



Prediction of epidemics dynamics on networks with partial differential equations: A case study for COVID-19 in China

Ru-Qi Li(李汝琦), Yu-Rong Song(宋玉蓉), and Guo-Ping Jiang(蒋国平)

Citation: Chin. Phys. B, 2021, 30 (12): 120202. DOI: 10.1088/1674-1056/ac2b16

Journal homepage: <http://cpb.iphy.ac.cn>; <http://iopscience.iop.org/cpb>

What follows is a list of articles you may be interested in

Second order conformal multi-symplectic method for the damped Korteweg-de Vries equation

Feng Guo(郭峰)

Chin. Phys. B, 2019, 28 (5): 050201. DOI: 10.1088/1674-1056/28/5/050201

Improved control for distributed parameter systems with time-dependent spatial domains utilizing mobile sensor—actuator networks

Jian-Zhong Zhang(张建中), Bao-Tong Cui(崔宝同), Bo Zhuang(庄波)

Chin. Phys. B, 2017, 26 (9): 090201. DOI: 10.1088/1674-1056/26/9/090201

Group solution for an unsteady non-Newtonian Hiemenz flow with variable fluid properties and suction/injection

H. M. El-Hawary, Mostafa A. A. Mahmoud, Reda G. Abdel-Rahman, Abeer S. Elfeshawey

Chin. Phys. B, 2014, 23 (9): 090203. DOI: 10.1088/1674-1056/23/9/090203

Non-probabilistic solutions of imprecisely defined fractional-order diffusion equations

S. Chakraverty, Smita Tapaswini

Chin. Phys. B, 2014, 23 (12): 120202. DOI: 10.1088/1674-1056/23/12/120202

A meshless scheme for partial differential equations based on multiquadric trigonometric B-spline quasi-interpolation

Gao Wen-Wu, Wang Zhi-Gang

Chin. Phys. B, 2014, 23 (11): 110207. DOI: 10.1088/1674-1056/23/11/110207

Prediction of epidemics dynamics on networks with partial differential equations: A case study for COVID-19 in China*

Ru-Qi Li(李汝琦)¹, Yu-Rong Song(宋玉蓉)², and Guo-Ping Jiang(蒋国平)^{2,†}

¹School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

²College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

(Received 17 July 2021; revised manuscript received 13 September 2021; accepted manuscript online 29 September 2021)

Since December 2019, the COVID-19 epidemic has repeatedly hit countries around the world due to various factors such as trade, national policies and the natural environment. To closely monitor the emergence of new COVID-19 clusters and ensure high prediction accuracy, we develop a new prediction framework for studying the spread of epidemic on networks based on partial differential equations (PDEs), which captures epidemic diffusion along the edges of a network driven by population flow data. In this paper, we focus on the effect of the population movement on the spread of COVID-19 in several cities from different geographic regions in China for describing the transmission characteristics of COVID-19. Experiment results show that the PDE model obtains relatively good prediction results compared with several typical mathematical models. Furthermore, we study the effectiveness of intervention measures, such as traffic lockdowns and social distancing, which provides a new approach for quantifying the effectiveness of the government policies toward controlling COVID-19 via the adaptive parameters of the model. To our knowledge, this work is the first attempt to apply the PDE model on networks with Baidu Migration Data for COVID-19 prediction.

Keywords: partial differential equations, intervention measures, Baidu Migration Data, COVID-19 prediction

PACS: 02.30.Jr, 88.10.gc, 02.60.Ed, 05.10.-a

DOI: 10.1088/1674-1056/ac2b16

1. Introduction

The rapid development of complex network theory has effectively enhanced people's understanding of the transmission mechanisms of the outbreak of infectious diseases. There is a significant body of research on the spread of infection diseases on complex networks with various approaches including statistical and mathematical models to understand and predict infectious scales over a time period in complex networks.^[1–3] Mathematical modeling based on epidemiological characteristics is playing an important role in the research on the spread of infectious diseases.

Recently, in response to the current COVID-19 pandemic, scholars have proposed many epidemic models to analyze and predict the impact of COVID-19 in the outbreak area.^[4–7] For example, Tao *et al.*^[8] relied on the classical SIR model to evaluate the basic reproductive number of COVID-19 outbreaks in China and the United States. de Sousa *et al.*^[9] used SEIR model to simulate different intervention scenarios in Brasilia for predicting the growth trend of cases. The extended classical model also includes generalized growth models^[10] or age-stratified model.^[11] Although the smooth prediction curve produced by this type of ordinary differential equation (ODE) model can reflect the overall trend of COVID-19 cases, the fluctuation in a small range of cases under the overall increase

or decrease trend cannot be described in detail due to the factors of policy or human activities, which leads to the relatively low prediction accuracy.

In this paper, we propose a novel computational model for studying the spread of epidemic on networks based on partial differential equations (PDE). The proposed model captures epidemic diffusion along the edges of a network and extends the PDE models on social networks in Ref. [12] where PDE models are used to describe information diffusion and epidemic spread between clusters of the underlying networks. Many spatio-temporal models describe the transmission characteristics of infectious diseases by PDEs, and capture spatio-temporal information for infectious diseases.^[13–16] For example, Wang *et al.*^[17] applied the PDE model to analyze and predict regional COVID-19 in Arizona. The PDE model we develop in the present work focuses on the spread of COVID-19 across a city network and incorporates the factors of city traffic lockdowns and social distancing. The PDE model provides a new way for quantifying the effectiveness of the government policies toward controlling COVID-19 cases.

In addition, the difference between our model and the existing PDE models is that we use real-time human mobility data provided by Baidu,^[18] which includes the travel intensity within Wuhan and the scale of population migration from Wuhan to other cities in China. The use of population outflow

*Project supported by the National Natural Science Foundation of China (Grant Nos. 61672298, 61873326, and 61802155) and the Philosophy Social Science Research Key Project Fund of Jiangsu University (Grant No. 2018SJZDI142).

†Corresponding author. E-mail: jianggp@njupt.edu.cn

data from epidemic areas and epidemiological data can further demonstrate the role of population movement in the early spread of COVID-19 in various cities and how these control measures affect the spread from Wuhan to other cities. Although a large number of literatures have applied partial differential equations to the spatio-temporal propagation modeling of epidemiology, to our knowledge, this paper is the first attempt to combine the Baidu migration data and the PDE model to predict COVID-19 and evaluate how the local intervention has affected the spread of COVID-19.

