

Spread of Covid-19: Estimation and Simulation

Table of Contents

- 1 Introduction
- 2 Literature Review
- 3 Deterministic Model for Disease Spreading
 - 3.1 SIR model
 - 3.2 SEIR model
- 4 Disease Spreading on Network
 - 4.1 Considering Transmissibility for Outbreak Analysis
 - 4.2 Scale-free Network
- 5 Experiments
 - 5.1 Deterministic Model based on Real Data
 - 5.2 Network Model Estimation
 - 5.3 Scale-free Network Simulation
- 6 Conclusion
- 7 Reference

1 Introduction

Covid-19 virus has been spreading all over the world for several months. To figure out the pattern behind the disease spreading can not only help us predicting the future trend, but also help others reducing the probability from getting disease. The fundamental reason why the disease can be spreaded is because people has some connections. If we think of the social network as a graph composed of vertices and edges, the disease spreading would be the events happening in this graph.

According to the case in real world, the probability of knowing each person in a society between one to the others can be hugely different. A random graph would not be the case to mimic a social network. Instead, some vertices should have much greater connectivity than the others, where a scale-free graph can be used to well express the social network. If we calculate the distribution of the degree of each vertex, we may find that the degree counted in a scale-free graph follows power-law distribution. Therefore, the following study will mainly focus on power-law distribution and the repated estimations.

2 Literature Review

The mathematics modeling of epidemic spreading has largely been studied. [M. Choisy,2007] comprehensively review the dynamics model of infectious diseases. The classic model of epidemic spreading is Susceptible-Infectious-Removed (SIR) model, which is based on the law of mass action [Hethcote,2000]. The three most common compartment that exist in SIR epidemic model are called 'Susceptible', 'Infected' and 'Removed' [W. O. Kermack, 1927], [Hethcote,2000]. , in which the 'Susceptible' (S) capturing individuals who are susceptible to becoming infected The 'Infected' (I) representing individuals who are infected, but able to recover, and 'Removed' (R) compartment representing individuals who are no longer susceptible or unfortunately pass away. [Hethcote & Yorke 1984; Ghani et al. 1997; Keeling 1997] construct more complex structure within the population, mostly classifying the population into several subgroups and mixing different rates between S, I and R. Specifically, one of the variations of the SIR model that include 'Exposed' (E) compartment, representing individuals infected by the disease but cannot yet transmit it, is called SEIR model.

The SEIR model is one of the paradigmatic models for the spreading of influenza-like illnesses. [Small and Tse ,2005] study the evolution of the Severe Acute Respiratory Syndrome (SARS) by deterministic and stochastic versions of SEIR model with different social settings. Two of the classic variations of SEIR are SEIRS [XIN JQ, WANG WJ , 2009] and SEIS [Krone, 1999]. The former simulates the eventual waning of the immunization of recovered individuals and the latter considers no immunization after recovery.

Network structure assign to each individual a finite set of enduring contact, simulating the possible transmit infection between people in different form of interaction, such as social interaction and transportation. Networks thus capture the epidemiological detail of disease spreading. [Matt J. Keeling 2005] reviews the epidemiological theory and network theory, suggesting how the combine of two fields can improve the understanding of epidemic evolution, in which the network forms a contact background to the dynamic transmission. [M.E.J. Newman, 2001] deeply depict the theory of random graphs with arbitrary degree distributions, and demonstrate that the highly accurate prediction can be obtained with appropriate distributions of vertex degree. [Granell, 2013], simulate the epidemic by a network of persistent real contacts and use Microscopic Markov Chain Approach (MMCA) to reveal the evolution of the epidemic over period of time.

In the real case of epidemic spreading, however, the super-spreaders are extremely important in disease dynamic transmission, considering super-spreaders into networks is essential if we aim to capture the complexities of spread. Scale free network, a network that the degree of relative commonality of vertices far exceeding the average, is widely be studied. [Albert-László Barabási, 1999] started the interest in scale-free model, finding that the network with several highest-degree nodes (often called 'hubs') characterize a power-law distribution of the number of links connecting to a node. In SARS epidemic, [Riley, 2003] reveal super-spreaders are the significant reason of all infectious cases, and this result considerably reinforced a standard public-health guideline – targeted vaccination, revealing the analysis in network do facilitate the epidemic policy.

During the outbreak of COVID-19, proven as a group of SARS-like coronaviruses [Hu, 2018] raised intense attention internationally. Researchers concerned the feasibility of controlling the outbreak by the execution of physical distancing and isolation of cases [Hellewell J 2020]. Similarly, [S Lin, 2020] depict the relation between the intensity of inner city and possibly spreading of the epidemic based on of transmission mechanism.

3 Deterministic Model for Disease Spreading

Typical deterministic model for disease spreading includes SIR mode and SEIR model.

3.1 SIR model

The general SIR model takes s as the ratio of susceptible people, i as the infected people, r as the recovered people. Then we have:

$$\begin{aligned}\frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \\ \frac{dr}{dt} &= \gamma i\end{aligned}$$

Here, β is the rate of transmission (transmissions per S-I contact per time), γ is the rate of recovery (inverse of infectious period). We could solve these equation and get:

$$s = s_0 e^{-\frac{\beta}{\gamma} r}$$

With $s + i + r = 1$, we could drive:

$$\begin{aligned}\frac{dr}{dt} &= \gamma(1 - r - s_0 e^{-\frac{\beta}{\gamma} r}) \\ t &= \frac{1}{\gamma} \int_0^r \frac{du}{1 - u - s_0 e^{-\frac{\beta}{\gamma} u}}\end{aligned}$$

With the above equations, we could fit the lines of the rate of susceptible, infectious and recovered people and get the parameter estimation of γ and β .

