line represents the result of the low centrality. The two simulations for each centreality have the same total number of infected people, and peak level of infection. The difference is the time that reaches to the peak, starts a large infection, and starts to recover. We listed the result of the degree centreality (left) and between centreality (right).

3.5 SEIR Model

Based on the knowledge we have about COVID-19, the SEIR model is closer to the characteristics of spreading. As a result, we did a lot of simulations in the SEIR model.

By plotting out the trend of susceptible (S), exposed (E), infected (I) and recovered (R). We want to see how different parameters affect the spreading trend. Specifically, we focused on the isolation stage (E or I), isolation length (7 or 14 days), isolation delay (2-5 days), and probability of exposure during isolations.

Here are a couple hypotheses we have:

● Isolation starts in I is a good start point, but since there are 6 incubation periods, it might be too late.
● Isolation start in E should be able to control the spreading trend earlier and better
● Longer the delay for isolation, worse the spreading will get. If the delay is long enough (>5 days), there is no difference than Isolation start in I
● If we have a low probability of exposure during isolation, the spreading trend will be under control even faster.
● Start with high centrality node will accelerate the sp

3.5.1 Simulation Using Fixed Number of Nodes

We did 100 and 1000 nodes random network simulation of the SEIR model using networkX [10]. For both fixed numbers of nodes, the trend is about the same. The only difference is 1000 nodes provided a smoother line plot for all susceptible (S), exposed (E), infected (I) and recovered (R) numbers during the 45 day duration.
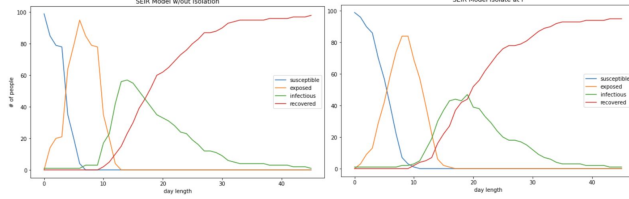
The plot above shows the trend of no isolation (left) and isolation at I (right). We can see that isolate at I is not good enough to help with the pandemic. We can see that the exposure line (yellow) and infected line (green) are less left skewed, which means that the isolation during the I stage did slow down the spreading at the beginning. However, because of the 6 days of not doing anything, the majority of the people will get infected at the end of 45 days.
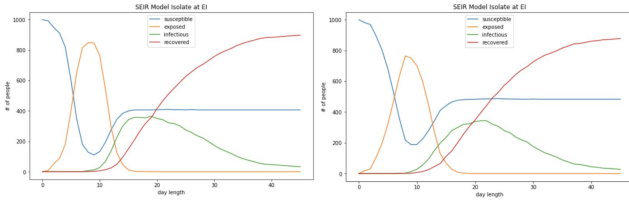

Those 2 plots showed the trend if isolation took place in the whole E and I period. The left one is isolation duration 7 days and the right one is isolation duration for 14 days. We can see that even though the isolation probability for exposed people is only 50%, it is significantly different from the ones that don't have isolation or isolation started at I only. Compare 7 days isolation and 14 days isolation, we can see that it will slow down the exposure and infection trend even more and there will be more people remain susceptible (blue) at the end.

### 3.5.2 Simulation Using Real World Data

For the real world simulation, we used *Caltech36.txt* from Facebook100 [4]. This dataset contains about 770 nodes. Before we run the simulation, starting with a random node in the real world data from Facebook100 [4], we want to see whether starting with a high centrality node will affect the spreading trend, so we did a centrality analysis using the real world network data.
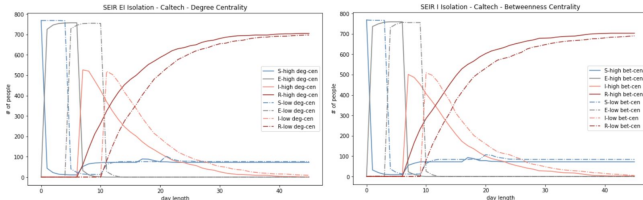
3.5.2.1 Centrality Analysis

