Applications of multisource data-based dynamic modeling to cell-cell signaling and infectious disease spreading

Chen, D.

Chapter 7

Determining travel fluxes in epidemic areas

Daipeng Chen, Yuyi Xue, Yanni Xiao[1]

Abstract

Infectious disease attacks humans from time to time and threatens the lives of people all around the world. An important strategy to prevent the spatial spread of infectious disease is to restrict population mobility. As the epidemic threat is reduced, when the travel restrictions can be lifted, and how to organize orderly travel movement become critical. In this study, we define a novel diffusion distance derived from the estimated mobility network. Based on such diffusion distance a simple model was proposed to describe the spatiotemporal dynamics of infectious disease. Next we modeled the mobility restrictions and the exiting strategy by extending this simple model with a deterministic drift process of the population. We consequently develop a multi-source data fusion method to determine the drift process of population in epidemic areas. By solving two optimizations, we first select available subregions in epidemic areas, and then provide solutions to initiate new travel flux between these subregions. To test our method, we analyzed the multi-source data from mainland China and proposed a scheme triggering new travel flux during the first wave of COVID-19 in the selected 29 cities in mainland China. The testing predictions in these selected cities show that

[1]This chapter is based on Chen et al. (2021) "Determining travel fluxes in epidemic areas". PLoS Computational Biology, 17(10), e1009473.

reopening the borders in accordance with our proposed travel flux reduces the risk of a second outbreak of COVID-19 in these cities. This finding provides a methodology for re-starting travel movement of population during the weakening spread stage of the epidemic.

7.1 Introduction

In December 2019, the first COVID-19 case was reported in Wuhan, China [373]. The coronavirus SARS-CoV-2 then quickly spread out nationally [374] and internationally [375]. To prevent the spread of this novel virus, unprecedented measures such as travel restriction, quarantine and isolation followed by contact tracing were introduced by the Chinese government [376]. Most notably, strict travel restrictions were imposed in Wuhan and nearby cities from 23 January 2020. Subsequently, the highest level of alert and responses to the public health emergency were activated in mainland China [376]. The restrictions on mobility were also imposed in many other countries around the world to mitigate the spatial spread of SARS-CoV-2 [377]. Several studies have evaluated the effect of travel restrictions on the spread of SARS-CoV-2 [378] and have demonstrated that the lockdown imposed on Wuhan combined with measures to control movements in and out of the city reduced the number of domestic [379] and international [380] infections.

In epidemiology, the restriction on population movement into or out of a specific area such as a community, a city or a country is called cordon sanitaire [381]. In addition to the response to recent COVID-19 [382], cordons sanitaire had been established to stop the spread of other infectious diseases such as the 2001 foot and mouth disease epidemic in Britain [195], the 2003 SARS epidemic in China [196], the 2009 H1N1 flu epidemic in Mexico [20] and the 2014 Ebola epidemic in West Africa [197]. Reduced mobility between areas plays a potentially important role in decreasing the spread of infectious disease [383, 384], but it also has a negative effect on social health and economic development [385, 199]. Once infectious disease or diseases are under control or successfully contained in some specific areas, the reopening of borders has been put on the agenda. However, it is not clear how to balance the trade-offs between inter-regional mobility and the risk of potential new outbreaks. Determining where and when mobility restrictions can be relaxed and designing a new orderly travel movement fall within the scope of this study.

In the most extreme case, the cordon will not be lifted until the infectious disease is extinguished [359], but it is not suitable for curbing a large-scale emergent epidemic

[386], and hence many countries are seeking a balance between reopening and preventing the epidemic from rebounding. A number of studies have formulated mathematical models with population mobility to investigate the effect of massive movements on the spread of infectious disease [184, 387, 388, 389]. However, few studies have focused on the trade-off between mobility restrictions and the spread of pathogen. Preparing for a responsible lockdown exit strategy [390] has been extremely important not only for the COVID-19 epidemic [391], but also for future pandemic outbreaks. Most modeling studies integrating mobility data and the spatial spread of infectious disease are based on meta-population models [185]. Combining such model with mobility data, Linka et al. [193] provided an exit strategy for the travel ban against COVID-19 in Canada. In their study, they made reopening forecasts by updating the fraction by which the baseline movement matrix was multiplied and running the model repeatedly. Once some acceptable simulations were obtained, the corresponding movement matrixes were translated into suggestions to guide population mobility. Based on the same model, Ruktanonchai et al. [192] assessed the impact of coordinated COVID-19 exit strategies across Europe. Similar to the work of Linka et al. [193], they ran their model under different parameter sets and initial states and found that appropriate coordinated exit strategies greatly improved the possibility of curbing the spread of COVID-19 in Europe.