Also, we could drive their iteration form with $S = sN$, $I = iN$, $R = rN$ and let c be the number of people one contact:

$$\begin{aligned}S_n &= S_{n-1} - \frac{c\beta I_{n-1} S_{n-1}}{N} \\ I_n &= I_{n-1} + \frac{c\beta I_{n-1} S_{n-1}}{N} - \gamma I_{n-1} \\ R_n &= R_{n-1} + \gamma I_{n-1}\end{aligned}$$

3.2 SEIR model

The general SEIR model takes s as the ratio of susceptible people, e as the exposed people, i as the infected people, r as the recovered people. Then we have:

$$\begin{aligned}\frac{ds}{dt} &= -\beta si \\ \frac{de}{dt} &= \beta si - \alpha e \\ \frac{di}{dt} &= \alpha e - \gamma i \\ \frac{dr}{dt} &= \gamma i\end{aligned}$$

Here, β is the rate of transmission (transmissions per S-I contact per time), γ is the rate of recovery (inverse of infectious period), α is the rate of progression (inverse of incubation period). We could solve these equation and get:

$$s = s_0 e^{-\frac{\beta}{\gamma} r}$$

With $s + e + i + r = 1$, we could drive:

$$\begin{aligned}\frac{dr}{dt} &= \gamma(1 - e - r - s_0 e^{-\frac{\beta}{\gamma} r}) \\ t &= \frac{1}{\gamma} \int_0^r \frac{du}{1 - e - u - s_0 e^{-\frac{\beta}{\gamma} u}}\end{aligned}$$

With the above equations, we could fit the lines of the rate of susceptible, infectious and recovered people and get the parameter estimation of γ and β .

Also, we could drive their iteration form with

$S = sN, E = eN, I = iN, R = rN$ and let c be the number of people one contact:

$$\begin{aligned}S_n &= S_{n-1} - \frac{c\beta I_{n-1} S_{n-1}}{N} \\ E_n &= E_{n-1} + c\beta I_{n-1} S_{n-1} N - \alpha E_{n-1} \\ I_n &= I_{n-1} + \alpha E_{n-1} - \gamma I_{n-1} \\ R_n &= R_{n-1} + \gamma I_{n-1}\end{aligned}$$

4 Disease Spreading on Network

We consider the number of connection in a network and build up a model to describe the epidemic disease spreading on network. First, we could consider a network with n nodes and random edges and k be the degree of a node which means the number of connections each person has with others. The distribution of the network could be described a degree distribution p_k , and $G_0(t)$ is the moment generating function of k .

According to the definition of MGF,

$$G_0(t) = E(e^{tk}) = \sum_{k=1}^{\infty} p_k e^{tk} = \sum_{k=1}^{\infty} p_k x^k \quad (\cdot: x = e^t)$$

According to the normalization property, $G(1) = \sum_{k=1}^{\infty} p_k = 1$

We could find p_k by getting the function of $G_0(t)$, since $p_k = \frac{d^k}{dX_k} \frac{G_0(x)}{k!} \big|_{x=0}$

We could the r -th moments of p_k by

$$z_r = E[k^r] = \sum_{k=1}^{\infty} k^r p_k = \left[\left(x \frac{d}{dx} \right)^r G_0(x) \right] \big|_{x=1}, \text{ for example, the first moment } z = G'_0(1).$$

For $Y = \sum_{i=1}^m k_i$, $k_i \sim p - k$, the MGF of sum of degree of m independent nodes is $G_Y(x) = [G_0(x)]^m$

If node c has k_c edges, it is k_c times as likely connect to A as B is, thus $p'_k \propto kp_k$, $\sum p'_k = c \sum kp_k = 1$, so $p'_k = \frac{kp_k}{\sum_j p_j} = \frac{kp_k}{z}$. The MGF of the original and first neighbors is $G'_1(x) = \sum p'_k x^k = \sum \frac{kp_k}{\sum_j p_j} x^k = x \frac{G'_0(x)}{G'_0(1)}$.

$$\begin{aligned} \text{For excess degree, } q_k &= \frac{(k+1)p_{k+1}}{\sum_j p_j} \\ &= \frac{(k+1)p_{k+1}}{z} \\ G_1(x) &= \sum_{k=1}^{\infty} q_k x^k = \sum_{k=1}^{\infty} \frac{(k+1)p_{k+1} x^k}{z} \\ &= \sum_{k'=0}^{\infty} \frac{k' p_{k'} x^{k'-1}}{z} \\ &= \frac{G'_0(x)}{G'_0(1)} \end{aligned}$$

The MGF for the second neighbor is

$G^{(2)} = \sum p_k G^{(2)}(x|k) = \sum p_k [G_1(x)]^k = G_0(G_1(x))$. Similarly, the number of m -th nearest neighbor is $G^{(m)}(x) = G^{(m-1)}(G_1(x))$ for $m > 2$. Thus, the average number of n -th neighbor is $z_n = G^{(m)'}(x) \big|_{x=1} = G'_0(1) * (G'_1(1))^{(m-1)}$.

4.1 Considering Transmissibility for Outbreak Analysis

Let r_{ij} be the rate of disease-causing contacts between i and j . τ_i be the time duration that i remains infective. T_{ij} be the probability of transmission which is unknown but should be a function of r_{ij} and τ_i .

