

# Network graph based SEIR\_H model

Yiming Ren, Chengkai Jiang and Qihong Liu  
 Department of Electrical and Computer Engineering,  
 University of California San Diego  
 9500 Gilman Dr., La Jolla, CA 92093  
 {y5ren, cljiang, qil011}@ucsd.edu  
 A{59005465, 59005010, 59004984}

**Abstract**—Our project explored three aspects in modelling the Covid-19 pandemic. Firstly, we designed a network SEIR model that receives a customized graph from scratch. It also simulates the hospitalization situation with exploiting the individuality of network model. Individuality means the model’s ability of dealing with every single individual. The result is then compared with the Mean-field SEIR method. Secondly, we explored the effects of specifying mortality by age. The specification of age-related attributions improves the performance of our model, making it closer to reality. Lastly, we attempted to reproduce the real Covid-19 data by using a more sophisticated model.

## I. INTRODUCTION

### A. Motivation

The graph based SEIR model and the Mean-field SEIR model are distinct. In a Mean-field SEIR model, we consider every individual as a part of cluster, in which everyone is equally possible to interact with the other and is totally same as the others, regarding any attribute like their age or immunity. However, in order to pursue a more realistic simulation of *Covid19*, it is important to also consider the heterogeneity and stochasticity. In this way, the network random graph comes to be a suitable tool to generate a representation of stochasticity of a population. In the paper, we only pay attention to the heterogeneity of a network SEIR model. As the individuality of contact network, we could pair each node to an individual, and their edges to the connections between individuals. We could also assign each individual a bunch of specified attributes to underline heterogeneity. With those functions, we could distinguish individuals by their ages or remove their connections to their neighbors if they are tested positive (hospitalized), and hence make the simulation more pragmatic.

### B. Literature review

The SEIR model is widely used to characterize the epidemic dynamics and to predict possible contagion

scenarios. Since it is based on a series of nonlinear ordinary differential equations and the influenza virus is transmitted through person to person contact, integrating a social network with it can help us understand the spread of the epidemic more precisely. In [1], it has been shown that the compartment-based models are subject to overestimating the scale of the spread. They proposed a graph-based model that can precisely represent how different non-pharmaceutical intervention (NPI) strategies affect the pandemic. Fatima et al. adopted the Small World (SW) model introduced by Watts and Strogatz, which has small diameter and a high clustering coefficient. These two properties lead to a close approximation of many real world networks. It was demonstrated that this SEIR-SW model can provide a useful early prediction of the epidemic peak.

### C. Report outline

The second section discusses the theories and parameters involved in SEIR models (A-D). Then the method of specifying mortality by age is introduced (E). Finally, explanations of the choice of data set for Covid-19 are given (F). The results for all three parts of our project are illustrated and analyzed in the third section. In the last section, the conclusion is given.

## II. MODEL AND NOTATION

### A. Network SEIR\_H model

The main difference between the network SEIR model and the Mean-field SEIR model is how they calculate the population in each state. In a Mean-field SEIR model, we do calculations on the populations directly following the equations. The equations are designed to accurately predict the number of population from the population in the previous state. Regardless of individuality, we only pay attention to the numbers. By contrast, in a network SEIR model, we do not use equations to calculate the number; instead, we collect the rate of state transferring from online statistics and use the rates to calculate the

propensity of changing an individual's state. After that, we count the total number of people in each state as the population in that state.

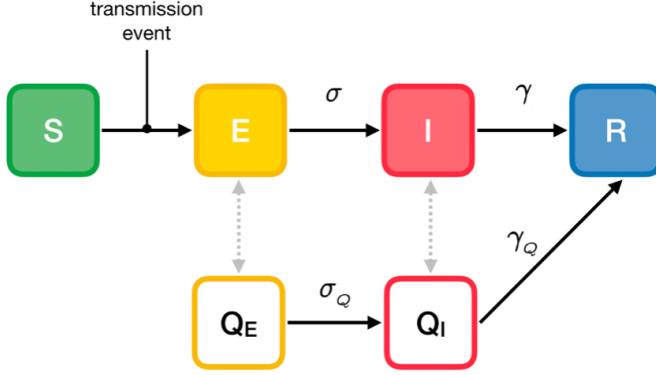


Fig. 1: flow chart of network SEIR model

Figure 1 shows the workflow of a network SEIR\_H model, where Q means the status of being hospitalized. Starting from state S, each node can be in one of the 6 states. All parameters of calculations are shown in the parameter II-A4 section. The subsections introduce the principles of determining the transmission between states.