Although meta-population models have provided a rich and nuanced perspective on predicting epidemic spread [392] and evaluating the effectiveness of interventions on population mobility [393], they would become highly complex and difficult to compute if more spatial heterogeneity was introduced [391]. Another limitation of the previous studies comes from the methods that are based on adjusting parameters or updating initial states of models. This approach generally only provides some suggestions on the order [192] or degree [193] of reopening in some epidemic areas, and does not provide relatively detailed information on where and when the mobility restrictions can be relaxed, nor answer what kind of population movement pattern among various areas is optimal. To address this challenge, Gösgens et al. [194] developed a method to determine the optimal travel flux in epidemic areas. Given the time horizon and regional division, Gösgens et al. [194] proposed an optimal function which is positively related to population mobility rates between different regions and negatively related to infection cases in each region. Maximizing this optimization function suggests a mobility strategy in epidemic areas. However, their method is computationally expensive because it involves repeatedly updating regional divisions and solving a meta-population epidemiological model.

To solve the difficulties induced by the meta-population model, by defining a diffusion distance, we reduced the complex spatiotemporal spread of pathogen into a simple wave mode, and consequently formulated a reaction diffusion equation to describe such spread. In this study, we further extended the reaction diffusion system to a reaction diffusion-drift equation, where the drift term corresponds to a deterministic travel flux. Instead of testing different movement patterns, here we develop a novel multi-source data fusion method that can integrate the model with data to calculate the optimal travel flux in epidemic areas. Unlike the work of Gösgens et al. [194], our objective function and constraints give a convex optimization, which means that the global optimum is available. As a case study, we test this method with data from mainland China where a strict lockdown against COVID-19 was implemented in early 2020. We obtained the specific timings for reopening and determined the travel flux between these cities that would allow them to reopen as well as keeping the epidemic from rebounding.

7.2 Materials and Methods

7.2.1 Data

The migration indexes were collected from a website http://qianxi.baidu.com which is an open big data platform including the daily immigration rate and emigration rate of each city in Mainland China. These indices are obtained by dividing the number of people moving out (in) from the current city by the total number of people moving out (in) from all cities in Mainland China during the same period. Specifically, a greater migration index in a city is associated with its more outbound (inbound) events by rail, air and road traffic. We selected the cities with the top 25 immigration rate or top 25 emigration rate as the objects of our case study. Although the migration index quantifies the movement of a population, it does not track the movement directions and trajectories of individuals. In other words, the migration index ranks the mobility of population in different cities but does not provide a quantity for the population flow from one city to another.

The population flow from one city to another is derived from other data indirectly. Generally, the population flow from one region to another depends on many factors [394] such as the local population size [395, 396], economic level [397] and geographic distance between them [398]. In this study, the per capita Gross Domestic Product (GDP) (corresponding to the local economic level) and the number in the resident

population (the local population size) are obtained from the China Statistical Yearbook 2019 and are summarized in Table 7.1. The geographic distance between each pair of cities was collected from the Baidu Maps (https://map.baidu.com). Based on these data, we estimated the relative population flow between the selected cities with high migration indices for further analysis.

We obtained the number of daily confirmed COVID-19 cases and the cumulative number of confirmed cases between 23 January and 24 February from the National Health Commission of the People's Republic of China [399]. Although there will be a time lag in reporting, it is believed that the number of confirmed cases reflect the true epidemic situation in each city.