Assume $r_{ij} \sim P_r(r)$, $\tau_i \sim P_\tau(\tau)$ and they are independent from each other. Within ξ_t time interval, i does not infect j with probability $1 - r_{ij}\xi_t$. so in all $\frac{T_i}{\xi_t}$ time interval, the probability of Not getting the disease is $(1 - r_{ij}\xi_t)^{\frac{T_i}{\xi_t}}$.

$$\begin{aligned} P[i \text{ does not infect } j] &= 1 - T_{ij} \\ &= \lim_{\xi_t \rightarrow 0} (1 - r_{ij}\xi_t)^{\frac{T_i}{\xi_t}} \\ &= e^{-r_{ij}\tau_i} \\ \therefore T_{ij} &= 1 - e^{-r_{ij}\tau_i} \end{aligned}$$

$$\begin{aligned} \text{The average transmissibility is } T &= E[T_{ij}] \\ &= 1 - E[e^{-r_{ij}\tau_i}] \\ &= 1 - \int_0^\infty \int_0^\infty e^{-rt} P_r(r) P_\tau(\tau) dr d\tau \end{aligned}$$

Then we consider the MGF as a function of transmissibility in which a node with k links, each is accessibly transmitted(occupied) with probability T .

$$\begin{aligned} P_m^0 &= P[m \text{ of } k \text{ links occupied}] = C_m^k T^m (1 - T)^{k-m} \\ G_0(x; T) &= E[x^m] = E[E[x^m | k]] = \sum_{k=1}^\infty E[x^m | k] p_k \\ &= \sum_{k=1}^\infty \sum_{m=0}^k x^m P_m^k P_k \\ &= \sum_{k=1}^\infty \sum_{m=0}^k C_m^k T^m (1 - T)^{k-m} x^m P_k \\ &= \sum_{k=1}^\infty \sum_{m=0}^k C_m^k (xT)^m (1 - T)^{k-m} P_k \\ &= \sum_{k=1}^\infty (xT + 1 - T)^k P_k \\ &= G_0(1 + (x - 1)T) \\ \text{And } G_0(x; 1) &= G_0(x) \\ G_0(1; T) &= G_0(1) \\ G'_0(1; T) &= T G'_0(1) \\ \text{Similarly, } G_1(x; T) &= G_1(1 + (x - 1)T) \\ G_1(x; 1) &= G_1(x) \\ G_1(1; T) &= G_1(1) \\ G'_1(1; T) &= T G'_1(1) \end{aligned}$$

Let $P_s(T)$ be the distribution of outbreak size s , $H_0(x; T)$ is the corresponding MGF. By definition, $H_1(x; T)$ is the MGF for the outbreak size reached by following a random edge.

$$H_0(x; T) = \sum_{k=1}^{\infty} P_s(T) x^k$$

$$H_1(x; T) = E[x^{s(T)}] = E[E[x^{s(T)} | k]] = \sum_{k=1}^{\infty} E[x^{ks(T)+1} | k] q_k$$

$$= x \sum_{k=1}^{\infty} H_0(x; T)^k q_k$$

$$= x G_1(H_1(x; T); T)$$

$$\text{Similarly } H_0(x; T) = x G_0(H_1(x; T); T)$$

$$\text{The probability of outbreak size is } P_s(T) = \frac{1}{s!} \frac{d^s H_0}{d^s s} \Big|_{x=0}$$

$$= \frac{1}{2\pi i} \int \frac{H_0(\xi; T)}{\xi^{s+1}} d\xi$$

where the integral is over a unit circle

$$H'_1(1; T) = (x G_1(H_1(x; T); T))'$$

$$= G_1(H_1(x; T); T) + x G_1(H_1(x; T); T)' H'_1(x; T)$$

$$= 1 + G'_1(1; T) H'_1(1; T)$$

$$\therefore H'_1(1; T) = \frac{1}{1 - G'_1(1; T)}$$

$$\text{The average outbreak size is } E[s] = H'_0(1; T) = (x G_0(H_1(x; T); T))'$$

$$= G_0(H_1(x; T); T) + x G_0(H_1(x; T); T)' H'_1(x; T)$$

$$= 1 + G'_0(1; T) H'_1(1; T)$$

$$= 1 + \frac{G'_0(1; T)}{1 - G'_1(1; T)}$$

$$= 1 + \frac{T G'_0(1)}{1 - T G'_1(1)}$$

$$\text{The outbreak threshold is when } T G'_1(1) = 1$$

$$T_c = \frac{1}{G'_1(1)} = \frac{G'_0(1)}{G''_0(1)}$$

4.2 Scale-free Network Case

A scale-free network is a network whose degree distribution follows a power law distribution. This type of network has three important features: 1) has super nodes which have many more connections than others and this is called "preferential attachment"; 2) the number of links connecting to a node follows a power-law distribution or clustering coefficient distribution; 3) the degree distribution has a heavy tail. Since many research find that social network is a specific kind of scale-free network, we here study the spread of covid-19 in a scale-free network and conduct a experiment based on Shenzhen data later.