All centrality analysis used the SEIR model that isolated both stage E and I. So for below analysis, the only changing parameter is the start node, which is the first node we picked categorized as "infected" to start the spreading.
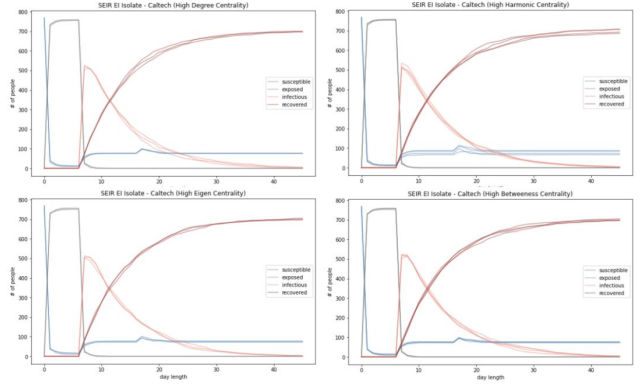
The 4 plots above showed how the spreading will go for all S, E, I and R. We can see that if the start node for the simulation is any high degree, harmonic, eigenvector and betweenness centrality, the spreading trend remains the same that there will be a large number of people exposed really fast at the beginning, followed by the peak of the infection curve.
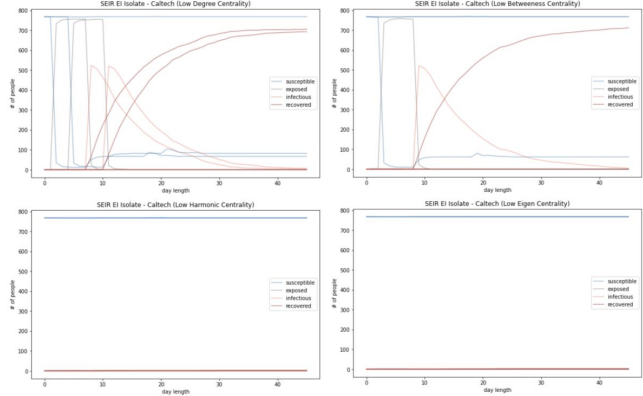
As for the 4 plots of low centrality start node plots above (degree: left-up; betweenness: right-up; harmonic: left-lower eigenvector: right-lower), we can see that it is different from high centrality start nodes. We can see that some of the low centrality nodes are still able to start the spreading similar to high centrality nodes, but some low centrality nodes are too isolated that the spread will not even start. Those extreme cases might happen when a node doesn't have any node connected or it is only connected to 1 or 2 other nodes in the whole network, but they are totally separated from the biggest component.


Because of the analysis above, we compared high-degree centrality (left) and high-betweenness centrality (right). In the plots, solid lines are the spreading start with a high centrality node and the dashed lines are the ones started with a low centrality node We can see that if we started at a low centrality node, it

doesn't change the general trend of spreading but delayed the pandemic for a couple days.

3.5.2.2 SEIR Simulation

Based on the centrality analysis for the network, we know that the spreading trend of COVID19 is not super highly correlated with picking a high or low centrality node to start with. It is true that some extreme cases that pick a node that is totally separate with the rest network will not start the spreading trend, but in general, our simulation using real work network data can just randomly pick 1 node to start with.
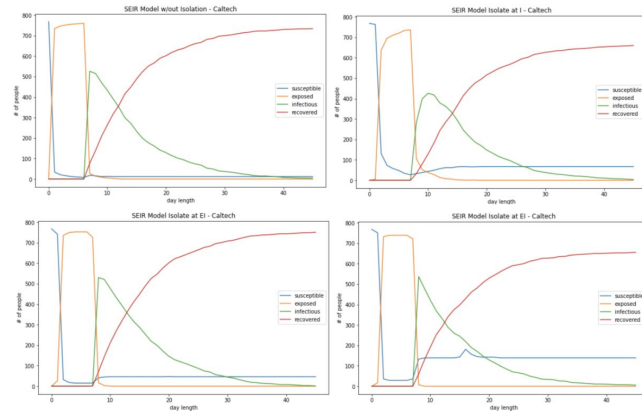
The 4 plots above are the simulations used the probability mentioned in 3.4. However we can see that even we did simulation of no isolation (left upper), isolation at I (right upper), isolation at E and I with 2 day delay (left lower), and isolation at e and I without delay (right lower). The result doesn't have a big change compared to the 1000-node model in 4.4.1. The reason why this is happening is because the average degree for facebook network is really high, which is about 50. So the susceptible people decrease super fast on the first day for no isolation models, and decrease really fast for the beginning 2-3 days for the ones with isolations.