1) *Susceptible to Exposed*: To determine the transmission from S to E, we make nodes in S contact with others in 2 ways. Each time a node makes a contact, it has probability  $p$  to contact with its neighbor or probability  $1-p$  to contact to a random individual globally. The propensity of state transferring is decided by the susceptibility  $\alpha$  of the node in S and the transmissibility  $\beta$  of the contacted node if it is in I. When the node contacting with its neighbor, the propensity of transmission is equal to the product of its susceptibility  $\alpha$  and the sum of the transmissibility  $\beta$  of its infected neighbor. This means that the propensity of one node transferring from S to E is proportional to the number of its infected neighbors. The calculation of propensity of a global contact is similar. We use the number of all infected nodes in the network to multiply the transmissibility  $\beta$  and the susceptibility  $\alpha$ . Then, we compare a random number with the propensity to determine if the transmission happens.

While calculating the propensity, if an infected node is hospitalized, we make its transmissibility  $\beta$  smaller ( $\beta_Q$ ), to represent a weaker infectiousness.

2) *Exposed to Infectious, Infectious to Recovered, and Infectious to Death*: The rate of progression to infectiousness is represented by the parameter  $\sigma$ . The propensity of transferring from E to I is equal to the

$\sigma$ . If a node is in state E, we calculate its progress of being infected to determine if it is infected at this time.

Similarly, the rate of recovery is the same as  $\gamma$ , which tells one's speed of recovering from the disease. Once the node is recovered, we considered it being permanently immunized.

When determining death cases, we compare a random number to the possibility of death  $f$ , considering that unhospitalized infected people are not able to receive enough medical treatment, and therefore their death are unpredictable. If an individual is considered deceased, the node of it will be permanently removed from the network.

3) *Test and Hospitalization*: People in E or I will receive tests. When building the model, we assume that all tests are absolutely accurate and all nodes in E or I will eventually be tested positive if they keep in these two states. The rates of testing people in E or I are equal to the parameters  $\theta_E$  and  $\theta_I$ . We increment the progress of test by  $\theta$  each time and move the node to its corresponding hospitalized states once the test is finished. While the node is in hospitalized states, its transmissibility is smaller ( $\beta_Q$ ), its rate of progression to infectiousness  $\sigma_Q$  is faster, and its rate of recovery  $\gamma_Q$  is faster.

Unlike the infected nodes, we make the hospitalized nodes well hospitalized. With enough medical treatment, their rate of recovery  $\gamma_Q$  could increase largely, and their death would not be unpredictable because the medical diagnosis can tell the patient's condition. In this way, we use the rate of death for hospitalized infected individuals  $\mu_I$  to calculate a node's progress to death.

If both progresses of 2 transmission reach the destination, for example, the progresses of testing and infect for a node in E both reach 1, we compare the 2 numbers and select the larger one to decide the transmission.

Since the parameters are given by the user, the user can simulate various society phenomena under pandemic. For example, to simulate the areas where the medical techniques and infrastructures are advanced, the user could slightly turn up the rate of test and rate of hospitalized recovery. Or for areas that is sparsely populated, the user could reduce the edges between the nodes to represent a lower connectivity. Moreover, if the user sets the rate of testing larger than the rate of progression to infectiousness, the nodes in E will more likely to go to hospitalized states and get hospitalized on time before infected.