For the degree distribution to be power-law distribution, $p_k = Ck^{-\nu}$. For normalization property, $G(1) = \sum_{k=1}^{\infty} Ck^{-\nu} = 1$ so $C = \frac{1}{\sum_{k=1}^{\infty} k^{-\nu}}$.

$$\begin{aligned} G_0(x) &= \sum_{k=1}^{\infty} Ck^{-\nu} x^k = \frac{\sum_{k=1}^{\infty} k^{-\nu} x^k}{\sum_{k=1}^{\infty} k^{-\nu}} \\ z_1(x) &= G'_0(1) = \frac{\sum_{k=1}^{\infty} k * k^{-\nu} x^{k-1}}{\sum_{k=1}^{\infty} k^{-\nu}} \Big|_{x=1} \\ &= \frac{\sum_{k=1}^{\infty} k^{-\nu+1}}{\sum_{k=1}^{\infty} k^{-\nu}} \\ &= \frac{\xi(\nu-1)}{\xi(\nu)} \\ z_2(x) &= G''_0(1) = \frac{\sum_{k=1}^{\infty} k * (k-1) * k^{-\nu} x^{k-2}}{\sum_{k=1}^{\infty} k^{-\nu}} \Big|_{x=1} \\ &= \frac{\sum_{k=1}^{\infty} k^{-\nu+2} - k^{-\nu+1}}{\sum_{k=1}^{\infty} k^{-\nu}} \\ &= \frac{\xi(\nu-2) - \xi(\nu-1)}{\xi(\nu)} \end{aligned}$$

Assume $r_{ij} \sim P_r(r) = \frac{1}{r_m} \ 0 \leq r \leq r_m, \tau_i \sim P_\tau(\tau) = \frac{2\tau}{\tau_m^2}$

$$\begin{aligned} \text{The average transmissibility is } T &= 1 - \int_0^{\infty} \int_0^{\infty} e^{-rt} P_r(r) P_\tau(\tau) dr d\tau \\ &= 1 - \frac{2}{r_m^2 \tau_m^2} (e^{-r_m \tau_m} - 1 + r_m \tau_m) \end{aligned}$$

$$\text{In this case, } T_c = \frac{G'_0(1)}{G''_0(1)} = \frac{\xi(\nu-1)}{\xi(\nu-2) - \xi(\nu-1)}$$

$$\text{Average outbreak size } E[s] = 1 + \frac{T(\xi(\nu-1))^2}{\xi(\nu)[(T+1)\xi(\nu-1) - T\xi(\nu-2)]}$$

Here, we know the number of first neighbor and second neighbor, we would estimate the parameter in the power-law distribution by method of moment and maximizing log likelihood.

For sampling distribution of continuous power law

$$p_k = \frac{\nu-1}{k_{min}} \left(\frac{k}{k_{min}} \right)^{-\nu} \text{ where } c = \frac{\nu-1}{k_{min}^{-\nu+1}}$$

$$\text{For } m\text{-th moment } z_m = \int_{x_{min}}^{\infty} x^m \frac{\nu-1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\nu} = \frac{\nu-1}{\nu-1-m} x_{min}^m$$

$$\text{For method of moments } E[x] = \frac{\nu-1}{\nu-2} x_{min}$$

$$E[x^2] = \frac{\nu-1}{\nu-3} x_{min}^2$$

$$\text{var}(x) = \frac{\nu-1}{(\nu-3)(\nu-2)} x_{min}^2$$

$$\therefore \hat{x}_{min} = \frac{\sum x_i^2 (\nu-3)}{\sum x_i (\nu-2)}$$

$$\therefore \hat{\nu} \text{ satisfies } \frac{(\nu-2)^2}{(\nu-1)(\nu-3)} = n \frac{\sum x_i^2}{(\sum x_i)^2}$$

$$\hat{\nu} = \frac{-2(\sum x_i)^2 + 2n \sum x_i^2 \pm \sqrt{-n(\sum x_i)^2 \sum x_i^2 + n^2 (\sum x_i^2)^2}}{-(\sum x_i)^2 + n \sum x_i^2}$$

For maximizing log-likelihood estimator

With continus assumption

$$p(x) = \prod_{i=1}^n \frac{\nu-1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\nu}$$

$$\log p(x) = n \log(\nu-1) - n \log x_{min} - \nu \sum_{i=1}^{\infty} \log \frac{x_i}{x_{min}}$$

$$\frac{\partial \log p(x)}{\partial x_{min}} = \frac{-n}{x_{min}} + \frac{\nu}{x_{min}} \sum_{i=1}^{\infty} \frac{1}{x_i} = 0$$

$$\frac{\partial \log p(x)}{\partial \nu} = \frac{n}{\nu-1} - \sum_{i=1}^{\infty} \log \frac{x_i}{x_{min}}$$

$$\nu = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1}$$

With discrete assumption

$$\nu = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min} - \frac{1}{2}} \right]^{-1}$$

$$\text{Let us assume } x_{min} = 1, \hat{\nu} = 1 + \frac{n}{\ln(\sum x_i)}$$

$$\begin{aligned} \hat{T}_c &= \frac{G'_0(1)}{G''_0(1)} = \frac{\xi(\nu-1)}{\xi(\nu-2) - \xi(\nu-1)} \\ &= \frac{\hat{\nu}-3}{\hat{\nu}-2} \frac{1}{\hat{x}_{min}} \end{aligned}$$