7.2.2 Mobility network

We show the distribution of the migration indexes of the selected cities in Fig 7.1A, and rank these cities according to the mean value of the migration index. These selected cities are mainly divided into labor-intensive export cities with large populations (Chongqing, Zhoukou, Langfang, etc.) or cities with developed economies (Beijing, Shanghai, Guangzhou, Shenzhen, etc.). Due to the Spring Festival, the emigration rate in these cities does appear to be higher than the immigration rate, but the cities ranked by two indices show a high degree of overlap (Fig 7.1A). Beijing and Guangzhou with the highest emigration rate also had the highest immigration rate. Thus, despite the effect of returning home during the Spring Festival, the population moving out of and into a city still has a weak symmetry. To compensate for the lack of directionality of the migration index, we estimate the direct population flow W_{ij} from any subregion Ω_i to subregion Ω_j $(i \neq j)$ with the gravity model [400]

$$W_{ij} = G\frac{M_i^{\sigma_1} M_j^{\sigma_2}}{e^{r_{ij}/r}}, \tag{7.1}$$

where G is a proportionality constant, M_i and M_j are the quantitative representation of 'mass' (e.g., the mean of population size and GDP) of the regions Ω_i and Ω_j, respectively. The constant r is the characteristic length that governs the decay of population flow with geographic distance r_{ij}. The parameters σ_1 and σ_2 tune the dependence of population flow to the regional mass. The estimation of parameters is derived from a global scale study by Balcan et al. [401].

To illustrate the reliability of our estimated results of population mobility, we compared the estimated relative outflow from Wuhan to another 295 cities in China

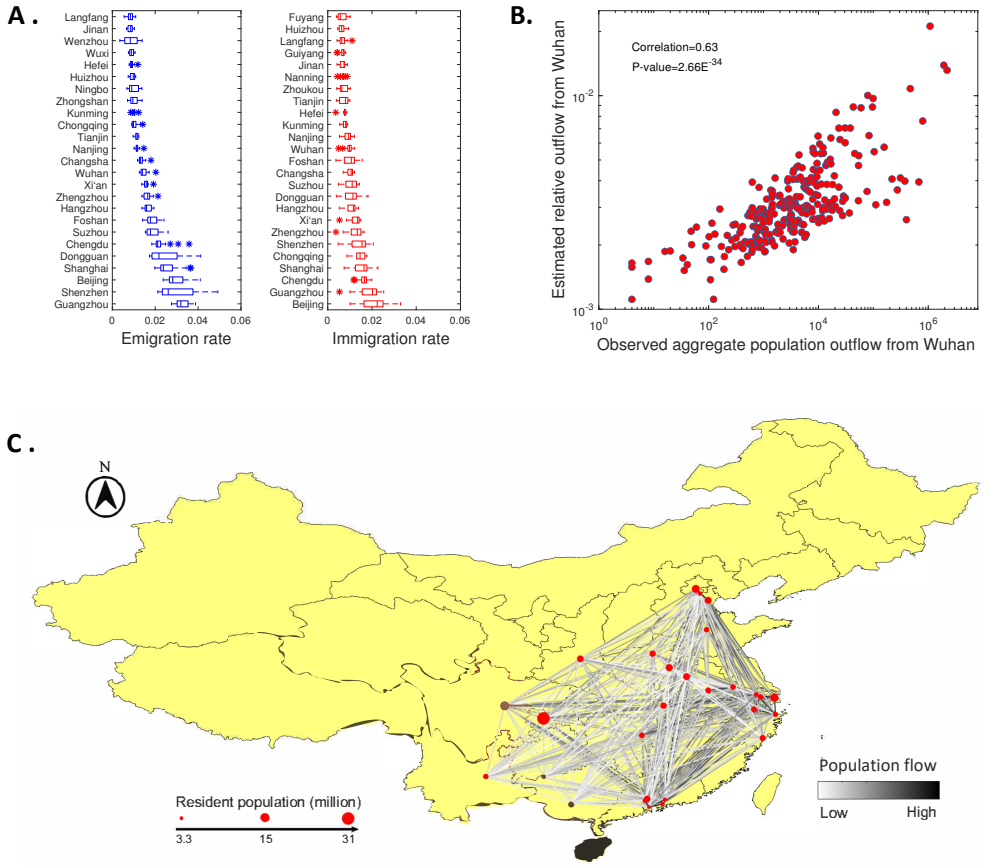


Figure 7.1: Mobility network among twenty-nine cities with the most active personnel movements in Mainland China in early 2020. (A) The selected cities are ranked by the value of the migration rate from 1 to 22 January. (B) The estimated relative outflow from Wuhan to other cities versus the observed aggregate population outflow. (C) The mobility network among these cities. The diameter of nodes indicates the size of the resident population in each city, and the gray scale indicates the relative population flow between two cities. The base layer of the map was created using public sources: https://data.humdata.org/dataset/china-administrative-boundaries.

