COVID-19 Outbreak Prediction Based on SEIQR Model

Yuankang Zhao[1], Yi He[2], Xiaosong Zhao[3]

1. College of Science, Wuhan University of Science and Technology, Wuhan, Hubei, China,430065
Email: ykzhao9@outlook.com
2. College of Biological Science, University of California, Davis. Davis, California, U.S., 95616
Email:fyhe@ucdavis.edu
3. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China, 430074
Email: zxs@hust.edu.cn

Abstract: The COVID-19 epidemic has caused great disruption to people's life in China as well as around the world. By using accurate and effective models to predict the development of the epidemic, authorities can better mitigate the disruption and prepare the people for the outbreak. This paper constructs a SEIQR model with a time-varying parameter that describes the declining reproduction rate. The model predicts a result of 24,318 (95% CI 13312-41054) confirmed cases in Wuhan before the lockdown, and a peak of daily increase in late February as the total confirmed cases reaches 63,315 (95% CI 28724-117566). The model further combines the prediction with a network migration model to analyze the risk of outbreak crosses the nation.

Key Words: COVID-19, SEIQR model, network model, epidemic prediction

1 Introduction

At the beginning of January 2020, a novel coronavirus outbreak was observed in Wuhan, China. The reproduction rate and the speed of the spread, as of the submission of this paper, has exceeded the 2003 SARS outbreak. Based on the preliminary researches and evaluations released by The Chinese Center for Disease Control and Prevention, Huanan Seafood Market was believed to be the source of the outbreak [1]. However, later research done by the Chinese Academy of Sciences suggested that the Huanan seafood market was only a conjugation of outbreak and the source of the virus remained unknown [2]. According to a release from the US health officials, community transmission of the COVID-19 has almost certainly occurred in the United States.[3] Subsequently, the virus is believed to cause outbreaks among other places in the world in the near future.(This model has done at February 20,2020)

As of February 24, 2020, the number of confirmed cases of COVID-19 in China has reached 77,269 and 46,607 in Wuhan, the city where the outbreak first occurred. The virus also claimed 1987 lives in Wuhan, more than the death toll of the entire SARS outbreak. The case data in Wuhan suggested that the transmission rate is higher than the SARS outbreak. Therefore, in the early stage of the epidemic, the prediction of the development of the epidemic is needed to help formulate policies to prevent the spread of the virus, and help people effectively prepare for the breakout.

At present, a lot of prediction models have been developed based on the actual situation of the virus transmission and the classical SEIR model. We made further development on the SEIR model to a SEIQR model and incorporated time steps. Furthermore, by using open-sourced population migration data, we apply this prediction model to the national level and estimate the propagation risk across the nation.

2 Model
2.1 Wuhan prediction model

The prediction model within the city of Wuhan is partly based on an improved version of the SEIR model developed by Xun Ma [4], whose prediction model classifies each individual into four categories: "Susceptible" (S), "Exposed" (E), "Infected", "Removed" (R). We add a "Quarantined" (Q) subcategory, which refers to the infected individuals who are quarantined and pose minimal risk to the general public, and a "non-isolated Infected" (I) subgroup to the infected category. Susceptible category refers to uninfected individuals in the population as all age groups are susceptible for infection. Exposed category refers to the individuals during the incubation period and Removed category refers to the recovered or deceased individuals that were previously infected. Patients within Exposed and Infectious category are both contagious [5].

In this model, we assume the population is equally distributed in the city and the whole population is susceptible to the virus. At each time step, if a susceptible induvial have contact with an exposed individual, this individual has a chance of β_1 to become exposed. A susceptible individual has a chance of β_2 to become exposed when contact with a non-isolated patient. Each exposed individual has a chance of λ to transfer into the non-isolated infectious stage I, so the incubation period is $\frac{1}{\lambda}$ time step. The unquarantined will have the chance of θ to become quarantined, so it takes $\frac{1}{\theta}$ time step for an individual to be quarantined. We assume each quarantined patient has the chance of γ to become removed, so it takes

$\frac{1}{\gamma}$ time step to recover or death. We assume the removed individuals have antibody and will not be infected [6].

