

# A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons

Yi-Cheng Chen\*, Ping-En Lu<sup>†</sup>, Graduate Student Member, IEEE, Cheng-Shang Chang<sup>‡</sup>, Fellow, IEEE, and Tzu-Hsuan Liu<sup>§</sup>

Institute of Communications Engineering  
National Tsing Hua University  
Hsinchu 30013, Taiwan, R.O.C.

Email: \*yichengchen@gapp.nthu.edu.tw, <sup>†</sup>j94223@gmail.com, <sup>‡</sup>cschang@ee.nthu.edu.tw, <sup>§</sup>carinaliu@gapp.nthu.edu.tw

The latest version will be placed on this link:

[http://gibbs1.ee.nthu.edu.tw/A\\_TIME\\_DEPENDENT\\_SIR\\_MODEL\\_FOR\\_COVID\\_19.PDF](http://gibbs1.ee.nthu.edu.tw/A_TIME_DEPENDENT_SIR_MODEL_FOR_COVID_19.PDF)

**Abstract**—In this paper, we propose two mathematical models for analyzing and predicting the number of confirmed cases of COVID-19. Our first model is a time-dependent susceptible-infected-recovered (SIR) model that tracks two time series: (i) the transmission rate at time  $t$  and (ii) the recovering rate at time  $t$ . Our time-dependent SIR method is better than the traditional static SIR model as it can adapt to the change of contagious disease control policies such as city lockdowns. Moreover, it is also more robust than the direct estimation of the number of confirmed cases, as a sudden change of the definition of confirmed cases might result in a spike in the number of new cases. Using the data provided by the National Health Commission of the People's Republic of China (NHC) [1], we show that the one-day prediction errors for the numbers of confirmed cases are almost less than 3% except for the day when the definition of confirmed cases is changed. Also, the turning point, defined as the day that the transmission rate is less than the recovering rate, is predicted to be Feb. 17, 2020. After that day, the basic reproduction number, known as the  $R_0(t)$  value, is less than 1 if the current contagious disease control policies are maintained in China. In that case, the total number of confirmed cases is predicted to be around 80,000 cases in China under our deterministic model. One problem for the first model is that there are *undetectable* infected persons for COVID-19. According to the recent WHO data, only 87.9% of the confirmed cases of COVID-19 have a fever. If we use body temperature as a means to detect COVID-19 infected cases, then more than 10% of infected persons cannot be detected. To model this, we propose our second SIR model that has two types of infected persons: detectable infected persons (type I) and undetectable infected cases (type II). Type I infected persons have a lower transmission rate than that of type II infected persons (as type I infected persons can be isolated). Whether there is an outbreak in such a model is characterized by the spectral radius of a  $2 \times 2$  matrix that is closely related to the basic reproduction number  $R_0$ . We plot the phase transition diagram of an outbreak with respect to two parameters: the probability of a type II infected case and the transmission rate of type II infected cases. Our numerical results show that there are several countries, including South Korea, Italy, and Iran, that are above the percolation threshold curve, and they are on the verge of COVID-19 outbreaks on Mar. 2, 2020.

**Index Terms**—COVID-19, SARS-CoV-2, 2019-nCoV, Coronavirus, time-dependent SIR model, asymptomatic infection

## I. INTRODUCTION

At the beginning of December 2019, the first COVID-19 victim was diagnosed with the coronavirus in Wuhan, China. In the following weeks, the disease spread widely in China mainland and other countries, which causes global panic. The virus has been named “SARS-CoV-2,” and the disease it causes has been named “coronavirus disease 2019 (abbreviated “COVID-19”). There have been 80,151 people infected by the disease and 2,943 deaths until Mar. 2, 2020 according to the official statement by Chinese government. To block the spread of the virus, there are some strategies such as city-wide lockdown, traffic halt, community management, social distancing, and propaganda of health education knowledge that have been adopted by the government of China and other countries in the world.

Unlike the Severe Acute Respiratory Syndrome (SARS) and other infectious diseases, one problematic characteristic of COVID-19 is that there are asymptomatic infections (who have very mild symptoms). Those asymptomatic infections are unaware of their contagious ability, and thus get more people infected. The transmission rate can increase dramatically in this circumstance. According to the recent report from WHO [2], only 87.9% of COVID-19 patients have a fever, and 67.7% of them have a dry cough. If we use body temperature as a means to detect COVID-19 infected cases, then more than 10% of infected persons cannot be detected.

Due to the recent development of the epidemic, we are interested in addressing the following important questions for COVID-19:

- (Q1) Is it possible to contain COVID-19? Are the commonly used measures, such as city-wide lockdown, traffic halt, community management, social distancing, and propaganda of health education knowledge, effective in containing COVID-19?
- (Q2) If COVID-19 can be contained, when will be the peak of the epidemic, and when will it end?
- (Q3) How do the asymptomatic infections affect the spread of disease?

For (Q1), we analyze the cases in China and aim to predict how the virus spreads in this paper. Specifically, we propose using a time-dependent susceptible-infected-recovered (SIR) model to analyze and predict the number of infected persons and the number of recovered persons (including deaths). In the traditional SIR model, it has two time-invariant variables: the transmission rate  $\beta$  and the recovering rate  $\gamma$ . The transmission rate  $\beta$  means that each individual has on average  $\beta$  contacts with randomly chosen others per unit time. On the other hand, the recovering rate  $\gamma$  indicates that individuals in the infected state get recovered or die at a fixed average rate  $\gamma$ . The traditional SIR model neglects the time-varying property of  $\beta$  and  $\gamma$ , and it is too simple to precisely and effectively predict the trend of the disease. Therefore, we propose using a time-dependent SIR model, where both the transmission rate  $\beta$  and the recovering rate  $\gamma$  are functions of time  $t$ . Our idea is to use machine learning methods to track the transmission rate  $\beta(t)$  and the recovering rate  $\gamma(t)$ , and then use them to predict the number of the infected persons and the number of recovered persons at a certain time  $t$  in the future. Our time-dependent SIR model can dynamically adjust the crucial parameters, such as  $\beta(t)$  and  $\gamma(t)$ , to adapt accordingly to the change of control policies, which differs from the existing SIR and SEIR models in the literature, e.g., [3], [4], [5], [6], and [7]. For example, we observe that city-wide lockdown can lower the transmission rate substantially from our model. Most data-driven and curve-fitting methods for the prediction of COVID-19, e.g., [8], [9], and [10] seem to track data perfectly; however, they are lack of physical insights of the spread of the disease. Moreover, they are very sensitive to a sudden change in the definition of confirmed cases on Feb. 12, 2020. On the other hand, our time-dependent SIR model can examine the epidemic control policy of the Chinese government and provide reasonable explanations. Using the data provided by the National Health Commission of the People's Republic of China (NHC) [1], we show that the one-day prediction errors for the numbers of confirmed cases are almost less than 3% except for the day when the definition of confirmed cases is changed on Feb. 12, 2020.