To solve the problem mentioned above, to adapt to the high average degree problem, we lower the probability to get exposed to a quarter of the original (0.6) and remain the rest the same. And then we did the same 4 simulations again, which are the same simulations that isolation with no isolation (left upper), with isolation at I (right upper), with isolation at E and I with 2 day delay (left lower), and with isolation at e and I without delay (right lower).
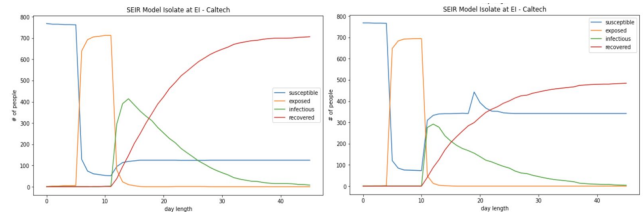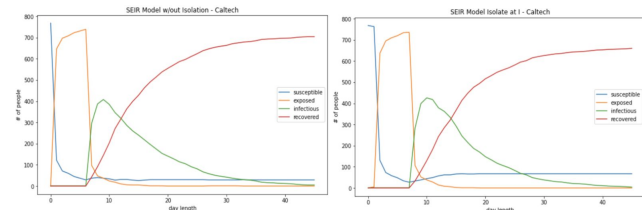
The plots shown above are the simulation with modified probability of getting exposed to mediate the problem that the average degree for the network *Caltech36.txt* is too high [4]. We can see that the simulation with isolation at both state E and I are really different from the ones before we modify the probability of exposure. We can see that if isolation 50% at E with 2 day delay and 100% at I with no delay, we will have about ⅙ of people will remain susceptible till the end. It's even better if there is no isolation delay at E. With only a 50% isolation rate at E, we will have just about half of the people remain susceptible till the end. For the right lower graph (isolation at EI without delay), the spike on the susceptible line (blue) is when after 14 days of isolation, we will still have a group of people not getting any symptoms even if they got exposed.

3.5.3   Exposure Probability Analysis

Like mentioned above, isolation in the SEIR model is mimicking the isolation rooms that every exposed people will be put in a room with maximum 4 points of contact, which is why 为p put in a grid network is able to mimic the isolation situation. If this type of isolation are also with What is also important is the prob With a fixed probabilities at each stage, fixed number of isolations days and default 2 day delay. X-axis is days, y-axis is probability of exposition during isolation (p_e_isolation), and z-axis is number of peoples for corresponding stages (S, E, I and R) over time.
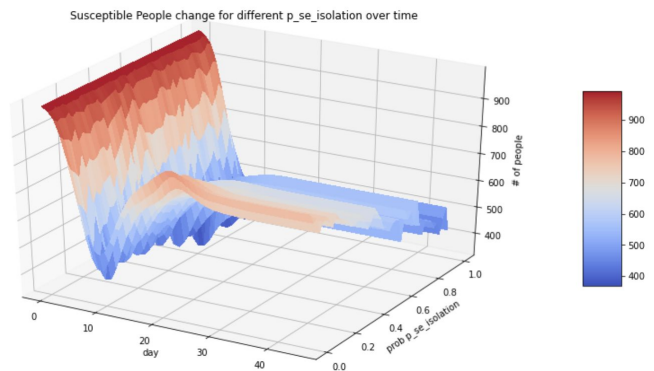
Susceptible with Different p_e_isolation

For susceptible plots, we can see that lower the p_e_isolatoin, more people will remain susceptible at the end of 45 days. What helps is that we can see that at about day 14 (end of isolation), the susceptible lines for low p_e_isolation bounced back. Lower the p_e_isolation, higher the increase. It makes sense because if the probability of getting exposed is nearly 0 during isolation, the ones that got exposed but not in the incubation period will not be exposed again.
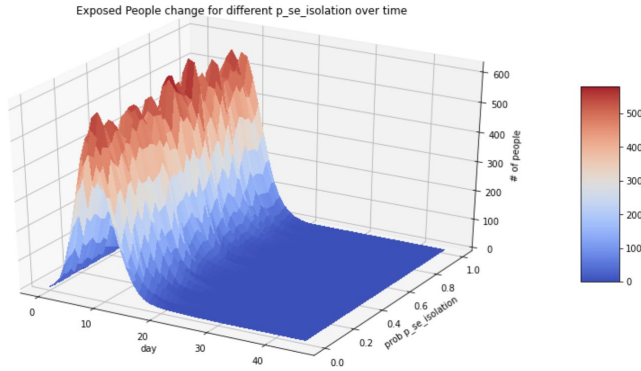
Exposure with Different p_e_isolation