The process can be visualized as the flow chart below:

The previous process can be represented as five differential equations:

$$\frac{ds(t)}{dt} = -\left(\frac{\beta_1(t)E(t)}{N} + \frac{\beta_2(t)I(t)}{N}\right)S(t) - Z(t) \quad (1)$$

$$\frac{dE(t)}{dt} = \left(\frac{\beta_1(t)E(t)}{N} + \frac{\beta_2(t)I(t)}{N}\right)S(t) + Z(t) - \lambda E(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \lambda E(t) - \theta I(t) \quad (3)$$

$$\frac{dQ(t)}{dt} = \theta I(t) - \gamma Q(t) \quad (4)$$

$$\frac{dR(t)}{dt} = \gamma Q(t) \quad (5)$$

S(t), E(t), I(t), Q(t), R(t) each represent the number of susceptible, exposed, infected, quarantined and removed individuals at time t; Z(t) is the number of zoonotic transfections from animal to human at time t. We assume zoonotic transfection is only observed before January 1st, 2020. The shutdown of Huanan Seafood Market, and there is no zoonotic transfection after this. Based on the research of Qun Li [7], we set the Z(t) before January 1st as 2 for each day; and N is the whole population. $\beta_i(t)$ is a time-varying parameter. We adopt the use of exponential decrease of transfection rate to describe interventions (Ex: people have the idea of wearing masks, government enhanced regulation within the city), the function can be represented as below:

$$\beta_i(t) = \begin{cases} \beta, & t < \tau \\ \beta e^{-k(t-\tau)}, & t \geq \tau \end{cases} \quad (6)$$

In formula (6), τ is the time that intervention take place. We take τ as January 23rd, the time the whole city starts quarantine. k is the decrease parameter. k fluctuates with the rate of increase in patient number in Beijing. We used Beijing's rate on patients in Wuhan at the beginning of the intervention as the city-wide deficiency in medical resources caused the initial statistical data to be inaccurate. So, under the same intervention condition, Beijing has a smaller patient size and is more suitable for calculating decrease parameter k.

In this paper, we use 4.7 days as the incubation period [8], 14 days as the removal period [8]. The θ in the model is directly associated with the abundance of medical resources and the general public's attitude toward the novel coronavirus. In the Wuhan area, the phenomenon of low hospital availability is observed during the peak outbreak period [9]. The extreme limited medical resources lessen the chance of infected patients to be quarantined even after being confirmed of the virus. These patients remain contagious during this period as the severe scarcity of medical resources last for days before being properly addressed by the government. With the influx of medical personnel and equipment, the effectiveness of the quarantine shows a significant increase [10]. To take account for this factor, we describe quarantine time $\frac{1}{\theta}$ as an exponential decrease model and combine it with Wuhan area's circumstances.

$$\frac{1}{\theta}(t) = \begin{cases} \frac{1}{\theta_0}, & t < \psi \\ \frac{1}{\theta_1 + (\theta_0 - \theta_1)e^{-k'(t-\tau)}}, & t \geq \psi \end{cases} \quad (7)$$

In the equation above, ψ is the medical resources restocking time, and θ_0 is the average quarantine rate before the restocking of medical resources, θ_1 is the average quarantine rate after the restocking of medical resources, that is, $\frac{1}{\theta_1}$ is the time from onset to isolation when medical conditions are sufficient.

2.2 Nationwide Model Prediction

Based on the Wuhan prediction model and Aleksa Zlojutro et al.'s study on network and infectious disease [11], we come up with a prediction mode on the propagation risk for each city. Since the outbreak of COVID-19 occurred during the Chinese lunar new year, the high volume of intercity trafficking caused the coronavirus to transmit rapidly nationwide. So, the priority was to contain the virus by using this network model to predict the outbreak. The fundamental of this model is based on the prediction model which consist of differential equation within Wuhan, with data on the influx and efflux of population. Essentially, we apply the Wuhan model to each node (i.e. city) on the network.