For (Q2), the basic reproduction number  $R_0$ , defined as the number of additional infections by an infected person before it recovers, is one of the commonly used metrics to check whether the disease will become an outbreak. In the classical SIR model,  $R_0$  is simply  $\beta/\gamma$  as an infected person takes (on average)  $1/\gamma$  days to recover, and during that period time, it will be in contact with (on average)  $\beta$  persons. In our time-dependent SIR model, the basic reproduction number  $R_0(t)$  is a function of time, and it is defined as  $\beta(t)/\gamma(t)$ . If  $R_0(t) > 1$ , the disease will spread exponentially and infects a certain fraction of the total population. On the contrary, the disease will eventually be contained. Therefore, by observing the change of  $R_0(t)$  with respect to time or even predict  $R_0(t)$  in the future, we can check whether certain epidemic control policies are effective or not. Using the data provided by the National Health Commission of the People's Republic of China (NHC) [1], we show that the turning point (peak), defined as the day that the basic reproduction number is less than 1, is predicted to be Feb. 17, 2020. Moreover, the disease

in China will end in about 6 weeks after its peak in our (deterministic) model if the current contagious disease control policies are maintained in China. In that case, the total number of confirmed cases is predicted to be around 80,000 cases in China under our (deterministic) model.

For (Q3), we propose our second SIR model that has two types of infected persons: detectable infected persons (type I) and undetectable infected cases (type II). Type I infected persons have a lower transmission rate than that of type II infected persons (as type I infected persons can be isolated). For such a model, whether the disease is controllable is characterized by the spectral radius of a  $2 \times 2$  matrix. If the spectral radius of that matrix is larger than 1, then there is an outbreak. On the other hand, if it is smaller than 1, then there is no outbreak. One interesting result is that the spectral radius of that matrix is larger (resp. smaller) than 1 if the basic reproduction number  $R_0$  is larger (resp. smaller) than 1. The curve that has the spectral radius equal to 1 is known as the percolation threshold curve in a phase transition diagram [11]. Using the historical data from Jan. 22, 2020 to Mar. 2, 2020 from WHO [2], we extend our study to some other countries, including Japan, Singapore, South Korea, Italy, and Iran. Our numerical results show that there are several countries, including South Korea, Italy, and Iran, that are above the percolation threshold curve, and they are on the verge of COVID-19 outbreaks on Mar. 2, 2020.

The rest of the paper is organized as follows: In Section II, we propose the time-dependent SIR model. We then extend the model to the SIR model with undetectable infected persons in Section III. In Section IV, we conduct several numerical experiments to illustrate the effectiveness of our two models. In Section V, we put forward some discussions and suggestions to control COVID-19. The paper is concluded in Section VI.

## II. THE TIME-DEPENDENT SIR MODEL

### A. Susceptible-infected-recovered (SIR) Model

In the typical mathematical model of infectious disease, one often simplify the virus-host interaction and the evolution of an epidemic into a few basic disease states. One of the simplest epidemic model, known as the susceptible-infected-recovered (SIR) model [11], includes three states: the susceptible state, the infected state, and the recovered state. An individual in the *susceptible state* is one who does not have the disease at time  $t$  yet, but may be infected if one is in contact with a person infected with the disease. The *infected state* refers to an individual who has a disease at time  $t$  and may infect a susceptible individual potentially (if they come into contact with each other). The *recovered state* refers to an individual who is either recovered or dead from the disease and is no longer contagious at time  $t$ . Also, a recovered individual will not be back to the susceptible state anymore. The reason for the number of deaths is counted in the recovered state is that, from an epidemiological point of view, this is basically the same thing, regardless of whether recovery or death does not have much impact on the spread of the disease. As such, they can be effectively eliminated from the potential host of the disease [12]. Denote by  $S(t)$ ,  $X(t)$  and  $R(t)$  the numbers of

susceptible persons, infected persons, and recovered persons at time  $t$ . Summing up the above SIR model, we believe it is very similar to the COVID-19 outbreak, and we will adopt the SIR model as our basic model in this paper.

In the traditional SIR model, it has two time-invariant variables: the transmission rate  $\beta$  and the recovering rate  $\gamma$ . The transmission rate  $\beta$  means that each individual has on average  $\beta$  contacts with randomly chosen others per unit time. On the other hand, the recovering rate  $\gamma$  indicates that individuals in the infected state get recovered or die at a fixed average rate  $\gamma$ . The traditional SIR model neglects the time-varying property of  $\beta$  and  $\gamma$ . This assumption is too simple to precisely and effectively predict the trend of the disease. Therefore, we propose the time-dependent SIR model, where both the transmission rate  $\beta$  and the recovering rate  $\gamma$  are functions of time  $t$ . Such a time-dependent SIR model is much better to track the disease spread, control, and predict the future trend.

### B. Differential Equations for the Time-dependent SIR Model

For the traditional SIR model, the three variables  $S(t)$ ,  $X(t)$  and  $R(t)$  are governed by the following differential equations (see, e.g., the book [11]):

$$\begin{aligned}\frac{dS(t)}{dt} &= \frac{-\beta S(t)X(t)}{n}, \\ \frac{dX(t)}{dt} &= \frac{\beta S(t)X(t)}{n} - \gamma X(t), \\ \frac{dR(t)}{dt} &= \gamma X(t).\end{aligned}$$

We note that

$$S(t) + X(t) + R(t) = n, \quad (1)$$

where  $n$  is the total population. Let  $\beta(t)$  and  $\gamma(t)$  be transmission rate and recovering rate at time  $t$ . Replacing  $\beta$  and  $\gamma$  by  $\beta(t)$  and  $\gamma(t)$  in the differential equations above yields

$$\frac{dS(t)}{dt} = \frac{-\beta(t)S(t)X(t)}{n}, \quad (2)$$

$$\frac{dX(t)}{dt} = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t), \quad (3)$$

$$\frac{dR(t)}{dt} = \gamma(t)X(t). \quad (4)$$

The three variables  $S(t)$ ,  $X(t)$  and  $R(t)$  still satisfy (1).

Now we briefly explain the intuition of these three equations. Equation (2) describes the difference of the number of susceptible persons  $S(t)$  at time  $t$ . If we assume the total population is  $n$ , then the probability that a randomly chosen person is in the susceptible state is  $S(t)/n$ . Hence, an individual in the infected state will contact (on average)  $\beta(t)S(t)/n$  people in the susceptible state per unit time, which implies the number of newly infected persons is  $\beta(t)S(t)X(t)/n$  (as there are  $X(t)$  people in the infected state at time  $t$ ). On the contrary, the number of people in the susceptible state will decrease by  $\beta(t)S(t)X(t)/n$ . Additionally, as every individual in the infected state will recover with rate  $\gamma(t)$ , there are (on average)  $\gamma(t)X(t)$  people recovered at time  $t$ . This is shown in (4) that

illustrates the difference of  $R(t)$  at time  $t$ . Since three variables  $S(t)$ ,  $X(t)$  and  $R(t)$  still satisfy (1), we have

$$\frac{dX(t)}{dt} = -\left(\frac{dS(t)}{dt} + \frac{dR(t)}{dt}\right),$$

which is the number of people changing from the susceptible state to the infected state minus the number of people changing from the infected state to the recovered state (see (3)).

### C. Discrete Time Time-dependent SIR Model

Due to the COVID-19 data is updated in days [1], we revise the differential equations in (2), (3), and (4) into discrete time difference equations:

$$S(t+1) - S(t) = \frac{-\beta(t)S(t)X(t)}{n}, \quad (5)$$

$$X(t+1) - X(t) = \frac{\beta(t)S(t)X(t)}{n} - \gamma(t)X(t), \quad (6)$$

$$R(t+1) - R(t) = \gamma(t)X(t). \quad (7)$$

Again, the three variables  $S(t)$ ,  $X(t)$  and  $R(t)$  still satisfy (1).