with the real aggregate population outflow (Fig 7.1B), where the real data come from the work of Jia et al. [374]. The estimated relative outflow from Wuhan show a significant correlation with the real data, and the correlation coefficient is 0.63 with the p-value less than 10^{-30}. It is worth noting that the existing data only focus on the population outflow from Wuhan, which can neither form a complete mobility network nor form cross-validation to train the gravity model. Here, we construct an undirected

mobility network (MN) with the edges weighted through the symmetrization of the estimated relative population flow (Eq. 7.1). That is,

$$A_{i,j} = \frac{W_{i,j} + W_{j,i}}{2\sum_{i,j} W_{i,j}}. \tag{7.2}$$

We visualize the undirected mobility network in Fig 7.1C. The undirected weighted network shows that there were high population flows (the darker edges in Fig 7.1C) between the cities with larger population sizes (bigger red dots) and the eastern cities with more developed economies in mainland China. This predicts that more SARS-CoV-2 carriers came to these cities than to others weakly connected to Wuhan before the strict lockdown, which implies that there would be more severe epidemics in these strongly connected cities. Later statistical analysis on the geo-temporal spread of COVID-19 in China [378, 399] fully support this conclusion. In short, it is believed that using the estimated population flow to characterize the connectivity between cities is as reliable as real data.

7.2.3 Diffusion distance

The complexity of human mobility, especially via air traffic and high-speed train, has reconstructed cities around the world into a high-dimensional spatial structure, which makes it increasingly difficult to model the spread of emerging infectious diseases. To simplify the spatial transmission of infectious disease, we first introduce the Diffusion maps theory [402].

Lemma 7.2.1 (Diffusion maps theory)**.** *Let $A(x, y)$ be a kernel function in a measure space $(X, \mathcal{A}, \mu)$. The function $A(x, y)$ satisfies $A(x, y) = A(y, x)$ and $A(x, y) \geq 0$. Then there is a map $\Phi(x)$ from the space X to an Euclidean space $\mathbb{R}^k$.*

The kernel function $A(x, y)$ is a symmetric, positivity-preserving kernel defined on the space X. Particularly, $A(x, y)$ would transform a mobility network (MN) into an undirected graph with edge weight $A(x, y)$ if X is a set of the nodes of the network. The quantity $P(x, y) = A(x, y)/\sum_y A(x, y)$ can be viewed as the probability for a random walker on X to make a step from the vertex x to vertex y. Naturally, we get the transition matrix P and the stationary distribution c of this Markov chain given by $c(x) = \sum_y A(x, y)/\sum_{x,y} A(x, y)$. Let $P^{(l)}(x, y)$ be the element of P^l (the power of matrix P) given a scale parameter l. We then define the diffusion distance $D_l(x, y)$

between any locations x and y in space X by

$$D_l(x,y) = \left(\sum_z \frac{1}{c(z)} \Big(P^{(l)}(x,z) - P^{(l)}(y,z) \Big)^2 \right)^{\frac{1}{2}}, \tag{7.3}$$

which $P^{(l)}(x,z)$ represents the probability of transition from x to z in l steps. For a sufficiently large l, the diffusion map $\Phi(x)$ embeds the points $\{x \in X\}$ into a low-dimensional Euclidean space $\mathbb{R}^k$ in an approximately isometric manner [403, 402]. That is to say, the diffusion distance $D_l(x,y)$ of any two points in X is approximately equal to the Euclidean distance $\|\Phi(x) - \Phi(y)\|$ between the points where they are mapped in $\mathbb{R}^k$.