2. COVID-19 outbreak in China

Large-scale population migration in the short term can spread from local outbreaks into widespread epidemics.^[19-22] On January 26, 2020, at a Wuhan press conference, the mayor stated that about 5 million people had left Wuhan due to the Spring Festival travel and epidemic factors.^[23] In order to control the development of the epidemic, China has implemented an unprecedented intervention strategy from January 23, 2020. The entire Wuhan City has been quarantined and strict travel restrictions have been implemented.^[24] Although epidemic prevention measures have been adopted in cities across China, a series of factors such as population movements, geographical and social characteristics may lead to different characteristics of the spread of COVID-19 in various cities.

Before introducing specific methods and assumptions, let us point out an important fact. The number of confirmed cases in various regions is largely affected by the outflow population before the closure of Wuhan. Table 1 shows the population outflow index of the top-50 cities with maximum outflow population from Wuhan during the two weeks before the closure and the cumulative number of confirmed cases as of February 25, 2020. Figure 1 shows the correlation between the population outflow index and the cumulative number of confirmed cases in the 50 cities. It can be seen that the relationship between them is almost linear, and the corresponding Pearson correlation coefficient is as high as 0.972.

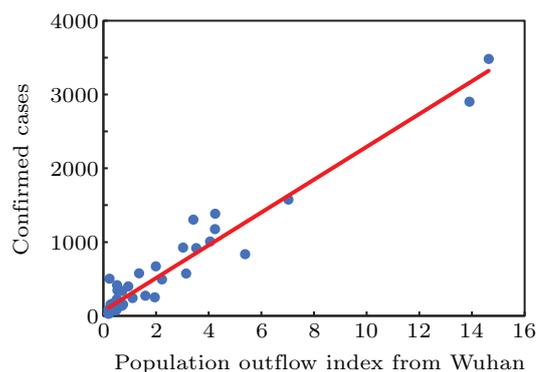


Fig. 1. The relationship between the cumulative number of confirmed cases in the top-50 cities with maximum population outflow index from Wuhan and corresponding population outflow index. Cities are represented by data points and the linear fitting is represented by the red line ($R^2 = 0.945$).

Table 1. The six columns respectively show the names of the top-50 cities with maximum outflow population from Wuhan during the two weeks before the closure, the population outflow index, the cumulative number of confirmed cases as of February 25, 2020.

City name	Outflow index	Confirmed cases	City name	Outflow index	Confirmed cases
Xiaogan	14.632	3482	Shenzhen	0.522	417
Huanggang	13.905	2904	Nanchang	0.513	229
Jingzhou	7.029	1576	Chengdu	0.505	143
Xianning	5.385	836	Anqing	0.492	83
Ezhou	4.248	1385	Zhoukou	0.473	76
Xiangyang	4.235	1175	Hefei	0.434	174
Huangshi	4.051	1008	Fuyang	0.366	169
Jingmen	3.526	920	Xian	0.363	120
Suizhou	3.418	1305	Changde	0.358	82
Xiantao	3.142	575	Shangqiu	0.354	91
Yichang	3.026	926	Nanjing	0.310	93
Tianmen	2.226	495	Yichun	0.282	106
Shiyan	1.991	671	Hangzhou	0.275	155
Enshi	1.951	251	Hengyang	0.258	48
Xinyang	1.593	274	Kunming	0.246	53
Chongqing	1.358	576	Wenzhou	0.230	504
Changsha	1.107	242	Liuan	0.222	69
Beijing	0.944	400	Quanzhou	0.218	47
Nanyang	0.744	155	Fuzhou	0.215	72
Shanghai	0.715	336	Nanning	0.206	55
Zhumadian	0.707	139	Luoyang	0.205	31
Zhengzhou	0.642	157	Guiyang	0.202	36
Jiujiang	0.563	118	Suzhou	0.201	87
Yueyang	0.557	156	Luohe	0.197	35
Guangzhou	0.531	346	Shijiazhuang	0.190	29

To further verify the ability of population mobility to predict the growth of COVID-19 cases, we conducted the Granger causality test. The significance of Granger causality test is to test whether the lag value (past information) of a variable has predictive effect on the explained variable. Considering the lagging effect of population mobility on case growth, we use the daily population outflow from Wuhan two weeks before Wuhan closure on January 23, 2020 and the confirmed cases summed across top-50 cities with maximum population outflow from Wuhan two weeks after Wuhan closure. Unit root test shows that the two time series are stationary. Select the default lag period of 2 for Eviews 8 software, and the results of Granger causality test are shown in Table 2. It can be seen that the concomitant probability of population migration does not Granger cause cases is less than 0.05, so it rejects the null hypothesis, that is, population migration is the Granger causality of cases, and population migration data has the ability to predict the confirmed case data, which also verifies the rationality of the PDE prediction model driven by population flow.

Table 2. Granger causality test results.

Pairwise Granger causality tests		
Null hypothesis	<i>p</i> -value	Decision
Population migration does not Granger cause cases	0.0428	Reject
Cases does not Granger cause population migration	0.0228	Reject

Based on the above observations, we focus on the China city network and choose eight cities with different population flows centered on Wuhan and use the proposed model to predict the spread of COVID-19 at a city level. In terms of China's per capita COVID-19 infection rate, Hubei province has a higher per capita infection rate. As shown in Fig. 2, we select Xiaogan and Jingzhou in Hubei province close to Wuhan. In the southern part of Hubei, we select Changsha and Guangzhou, which locate at a long distance away from Wuhan; similarly, in the eastern part of Hubei, Nanjing and Shanghai are chosen; in the western part, Chongqing is chosen; in the north part of China, Beijing is selected. Note that we are focusing on the overall flow of people in each city rather than individual tracking.

We use the city-level COVID-19 data provided by the China Data Lab on Harvard Dataverse for some time.^[25] To smooth the data, we use the seven-day rolling average of the cumulative city level data. In addition, we also incorporate the Baidu Migration Index into our model, which includes the Wuhan migration index, the traffic intensity within Wuhan, and the proportion of the population outflow index from Wuhan to each city from January 10, 2020 to March 15, 2020. Then, we calculate the daily change of input from Wuhan for each city by multiplying the Wuhan migration index by the proportion of moving in destinations from Wuhan. In the days before the Lunar New Year and the establishment

of quarantine measures, the flow of people leaving Wuhan began to increase and reached a peak on the eve of the Lunar New Year on January 24, 2020, and then quickly reduced to almost no activity. The travel ban seems to have blocked the movement in and out of Wuhan to celebrate the Chinese New Year and has minimized the further spread of SARS-COV-2 exported from Wuhan.