4) *Parameters*:

**G** The contact network random graph  
 **$\beta$**  Transmissibility of infected individuals  
 **$\sigma$**  rate of progression to infectiousness

**gamma** rate of recovery  
**alpha** susceptibility  
**mu\_I** rate of death for hospitalized infected individuals  
**f** probability of death for infected individuals  
**p** possibility of local contact  
**beta\_Q** Transmissibility of hospitalized infected individuals  
**sigma\_Q** rate of progression to infectiousness for hospitalized individuals  
**gamma\_Q** rate of recovery for hospitalized individuals  
**initI** initial infected population  
**theta\_E** rate of testing for exposed individuals  
**theta\_I** rate of testing for infected individuals

### B. Specified mortality by age

Since we treat individuals as nodes in a network graph, we are able to assign attributes to each individual as nodes. To make the model more pragmatic, we assign each node an age, and determine its own fatality according to its age. CDC presents the relative mortality rate for each age group corresponding to a specific age group in United States on February 18, 2021 [2]. It selects the age group of 5-17 as reference, and calculates the ratios of mortality rate of other age groups compared to the reference. Considering that the difference of mortality rates between group 0-4 and group 5-17 is trivial, we combine the two groups to make the statistic simpler. Then we change the reference group to the age group of 0-17. The hospitalization and death rate of other age groups relative to the rates of reference group are shown in the presentation. For example, the hospitalization rate of 50-64 is 25 times larger than the reference group, and the death rate of 50-64 is 440 times larger than the reference group. We use the relative ratios to calculate the mortality rate for all other groups.

*American Academy of Pediatrics (AAP)* provides the report of children status during pandemic in United States [3]. In the report, 49 states, NYC, DC, PR, and GU have reported age distribution of cases. Among them 43 states, NYC, PR, and GU have reported age distribution of mortality. Although states have different criteria of defining children, e.g. Florida reported children being 0-14 years old, while California reported children being 0-17 years old, we could simply summarize the children as between 0-17 years old. Because the children age ranges of different states are varied between 14 20 years old, the mean value 17 is a suitable choice to balance the reported cases. In the summary of child mortality data from November 5, 2020 to June 3, 2021, the percent of child cases resulting in death keeps being 0.01 %. Therefore, we choose 0.0001 as the mortality rate of the reference group.

We choose the official demographics report of United States government on April 1, 2020 [4] as the sample of age distribution. We calculate the percentage of population in each age group based on the data. The division of age groups is following the division of age groups in CDC mortality report [2]: 0-17, 18-29, 30-39, 40-49, 50-64, 65-74, 75-84, and 85+. The percentages of population for age groups are 25%, 17.5%, 15.2%, 14.7%, 15.6%, 6.5%, 4.4%, and 1.1% respectively. We then divide the range of [0,1) by the percentages, where each part represents an age group for all 8 parts.

When determining the age of each node, we use *random()* function to generate a random number between [0, 1). We find the part that the random number belongs to, and find its corresponding age group. Figure 2 shows the age distribution of 10k nodes generated from random numbers. In this way, we are able to assign the node a mortality rate by its age. We write the age and mortality rate as attribute to the node, which will be used in further calculation of death population.

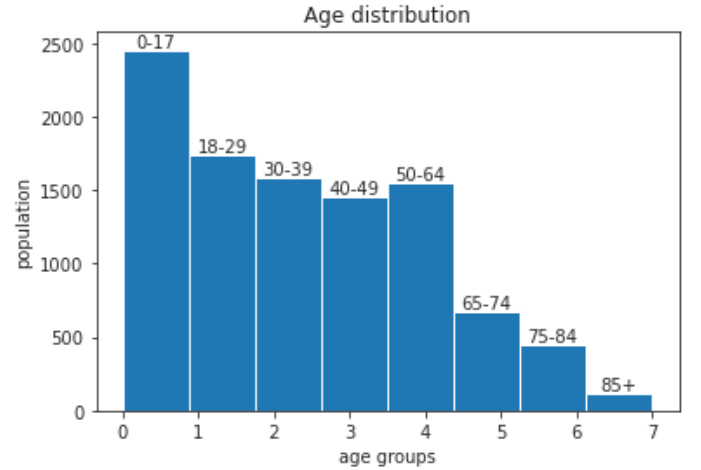


Fig. 2: Age distribution of 10k nodes

### C. Reproduce real Covid-19 data

First of all, we chose the number of nodes in the network to be 10,000, because the real-world population is large and networks with more than 10,000 nodes take significantly longer time to run. We chose the Covid-19 data from March 15th to May 15th in Austria, provided by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [5]. This choice is due to the fact that Austria has a relatively small population while the number of cases is a significant portion of it. It allows us to simulate our model with 10,000 nodes. For example, the number of cases in March 15th is roughly 11,300 in Austria. Since its population is about 8,859,000 and the number of nodes