Setting $A_{i,i} = 1$ ensures that the population flow, as defined by equation (7.2), remains consistent with the kernel function required in Lemma 7.2.1. Therefore, we can establish a diffusion distance between cities based on the population flow in a mobility network. By deriving a diffusion distance from the mobility network (MN), we replace the conventional geographic distance by this new metric and plot the diffusion distances from Wuhan to other selected cities in Fig 7.2. The length of polar paths connecting Wuhan to other cities in this pandemic invasion tree is the estimated diffusion distance. Interestingly, in the perspective of diffusion distance, Guangzhou is the closest city to Wuhan among the 29 cities, while Fuyang, which has the closest geographical distance to Wuhan, has become the farthest. Moreover, it follows from Fig 7.2A to 7.2D that the diffusion distance re-shapes the irregular and complicated spatiotemporal spread patterns of COVID-19 in the conventional geographic perspective [378, 399] into a regular, wavelike solution (i.e., the epidemic first reaches the regions closest to the initial outbreak city). Specifically, COVID-19 cases outside of Hubei Province are first confirmed in Beijing and Shenzhen, and the diffusion distance between Wuhan and them are also shorter than most other cities (Fig 7.2A). On 20 January 2020, confirmed cases were also reported in Shanghai and Huizhou. These two cities are in the top 11 cities with the smallest diffusion distance from Wuhan (Fig 7.2B). On 21 and 22 January 2020, COVID-19 cases were confirmed in the cities with larger diffusion distance from Wuhan (Fig 7.2C and 7.2D). From this sequence of panels, we find that the shorter the diffusion distance to Wuhan, the earlier the confirmed cases appear, which implies that to a great extent, the diffusion distance can reflect the order of disease spread from the epicenter to other cities. This provides the evidence that our redefined diffusion distance reshapes the spatial spread of the SARS-CoV-2 into a wave-like solution, which led us to connect the spread pattern

of COVID-19 with the classical reaction diffusion system [404] in the perspective of diffusion distance.

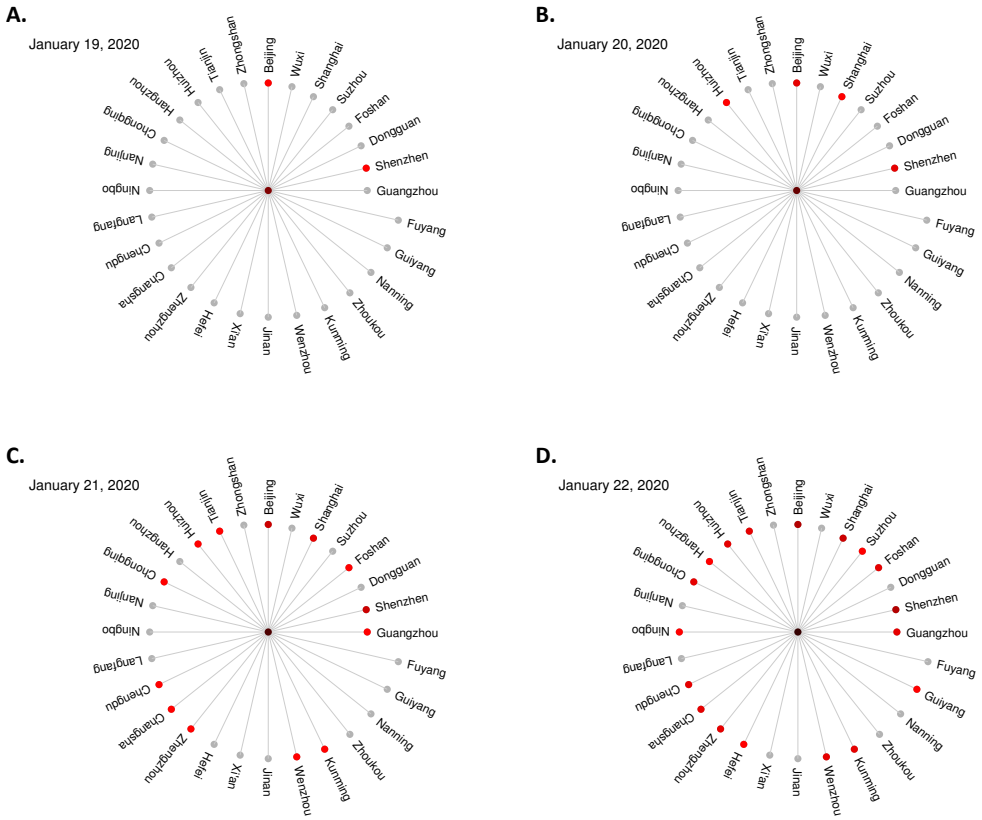


Figure 7.2: Epidemic invasion trees rooted in Wuhan. The sequence (from 19 to 22 January 2020) of panels depicts the early spread of the SARS-CoV-2 infection in 29 cities. The length of the polar path represents the diffusion distance from Wuhan to each city. The cities with confirmed infected individuals are marked in red, and the prevalence is reflected by the redness of the nodes.