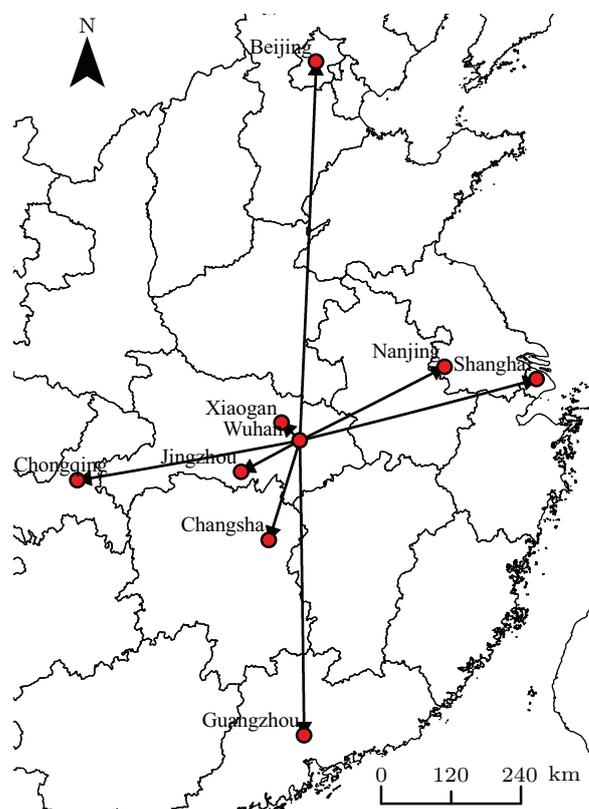


Fig. 2. Geographical distribution of nine cities.

3. A new prediction framework of epidemic on networks

The efficiency, speed, and coverage of modern transportation networks play a significant role in the spread of epidemics. As a result, we focus on the spread of epidemics along the edges of a network which often represent roads or traffic routes.

From a mathematical point of view, the spread of infectious diseases can be modeled by diffusive differential equations. Given a network can be presented by a graph. To be more precise we consider a finite, simple, connected graph G with m edges e_1, \dots, e_m and n vertices v_1, \dots, v_n . The edge j is isomorphic to the interval $[0, 1]$ and spatial variable x on it, where $x = 0, 1$ represent the two ends. Here we only consider undirected graphs and the two ends of an edge can be either 0 or 1. To capture the spread of epidemic between nodes on the network, a system of diffusion equations are proposed to predict the infection on the networks. On each edge

$e_j, j = 1, \dots, m$, we have

$$\begin{cases} \frac{\partial u_j}{\partial t}(t, x) = c_j(t) \frac{\partial^2 u_j}{\partial x^2}(t, x) + F_j(t, x, u), \\ u_j(x, 0) = \psi_j(x), \quad 0 < x < 1, \\ |u_j(t, \cdot) - u_i(t, \cdot)| \leq \varepsilon, \\ \frac{\partial u_j}{\partial x}(t, \cdot) = 0, \quad t > 1. \end{cases} \quad (1)$$

If a vertex is connected to multiple edges, $u(t, \cdot)$ at the vertex is the average of $u(t, \cdot)$ computed from the edges. Here we think of $u_j(t, x)$ as a function defined on the j -th edge e_j which indicates the infection level edge $e_j, j = 1, \dots, m$. and impose that, for each $t \geq 0$, $(u_1(t, \cdot), \dots, u_m(t, \cdot))$ is continuous on the graph, i.e., that the values of $u_j(t, \cdot)$ are compatible at the vertices if we assume that the constant ε is sufficient small. Here $F_j(t, x, u)$ can be viewed as local growth of the infection and $\psi_j(x)$ is the initial function on each edge. The existence and uniqueness of such evolutionary systems on networks have been studied by many authors.^[26,27]

From a computational point of view, model (1) is computationally efficient as it allows us to compute each edge concurrently. Model (1) is also an extension of PDE models on social networks in Ref. [12] where the entire networks or some clusters are embedded into a Euclidean space and PDE models are used to describe information diffusion and epidemic spread between clusters of the underlying networks. If a network exhibits significant clusters, then model (1) can be applied to its cluster networks for taking the advantage of spatio-temporal modeling on networks. In this paper, we will use the spatio-temporal modeling framework to examine the outbreak of COVID-19 at 9 major cities in China.

4. PDE model for China city network

Based on the prediction framework (1), we propose a specific PDE model to analyze the spatio-temporal dynamics of COVID-19 cases in the China city network. We now focus on each edge which links the Wuhan and other eight cities as shown in Fig. 3 where only the edge between Wuhan and Xiangan is shown. The spatio-temporal model can describe the two transmission processes of COVID-19: the local process (local cases caused by the community transmission within a city) and the global process (imported cases caused by cross-border transmission between cities).

Let $C(x, t)$ represent the number of the COVID-19 cases in a China city x at time t . Through the combination of the two processes in Fig. 3, the differential dynamic equation of the COVID-19 prediction model is constructed as follows:

$$\begin{cases} \frac{\partial C(x, t)}{\partial t} = \frac{\partial}{\partial x} \left(d(x) \frac{\partial C(x, t)}{\partial x} \right) \\ \quad + r(t)l(x)a(x, t - 10)C(x, t) - c \frac{C(x, t)}{k + C(x, t)}, \\ C(x, 1) = \psi(x), \quad 0 < x < 1, \\ \frac{\partial C}{\partial x}(0, t) = \frac{\partial C}{\partial x}(1, t) = 0, \quad t > 1, \end{cases}$$

The term $\frac{\partial}{\partial x} \left(d(x) \frac{\partial C(x, t)}{\partial x} \right)$ describes the transborder diffusion of COVID-19 cases between different China cities, where $d(x)$ is the transmission rate of COVID-19 across different cities. $\frac{\partial}{\partial x} \left(d(x) \frac{\partial C(x, t)}{\partial x} \right)$ has been widely used for indicating the spatial transmission of epidemics.^[13,28] Here we assume $d(x)$ to be constant, i.e., $d(x) \equiv \exp(-d_0) > 0$.