Because p_e_isolation will not affect the first round of exposure before isolation happens, so the Exposure plot above is not affected by the probability of exposure.
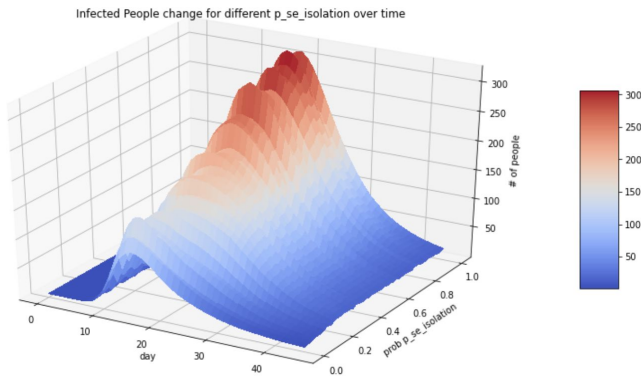
Infected Trend with different p_e_isolation

The trend of infected numbers for different p_e_isolation is very obvious that with a low probability of exposure during isolation period, the infected number over time will be decreased validly that the peak of the infectious number of people is only about one fourth to one third of the ones with a higher exposure rate. Infected trend plot correspond to the susceptible plot that over the 45 days duration, a low exposure rate during isolation will be able to make the spreading trend under control.
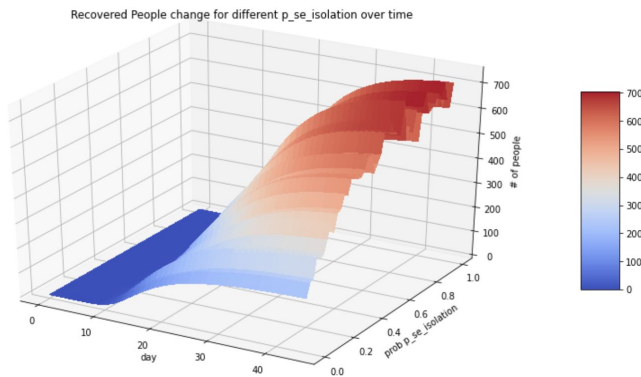
Recovery Trend with different p_e_isolation

When exposure probability is low, the recovery number is also low. It is because when the probability is low, there is not as many people will get infected compared to a high probability, so less people will even reach the recovery stage.

4. DISCUSSIONS

4.1 Math Model

4.1.1 Future predictions

Since the control intensity of each stage is different, it can only better fit the initial stage of the epidemic, and can provide certain theoretical support for making future epidemic intervention decisions. Even though we optimized our parameter using curve fit and scipy functions that getting the curve fits the current data we have. However, we would imagine that the model will inevitably be different from reality, and there are lots more factors that we need to consider such as asymptomatic population and different testing method, and even for different countries, since the policy is different and the date applying policies are different, then our model can lead to a certain deviation in the analysis and prediction results. The limitations of SEIR are quite obvious, such as not considering changes in external dynamics.

4.2 SIR Model

Although the SIR model is not the one closest to the COVID-19 spreading. However, we can still find some results from the SIR model. From the SIR model, we can conclude that immediate isolation is a valid way to control the spreading of COVID-19. The reducing time between getting infected and isolation has a significant contribution to controlling the spreading. The less time, the better the pandemic is to be controlled. For our SIR model, we assumed that infected people stayed at home for isolation. Except staying at home, we can use masks to protect yourself away from the COVID-19 as well.

4.2.1 Spreading Process

We considered that infected people can spread the COVID-19 before they know they get infected. In reality, people need to spend some time to find out whether they get infected or not. Therefore, we simulated delayed isolation models on different days.

The spreading of the SIR model is far away from reality. We only consider the spread between friends. But in reality, spread can emerge not only between the people who know each other, but also between strangers. For example, people who shop in the supermarkets don't really know each other. For the SIR model, we do not consider the spreading between strangers.

In addition, we assumed that everyone stays at home if he/she gets infected. Staying at home is just a suggestion. This suggestion does not have a powerful constraint. Parts of infected people can break the rule without any punishment. The SIR model does not consider the probability of people breaking the rule. But the SEIR

model is considered that. From this point, it reflects that the SEIR model is more reality than the SIR model.