$$E[\hat{s}] = 1 - \frac{(\hat{\nu}-1)(\hat{\nu}-3)}{\hat{\nu}-2}$$

5 Experiments

5.1 Deterministic Model based on Real Data

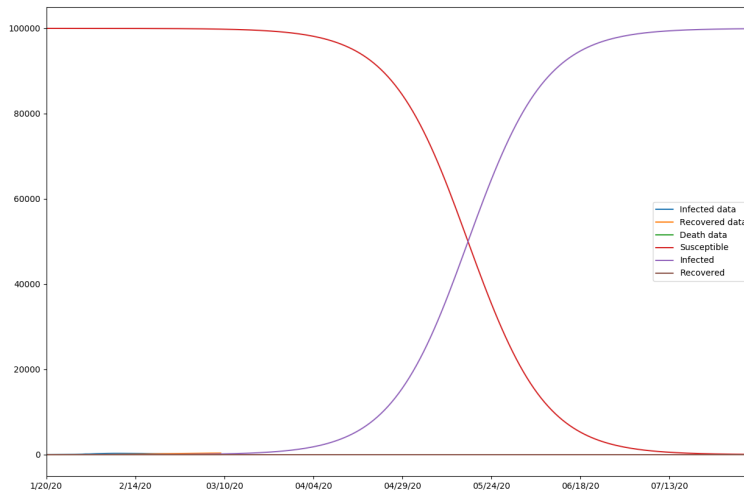
Here, we used the data from January 2nd 2020 to May 9th 2020 of Shenzhen, including the cumulative confirmed cases, cumulative cured cases and cumulative dead cases. Here, we apply SIR model and fit the data of Shenzhen to get the s the ratio of susceptible people, i as the infected people, r as the recovered people. Here, we fit the model by minimize the weighted MSE of the solution of the derivative equations for s and r . We initialized the algorithm that total number of people is 10000 and infected number of people is 2.

$$\begin{aligned} \text{minimize} \quad & \alpha_1 \left(\sum_{i=1}^n (\hat{s}_i - s_i)^2 \right) + (1 - \alpha_1) \left(\sum_{i=1}^n (\hat{r}_i - r_i)^2 \right) \\ \text{Set } & \alpha_1 = 0.1 \end{aligned}$$

The results are $\beta=0.00000091$, $\gamma=0.00000024$, $r_0: 3.81126064$

Also, we draw the graph with 150 days of prediction. The graph shows that the number of infected people would be more than the number of susceptible people at around the early of May if we didn't take any measure. the R_0 is 3.8 and is over 1 which means that covid-19 would spread to the whole group of people in this situation according to our model fitting.

Virus spread along time(SIR)



For SEIR model, since we could not get the data of exposed people, we simply assume that the exposed people is a static ratio of the confirmed cases. Here, we set it as 0.5 to get the model fit parameters.

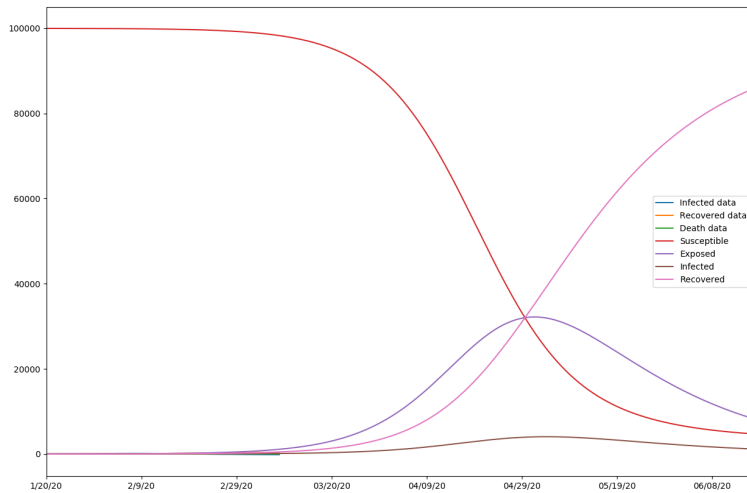
$$\text{minimize } \alpha_1 \left(\sum_{i=1}^n (\hat{s}_i - s_i)^2 \right) + \alpha_2 \left(\sum_{i=1}^n (\hat{e}_i - e_i) \right) + (1 - \alpha_1 - \alpha_2) \left(\sum_{i=1}^n (\hat{r}_i - r_i) \right)$$

Set $\alpha_1 = 0.1, \alpha_2 = 0.1$

The results are $\text{beta}=0.00001423$, $\text{alpha}=0.05080899$, $\text{gamma}=0.40000000$, $r_0: 0.00003558$.

By looking at the real line and 150 days' predicted trend of the virus spread we found that when susceptible ratio of people declines and the infected people increases the exposed people who are not infected but infective will increase to a peak and decrease there after which is very similar to the virus spread situation recently.

Virus spread along time(SEIR)