In the beginning of the disease spread, the number of confirmed cases is very low, and most of the population are in the susceptible state. Hence, for our analysis of the initial stage of COVID-19, we assume  $\{S(t) \approx n, t \geq 0\}$ , and further simplify (5), (6), and (7) as follows:

$$S(t+1) - S(t) = -\beta(t)X(t), \quad (8)$$

$$X(t+1) - X(t) = \beta(t)X(t) - \gamma(t)X(t). \quad (9)$$

From the difference equations above, one can easily derive  $\beta(t)$  and  $\gamma(t)$  of each day. From (7), we have

$$\gamma(t) = \frac{R(t+1) - R(t)}{X(t)}. \quad (10)$$

Using (7) in (9) yields

$$\beta(t) = \frac{[X(t+1) - X(t)] + [R(t+1) - R(t)]}{X(t)}. \quad (11)$$

Given the historical data from a certain period  $\{X(t), R(t), 0 \leq t \leq T-1\}$ , we can measure the corresponding  $\{\beta(t), \gamma(t), 0 \leq t \leq T-2\}$  by using (10) and (11). With the above information, we can use machine learning methods to predict the time varying transmission rates and recovering rates.

### D. Tracking Transmission Rate $\beta(t)$ and Recovering Rate $\gamma(t)$ by Ridge Regression

In this subsection, we track and predict  $\beta(t)$  and  $\gamma(t)$  by the commonly used Finite Impulse Response (FIR) filters in linear systems. Denote by  $\hat{\beta}(t)$  and  $\hat{\gamma}(t)$  the *predicted* transmission rate and recovering rate. From the FIR filters, they are predicted as follows:

$$\begin{aligned}\hat{\beta}(t) &= a_1\beta(t-1) + a_2\beta(t-2) + \cdots + a_J\beta(t-J) + a_0 \\ &= \sum_{j=1}^J a_j\beta(t-j) + a_0,\end{aligned} \quad (12)$$

$$\begin{aligned}\hat{\gamma}(t) &= b_1\gamma(t-1) + b_2\gamma(t-2) + \cdots + b_K\gamma(t-K) + b_0 \\ &= \sum_{k=1}^K b_k\gamma(t-k) + b_0,\end{aligned} \quad (13)$$

where  $J$  and  $K$  are the orders of the two FIR filters ( $0 < J, K < T - 2$ ),  $a_j, j = 0, 1, \dots, J$ , and  $b_k, k = 0, 1, \dots, K$  are the coefficients of the impulse responses of these two FIR filters.

There are several widely used machine learning methods for the estimation of the coefficients of the impulse response of an FIR filter, e.g., ordinary least squares (OLS), regularized least squares (i.e., ridge regression), and partial least squares (PLS) [13]. In this paper, we choose the ridge regression as our estimation method that solves the following optimization problem:

$$\min_{a_j} \sum_{t=J}^{T-2} (\beta(t) - \hat{\beta}(t))^2 + \alpha_1 \sum_{j=0}^J a_j^2, \quad (14)$$

$$\min_{b_k} \sum_{t=K}^{T-2} (\gamma(t) - \hat{\gamma}(t))^2 + \alpha_2 \sum_{k=0}^K b_k^2, \quad (15)$$

where  $\alpha_1$  and  $\alpha_2$  are the regularization parameters.

### E. Tracking the Number of Infected Persons $\hat{X}(t)$ and the Number of Recovered Persons $\hat{R}(t)$ of the Time-dependent SIR Model

In this section, we show how we use the two FIR filters to track and predict the number of infected persons and the number of recovered persons in the time-dependent SIR model. Given a period of historical data  $\{X(t), R(t), 0 \leq t \leq T - 1\}$ , we first measure  $\{\beta(t), \gamma(t), 0 \leq t \leq T - 2\}$  by (10) and (11). Then we solve the ridge regression (with the objective functions in (14) and (15) and the constraints in (12) and (13)) to learn the coefficients of the FIR filters, i.e.,  $a_j, j = 0, 1, \dots, J$  and  $b_k, k = 0, 1, \dots, K$ . Once we learn these coefficients, we can predict  $\hat{\beta}(t)$  and  $\hat{\gamma}(t)$  at time  $t = T - 1$  by the trained ridge regression in (12) and (13).

Denote by  $\hat{X}(t)$  (resp.  $\hat{R}(t)$ ) the predicted number of infected (resp. recovered) persons at time  $t$ . To predict  $\hat{X}(t)$  and  $\hat{R}(t)$  at time  $t = T$ , we simply replace  $\beta(t)$  and  $\gamma(t)$  by  $\hat{\beta}(t)$  and  $\hat{\gamma}(t)$  in (7) and (9). This leads to

$$\hat{X}(T) = (1 + \hat{\beta}(T - 1) - \hat{\gamma}(T - 1))X(T - 1), \quad (16)$$

$$\hat{R}(T) = R(T - 1) + \hat{\gamma}(T - 1)X(T - 1). \quad (17)$$

To predict  $\hat{X}(t)$  and  $\hat{R}(t)$  for  $t > T$ , we estimate  $\hat{\beta}(t)$  and  $\hat{\gamma}(t)$  by using (12) and (13). Similar to those in (16) and (17), we predict  $\hat{X}(t)$  and  $\hat{R}(t)$  as follows:

$$\hat{X}(t + 1) = (1 + \hat{\beta}(t) - \hat{\gamma}(t))\hat{X}(t), \quad t \geq T, \quad (18)$$

$$\hat{R}(t + 1) = \hat{R}(t) + \hat{\gamma}(t)\hat{X}(t), \quad t \geq T. \quad (19)$$

The detailed steps of our tracking/predicting method are outlined in Algorithm 1.

We note that this *deterministic* epidemic model is based on the mean-field approximation for  $X(t)$  and  $R(t)$ . Such an approximation is a result of the law of large numbers. Therefore, when  $X(t)$  and  $R(t)$  are relatively small, the mean-field approximation may not be as accurate as expected. In those cases, one might have to resort to stochastic epidemic models, such as Markov chains.

---

### ALGORITHM 1: Tracking Discrete Time Time-dependent SIR Model

---

- Input:**  $\{X(t), R(t), 0 \leq t \leq T - 1\}$ , Regularization parameters  $\alpha_1$  and  $\alpha_2$ , Order of FIR filters  $J$  and  $K$ , Prediction window  $W$ .
- Output:**  $\{\beta(t), \gamma(t), 0 \leq t \leq T - 2\}$ ,  $\{\hat{\beta}(t), \hat{\gamma}(t), t \geq T - 1\}$ , and  $\{\hat{X}(t), \hat{R}(t), t \geq T\}$ .
- 1: Measure  $\{\beta(t), \gamma(t), 0 \leq t \leq T - 2\}$  using (11) and (10) respectively.
  - 2: Train the ridge regression using (14) and (15).
  - 3: Estimate  $\hat{\beta}(T - 1)$  and  $\hat{\gamma}(T - 1)$  by (12) and (13) respectively.
  - 4: Estimate the number of infected persons  $\hat{X}(T)$  and recovered persons  $\hat{R}(T)$  on the next day  $T$  using (16) and (17) respectively.
  - 5: **while**  $T \leq t \leq T + W$  **do**
  - 6:   Estimate  $\hat{\beta}(t)$  and  $\hat{\gamma}(t)$  in (12) and (13) respectively.
  - 7:   Predict  $\hat{X}(t + 1)$  and  $\hat{R}(t + 1)$  using (18) and (19) respectively.
  - 8: **end while**
- 

### III. THE SIR MODEL WITH UNDETECTABLE INFECTED PERSONS

According to the recent report from WHO [2], only 87.9% of COVID-19 patients have a fever, and 67.7% of them have a dry cough. This means there exist asymptomatic infections. Recent studies in [7] and [14] also pointed out the existence of the asymptomatic carriers of COVID-19. Those people are unaware of their contagious ability, and thus get more people infected. The transmission rate can increase dramatically in this circumstance.