According to Aleksa Zlojutro's research [11], a multi-commodity network flow model is established first. We denote Π_{ij} the set of paths from $i \in V$ to $j \in V$, and denote $f_{ij,t}^k$ the average passenger from $i \in V$ to $j \in V$ by the path by the path $k \in \Pi_{ij}$ at time t. We denote $\Gamma_i^- = \{j \in V : \exists k \in \Pi_{ij}, t \in T, f_{ji,t}^k > 0\}$ and $\Gamma_i^+ = \{j \in V : \exists k \in \Pi_{ij}, t \in T, f_{ij,t}^k > 0\}$ the set of nodes which connects with $i \in V$. Path $k \in \Pi_{ij}$ is a sequence of nodes shows the path from i to j, ie k = {i, n1, n2, …, j}. We assume that $f_{ij,t}^k = f_{ji,t}^k$. Combined with the former model, the new epidemic model is summarized in below.

$$S_{i,t+1} = -\left(\frac{\beta_{1,i,t}E_{i,t+1}}{N} + \frac{\beta_{2,i,t}E_{i,t+1}}{N}\right)S_{i,t} + \sum_{j \in \Gamma_i^-}\sum_{k \in \Pi_{ji}} S_{ji,t}^k - \sum_{j \in \Gamma_i^+}\sum_{k \in \Pi_{ij}} S_{ij,t}^k \quad (8)$$

$$E_{i,t+1} = \left(\frac{\beta_{1,i,t}E_{i,t+1}}{N} + \frac{\beta_{2,i,t}E_{i,t+1}}{N}\right)S_{i,t} - \lambda E_{i,t} + \sum_{j \in \Gamma_i^-}\sum_{k \in \Pi_{ji}} E_{ji,t}^k - \sum_{j \in \Gamma_i^+}\sum_{k \in \Pi_{ij}} E_{ij,t}^k \quad (9)$$

$$I_{i,t+1} = \lambda E_{i,t} - \theta I_{i,t} + \sum_{j \in \Gamma_i^-}\sum_{k \in \Pi_{ji}} I_{ji,t}^k \left(\prod_{p \in k\{j\}}(1 - x_{p,t})\right) - \sum_{j \in \Gamma_i^+}\sum_{k \in \Pi_{ij}} I_{ij,t}^k \quad (10)$$

$$Q_{i,t+1} = \theta I_{i,t} - \gamma Q_{i,t} \quad (11)$$

$$R_{i,t+1} = \gamma Q_{i,t} + \sum_{j \in \Gamma_i^-}\sum_{k \in \Pi_{ji}} R_{ji,t}^k - \sum_{j \in \Gamma_i^+}\sum_{k \in \Pi_{ij}} R_{ij,t}^k \quad (12)$$

The symbols $S_{ji,t}^k$, $E_{ji,t}^k$, $I_{ji,t}^k$ and $R_{ji,t}^k$, represent passengers flows on (i, j) with destination k at time step t.

The number of passengers of each stage is be calculated as $S_{ij,t}^k = f_{ij,t}^k \frac{S_{i,t}}{N_i}$.

To control the outbreak of the coronavirus, checkpoints are widely set at transportation hubs all over China. Therefore, we denote $x_{I,t} \in [0, 1]$ the control rate at node i and time t.

During the application of the model, we use provinces and province-level municipalities as nodes (i.e. element of set V). For the population data and the volume of passenger traffic $f_{ij,t}^k$, we use Baidu's migration data, a product based on location service. We adjust the migration volume according to the lunar new year, and we keep the other data the same as Wuhan model.

3 Result

3.1 Wuhan Area

Because of the limitation of the diagnostic techniques and medical resources, the infection data was delayed and there was a lack of accuracy at the beginning of the outbreak. For the same reason, we cannot incorporate the data collected before the lockdown of the city into our model. We use Joseph T Wu et al's study [12] on the basic reproductive number R_0 and found out the R_0 is 2.68 (95% CI 2.47-2.86). After conversion, we can calculate the transfection rate β, then insert this rate back to the model and we find out that the infected population before the quarantine (including patient during the incubation period) is around 24318 (95% CI 13312-41054), and 11137 of them have already displayed symptom.
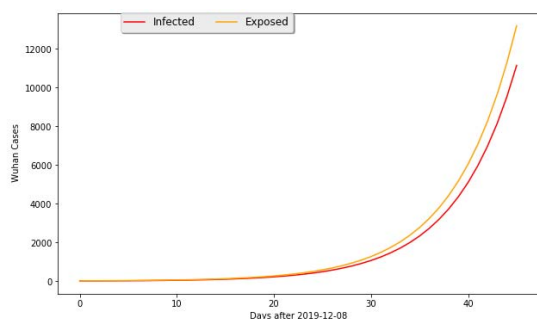