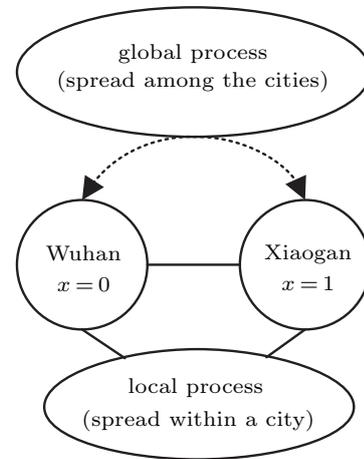


Fig. 3. Spread of epidemic between one of the edges of the China city network.

Recall that the mathematical expression of population growth in Ref. [28], such as bacteria and tumors, $r(t)l(x)a(x, t - 10)C(x, t)$ denotes the local diffusion caused by the second generation reproduction of infected cases within each China city x at time t .

(i) The $r(t) > 0$ is the infection rate that changes with time t within a city, and it increases with the increase of confirmed cases $C(x, t)$. Therefore, $r(t)$ can be expressed as the function $r(t) = g(b_0 + b_1 t)$ and $g(u) = 1/(1 + \exp(-u))$ with parameters $b_0 > 0, b_1 > 0$ for describing the pattern.

(ii) The $l(x)$ is a dimensionless location function that indicates different cities in China's city network, and it is built through a cubic spline interpolation, which satisfies $l(x_i) \equiv l_i, i = 0, 1$, where x_i represents city i .

(iii) The $a(x, t - 10)$ is the population outflow data from Wuhan to Xiaogan (or one of the China cities) when $x = 1$, which will lead to the increase of COVID-19 cases. $t - 10$ represents the lag period in which the population movement affects the epidemic. Here we assume that it takes ten days for the effect of population movement on changes in confirmed cases, which covers the incubation period of most cases

simultaneously.^[29,30] When $x = 0$, $a(x, t - 10)$ is the traffic intensity within Wuhan.

The term $\frac{C(x,t)}{k+C(x,t)}$ is the Michaelis–Menten function, which approaches 1 as $C(x,t)$ increases; $c > 0$ denotes the maximum decay rate of COVID-19 cases due to human efforts. k is the number of COVID-19 cases at which the decay rate is $c/2$. Actually, this type of function is widely used to describe human behavior changes such as wearing masks and keeping social distancing to curb epidemics in Ref. [31]. As a result, we can analyze the effect of personal preventive measures on the potential reduction of COVID-19 via adjusting the parameter c .

Here d_0 , k , c and parameters of $r(t)$, $l(x)$ are determined by the collected COVID-19 data with the best fitting function of Matlab.

We focus on the spread of COVID-19 on the edges between Wuhan and one of the eight cities. For simplicity, Neumann boundary condition $\frac{\partial C}{\partial x}(0, t) = \frac{\partial C}{\partial x}(1, t) = 0$, $t > 1$ is applied in Ref. [27].

Initial function $\psi(x)$ represents the initial counts of COVID-19 cases between Wuhan and one of the eight cities,

i.e., $\psi(x) = C(x, 1)$, which can be constructed from the historical case data by cubic spline interpolation.

5. Prediction process and evaluation

To forecast the COVID-19 of a given day, we use population mobility data and historical COVID-19 case data from nine cities to solve the inversion problem of PDE model (2). In the current work, we use the training data set of seven days for predicting the COVID-19 cases of the following day, i.e., days 1–7, 2–8, ..., are used as the training data, and we predict the number of COVID-19 cases for the following days 8, 9, ..., respectively. Specifically, we adopt a hybrid approach: First a tensor train global optimization^[32] is used to search the parameter space for locating the starting points, and then Nelder–Mead simplex method^[33] is used to further explore the local optimization, where the Nelder–Mead simplex method depends on the `fminsearch` function in MATLAB (version R2020a). Once the model parameters are determined, we apply the fourth-order Runge–Kutta method to solve the PDE for one-step forward prediction in time dimension numerically.