7.2.4 Reaction diffusion-drift equation

For an epidemic outbreak in the space X, we let $p(t, x)$ denote the density of infectious individuals who can move freely and contact others at location $x \in X$ at time t. In the perspective of diffusion distance (Eq. 7.3), the spatial spread of infectious disease is approximated as a wave pattern (see Fig 7.2). A similar result is observed in the context of effective distance [405]. Therefore, the evolution of density for infectious

individuals can be modeled by a simple reaction-diffusion equation:

$$\frac{\partial p(t,x)}{\partial t} = \gamma \Delta p(t,x) + R(t,x)p(t,x). \tag{7.4}$$

Equation (7.4) states that, in each small region, the rate of change in the density of infectious individuals (left-hand side) is balanced by the physical movements with constant diffusion rate γ and the net reproduction of infectious individuals. Note that the net growth rate $R(t,x)$ is formulated as $R(t,x) = \Lambda(t,x) - \Gamma(t,x)$ with the incidence rate $\Lambda(t,x)$ and removal rate $\Gamma(t,x)$. The incidence rate is the average number of adequate contacts with susceptibles of a typical virus carrier per unit time [406], and is given by

$$\Lambda(t,x) = \beta C(t,x)\frac{s(t,x)}{n(t,x)}, \tag{7.5}$$

where β is the transmission probability, $C(t,x)$ is the average contact rate of one person with other individuals, and depends heavily on the intensity of intervention measures. $s(t,x)$ and $n(t,x)$ denote the density of susceptible individuals and the total number of free-moving individuals, respectively.

Facing severe epidemics (e.g., COVID-19), countries around the world would adopt measures such as entry-exit testing and regional lockdown to curb the spatial spread of the pandemic [377]. From a mathematical point of view, this type of intervention introduces a potential function $U = -\gamma \ln(p)$ to counteract diffusion and to maintain a non-uniform density $p(t,x)$ in space X. The mobility of infectious individuals is restricted by these interventions since $\gamma \Delta p + \nabla.(p \nabla U) = 0$. The research aim we pursue in this study is to design a reasonable lockdown exit strategy. To achieve this goal, we introduce another unknown potential function V to counteract the effect of function U and obtain a new function $F = U + V$. Consequently, we get a new reaction diffusion-drift equation [407, 408], which has the following form:

$$\frac{\partial p}{\partial t} = \gamma \Delta p + \nabla.\big(p \nabla F\big) + Rp. \tag{7.6}$$

The second term of the equation (7.6) represents the spatial movements of individuals induced by the deterministic travel flux $-\nabla F$ in X. Obviously, the equation (7.6) describes the epidemic dynamics under complete lockdown if $V = 0$ (or, $\nabla V = 0$), and it would be equation (7.4) without any regional lockdown if $V = \gamma \ln(p)$. What we want to do here is to find an optimal V to lift the strong lockdown without causing a second outbreak in the epidemic area.

7.2.5 Lockdown exit strategy

Lifting travel restrictions and restarting travel flux may cause an increase in the contact rate $C(t,x)$ in the selected subregions. We denote ϵ to quantify the variation in the contact rate if a new travel flux is initiated. Thus, the net growth rate $R(t,x)$ was extended to be a behavior-related growth rate $R(t,x,\epsilon) = \Lambda(t,x,\epsilon) - \Gamma(t,x)$ with

$$\Lambda(t,x,\epsilon) = \beta\Big((1-\epsilon)C(t,x) + \epsilon C(0,x)\Big)\frac{s(t,x)}{n(t,x)}. \tag{7.7}$$

Here $C(0,x)$ represents the normal contact rate at location x. Therefore, the lifting of travel restrictions has no effect on the net growth rate if $\epsilon = 0$, and the behavior of local individuals has completely returned back to normal if $\epsilon = 1$. Although the incidence rate $\Lambda(t,x)$ also depends on interventions such as testing, contact tracing and isolation [409], we only introduce an adjustment factor ϵ to summarize the variations here without considering the details because the purpose of this study is to explore the lockdown exit strategy.