of our network is 10,000, after mapping into our network, the number of cases can be calculated as

$$11300/(8859000/10000) \approx 13$$

The start date is chosen to be March 15th as before March there is nearly any cases. The end date is chosen to be 60 days after so that we are attempting to reproduce the Covid-19 data in the two-month period of rapid increasing. Another reason for choosing Austria is that during that time, with a travel restriction in Europe, there is less migration movement.

The model we used is a sophisticated model with realistic network properties as well as age-stratification, households, and communities that are calibrated to demographic statistics for a population of interest [6]. Another design is that when individuals enter hospitalize states, a separate graph can be used to define their changed parameters and interactions. In this separated graph, the quarantined nodes drop some of their edges to represent a status of being isolated. In addition, it has considered the effects of testing, tracing and isolation interventions on disease transmission.

### III. RESULTS AND ANALYSIS

#### A. Graph based SEIR

Because we only pursue functionality instead of reality when testing our Network SEIR\_H model, we didn't use an official dataset. In fact, the parameters the user passes in are hard to be collected from the official data of Covid19. For example, it is hard to estimate the possibility of death of an unhospitalized but infected person. To pursue reality, we have to take some mathematical or statistical method to get accurate parameters, but that is not in our field. Therefore, we only use a reasonable dataset to test our model, and in the following sections, we will try different datasets on a well implemented model to try to make a perfect simulation of Covid19 pandemic.

Considering some facts of Covid19, we make the rate of tests *theta* faster than the rate of progression to infectiousness *sigma*. We make the inverse of the rate of recovery *gamma* to be proportional to the transmissibility *beta*, meaning that weaker people tends to have a higher transmissibility. We need a certain amount of infected people to spread the virus, so we set the initial infected population *initI* to be 1/100 of initial population. We use a low possibility of local contact *p* to create a highly migrated crowd, and we set the possibility of death to 0.04, which is similar to the death rate of Covid19. We test our model for 100 days on 10k nodes.

```
S= 9460
E= 0
I= 0.0
R= 523
F= 14
Q_E= 3
Q_I= 0
N= 9986
```

Fig. 3: Population distribution on the last day of the network SEIR\_H model

Figure 3 shows the number of people in each state after 100 days, where N means the number of remained nodes in the graph. Since 14 nodes are labeled as deceased, the last number of nodes is 9986. Since we set the rate of tests *theta* as 0.5, which is far larger than the rate of progression to infectiousness *sigma*, almost all exposed individuals are expected to be hospitalized in 2 days before they are infected, and therefore we get no new infected cases during the simulation. Also because the rate of recovery for hospitalized individuals are larger than the rate of death for hospitalized infected individuals, once the individual is hospitalized, the individual is expected to always recover in the future. In this way, the death cases only occur in the first 2 days, and all infected cases are eliminated in the first 2 days. To create a more complicated scenario, the user should adjust the parameters according to the cases.

```
S = 9853.707202838881
E = -2.451324854968405e-07
I = 1.762449463723724e-07
Q_E = 5.388539253436661e-07
Q_I = 0.00845961849077571
R = 143.4160167378982
F = 2.8683203347579647
```

Fig. 4: Population distribution on the last day of the Mean-field SEIR model

Compared to the result of the network SEIR\_H model, the result of the Mean-field SEIR model is more restrained. It makes more population maintaining the susceptibility and less recovered cases. This is due to the difference of definition of hospitalization. In the Mean-field SEIR model we used for comparison, the hospitalization also brings isolation to the hospitalized one, which is in turn totally disconnected to the others. In other words, it is removed from the cluster until it is recovered. Whereas in our network SEIR\_H model,