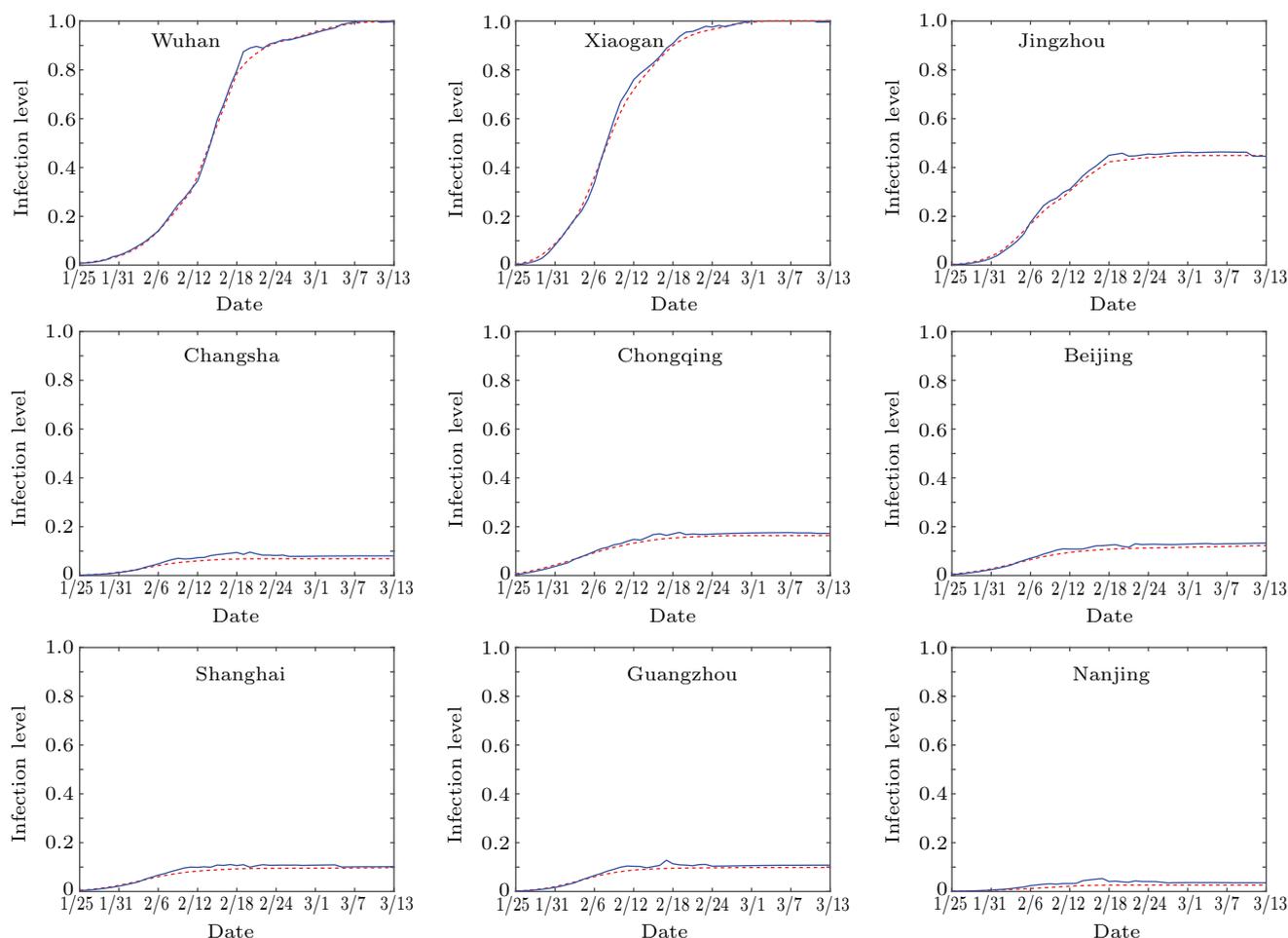


Fig. 4. Predicted versus actual cases growth in nine cities from January 25, 2020 to March 13, 2020. Blue lines: predicted data. Red lines: actual data.

Figure 4 shows the cumulative COVID-19 case prediction for nine cities in China from January 25, 2020 to March 13, 2020, and the data is normalized to between [0,1]. Due to the large gap in case numbers between Wuhan and other cities, Wuhan's data is based on the maximum value of Wuhan, and the other eight cities are based on the maximum value of Xiaogan. Obviously, there is a consistent trend between the predicted cases and the actual data. The estimated parameters are different for each time step. Using the bootstrap method, here we only provide the population mean, confidence interval and *p*-value of the estimated parameters in the last prediction for the nine city on March 13, as shown in Table 3.

Table 3. Parameter estimates of the PDE model on March 13, 2020.

Parameter	Mean	95%CI	<i>p</i>
<i>d</i> ₀	4.42290	4.02015–4.69566	<0.01
<i>k</i>	2.74189	1.70209–3.98324	<0.05
<i>b</i> ₀	0.50123	0.49978–0.50390	<0.01
<i>b</i> ₁	0.24995	0.24991–0.24998	<0.01
<i>c</i>	0.76990	0.71852–0.81780	<0.01
<i>l</i> ₀	0.48102	0.35160–0.58744	<0.01
<i>l</i> ₁	0.17872	0.11805–0.25193	<0.01

As we can see from Fig. 4, after two key dates, January 23 (Wuhan was quarantined) and January 24 (Lunar New Year's Eve, outbound holiday travel is typically completed before this evening), the confirmed cases of COVID-19 in several cities began to increase rapidly and eventually reached saturation. On February 11, the COVID-19 control headquarters of Wuhan made an announcement to further strengthen source control and minimize the flow of people. As a result, the city's communities were closed for management from that time. About a week later, on February 18, we can see the effect of this measure. The number of confirmed cases had changed from an exponential increase to a logarithmic increase, and the growth rate had slowed down significantly.

Figure 4 also indicates that Xiaogan has the largest case number next to Wuhan. On February 5, Xiaogan, which had a population of more than 2 million less than Huanggang, surpassed Huanggang's cumulative number of COVID-19 confirmed cases, becoming second in China, second only to Wuhan. The investigation found 11524 feverish persons, 4509 of whom returned from Wuhan. As of January 31, among the confirmed and suspected cases, returnees from Wuhan accounted for 60% and 66%, respectively. On February 14, the Xiaogan Municipal Government decided to comprehensively upgrade the prevention and control measures throughout the city: all communities would be fully closed for management. After the implementation of this measure lasted for about a week, the number of confirmed cases gradually approached saturation. Although the permanent population of Jingzhou is 650000 more than that of Xiaogan, the inflow population from Wuhan in Jingzhou is only half of that in Xiaogan. This

may be the main reason why Jingzhou's infection level did not reach the top line. Cities in other provinces have basically maintained the infection level below 0.2, which may be the common result of lower inflow from Wuhan, staying at home, and wearing masks.

We evaluate the prediction performance of the proposed model via comparing predicted cases with actual cases in nine cities. Here the prediction accuracy is defined as $A = 1 - |x_{\text{real}} - x_{\text{predict}}|$, where x_{real} is the actual COVID-19 cases at every time point and x_{predict} is the predicted cases. The average prediction accuracy of nine cities is above 95%. Obviously, the PDE prediction model based on population movement has good prediction capabilities during the active COVID-19 epidemic.

6. Prediction comparison of mathematical models

6.1. Statistical model

In this section, we compare the stochastic PDE model and a statistical ARIMA model^[34] which is most suitable for univariate time series modelling. The ARIMA model contains three parameters, namely, *p*, *d* and *q*, where *p* is the lag of the time series used in the prediction model, *d* is the difference number when the time series is stationary, and *q* is the lag of the prediction error. Mathematically the formula for ARIMA could be represented as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (2)$$

where y_t and ε_t are the actual data and the error terms at time *t*, ϕ_i and θ_i ($i = 1, 2, 3, \dots$) are model coefficients, and *c* is a constant.

To facilitate model comparison, we use daily new cases from January 20, 2020 to February 29, 2020 summed across the above 8 cities excludes Wuhan for predicting new cases one day ahead. The augmented Dicky Fuller test (ADF) shows that the time series is stationary after the first-order difference. The parameters *p* and *q* can be determined via the autocorrelation function (ACF) and partial autocorrelation function (PACF). The coefficients ϕ_i and θ_i of model (2) are evaluated via the least square method using Eviews 8 software, and the optimal model is selected in accordance with the relative minimum of AIC (Table 4). Therefore, the ARIMA(1,1,1) model can describe the time series.