4.2.2   Isolation

In the SIR model, the isolation means staying at home. In the US, if people have light symptoms, doctors only suggest the patients stay at home. Only people who have heavy symptoms need to stay in the hospital. From this point, we assumed that people stay at home when they get infected.

We assumed that people could not leave home if someone who is living in the same building with them got infected. Not only infected people can not leave their house but also his neighborhood can not leave their house. In reality, people can go out no matter his/her neighborhood getting infected or not. The reason why is that the message of which person gets infected is private. His/her neighborhood can not get to know which person gets infected.

4.3   SEIR Model

Among all 3 models, the SEIR model is the one closest to the COVID-19 spread in reality because of the "E - Exposure" stage. As we know that once a person gets exposed to COVID-19, she/he might be in the incubation period, which means they don't show any symptoms but have the ability to spread the disease to other people in contact. However, our simulation models are still more on the theory end because we have an upper limit of population in the network, the spreading process is hard to be concluded using a couple fixed probability, homonic type of spreading events, and 1 type of isolation method for all the exposed nodes with a default number of delays. Fortunately, using the models we have, we are able to provide useful recommendations. Also by adding in more parameters, uncertainties by randomizing, we should be able to make our model get closer to reality.

4.3.1   Spreading Process

So far all the simulations we had are still more on the theoretical level. Though our SEIR model is the best fit model for the spread of COVID-19, it is still far away from reality. First we have a fixed number of people, that is why we were able to show a full circle of spreading. In reality, many countries right now are facing the second round of cases increasing, and people are able to get re-infected, which are not included in the simulation we've done.

In addition, simulation with superspreading and normal spreading will make our modal closer to reality. Averaging out the exposure rate is able to simulate the exposure, infectious stages well, but when we reach an isolation period, isolating the source node of superspreading events will be helpful for the pandemic control.

4.3.2   Isolation

Using a grid network might be one of the most ideal methods for isolation. Beyond that a 2 day delay before put in isolation is also more ideal than reality.

Why grid isolation is probably one of the best isolation we can do in reality. First, we assumed that a person wouldn't be quarantining themselves at home because in our simulation, once people are exposed, we take them out from the original network with 2 days delay and put them in the grid network. Second, we assume that the maximum number of people they will be in contact with in isolated quarantine is 4. Lastly, we set the isolation duration to 14 days. In some countries, 14 days is also a recommendation that the actual length of quarantine is likely to be shorter than 2 weeks.

Delay length for our model is set to 2, we can see that when we switch from 2-day delay to 0, there was a significant decrease in infected populations over time. However, 2-day delay is still an ideal number we can have in reality because sometimes, it is very hard to know whether you got exposed or not. Thus, isolating 50% of exposed people is also beyond possible for the majority of the time.

5.   CONCLUSION

After all simulations, the SEIR model is the one closer to how COVID-19 is spreading. Built up from the math model and the SIR model, the SEIR model was able to take a lot of factors into account like incubation period, isolation probability, isolation delay and isolation duration. We were able to confirm that all the hypotheses from 3.5 are true that starting at stage E 50% is better than isolate at stage I. 14 days isolation works better than 7 days that more people remain susceptible till at the end of 45 days. We also showed that beside some extreme cases where start nodes have no edge, centrality doesn't affect the general spreading trend, but only delays the beginning of a large amount of spreading or pandemic.

Even though our simulation are still too far away from the reality, we are able to provide recommendations based on the results we get above:

- Isolation after confirming infected is needed, but too late to slow down the spreading
- Isolation starting at exposure is very important, once a person knows she/he has been exposed to the virus, do self isolation as soon as possible. The SIR model shows the same results. The time between getting infected and isolation is significant. We need to reduce that time. But it is hard for operations. Therefore, isolation starting at exposure is necessary.
- 14 days isolation after knowing get exposed or self quarantine are able to slow down the spreading and put pandemic under better control
- Social distancing is an important policy which reduces the number of contact per unit time
- Continue wearing masks which helps reduce the probability of infection.
- Strengthen personal precautions and hygiene, don't attend large gathering events.

FUTURE WORK

In the future, when better datasets are available, we will be able to check whether our model fits close to the real world COVID-19 data and be able to observe if temperature has any influence to

COVID-19.. Above that, we should also propose a way of changing probabilities as days increase in our simulation since new policies are established over time.