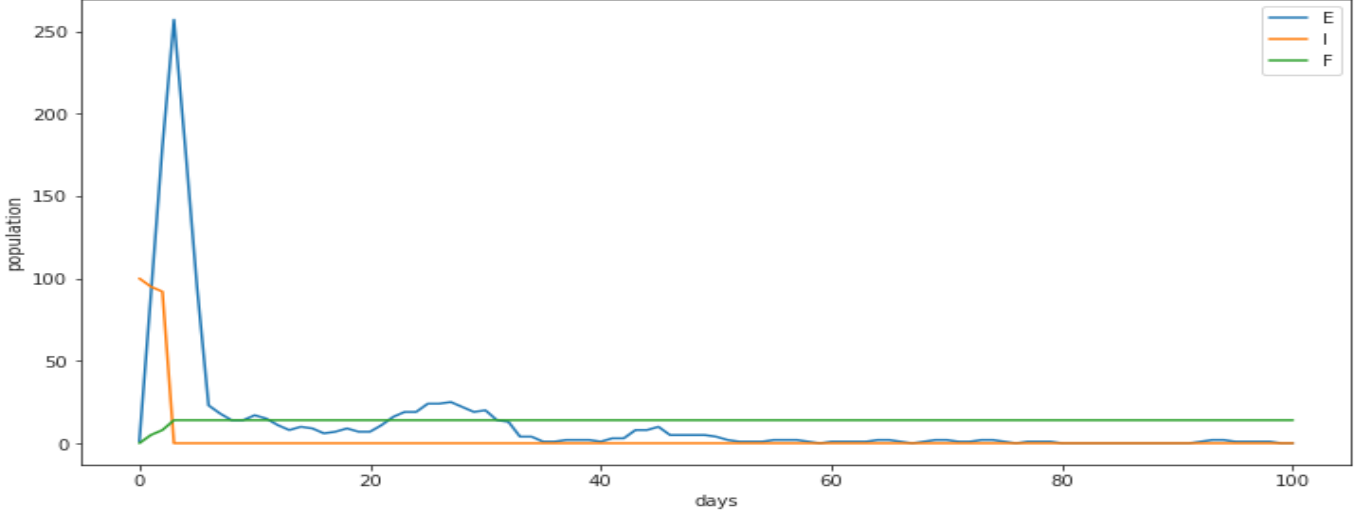


Fig. 5: Plot of E,I,F population for network SEIR\_H model. Blue for E, orange for I, and green for F

we only reduce the transmissibility of the hospitalized individuals, so we could expect more exposed cases. Furthermore, the Mean-field SEIR model is more complicated on calculating the propensity and progress. The parameters passed to the model are not simply the rates of progress; rather, the model applies complex equations on the calculations. Since the Mean-field SEIR model do not pay attention to the individuals, it ignores the impact of adjacency on the transmissibility. However, in the real world, plenty of reported Covid19 cases are from indoor or neighboring contact. The intensity of neighboring contact is not reflected in the Mean-field

SEIR model, and thus further reduce the number of reported cases.

As discussed above, in our network SEIR\_H model, the number of infection and death cases would not change after the second day. This is clarified in the Figure 5. The trend of change of exposed cases is also printed. Normally, the curve of the number of reported cases during a pandemic is similar to a normal distribution, which corresponds to the first several days of simulation. In the following days, although most of people are hospitalized, there are still a few newly exposed cases to smooth the curve. By contrast to Figure