Table 4. Candidate statistical models with different order *p* and *q*. The chosen model is shown in bold face.

Model	ARIMA(2,1,2)	ARIMA(2,1,1)	ARIMA(1,1,2)	ARIMA(1,1,1)
AIC*	12.038	12.015	11.985	11.935

*Akaike information criterion.

6.2. ODE model

Ordinary differential system has been widely used in the study of epidemic spread. Here we evaluate the prediction results by fitting an SEIR model^[35] to the total number of daily new confirmed cases from eight cities. The model is shown as

$$\begin{cases} \frac{dS(t)}{dt} = -\frac{\beta S(t)(E(t)+I(t))}{N}, \\ \frac{dE(t)}{dt} = \frac{\beta S(t)(E(t)+I(t))}{N} - \sigma E(t), \\ \frac{dI(t)}{dt} = \sigma E(t) - \gamma I(t), \\ \frac{dR(t)}{dt} = \gamma I(t), \end{cases} \quad (3)$$

where S , E , I , and R are the number of susceptible, exposed (latent), infectious, and removed individuals on day t . N is the total population in eight cities. β , σ , and γ are the transmission rate, the daily probability of an exposed individual becoming infectious, and the daily probability of recovery or death of an infected individual. Prior estimates of β , σ , and γ are derived from epidemiological surveys^[32] and the values are 0.6, 0.163, and 0.215, respectively. The initial number of infected is $I(0) = 1$. Model fitting is performed using the Monte Carlo method with the MATLAB (version R2020a).

6.3. PDE model with noise interference

In social network, the disturbance of infectious disease dynamics mainly comes from the change of network topology due to human activities. Considering the influence of environmental fluctuation, we establish the stochastic model by introducing the additive noise term into the population mobility-driven PDE. Here we assume that the daily migration index of city x is interfered and the interference could be described as additive noise. Therefore, we introduce randomness into the deterministic model (2) by perturbing population mobility item $a(x, t - 10)$ by $a(x, t - 10) + \sigma \zeta(x, t - 10)$ to capture the following stochastic PDE model:

$$\begin{cases} \frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x} \left(d(x) \frac{\partial C(x,t)}{\partial x} \right) + r(t)l(x)(a(x,t-10) + \sigma \zeta(x,t-10))C(x,t) - c \frac{C(x,t)}{k+C(x,t)}, \\ C(x,1) = \psi(x), \quad 0 < x < 1, \\ \frac{\partial C}{\partial x}(0,t) = \frac{\partial C}{\partial x}(1,t) = 0, \quad t > 1, \end{cases} \quad (4)$$

where $\sigma \zeta(x, t - 10)$ is a Gaussian white noise, which is characterized by

$$\langle \zeta(t - 10) \rangle = 0, \quad \langle \zeta(t - 10) \zeta(t' - 10) \rangle = \delta(t - t'),$$

and $\langle \cdot \rangle$ represents the overall average, $\delta(\cdot)$ is the Dirac- δ function, σ is the noise intensity which describes the stability of population mobility, and $\sigma = 1$.

Meanwhile, we use the daily new cases data from January 20, 2020 to February 29, 2020 summed across the above 8 cities excludes Wuhan city and the corresponding population outflow from Wuhan to train our PDE model with noise interference. To ensure the robustness of the conclusions, we also evaluate bootstrap estimates of the population mean, confidence interval and p -value of the model parameters (Table 5).

Table 5. Parameter estimates of the PDE model.

Parameter	Mean	95% CI	p
d_0	2.5093	2.4859–2.5410	< 0.01
k	51.1013	50.7522–51.4327	< 0.01
b_0	0.5004	0.4929–0.5077	< 0.01
b_1	0.2530	0.2501–0.2561	< 0.01
c	0.4134	0.3389–0.4931	< 0.01
l_0	0.5445	0.4824–0.6056	< 0.01
l_1	0.5093	0.5005–0.5207	< 0.01

6.4. Results comparison

Figure 5 shows the prediction curves of several mathematical models from January 20, 2020 to February 29, 2020. As seen from Table 6, the prediction performance of PDE model is better than ARIMA model in the view of mean absolute error (MAE) and mean absolute percent error (MAPE), but is slight worse in the view of root mean squared error (RMSE). Compared with ODE model, the prediction performance of PDE model is better in the view of RMSE and MAE. Significantly, our model is in the social network to naturally describe the relative distribution of cases across different geographic areas over time, which is fundamentally driven by the population outflow from Wuhan. Secondly, the proposed model involves community transmission within a city, cross-border transmission between cities and variable human behavior, which well simulates the impact of government intervention and personal prevention on COVID-19 cases, as shown in the next section.

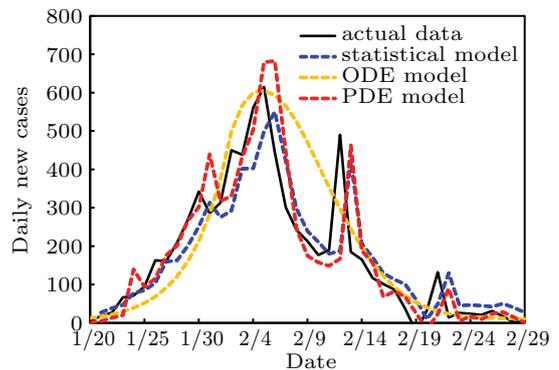


Fig. 5. Predictions of COVID-19 daily new cases from January 20, 2020 to February 29, 2020.

Table 6. The statistical comparison of the predicted results of mathematical models.

Prediction results	Statistical model	ODE model	PDE model
RMSE	87.68	107.35	93.17
MAE	60.99	72.40	55.86
MAPE	138.20	67.45	73.97

7. Effects of lockdown and social distancing

To measure the effectiveness of interventions such as lockdown and social distancing on reducing COVID-19, we predict the COVID-19 cases with a decreased a -value which represents the effect of the traffic within Wuhan and the out-flow from Wuhan, and an increased c -value which represents the effect of social distancing. It is general believed that sparse traffic flow and appropriate social distancing can effectively reduce COVID-19 cases. Specifically, based on the fact that the parameters in the model (2) remain unchanged, we multiply the a -value by 0.2 and the c -value by 2, which means that reduction of traffic flow from Wuhan by 80% and each city

add double effort for social distancing.