Superspreading events will also be a good parameter to add in especially considering isolation is able to slow down the spreading. If we are able to put in the superspreading parameter, our model will be closer to reality. In addition, in both the spreading process and isolation process, there are a lot of uncertainties that can be tilted by randomizing. Based on analysis of the dataset we might have access in the future, we can add in more parameters in the isolation period for the SEIR model.

We should also consider asymptomatic population, in SEIR setting, the difference between asymptomatic and exposed is that asymptomatic population may be mistaken for being healthy and not taking any coercive measures, including testing and no quarantine, in this situation, the asymptomatic's transmission capacity is greater than the exposed population.

In the mathematical simulation of SIR and SEIR, by status R, we include both recovered and deceased population, even though through the visualization for the US dataset, the fraction of deceased population seems to be really small, but we should add another parameter of probability of fatality and deceased population.

REFERENCES

[1] Althouse, Benjamin M., et al. "Superspreading Events in the Transmission Dynamics of SARS-CoV-2: Opportunities for Interventions and Control." *PLOS Biology*, Public Library of Science, Retrieved November 12, 2020, journals.plos.org/plosbiology/article?id=10.1371%2Fjournal.pbio.3000897.

[2] COVID-19 in USA, Number of Novel Corona Virus 2019 cases in USA. Retrieved October 17, 2020 https://www.kaggle.com/sudalairajkumar/covid19-in-usa

[3] Data Science for COVID-19 (DS4C). DS4C: Data Science for COVID-19 in South Korea. Kaggle. Retrieved September 29, 2020, https://www.kaggle.com/kimjihoo/coronavirusdataset

[4] Facebook100. Index of Complex Networks. Retrieved September 10, 2020, https://icon.colorado.edu/#!/networks

[5] Fanelli, D., Piazza, F., (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. Retrieved October 08, 2020, from https://doi.org/10.1016/j.chaos.2020.109761

[6] Firth, J., Hellewell, J., Klepac, P., Kissler, S., Kucharski, A., & Spurgin, L. (2020, August 07). Using a real-world network to model localized COVID-19 control strategies. Retrieved October 09, 2020, from https://www.nature.com/articles/s41591-020-1036-8

[7] Y. -C. Chen, P. -E. Lu, C. -S. Chang and T. -H. Liu, "A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons," in IEEE Transactions on Network Science and Engineering, doi: 10.1109/TNSE.2020.3024723. Retrieved October 05, 2020, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9200529

[8] Wang, P., Lu, J. A., Jin, Y., Zhu, M., Wang, L., & Chen, S. (2020). Statistical and network analysis of 1212 COVID-19 patients in Henan, China. International Journal of Infectious Diseases. Retrieved October 06, 2020, from https://reader.elsevier.com/reader/sd/pii/S1201971220302678?token=EBF672F7C4EAA5E46EF47EC5784BFBB9ED1AF3B8CD70A539C66D3BBC85518021FDA092FA497250DF8D4B4189E6C5B27C

[9] Parnell, D., Widdop, P., Bond, A., & Wilson, R. (2020). COVID-19, networks and sport. Managing Sport and Leisure, 1-7. Retrieved October 09, 2020, from https://www.tandfonline.com/doi/full/10.1080/23750472.2020.1750100?casa_token=6nP7uRnnM_kAAAAA%3Abc1323hddDAdbMUnOQZVCd0TT6ex9vmydGTQqvY-C7aFHINVIVrGfFHoop_Mmbza9YCKdkzCSA8n#aHR0cHM6Ly93d3cudGFuZGZvbmxpbmUuY29tL2RvaS9wZGYvMTAuMTA4MC8yMzc1MDQ3Mi4yMDIwLjE3NTAxMDA/bmVlZEFjY2Vzcz10cnVlQEBAMA==

[10] Software for Complex Networks¶. (2020, August 22). Retrieved December 11, 2020, from https://networkx.org/documentation/stable/index.html

[11] World Health Organisation. (2020).Who, Coronavirus disease (COVID-19) outbreak.(2020). Retrieved October 07, 2020, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[12] "Global COVID-19 Tracker & Interactive Charts: Real Time Updates & Digestable Information for Everyone." *1Point3Acres*, coronavirus.1point3acres.com/.

APPENDIX

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. All the work has been done by individual group members.