Figure1 Forecast COVID-19 Outbreak in Wuhan between 2019-12-08 and 2020-1-23 using estimated data (Running by R_0 =2.68)

After January 23rd, the infected population displays an increase and then decrease trend. The prediction peak would arrive around February 21st. At this peak, the infected population would be 45077 (95% CI 20442-89231). If we include all the exposed cases, the infected population would be 63315 (95% CI 28724-117566). The prediction matches the real patient count until February 13th.
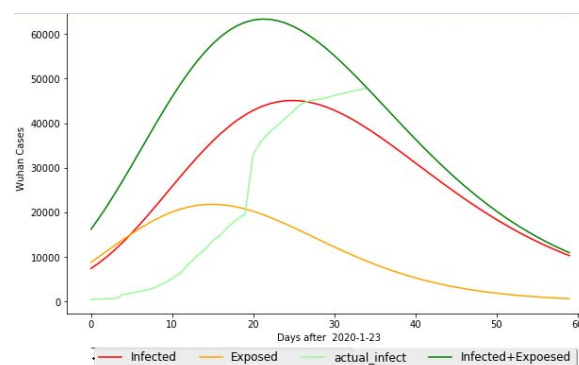

Figure 2 Forecast COVID-19 Outbreak in Wuhan after 2020-1-23 using estimated data (Running by initial R_0 =2.68)

Combine with the official statistic data [13] (shown as the pale green line in the Fig 2), the peak of increase in patient number will be from February 23nd to March 2nd, and the epidemic will reach a relative low level at the beginning of April.

3.2 Nationwide situation

We use Baidu migration data as the flow between each node. However Baidu migration data only provide the proportion of influx and efflux [14], so we cannot get the specific data of population migration between each node, and we can only calculate the influx risk base on the proportion. In this paper, we only use the population migration data before January 23rd as the local government implements quarantine measures to shut down all the transportation in and out of the city. This measure effectively eradicates the efflux of patient. We come up with this result.

Figure 3 National risk of COVID-19 outbreak (the geographic information is from Bing)

In Figure 3, the darker red areas indicate the higher risk of an outbreak. Our results show that the majority of the cities within Hubei province have high risks of outbreak and personal protection is needed for these areas. Guangdong province, Zhejiang province, Shanghai, Beijing as well as Hunan province, Hebei province all have elevated chances of outbreak. Other cities that experience small amount of influx of people such as Chengdu, Chongqing and Xiamen are rated as low risks

compare to previous areas. However, preventative measures are still needed as the chance of outbreak still exists. With more specific migration data, it can also predict relative accurate patient number at each city during the beginning of the outbreak and give early warnings to authorities.

4　Limitation

4.1 Data Limitation

The delay of patient data at the beginning stage cause the lack of accuracy of the prediction at the beginning as early data plays a determinative role in estimating the reproduction rate. Instead, this prediction adopts the R_0 calculated by Hongkong University to help eliminate the error [12]. In our prediction model, the open sourced migration data is only accurate to be on the city level, so our accuracy is also limited.

4.2 Method Limitation

When representing the decrease of reproduction rate, we cannot include the effect of public health interventions as the reproduction rate changes dramatically in response to the radically enforcement of mandatory mask-wearing policies and the traveling restrictions. The endorsement of same parameter k across cities may also introduce errors as preventative policy may differ from city to city.

5　Discussion

Compared to the basic SEIR model, our prediction more accurately represented the real situation of the COVID-19 outbreak. The lack of proper medical resources and public attention at the start of the epidemic resulted in a prolonged waiting period before patients could check into the hospital. This prolonged period combined the highly contagious nature of the virus caused a colossal outbreak and an elevated reproduction rate. This model incorporates a dynamic reproduction rate and network migration for the prediction of the patient number. The reproduction rate is a relatively large number at the beginning as no preventative measure is taken. The case propagates in a logarithmic pattern. However, as the epidemic progress, the general public as well as the authorities are taking measures to stop the propagation of the virus. The lockdown of the city and the mandatory mask-wearing greatly reduces the risk of getting an infection; and a fixed reproduction rate cannot reflect the change in risks. By changing the reproduction rate in response to the change of policy, the model provides a more accurate prediction over the course of the outbreak. With the inclusion of more data, we can also predict the approximate ending time of the epidemic.