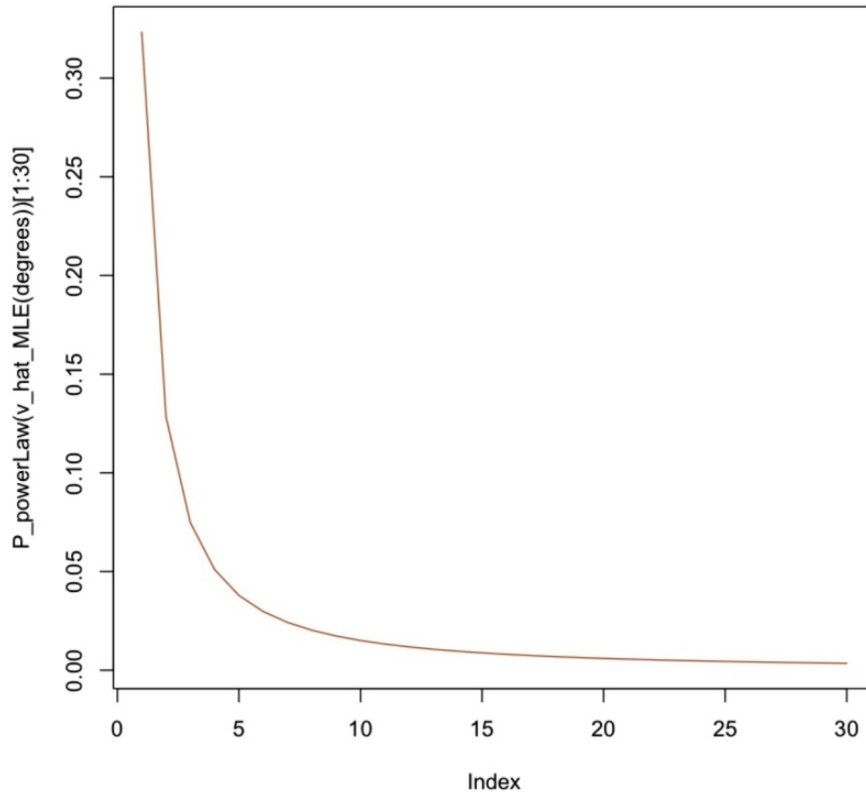
5.2 Network Model Estimation

Here, we use the data from January 2nd 2020 to May 9th 2020 of Shenzhen, including the cumulative confirmed cases, cumulative cured cases, cumulative dead cases for SIR model and SEIR model parameter estimation and we pop the number of total 1st neighbors and the number of total 2nd neighbors into the distribution and imply the simple cases of power-law distribution for network simulation

To predict the data using power-law distribution, the first step we should do is to estimate the parameter of the distribution. Maximum likelihood estimation was applied in our case and we get:

To predict the data using power - law distribution, the first step we should do is to estimaate the parameter of the distribution. Maximum likelihood estimation was applied in our case and we get:

$$P_k = C \cdot k^{-v} \rightarrow \hat{v}_{MLE} = 1.33241$$



However, the power - law distribution is not borned to be a normalized distribution. The distribution should be normalized by ourselves so that ξ function is defined, which is equal to $\frac{1}{C}$. In our case based on the estimation, we have:

$$\xi(\hat{v}_{MLE}) = 3.09270037128255$$

Neighbor Estimation

Neighbor distribution can be easily obtained by calculating moment generating function(MGF) where:

$$G^{(neighbor)} = G_0(G_1(\cdots G_1(x)\cdots)) \quad \dots \text{where} \quad G_1(x) = \frac{G'_0(x)}{G'_0(1)}$$

Mind that MGF is a helper to make the calculation process easier. The output value of MGF contains no meanings at all. By getting the above function, we can derivate the function so that the average number of n neighbors can be estimated.

$$Z_m = \frac{d}{dx} G^{(neighbor)} \Big|_{x=1} = G'_0(1) [G'_1(1)]^{m-1} \rightarrow Z_2 = 1321.40592$$

The average 2nd neighbor in real data is 965.66, which is approximately 30 % less than our estimation.

Transmissibility : T

The transmissibility can also be obtained by MGF such that:

$$T_c = \frac{G'_0(1)}{G''_0(1)} = 0.0123764$$

which means that the disease will outbreak as T goes beyond this threshold.

Average Outbreak Size : $\mathbb{E}(s)$

According to the public data, we know that $T = 0.22$. To substituting the following function, we get:

$$\mathbb{E}(s) = 1 + \frac{T \cdot G'_0(1)}{1 - T \cdot G'_1(1)} = 0.7855$$

meaning that 78 % of the whole population in the graph would be affected.

5.3 Scale-free Network Simulation

We conducted network simulation using Netlogo and constructed a scale-free network using Netlogo extension nw. The simulation is aimed to do parameter sensitivity test to explore the range of the outcome if we consider different values of the parameters. Here, we test the following parameters in the form of [minimum, step, maximum] in the model which run 500 times with 2000 agents and check the tick every 50 ticks.

PARAMETERS	RANGES
initial-outbreak-size	[1 1 5]
recovery-rate	[1 10 100]
virus-spread-rate	[1 10 100]

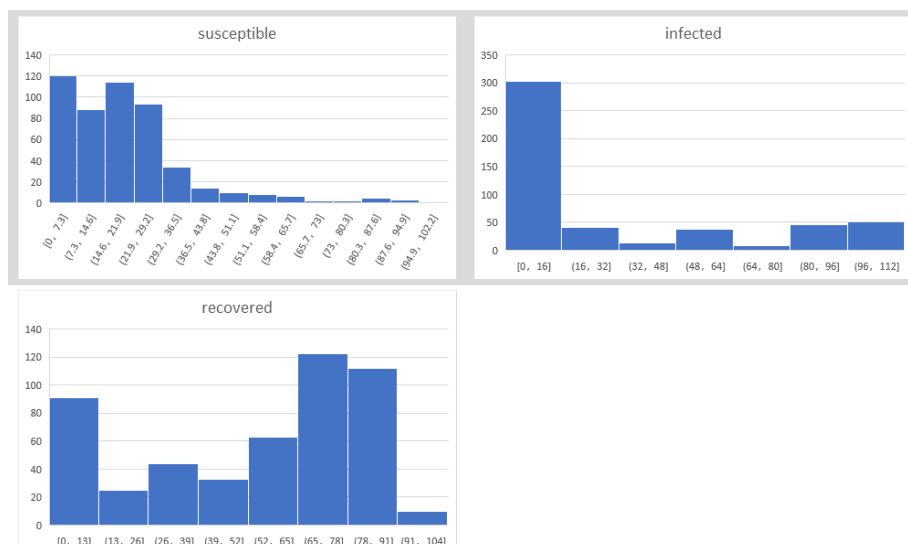
By running the model multiple times, we could get the average result of 500 time series data and calculate the *minimum, maximum and mean* results of the ratio of susceptible people, the ratio of infected people, the ratio of recovered people.