To take the undetectable infected persons into account, we propose the SIR model with undetectable infected persons in this section. We assume that there are two types of infected persons. The individuals who are detectable (with obvious symptoms) are categorized as type I infected persons, and the asymptomatic individuals who are undetectable are categorized as type II infected persons. For an infected individual, it has probability  $w_1$  to be type I and probability  $w_2$  to be type II, where  $w_1 + w_2 = 1$ . Besides, those two types of infected persons have different transmission rates and recovering rates, depending on whether they are under treatment or isolation or not. We denote  $\beta_1(t)$  and  $\gamma_1(t)$  as the transmission rate and the recovering rate of type I at time  $t$ . Similarly,  $\beta_2(t)$  and  $\gamma_2(t)$  are the transmission rate and the recovering rate for type II at time  $t$ .

#### A. The Governing Equations for the SIR Model with Undetectable Infected Persons

Now we derive the governing equations for the SIR model with two types of infected persons. Let  $X_1(t)$  (resp.  $X_2(t)$ ) be the number of type I (resp. type II) infected persons at time  $t$ . Similar to the derivation of (6), (7) in Section II-C, we assume that  $\{S(t) \approx n, t \geq 0\}$  in the initial stage of the epidemic



and split  $X(t)$  into two types of infected persons. We have the following difference equations:

$$X_1(t+1) - X_1(t) = \beta_1 X_1(t) w_1 + \beta_2 X_2(t) w_1 - \gamma_1 X_1(t), \quad (20)$$

$$X_2(t+1) - X_2(t) = \beta_1 X_1(t) w_2 + \beta_2 X_2(t) w_2 - \gamma_2 X_2(t), \quad (21)$$

$$R(t+1) - R(t) = \gamma_1 X_1(t) + \gamma_2 X_2(t), \quad (22)$$

where  $\beta_1, \beta_2, \gamma_1$ , and  $\gamma_2$  are constants. It is noteworthy that those constants can also be time-dependent as we have in Section II. However, in this section, we set them as constants to focus on the effect of undetectable infected persons. Rewriting (20) and (21) in the matrix form yields the following matrix equation:

$$\begin{bmatrix} X_1(t+1) \\ X_2(t+1) \end{bmatrix} = \begin{bmatrix} 1 + \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & 1 + \beta_2 w_2 - \gamma_2 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix},$$

where  $w_2 = 1 - w_1$ . Let  $\mathbf{A}$  be the transition matrix of the above system equations, i.e.,

$$\mathbf{A} = \begin{bmatrix} 1 + \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & 1 + \beta_2 w_2 - \gamma_2 \end{bmatrix}. \quad (23)$$

It is well-known (from linear algebra) such a system is stable if the spectral radius (the largest absolute value of the eigenvalue) of  $\mathbf{A}$  is less than 1. In other words,  $X_1(t+1)$  and  $X_2(t+1)$  will converge gradually to finite constants when  $t$  goes to infinity. In that case, there will not be an outbreak. On the contrary, if the spectral radius is greater than 1, there will be an outbreak, and the number of infected persons will grow exponentially with respect to time  $t$  (at the rate of the spectral radius).

To further examine the stability condition of such a system, we let

$$R_0 = w_1 \frac{\beta_1}{\gamma_1} + w_2 \frac{\beta_2}{\gamma_2}. \quad (24)$$

Note that  $R_0$  is simply the basic reproduction number of a newly infected person as an infected person can further infect on average  $\beta_1/\gamma_1$  (resp.  $\beta_2/\gamma_2$ ) persons if it is of type I (resp. type II) and that happens with probability  $w_1$  (resp.  $w_2$ ). In the following theorem, we show that there is no outbreak if  $R_0 < 1$  and there is an outbreak if  $R_0 > 1$ . Thus,  $R_0$  in (24) is known as the percolation threshold for an outbreak in such a model [11].

**Theorem 1.** *If  $R_0 < 1$ , then the spectral radius of  $\mathbf{A}$  in (23) is less than 1 and there is no outbreak of the epidemic. On the other hand, if  $R_0 > 1$ , then the spectral radius of  $\mathbf{A}$  in (23) is larger than 1 and there is an outbreak of the epidemic.*

**Proof.** (Theorem 1)

First, we note that  $\gamma_1$  and  $\gamma_2$  are recovering rates and they cannot be larger than 1 in the discrete-time setting, i.e., it takes at least one day for an infected person to recover. Thus, the matrix  $\mathbf{A}$  is a positive matrix (with all its elements being positive). It then follows from the Perron-Frobenius theorem that the spectral radius of the matrix is the larger eigenvalue of the  $2 \times 2$  matrix.

Now we find the larger eigenvalue of the matrix  $\mathbf{A}$ . Let  $\mathbf{I}$  be the  $2 \times 2$  identity matrix and

$$\tilde{\mathbf{A}} = \mathbf{A} - \mathbf{I}. \quad (25)$$

Then

$$\tilde{\mathbf{A}} = \begin{bmatrix} \beta_1 w_1 - \gamma_1 & \beta_2 w_1 \\ \beta_1 w_2 & \beta_2 w_2 - \gamma_2 \end{bmatrix}. \quad (26)$$

Let

$$z_1 = \beta_1 w_1 - \gamma_1 + \beta_2 w_2 - \gamma_2, \quad (27)$$

and

$$z_2 = \beta_1 w_1 \gamma_2 + \beta_2 w_2 \gamma_1 - \gamma_1 \gamma_2. \quad (28)$$

It is straightforward to show that the two eigenvalues of  $\tilde{\mathbf{A}}$  are

$$\lambda_1 = \frac{1}{2}(z_1 + \sqrt{z_1^2 + 4z_2}), \quad (29)$$

and

$$\lambda_2 = \frac{1}{2}(z_1 - \sqrt{z_1^2 + 4z_2}). \quad (30)$$

Note that  $\lambda_1 \geq \lambda_2$ . In view of (25), the larger eigenvalue of the transition matrix  $\mathbf{A}$  is  $1 + \lambda_1$ .

If  $R_0 < 1$ , we know that  $z_2 < 0$ ,  $w_1 \frac{\beta_1}{\gamma_1} < 1$ , and  $w_2 \frac{\beta_2}{\gamma_2} < 1$ . Thus, we have from (27) that  $z_1 < 0$ . In view of (29), we conclude that

$$\lambda_1 < \frac{1}{2}(z_1 + |z_1|) = 0.$$

This shows that  $1 + \lambda_1 < 1$  and the spectral radius of  $\mathbf{A}$  is less than 1.