Figure 6 describes COVID-19 case prediction under measures of lockdown and social distancing in Wuhan, Xiaogan and Jingzhou from January 25, 2020 to March 13, 2020, and demonstrates that the final confirmed cases under such double interventions may be 10% to 19% less than the actual cases. Therefore, our results validate that active intervention can effectively inhibit the growth of COVID-19 and curb the wider spread of emerging pathogens. In fact, China’s multifaceted response may have prevented worse situation from happening, which also provides useful information for the rest of the world.

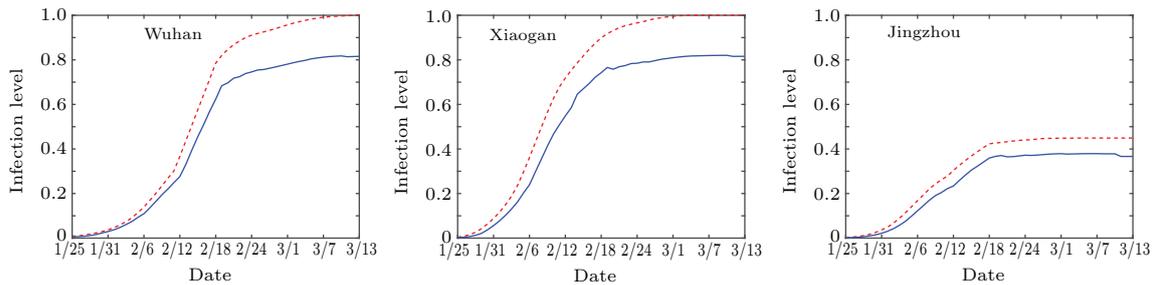


Fig. 6. Effect of interventions on case growth in three cities of Hubei province from January 25, 2020 to March 13, 2020. Blue lines: predicted cases. Red lines: actual cases.

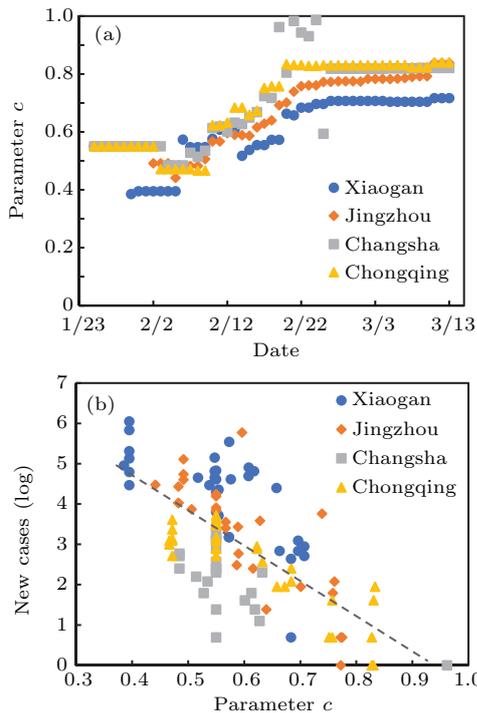


Fig. 7. The relationship between parameter c and new cases by city from January 25, 2020 to March 13, 2020.

Figure 7 illustrates the relationship between estimated parameter c and new confirmed cases for Xiaogan, Jingzhou, Changsha and Chongqing from January 25, 2020 to March 13, 2020. We see that the c -value of Fig. 7(a) for the four cities increases over time and eventually tends to saturated, which indicates that personal prevention awareness is constantly improving. Therefore, the relationship between the daily new

cases and the parameter c in the four cities in Fig. 7(b) shows a evident negative correlation, and the Pearson coefficients are -0.706 , -0.786 , -0.667 and -0.864 , respectively. In fact, these parameters provide a new way for quantifying the effectiveness of the government policies toward controlling COVID cases. For example, the curve of Fig. 7(a) for Xiaogan is lower than that of Jingzhou, Changsha and Chongqing, which indicates that the mitigation policies at Xiaogan is less effective than others. This could be one of the reasons why Xiaogan has more COVID-19 cases than the other cities.

Unfortunately, on July 20, 2021, Nanjing broke the “zero increase” of local COVID-19 cases, and the strain that caused the Nanjing epidemic was Delta virus.^[36] The continuous variation of SARS-CoV-2 results in various transmission characteristics. For this we estimate the characteristics of the Delta virus via fitting PDE model to confirmed cases in Nanjing from July 20, 2021 to August 12, 2021. As shown in Table 7, the results suggest that compared with the initial epidemic wave in Nanjing in early 2020, the personal prevention awareness $c_{2021} = 0.350$ is lower than $c_{2020} = 0.627$. Indeed, Chinese cities including Nanjing implemented unprecedented interventions for controlling the outbreak of COVID-19 in early 2020. In addition, we find that the fitted transboundary transmission rate $d(x)_{2021} = \exp(-2.60448) = 0.074$ is much higher than that of $d(x)_{2020} = \exp(-4.80041) = 0.008$, which indicates that Delta strain is more infectious. And, indeed, a recent study shows that the relative viral loads of cases infected with the Delta variant are 1260 times higher than those

for the 2020 infections with clade 19A/19B viruses.^[37] As a result, our comparative analysis demonstrates the effectiveness

and scalability of the PDE-based prediction model in diverse modes of infectious diseases.

Table 7. Model parameter estimates of the second wave and the initial epidemic wave in Nanjing under the same epidemic time (24 days, duration of the second epidemic wave).

Parameter	d_0	k	b_0	b_1	c	l_0	l_1
Value ₂₀₂₁	2.60448	36.73040	0.51177	0.25663	0.34972	0.68641	0.51742
Value ₂₀₂₀	4.80041	20.74381	0.50008	0.25006	0.62726	0.33376	0.08795

8. Discussion

In this paper, we developed a new computational network model based on PDEs to characterize the spread of COVID-19 in China and to predict COVID-19 cases. The proposed model incorporates Baidu migration data to capture the spread of epidemics along the edges of the city network. To ensure high COVID-19 prediction accuracy, we used historical case data, and included the human active interventions such as city lockdowns and social distancing into the model. As a result, our model can demonstrate the spread of COVID-19 in China and the reduction of the transmission of COVID-19 under the dual efforts of lockdown and social distancing, which provides a new way for quantifying the effectiveness of the government policies toward controlling COVID-19.