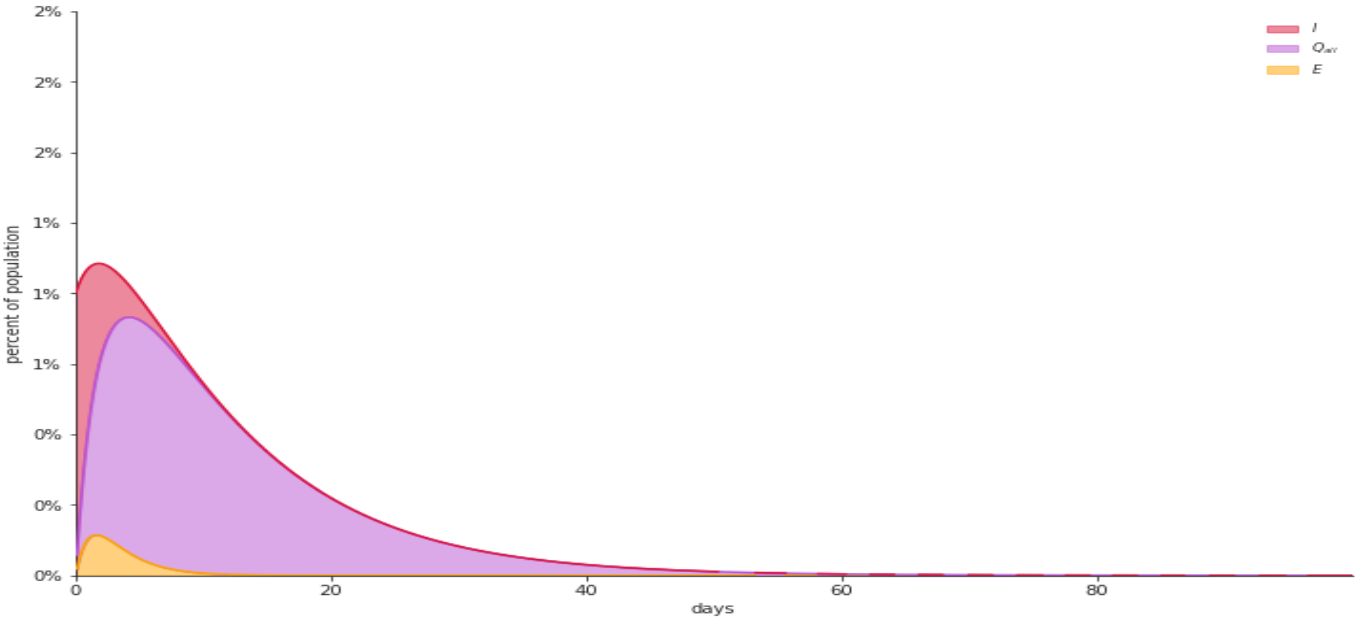


Fig. 6: Plot of E,Q,I population for Mean-field SEIR model. Yellow for E, red for I, and purple for Q



6, our model fails to simulate a real Covid19 curve, but again, a successful simulation requires complex mathematical and statistical method to collect data from the real world. The parameters should not be set only from guess and speculation, because our model does not use any equations to calculate the population.

### B. Age Specification

For age specification experiment, by assigning different mortality rate to various age range, we finally obtain the exposed population curve and infected population curve within the simulation of the SEIR model as Figure 7 and 8.

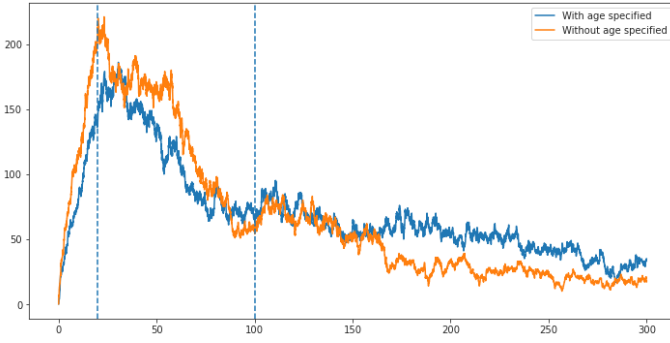


Fig. 7: Exposed Population with Age Specification

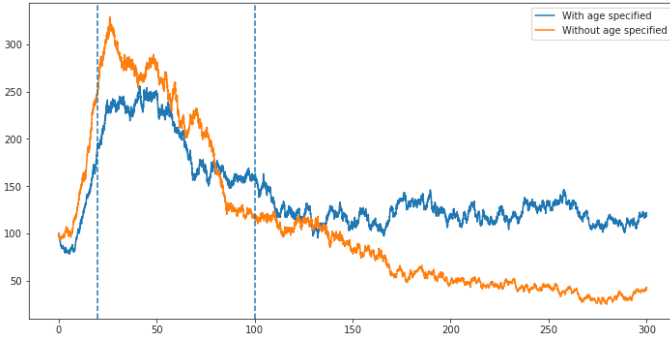


Fig. 8: Infected Population with Age Specification

As it can be seen that with specifying the age and the related attributes like mortality, there is an obvious curve change both in exposed population, what is, the percentage of exposed and infected population without age specification are higher at the beginning, but decrease rapidly after social-distance policy. This can be explained that without age-specification mortality, those who only occupy a smaller percentage of the total population but has larger mortality rate, like the elder and young children, should influenced the overall infection little. However, if we do not specify the age but merely average the mortality rate from all age range, such influence will

be enlarged, thus making the simulation and projection untrustworthy.