On the other hand, if  $R_0 > 1$ , then  $z_2 > 0$  and we have from (29) that

$$\lambda_1 > \frac{1}{2}(z_1 + |z_1|) \geq 0.$$

This shows that  $1 + \lambda_1 > 1$  and the spectral radius of  $\mathbf{A}$  is larger than 1. ■

The relation between the system parameters and the phase transition will be shown in the next section.

## IV. NUMERICAL RESULTS

### A. Dataset

In this section, we analyze and predict the trend of COVID-19 by using our time-dependent SIR model in Section II and the SIR model with undetectable infected persons in Section III. For our analysis and prediction of COVID-19, we collect our dataset from the National Health Commission of the People's Republic of China (NHC) daily Outbreak Notification [1]. NHC announces the data as of 24:00 the day before. We collect the number of confirmed cases, the number of recovered persons, and the number of deaths from Jan. 15, 2020 to Mar. 2, 2020 as our dataset. The confirmed case is defined as the individual with positive real-time reverse transcription polymerase chain reaction (rRT-PCR) result. It is worth noting that in the Hubei province, the definition of the confirmed case has been relaxed to the clinical features since Feb. 12, 2020, while the other provinces use the same definition as before.

### B. Parameter Setup

For our time-dependent SIR model, we set the orders of the FIR filters for predicting  $\beta(t)$  and  $\gamma(t)$  as 3, i.e.,  $J = K = 3$ . The stopping criteria of the model is set to  $X(t) \leq 0$ . Since the numbers of infected persons before Jan. 27, 2020 are too small to exhibit a clear trend (which may contain noises), we only use the data after Jan. 27, 2020 as our training data for predicting  $\beta(t)$  and  $\gamma(t)$ .

We use the scikit-learn library [15] (a third-party library of Python 3) to compute the ridge regression. The regularization parameters of predicting  $\beta(t)$  and  $\gamma(t)$  are set to 0.03 and  $10^{-6}$  respectively. Since the transmission rate  $\beta(t)$  is nonnegative, we set it to 0 if it is less than 0. Then, we use Algorithm 1 to predict the trend of COVID-19.

### C. Time Evolution of the Time-dependent SIR Model

In Figure 1, we show the time evolution of the number of infected persons and the number of recovered persons. The circle-marked solid curves are the real historical data by Mar. 2, 2020, and the star-marked dashed curves are our prediction results for the future. The prediction results imply that the disease will end in 6 weeks, and the number of the total confirmed cases would be roughly 80,000 if the Chinese government remains their control policy, such as city-wide lockdown and suspension of works and classes.

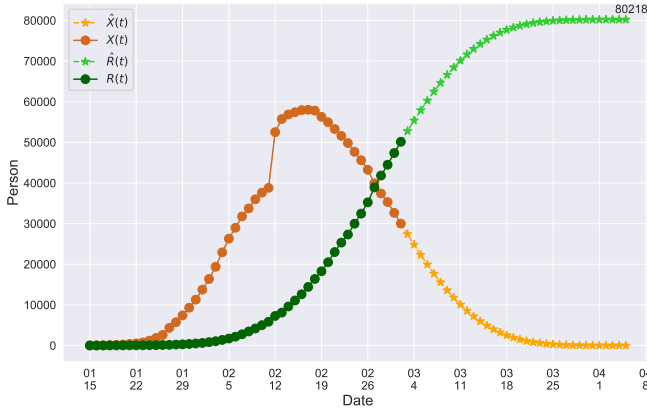


Figure 1: Time evolution of the time-dependent SIR model of the COVID-19. The circle-marked solid curve with dark orange (resp. green) color is the real number of infected persons  $X(t)$  (resp. recovered persons  $R(t)$ ), the star-marked dashed curve with light orange (resp. green) color is the predicted number of infected persons  $\hat{X}(t)$  (resp. recovered  $\hat{R}(t)$  persons).

In Figure 2, we show the measured  $\beta(t)$  and  $\gamma(t)$  from the real historical data. We can see that  $\beta(t)$  decreases dramatically, and  $\gamma(t)$  increases slightly. This is a direct result of the Chinese government that tries to suppress the transmission rate  $\beta(t)$  by city-wide lockdown and traffic halt. On the other hand, due to the lack of effective drugs and vaccines for COVID-19, the recovering rate  $\gamma(t)$  grows relatively slowly. Additionally, there is a definition change of the confirmed case on Feb. 12, 2020 that makes the data related to Feb. 11, 2020 have no

reference value. We mark these data points for  $\beta(t)$  and  $\gamma(t)$  with the gray dashed curve.

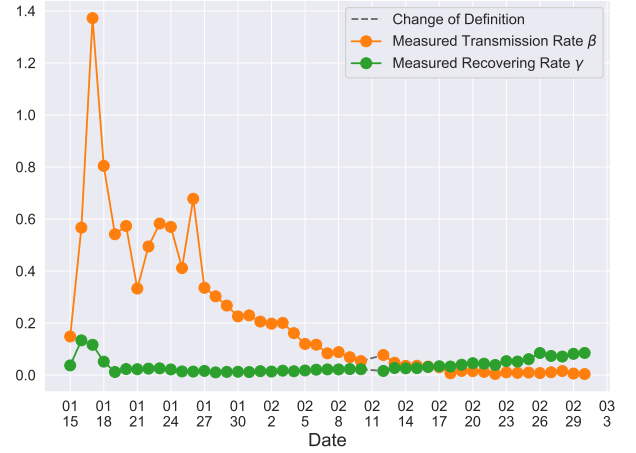


Figure 2: Measured transmission rate  $\beta(t)$  and recovering rate  $\gamma(t)$  of the COVID-19 from Jan. 15, 2020 to Feb. 19, 2020. The two curves are measured according to (11) and (10) respectively.

In an epidemic model, one crucial question is whether the disease can be contained and the epidemic will end, or whether there will be a pandemic that infects a certain fraction of the total population  $n$ . To answer this, one commonly used metric is the basic reproduction number  $R_0$  that is defined as the average number of additional infections by an infected person before it recovers. In the classical SIR model,  $R_0$  is simply  $\beta/\gamma$  as an infected person takes (on average)  $1/\gamma$  days to recover, and during that period time, it will be in contact with (on average)  $\beta$  persons. In our time-dependent SIR model, the basic reproduction number  $R_0(t)$  is a function of time, and it is defined as  $\beta(t)/\gamma(t)$ . If  $R_0(t) > 1$ , the disease will spread exponentially and infects a certain fraction of the total population  $n$ . On the contrary, the disease will eventually be contained. Therefore, by observing the change of  $R_0(t)$  with respect to time or even predicting  $R_0(t)$  in the future, we can check whether certain epidemic control policies are effective or not.

In Figure 3, we show the measured basic reproduction number  $R_0(t)$ , and the predicted basic reproduction number  $\hat{R}_0(t)$ . The blue circle-marked solid curve is the measured  $R_0(t)$  and the purple star-marked dashed curve is the predicted  $\hat{R}_0(t)$  (from Feb. 15, 2020). It is clear that  $R_0(t)$  has decreased dramatically since Jan. 28, 2020, and it implies that the control policies work in China. More importantly, it shows that the turning point is Feb. 17, 2020 when  $\hat{R}_0(t) < 1$ . In the following days after Feb. 17, 2020,  $X(t)$  will decrease exponentially, and that will lead to the end of the epidemic in China. Our model predicts precisely that  $R_0(t)$  will go less than 1 on Feb. 17, 2020 by 3 days in advance (Feb. 14, 2020). The results show that our model is very effective in tracking the characteristics of  $\beta(t)$  and  $\gamma(t)$ .