Compares to the research conducted at the PKU Visual Computing and Learning Lab [15], the PKU's model describes the outbreak of the epidemic more accurately by adding suspected and confirmed patients at the isolation point. However, the large number of parameters required by this model limits its practicality. Some of which are difficult to obtain nor can be fitted by actual data,

especially in the early and middle stages of epidemic development when statistical resources are scarce. Therefore, only the parameters used in the pre-study [16] can be used, and the exact number of cases in this outbreak cannot be accurately predicted.

6　Summary

This article focuses on the prediction of the COVID-19 epidemic by investigating the general situation of the epidemic in Wuhan and other parts of China. In view of the deficiencies in previous studies and the characteristics of this epidemic, improved methods are proposed: 1) According to the actual situation of the epidemic outbreak, the infected population in the traditional SEIR model is further subdivided into isolated infected category and non-isolated infected category; 2) Due to the impact of government policies on the outbreak, a SEIR model of dynamic reproduction rate is proposed; 3) The improved SEIR model is combined with a network model to predict the distribution of infected people across the country.

The results of the study show that 24,318 (95% CI 13312-41054) people are infected before Wuhan was lockdown, and that the number of infected people in Wuhan would reach a peak of 63,315 (95% CI 28724-117566) in late February. The network combined model shows that other cities in Hubei Province, Guangdong Province, Zhejiang Province, Shanghai, Beijing, and surrounding Hunan and Henan Provinces are all high risks areas.

The idea of infectious disease dynamics modeling proposed in this paper can be used to evaluate the current spreading situation, predict the spreading trend, and select effective prevention and control measures, etc. This is of great significance to suppress the spread of infectious diseases.

References

[1] The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, Li Q. An outbreak of NCIP (2019-nCoV) infection in China— Wuhan, Hubei Province, 2019–2020. China CDC Weekly 2020; 2: 79–80.

[2] Yu, Wen-Bin,Tang, Guang-Da,Zhang, Li,Corlett, Richard T.(2020).Decoding evolution and transmissions of novel pneumonia coronavirus using the whole genomic data. [ChinaXiv: 202002.00033]

[3] NBC News, Americans should prepare for coronavirus crisis in U.S., CDC says (2020). https://www.nbcnews.com/health/health -news/americans-should-prepare-coronavirus-spread-u-s-cdc-says-n 1142556

[4] Ma Xun, Ni Shunjiang, Shen Shifei. Risk assessment of Ebola virus transmission based on early surveillance cases [J]. Journal of Tsinghua University (Science and Technology), 2017 (08): 39-43.

[5] Guan W, Ni Z, Hu Y, et al. Clinical characteristics of 2019 novel coronavirus infection in China[J]. medRxiv, 2020.

[6] Sina News, Patients with neo-coronary pneumonia are very unl ikely to be infected, https://news.sina.cn/2020-02-22/detail-iimxx stf3507887.d.html (2020)

[7] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia[J]. New England Journal of Medicine, 2020.

[8] Yang Y, Lu Q, Liu M, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China[J]. medRxiv, 2020.

[9] Huanqiu News, Bed tension is still a short board, Available at: https://3w.huanqiu.com/a/882c30/3x0qFeH7e9S?p=4&agt=8, (2020)

[10] Wuhan Municipal Health Reply of Wuhan Municipal Health and Health Committee on Several Issues Concerned by Citizens. Available at:http://wjw.wuhan.gov.cn/front/web/showDetail/2020012309114, 2020

[11] Zlojutro, A., Rey, D. & Gardner, L. A decision-support framework to optimize border control for global outbreak mitigation. *Sci Rep* 9, 2216 (2019).

[12] Wu, Joseph T., Kathy Leung, and Gabriel M. Leung. "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study." *The Lancet* (2020).

[13] National Health Comssion of P.R.C, Update on the outbreak of new coronavirus pneumonia as of 24:00 on February 26 (2020)

[14] Baidu Corproation, Baidu Map Migration Data, Available at : http://qianxi.baidu.com(2020)

[15] Chen B, Ni X, Shi M, et al. Data visualization analysis and simulation prediction for the new crown epidemic Available at: https://github.com/NCP-VIS, (2020)

[16] Zhang J, Lou J, Ma Z, et al. A compartmental model for the analysis of SARS transmission patterns and outbreak control measures in China. Applied Mathematics and Computation, 2005, 162(2):909-924.