AVERAGE(%)	MIN	MAX	MEAN
susceptible	9.9720	97.0000	19.7972
infected	1.5560	78.2940	44.6495
recovered	0.0000	53.4540	35.5533
virus	1.5560	78.2940	44.6495

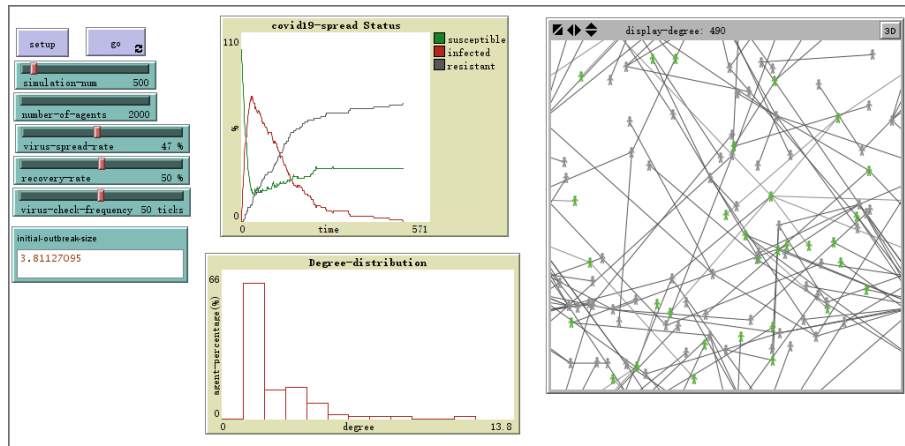
The results indicated that as the time passed by, the mean ratio of susceptible people, infected people and recovered people are around 19.80%, 44.65%, 35.55% respectively for the whole time period.

The simulation results show the experimental distribution of the ratio of the three types of people. FOR SIR model, the ratio of susceptible people and infected people were similar to the feature of power-law distribution while ratio of recovered people is not.

The experimental distribution of SIR



The network graph and screenshot of the SIR results



Similarly, for SEIR model we test the following parameters in the form of [minimum, step, maximum] in the model which run 500 times with 2000 agents and check the tick every 50 ticks and calculate the minimum, maximum and mean results of the ratio of susceptible people, exposed people, infected people, recovered people, and the ratio of the virus remaining in the system.

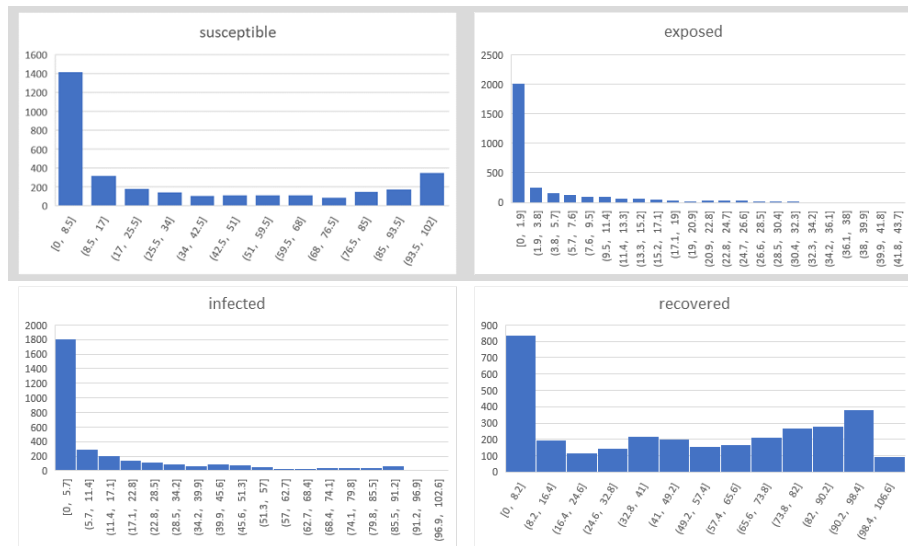
PARAMETERS	RANGES
initial-outbreak-size	[1 1 5]
recovery-rate	[1 10 100]
incubation-rate	[1 10 100]
virus-spread-rate	[1 10 100]

For SEIR model, we found that the mean ratio of exposed people was just 5.7% which was lower than other outputs. This correspond to our intuitive since in the case of Covid-19, the exposed people are not infected but infective and they would very much likely to transfer to infected people after the incubation period. With the long incubation period like 14 days, the ratio which is the 1 divide incubation period is smaller.