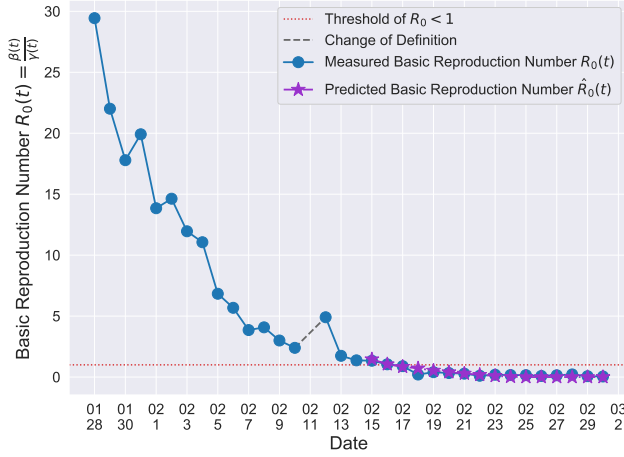


Figure 3: Basic reproduction number  $R_0(t)$  of the time-dependent SIR model of the COVID-19 in China. The circle-marked solid curve with blue color is the  $R_0(t)$  based on the given data from Jan. 27, 2020 to Feb. 20, 2020, the star-marked dashed curve with purple color is the predicted  $\hat{R}_0(t)$  based on the data from Jan. 27, 2020 to Feb. 15, 2020, and the dashed line with red color is the percolation threshold 1 for the basic reproduction number.

#### D. One-day Prediction

To show the precision of our model, we demonstrate the prediction results for the next day (one-day prediction) in Figure 4. It contains the predicted number of infected persons  $\hat{X}(t)$  (orange star-marked dashed curve), the predicted number of recovered persons  $\hat{R}(t)$  (green star-marked dashed curve), and the real number of infected and recovered persons (dark orange and dark green circle-marked solid curves) every day. The unpredictable days due to the change of the definition of the confirmed case on Feb. 12, 2020, are marked as gray. The predicted curves are extremely close to the measured curves (obtained from the real historical data). In this figure, we also annotate the predicted number of infected persons  $\hat{X}(t) = 27,433$  and the predicted number of recovered persons  $\hat{R}(t) = 52,785$  on Mar. 3, 2020.

We further examine our prediction accuracy in Figure 5. The error rates are all within  $\pm 3\%$  except for the predicted number of recovered persons  $\hat{R}(t)$  on Feb. 1, Feb. 3, and Feb. 5, 2020. The gray dashed curve stands for the unpredictable points due to the change of definition of the confirmed case. However, from the prediction results after Feb. 16, 2020, we find that our model can still keep tracking  $\beta(t)$  and  $\gamma(t)$  accurately and overcome the impact of the change of the definition.

#### E. Historical Data Analysis

In this subsection, we show the relationship between the epidemic prevention policies issued by the Chinese government and the historical data of the time-varying transmission rates.

As shown in Figure 2, the disease has been gradually controlled in China as time goes on. Excluding the small

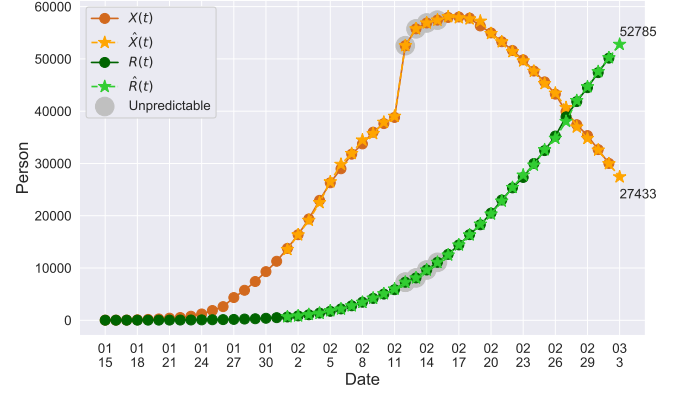


Figure 4: One-day prediction for the number of infected and recovered persons. The unpredictable points due to the change of definition of the confirmed case are marked as gray. The circle-marked solid curve with dark orange (resp. green) color is the real number of infected persons  $X(t)$  (resp. recovered persons  $R(t)$ ), the star-marked dashed curve with light orange (resp. green) color is the predicted number of infected persons  $\hat{X}(t)$  (resp. recovered persons  $\hat{R}(t)$ ).

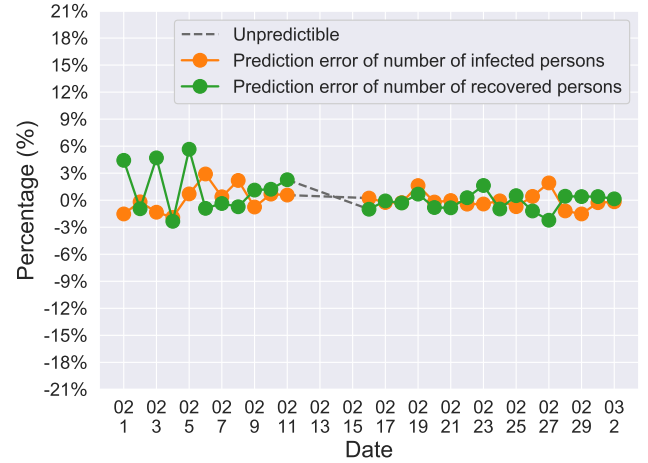


Figure 5: Errors of the one-day prediction of the number of infected and recovered persons. The unpredictable points due to the change of definition of the confirmed case on Feb. 12, 2020 are marked as the gray dash curve.

number of cases ( $X(t) < 500$ ) before Jan. 21, 2020 (that causes the curve to fluctuate a lot), it is notable that  $\beta(t)$  increases gradually then drops dramatically during Jan. 23, 2020 to Jan. 28, 2020, and it reaches the peak point on Jan. 26, 2020, which coincides with the trends of the moving out in Wuhan during the Chunyun (Spring Festival travel season) [16] in Figure 6. Especially, the emigration trend is almost the same as the  $\beta(t)$  during Jan. 23, 2020 to Jan. 28, 2020. We speculate that people rushed into the public transportation system when the announcement of the Wuhan city lockdown was out, which significantly increases the contact among people and speeds up the spread of the virus. As a result of that, the transmission rate  $\beta(t)$  increases substantially. As pointed out in Zhong's study [17], the median of the incubation period of COVID-19

is 3 days among 1099 valid confirmed cases, which makes the emigration trend aligns with  $\beta(t)$  if the extra 3 days are taken into account. Finally, the disease is gradually getting under control after the lockdown. The basic reproduction number  $R_0(t)$  is less than 1, i.e.,  $\beta(t) < \gamma(t)$  since Feb. 17, 2020.