Meanwhile, if we see the daily case report from CDC [7] in Figure 9, it is obvious that at the beginning of the epidemic, from 1st March to 31st May, which is same as the time we simulated, the daily case curve is much more similar to our age-specified SEIR model, with an rapid increasing before and a smooth and slow decrease rather than the rapid decrease as the non-age-specified situation. The similarity between the groundtruth and our simulation result strengthen the correctness of our adding age-specification to the model.

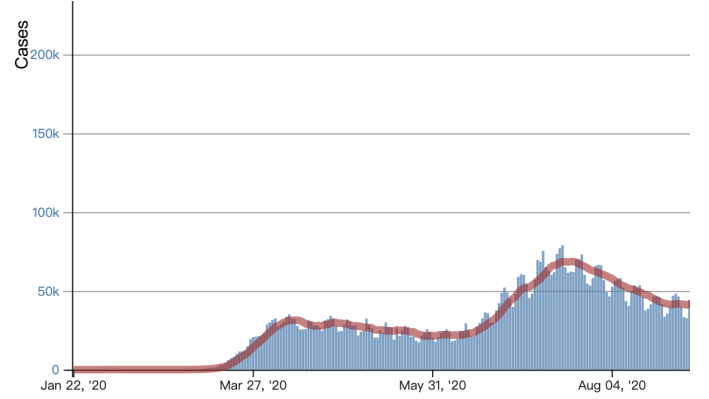


Fig. 9: Daily Infected Population from CDC

It cannot be ignored that the social distance starting at around day 20 plays an important role in decreasing the total exposed and infected population, both in simulation and reality. With such social distance, people from all age range would avoid some infections and contacts, thus making a reduce trend in following. The social distance policy has effected the epidemic regardless of the age-specification.

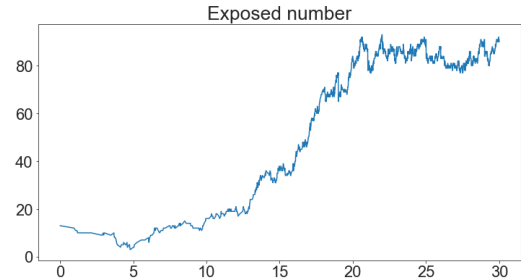


Fig. 10: Exposed number from March 15th to April 15th

### C. Covid-19 Simulation

We ran simulation for 30 days and compared the number of exposed with the ground truth cases. As

shown in Fig 11, on April 15th, the ground truth number of cases is 77,116 which is roughly 87 after mapped into our 10,000 nodes network. This matches the result obtained in Fig 10.

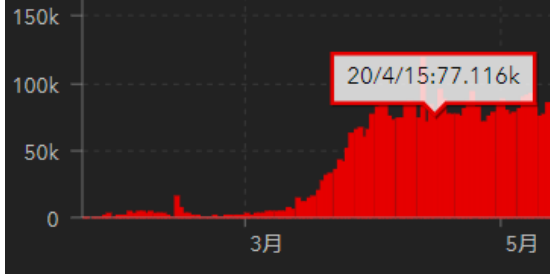


Fig. 11: Ground Truth: Cases from March 15th to April 15th

However, when we ran it for 60 days, there is a significant difference. As shown in Fig 13, on May 15th, the ground truth number of cases is 95,718 which is roughly 108 after mapped into our 10,000 nodes network. This is much larger than the number indicated in Fig 12, which is roughly 70 on the sixtieth day.