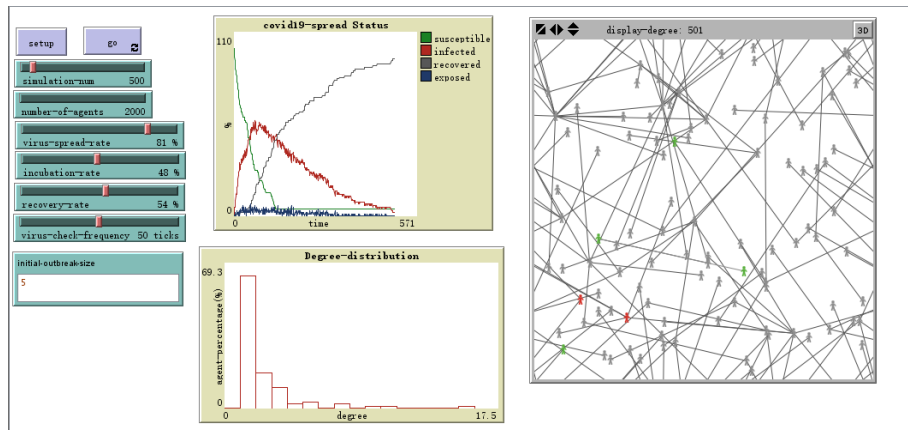
AVERAGE(%)	MIN	MAX	MEAN
susceptible	32.3918	97.8309	43.9674
exposed	0.0000	14.3373	5.7009
infected	1.2026	41.5529	21.8660
recovered	0.0000	46.8978	28.4640
virus	1.3938	48.6874	27.5569

The SEIR simulation result show that the ratio of the four types of people follows power law distribution and this proved that in the scale-free network system the experimental distribution for different types of people follow the distribution that is likely to be power-law.

The experimental distribution of SEIR



The network graph and screenshot of the SEIR results



6 Conclusion

3 epidemic models have been constructed accordingly. MGF starts from a micro perspective to observe the network and estimate numbers respectively. The probability of each node and the distribution of the connection of one selected node can be obtained. The given data is the resource for estimating the degree distribution. Once the parameter is estimated, we can deduce the situation of each node using different moment of MGF.

On the other hand, SIR and SEIR start from a macro perspective to estimate the overall situation of a network. The given data is the initial number of each state. The number of each state will be changed dynamically along the timeline. By solving a set of differential equations, we can obtain the trend of each state and summarize the result in a network.

According to $\mathbb{E}(s)$ given by MGF, we know that there are totally 0.7855 fraction of all population being affected by the epidemic, which is really closed to the the maximum infected fraction (0.78294) simulated by SIR model. As for the SEIR model, there is 1 additional state, which is the exposure state, included in the transition without any given data, the maximum infected rate could be influenced when 1 extra differential equation is involved in to calculation.

7 Reference

[M. Choisy,2007]

Choisy M, Guégan JF, Rohani P. Mathematical modeling of infectious diseases dynamics. Encyclopedia of infectious diseases: modern methodologies. 2007 Jul 17:379-404.

[Hethcote,2000]

Hethcote HW. The mathematics of infectious diseases. SIAM review. 2000;42(4):599-653.

[W. O. Kermack,1927]

W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," Proceedings of the Royal Society A, vol. 115, no. 772, pp. 700–721, 1927.

[Hethcote & Yorke 1984] Hethcote HW, Yorke JA. Lecture notes in biomathematics.

[Ghani et al. 1997]

Abdul-Ghani AS, Attwell PJ, Bradford HF. The protective effect of 2-chloroadenosine against the development of amygdala kindling and on amygdala-kindled seizures. European journal of pharmacology. 1997 May 12;326(1):7-14.

[Keeling, 1997]

Keeling MJ, Grenfell BT. Disease extinction and community size: modeling the persistence of measles. Science. 1997 Jan 3;275(5296):65-7.

[Small and Tse ,2005] Small M, Tse CK. Small world and scale free model of transmission of SARS. International Journal of Bifurcation and Chaos. 2005 May;15(05):1745-55.

[Krone, 1999] Krone SM. The two-stage contact process. The Annals of Applied Probability. 1999;9(2):331-51.

[XIN JQ, WANG WJ], 2009

XIN JQ, WANG WJ. An SIRS Epidemic Model with Stage Structure [J]. Mathematics in Practice and Theory. 2009;12.]

[Matt J. Keeling 2005]

Keeling MJ, Eames KT. Networks and epidemic models. Journal of the Royal Society Interface. 2005 Sep 22;2(4):295-307.

[M. E. J. Newman, 2001]

Newman ME. The structure of scientific collaboration networks. Proceedings of the national academy of sciences. 2001 Jan 16;98(2):404-9.

[Granell, 2013]

Granell C, Gómez S, Arenas A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical review letters*. 2013 Sep 17;111(12):128701.

[Albert-László Barabási, 1999]

Barabási AL, Albert R. Emergence of scaling in random networks. *science*. 1999 Oct 15;286(5439):509-12.

[Riley, 2003]

Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho LM, Lam TH, Thach TQ, Chau P. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science*. 2003 Jun 20;300(5627):1961-6.

[Hu, 2018]

Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, Yang L, Ding C, Zhu X, Lv R, Zhu J. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerging microbes & infections*. 2018 Dec 1;7(1):1-0.

[Hellewell J 2020]

Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday JD, Kucharski AJ, Edmunds WJ, Sun F, Flasche S. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*. 2020 Feb 28.

[Lin S, 2020]

Lin S, Qiao Y, Huang J, Yan N. Research on the Influence of Effective Distance Between Cities on the Cross-regional Transmission of COVID-19. *medRxiv*. 2020 Jan 1.