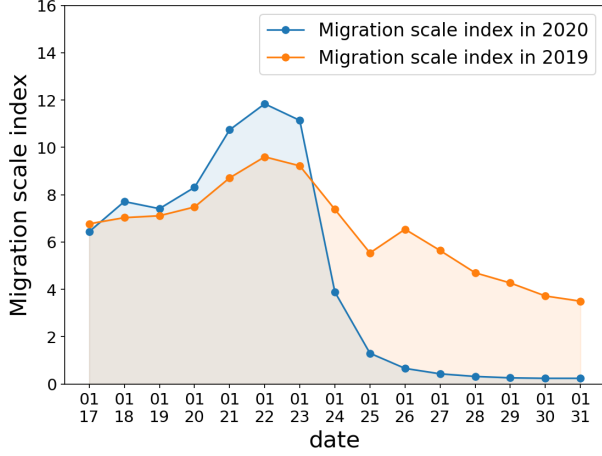


Figure 6: Trends of moving out during the Chunyun (Spring Festival travel season) in Wuhan city. The vertical axis represents the ratio between the number of people leaving the city and the resident population in Wuhan. The orange curve shows the ratio in 2019, while the blue curve shows the ratio in 2020. We redraw this figure by our-self from <https://qianxi.baidu.com/> [16].

#### F. The Effects of Type II Infected Persons

In this subsection, we show how undetectable (type II) infected persons affect the epidemic. In particular, we are interested in addressing the question of whether the existence of undetectable infected persons (type II) can cause an outbreak.

To carry out our numerical study, we need to fix some variables in the system of difference equations in (20)-(22). For the transmission rate (resp. recovering rate) of type I infected persons, i.e.,  $\beta_1$  (resp.  $\gamma_1$ ), we set it to be the measured  $\beta(t) = 0.00383$  (resp.  $\gamma(t) = 0.08493$ ) on Mar. 1, 2020 in China. The rationale behind this is that type I infected persons were detected and they were under treatment and isolation after Mar. 1, 2020 in China. Also, as there is no medicine for COVID-19, we may assume that these two types of infected persons have the same recovering rate, i.e.,  $\gamma_2 = \gamma_1$ . In view of the system equation in (20)-(22), there are still two free variables  $w_2$  and  $\beta_2$ , where  $w_2$  is the probability that an infected person is of type II and  $\beta_2$  is the transmission rate of type II.

In Figure 7, we illustrate how  $w_2$  and  $\beta_2$  affect the outbreak of the COVID-19. Such a figure is known as the phase transition diagram in [11]. The black curve in Figure 7 is the curve when the spectral radius of the transition matrix  $\mathbf{A}$  in (23) equals to 1. This curve represents the percolation threshold of COVID-19. If  $w_2$  and  $\beta_2$  fall above the black curve (in the orange zone), then there will be an outbreak. On the contrary, if  $w_2$  and  $\beta_2$  fall below the black curve (in the yellow zone), then there will not be an outbreak.

As shown in Figure 7, we would like to point out the importance of detecting an infected person. As long as more than 90% of those infected persons can be actually detected and properly isolated and treated, it is possible to contain the spread of the disease even if the transmission rate of type II infected persons, i.e.,  $\beta_2$ , is as high as 0.7. On the other hand, suppressing the transmission rate of type II infected persons can also be effective in controlling the disease while the detection rate is not that high. For example, wearing masks and washing hands can be an effective epidemic prevention mechanism to reduce  $\beta_2$ .

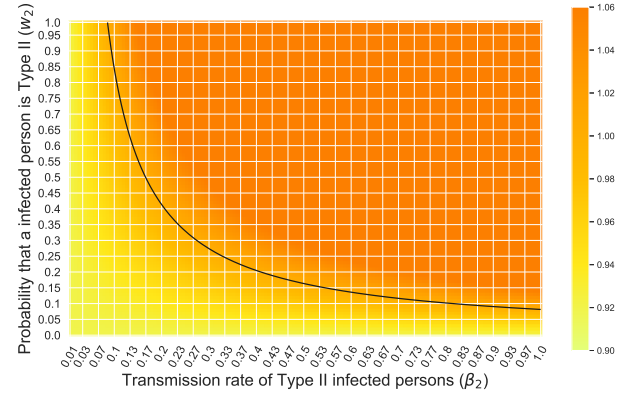


Figure 7: Phase transition diagram of an outbreak with respect to  $\beta_2$  and  $w_2$ . The black curve is the percolation threshold. The orange area means the disease will be an outbreak while the yellow area means the disease is under control.

In the following experiments, we extend our study to other countries, including Japan, Singapore, South Korea, Italy, and Iran. We collect the historical data from Jan. 22, 2020 to Mar. 2, 2020 from WHO [2]. For these countries, the transmission rates  $\beta(t)$  and the recovering rates  $\gamma(t)$  (measured from the time-dependent SIR model in Section II) during the initial period with a rapid increase of the number of confirmed cases can be viewed as  $\beta_2$  and  $\gamma_2$ . This is because during that period of time, there is no epidemic prevention intervention, and all the infected persons are basically not detected. It is interesting to note that different countries might have different  $\beta_2$  and  $\gamma_2$ . On the other hand, only 87.9% of COVID-19 cases have a fever from the report of WHO [2]. If we use body temperature as a means to detect an infected person, then only 87.9% of COVID-19 cases can be detected. For this, we set  $w_1 = 87.9\%$ .

With  $\beta_1$  and  $\gamma_1$  specified in the previous study for China, we plot the the phase transition diagram in Figure 8 in terms of the two variables  $\beta_2$  and  $\gamma_2$ . Again, the black curve is the curve when the spectral radius of the transition matrix  $\mathbf{A}$  in (23) equals to 1. Such a curve represents the percolation threshold of a COVID-19 outbreak. If  $\beta_2$  and  $\gamma_2$  fall above the black curve (in the orange zone), then there will be an outbreak. On the contrary, if  $\beta_2$  and  $\gamma_2$  fall below the black curve (in the yellow zone), then there will not be an outbreak. The countries with large confirmed cases, including Japan,



Singapore, South Korea, Italy, and Iran, are marked in Figure 8. From Figure 8, we observe that both Singapore and Japan are below the percolation threshold. On the other hand, both South Korea and Italy are above the percolation threshold, and they are on the verge of a potential outbreak on Mar. 2, 2020. These two countries must implement epidemic prevention policies urgently. Not surprisingly, on Mar. 10, 2020, the Italian government announces the lockdown and forbids the gatherings of people. It is worth mentioning that there are two marks for Iran in the Figure. The one above the percolation threshold is the one without adding the number of deaths into the number of recovered persons. The other one below the percolation threshold is the one that adds the number of deaths into the number of recovered persons. For some unknown reason, the death rate in Iran is higher than the other countries. The high death rate seems to prevent an outbreak in Iran.

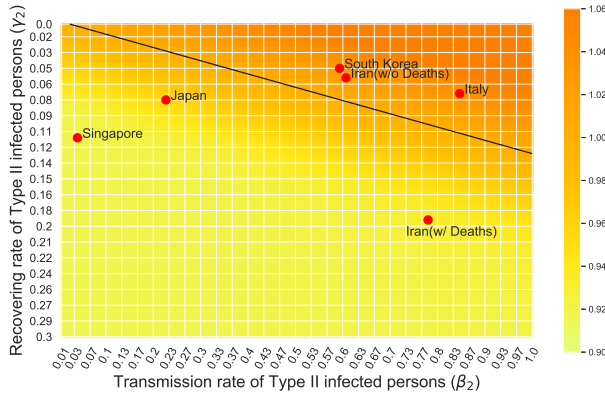


Figure 8: Phase transition diagram of an outbreak with respect to  $\beta_2$  and  $\gamma_2$ . The black curve is the percolation threshold. The orange area means the disease will be an outbreak, while the yellow area means the disease is under control.