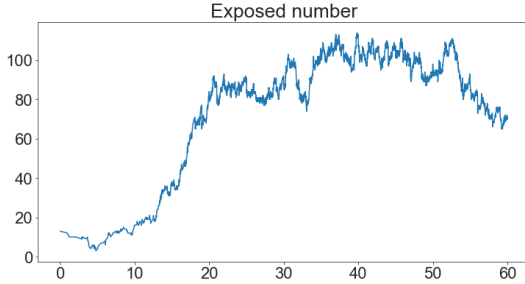


Fig. 12: Exposed number from March 15th to May 15th

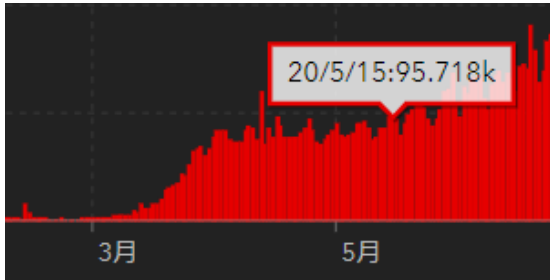


Fig. 13: Ground Truth: Cases from March 15th to May 15th

There are several problems with this simulation. Firstly, the nodes at the exposed state denotes people that are infected but not infectious, which means that they are only an uncertain portion of the infected people. Secondly, the ground truth cases are not equal to the

number of infected people as some of them might not be tested. Based on the results obtained, the difference between them could be caused by the fact that the hospitalization/distancing effects in the model overrides the real-world situation of hospitalization.

#### IV. CONCLUSION

In the paper, we elucidate the point of building a SEIR model with contact network random graph, and use its individuality to apply heterogeneity to each individual. We add a hospitalization function to reflect the human interference on the trend of pandemic. We test our model with a dataset from estimation and compare the result to a Mean-field SEIR model. Although the result is far from the normal because of exaggerated human force, we see the point of using individuality of contact network SEIR method to improve the basic SEIR model.

Because we are far from building a mature SEIR model, we choose to explore a well implemented SEIRS network model and try any improvement. First, we divide the populations to different age groups, assign different fatality rate to the groups, and compare the result between homogeneous fatality rate and heterogeneous fatality rates. This work also reveals the advantages of individuality of network SEIR model.

Finally we try to reproduce Covid19 data approaching to the reality by utilizing various functionalities of the model. We use the real dataset collected from official reports and compare them to the simulation result. We speculate the reason causing the difference and come up with potential improvements on the model. We can add a process of vaccine research and production, and adjust susceptibility according to the vaccination rate. We can also quarantine the healthy people to control the spread of virus. Another point is reflecting the virus mutation based on the infected population by either turn up the transmissibility or fatality or else. The features of the mutation are from the real world cases.

#### REFERENCES

- [1] A. Leung, X. Ding, S. Huang, and R. Rabbany, "Contact graph epidemic modelling of COVID-19 for transmission and intervention strategies," *CoRR*, vol. abs/2010.03081, 2020. arXiv: 2010.03081. [Online]. Available: <https://arxiv.org/abs/2010.03081>.
- [2] N. C. for Immunization and R. Diseases, *Risk for covid-19 infection, hospitalization, and death by age group*, <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html>, Accessed June 8, 2021.

- [3] A. A. of Pediatrics, *Children and covid-19: State-level data report*, <https://services.aap.org/en/pages/2019-novel-coronavirus-covid-19-infections/children-and-covid-19-state-level-data-report/>, Accessed June 3, 2021.
- [4] Wikipedia contributors, *Demographics of the united states — Wikipedia, the free encyclopedia*, [https://en.wikipedia.org/w/index.php?title=Demographics\\_of\\_the\\_United\\_States&oldid=1027005549](https://en.wikipedia.org/w/index.php?title=Demographics_of_the_United_States&oldid=1027005549), [Online; accessed 9-June-2021], 2021.
- [5] E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track covid-19 in real time,” *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533–534, 2020.
- [6] R. McGee, *Seirs+ model framework*, <https://github.com/ryansmcgee/seirsplus>, 2020.
- [7] Centers for Disease Control and Prevention, *Daily trends in number of covid-19 cases in the united states reported to cdc*, [https://covid.cdc.gov/covid-data-tracker/#trends\\_dailytrendscases](https://covid.cdc.gov/covid-data-tracker/#trends_dailytrendscases), [Online; accessed 9-June-2021], 2021.