In the present work, we have selected eight cities with different geographical locations and social characteristics centered on Wuhan. One may be able to use statistical models or ODE models on multiple patches to predict COVID-19 cases. Here, we choose PDEs based network modeling mainly because the approach can accurately predict the COVID-19 cases for these cities, and further quantify the impact of human behavior on changes of COVID-19 cases in the city network. In addition, it is found that 57% of the early cases reported outside Wuhan had travel history to Wuhan before Wuhan was quarantined (before January 23, 2020).^[22] This highlights that the outflow from Wuhan is the main driving force for the spread of COVID-19 in China. In this regard, the developed epidemic model framework with PDE incorporated the outflows from Wuhan is suitable for describing the process of spreading along the discoverer of the outbreak to other cities in the city network.

Our approach can be extended to other data set that captures population flows (e.g., train ticket or road traffic data) to assist policy decisions, such as the allocation of manpower and resources for specific geographic regions based on the predicted epidemic trend, and hoping to provide localized results to suppress the ongoing COVID-19 epidemic or other sudden outbreaks of infectious diseases in the future.

References

- [1] Newman M 2003 *SIAM Rev.* **45** 167
- [2] Newman M 2010 *Networks, An introduction* (Oxford: Oxford University Press)
- [3] Valba O, Avetisov V, Gorsky A and Nechaev S 2020 *Phys. Rev. E* **102** 010401
- [4] He S, Tang S and Rong L 2020 *Math. Biosci. Eng.* **17** 2792
- [5] Li Y, Zhao S, Luo Y, Gao D, Yang L and He D 2020 *Acta Phys. Sin.* **69** 090202 (in Chinese)
- [6] Yang Z, Zeng Z, Wang K, et al 2020 *J. Thorac. Dis.* **12** 165
- [7] Lai S, Ruktanonchai N W, Zhou L, Prosper O, Luo W, Floyd J R, Wesolowski A, Santillana M, Zhang C, Du X, Yu H, Tatem A J 2020 *Nature* **585** 410
- [8] Tao Y 2020 *Phys. Rev. E* **102** 032136
- [9] de Sousa L E, Neto P H O and Filho D A D S 2020 *Phys. Rev. E* **102** 032133
- [10] Maier B F and Brockmann D 2020 *Science* **368** 742
- [11] Prem K, Liu Y, Russell T W, Kucharski A J, Eggo R M and Davies N 2020 *Lancet Public Health* **5** e261
- [12] Wang H, Wang F and Xu K 2020 *Modeling information diffusion in online social networks with partial differential equations* (Cham, Switzerland: Springer)
- [13] Brauer F, Castillo-Chavez C and Feng Z 2019 *Mathematical Models in Epidemiology* (Cham, Switzerland: Springer)
- [14] Holmes E E, Lewis M A, Banks J E and Veit R R 1994 *Ecology* **75** 17
- [15] Zhu M, Guo X and Lin Z 2017 *Math. Biosci. Eng.* **14** 1565
- [16] Wang Y, Xu K, Kang Y, Wang H, Wang F and Avram A 2020 *Int. J. Environ. Res. Public Health* **17** 678
- [17] Wang H and Yamamoto N 2020 *Math. Biosci. Eng.* **17** 4891
- [18] Baidu qianxi. <https://qianxi.baidu.com/> [2021-7-15]
- [19] Chen S, Yang J, Yang W, Wang C and Barnighausen T 2020 *Lancet* **395** 764
- [20] Halloran M E, Vespignani A, Bharti N, Feldstein L R, Alexander K A, Ferrari M, Shaman J, Drake J M, Porco T, Eisenberg J N S, Del Valle S Y, Lofgren E, Scarpino S V, Eisenberg M C, Gao D, Hyman J M, Eubank S and Longini I M 2014 *Science* **346** 433
- [21] Jia J S, Lu X, Yuan Y, Xu G, Jia J and Christakis N A 2020 *Nature* **582** 389
- [22] Kraemer M U G, Yang C H, Gutierrez B, Wu C H, Klein B, Pigott D M, du Plessis L, Faria N R, Li R, Hanage W P, Brownstein J S, Layan M, Vespignani A, Tian H, Dye C, Pybus O G and Scarpino S V 2020 *Science* **368** 493
- [23] China News [2021-7-15]
- [24] Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer M U G, Li B, Cai J, Xu B, Yang Q, Wang B, Yang P, Cui Y, Song Y, Zheng P, Wang Q, Bjornstad O N, Yang R, Grenfell B T, Pybus O G and Dye C 2020 *Science* **368** 638
- [25] China Data Lab. 2021 *China COVID-19 Daily Cases with Basemap* [2021-7-15]
- [26] Arendt W, Dier D and Kramar Fijavz M 2014 *Appl. Math. Optim.* **69** 315
- [27] von Below J 1988 *J. Differ. Equ.* **72** 316
- [28] Murray J D 2002 *Photosynthetica* **40** 414
- [29] WHO 2019 *Novel Coronavirus (2019-nCoV) Situation Report - 7* [2021-9-12]
- [30] Li Q, Guan X, Wu P, et al 2020 *N. Engl. J. Med.* **382** 1199
- [31] Tang S, Yan Q, Shi W, Wang X, Sun X, Yu P, Wu J and Xiao Y 2018 *Environ. Pollut.* **232** 477
- [32] Oseledets I V 2011 *SIAM J. Sci. Comput.* **33** 2295
- [33] Lagarias J C, Reeds J A, Wright M H and Wright P E 1998 *SIAM J. Optim.* **9** 112
- [34] Sahai A K, Rath N, Sood V and Singh M P 2020 *Diabetes Metab. Syndr.* **14** 1419
- [35] Prem K, Liu Y, Russell T W, Kucharski A J, Eggo R M, Davies N, Jit M and Klepac P 2020 *Lancet Public Health* **5** e260
- [36] China News [2021-9-12]
- [37] Li B, Deng A, Li K, et al 2021 medRxiv: 2021.07.07.21260122 [Epidemiology]