## V. DISCUSSIONS AND SUGGESTIONS

In this section, we put forward some discussions and suggestions based on the observation made from the results of our system models. As of Mar. 2, 2020, it seems that COVID-19 has been gradually controlled in China since the prevention policies, such as city-wide lockdown have been issued in China. Although the policies such as traffic halt, small community management, and city-wide lockdown can effectively reduce the transmission rates  $\beta_1$  and  $\beta_2$ . However, these relatively extreme policies not only restrict the right of personal freedom but also affect the normal operation of society. These extreme policies forced several companies and factories to halt production, which impact all sectors of the economy. Therefore, to strike a balance between the prevention of disease and ensuring the normal operation of society is crucial, and it is important to suggest the so-called optimal control policies.

Certainly, the most effective way is to find anti-virus drugs or vaccines; however, it takes time. Hence, we should focus on controlling the transmission rate in this stage. Many countries

have developed comprehensive isolation systems to limit the spread of disease by confirmed cases, which can lower  $\beta_1$  effectively. Besides, from the experimental result in IV-F, managing, and controlling the asymptomatic infected persons are also important and urgent. Due to the limitation of the medical resource, we cannot put everyone under rRT-PCR test and detect all of the asymptomatic infected persons. However, we can track the travel history, occupation, contact, and cluster (TOCC) of the confirmed cases to narrow the range of the possible source of infection. In other words, by tracing the connection of social networks, it helps us to identify the most possible source of the infection. Taking care of the close contacts of the possible infection source can increase  $w_1$  (and thus decrease  $w_2$ ). Moreover, the propaganda of health education knowledge, wearing masks, washing hands, and other prevention policies can also reduce the transmission rate  $\beta_2$  of asymptomatic infected persons. We can observe the effectiveness of suppressing  $w_2$  and  $\beta_2$  to prevent an outbreak of the disease from Figure 7. Simultaneously, medical resources like the rRT-PCR test need to be allocated according to priorities. By doing so, the society can operate normally, which leads to the goal of optimal control. From the view of our system model, we regard the following four points as particularly important when preventing the disease from spreading:

1. Increasing the recovering rate  $\gamma$  can be helpful in controlling disease, but it depends on the standard of medical treatment, which we cannot control. Therefore, reducing the transmission rate is the first priority.
2. In addition to the rRT-PCR test, having a fever is a crucial symptom to identify the infected persons. Once confirmed, isolation is essential to avoid extra infection on society and lower the transmission rate  $\beta_1$ .
3. Asymptomatic infected persons may be identified by tracking the TOCC of detected cases and their social network connections. Quarantine of the persons who are suspected to be in contact with infected persons is also necessary in this stage to lower transmission rate  $\beta_2$ .
4. Propaganda of health education knowledge can reduce the transmission rate  $\beta_2$  substantially. For example, wearing masks in public and enclosed space, washing hands, avoiding touching your mouth, eyes, and nose, limiting the gatherings are good ways to not only protect ourselves from being infected by the asymptomatic infected persons but also avoid infecting others.

## VI. CONCLUSION AND FUTURE WORK

During the outbreak of COVID-19, we proposed the time-dependent SIR model that uses the FIR filters to simulate and predict the future trend of COVID-19. Our approach is different from the other time-invariant epidemic model in the literature. From the experimental results, we show that our model is able to predict the number of infected persons  $X(t)$  and the recovered persons  $R(t)$  within (almost) 3% of the error for the dataset collected from the National Health Commission of the People's Republic of China (NHC) [1]. Moreover, we are capable of tracking the characteristics of the transmission

rate  $\beta(t)$  and the recovering rate  $\gamma(t)$  with respect to time  $t$ , and precisely predict the future trend of the COVID-19 outbreak. Meanwhile, the effect of the Wuhan city lockdown can be observed in our model. Our model predicts precisely that  $R_0(t)$  will be less than 1 on Feb. 17, 2020 by 3 days in advance (Feb. 14, 2020). From our deterministic model, we estimate that COVID-19 will end in 6 weeks and infect roughly 80,000 people in China, given that the prevention policies in China remain unchanged. Moreover, considering the impact of asymptomatic infections in COVID-19, we proposed the SIR model with undetectable infected persons to analyze the percolation threshold for a COVID-19 outbreak when there are asymptomatic infections. In addition to our analysis for China, we further extended our study to other countries, including Japan, Singapore, South Korea, Italy, and Iran. Last but not least, based on the experimental results, some discussions and suggestions on epidemic prevention are proposed from the perspective of our models. In the future, we will extend our deterministic model by using stochastic models, such as the non-homogeneous Markov chain, to further improve the precision of the prediction results.

## REFERENCES

- [1] "Outbreak notification," Jan 2020. [Online]. Available: [http://www.nhc.gov.cn/xcs/yqtb/list\\_gzbd.shtml](http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml)
- [2] "Coronavirus disease (covid-19) outbreak," Jan 2020. [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [3] I. Nesteruk, "Statistics based predictions of coronavirus 2019-ncov spreading in mainland china," *MedRxiv*, 2020.
- [4] Y. Chen, J. Cheng, Y. Jiang, and K. Liu, "A time delay dynamical model for outbreak of 2019-ncov and the parameter identification," *arXiv preprint arXiv:2002.00418*, 2020.
- [5] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," *arXiv preprint arXiv:2002.06563*, 2020.
- [6] T. Zhou, Q. Liu, Z. Yang, J. Liao, K. Yang, W. Bai, X. Lu, and W. Zhang, "Preliminary prediction of the basic reproduction number of the wuhan novel coronavirus 2019-ncov," *Journal of Evidence-Based Medicine*, 2020.
- [7] B. F. Maier and D. Brockmann, "Effective containment explains sub-exponential growth in confirmed cases of recent covid-19 outbreak in mainland china," *arXiv preprint arXiv:2002.07572*, 2020.
- [8] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He *et al.*, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases*, 2020.
- [9] T. Zeng, Y. Zhang, Z. Li, X. Liu, and B. Qiu, "Predictions of 2019-ncov transmission ending via comprehensive methods," *arXiv preprint arXiv:2002.04945*, 2020.
- [10] Z. Hu, Q. Ge, L. Jin, and M. Xiong, "Artificial intelligence forecasting of covid-19 in china," *arXiv preprint arXiv:2002.07112*, 2020.
- [11] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [12] A. Y. Ng, A. X. Zheng, and M. I. Jordan, "Stable algorithms for link analysis," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 258–266.
- [13] B. S. Dayal and J. F. MacGregor, "Identification of finite impulse response models: methods and robustness issues," *Industrial & engineering chemistry research*, vol. 35, no. 11, pp. 4078–4090, 1996.
- [14] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens, "Estimating the generation interval for covid-19 based on symptom onset data," *medRxiv*, 2020.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [16] "Real-time big data report of covid-19 outbreak." [Online]. Available: <https://voice.baidu.com/act/newpneumonia/newpneumonia>
- [17] W.-J. Guan, Z.-Y. Ni, Y. Hu, W.-H. Liang, C.-Q. Ou, J.-X. He, L. Liu, H. Shan, C.-L. Lei, D. S. Hui *et al.*, "Clinical characteristics of 2019 novel coronavirus infection in china," *medRxiv*, 2020.