A good lockdown exit strategy should first ensure that the relaxation of travel restrictions would not cause a second outbreak of epidemics during the weakening spread stage. In other words, the changing rate in the density of infectious individuals $\partial p/\partial t$ should not be positive after a new deterministic travel flux $-\nabla V(t_0,x)$ is initiated at some time t_0. It is summarized by a mathematical expression:

$$\nabla.\big(p(t,x)\nabla V(t_0,x)\big) + R(t,x,\epsilon)p(t,x) \leq 0 \tag{7.8}$$

for any time $t \in [t_0, t_1)$, where t_1 is the end time of this travel flux. Here, the diffusion term is cancelled out by the vector field induced by potential function U.

I. Where and when

Considering the lockdown subregions $\{\Omega_1, \Omega_2, ..., \Omega_m\} \subset X$ and assuming that the growth rate $R(t,x,\epsilon)$ is homogeneous in each subregion $\Omega_i (i = 1, 2, ..., m)$. We integrated the inequality (7.8) on $\Omega = \bigcup_{i=1}^{m} \Omega_i$ and obtained

$$\sum_{i=1}^{m} \int_{\Omega_i} \nabla.\big(p(t,x)\nabla V(t_0,x)\big)d\mu(x) + \sum_{i=1}^{m} R(t,\Omega_i,\epsilon)I(t,\Omega_i) \leq 0, \tag{7.9}$$

where $I(t,\Omega_i) = \int_{\Omega_i} p(t,x)d\mu(x)$ is the number of infectious individuals in subregion Ω_i at time t. Without loss of generality, we assume that the lockdown exit strategy is

only implemented among the candidate subregions $\{\Omega_1, \Omega_2, ..., \Omega_m\}$, which means that the infectious individuals moving out from one candidate subregion must enter another one, so the total divergence remains zero. Consequently, $\sum_{i=1}^{m} R(t, \Omega_i, \epsilon) I(t, \Omega_i) \leq 0$ is a necessary condition for lifting the strict lockdown in Ω. According to Lemma 7.5.1 in Supplementary information, the subregions $\{\Omega_1, \Omega_2, ..., \Omega_m\}$ satisfying this necessary condition can be selected to initiate some new travel fluxes. We introduce a logical variable ξ_i here to indicate whether the lockdown exit strategy can be implemented in the subregion Ω_i and obtain the following zero-one programming,

$$\begin{aligned} \max \quad & H_1 = \sum_{i=1}^{m} \xi_i, \\ \text{s.t.} \quad & \sum_{i=1}^{m} R(t, \Omega_i, \epsilon) I(t, \Omega_i) \xi_i \leq 0, \\ & \xi_i \in \{0, 1\}. \end{aligned} \tag{7.10}$$

Where, $\xi_i = 1$ means that the subregion Ω_i can be selected to initiate a new travel flux at time t, and $\xi_i = 0$ means that the travel restrictions cannot be lifted at this point in time. Maximizing the objective function H_1 is to get more subregions where the mobility restrictions could be lifted. We can look at such $0-1$ integer linear programming (7.10) from two perspectives. On the one hand, we can solve programming (7.10) at any fixed time t_0 to get the available subregions. On the other hand, we can focus on the subregions to determine the available time.

II. How

Given the available subregions $\{\Omega_1, \Omega_2, ..., \Omega_m\}$, infectious density $p(t, x)$ and growth rate $R(t, x, \epsilon)$, we hope to determine the potential $V(t_0, x)$ by solving inequality (7.8). This is a big challenge, we solve it here based on the approximation theory of diffusion maps [403, 402], one important theory used in this chapter. Assuming that the individuals are homogeneous in each subregion, then $\{\Omega_1, \Omega_2, ..., \Omega_m\}$ can be regarded as a set of points sampled from the probability density function $f(t_0, x) = p(t_0, x)/I(t_0, X)$. Denote the kernel function

$$K_\delta(\Omega_i, \Omega_j) = \frac{1}{(\pi/\delta)^{k/2}} e^{-\delta D_l^2(\Omega_i, \Omega_i)}, \tag{7.11}$$

where $D_l(\Omega_i, \Omega_i)$ is the diffusion distance (7.3) between Ω_i and Ω_j, k is the dimension of the embedding space $\mathbb{R}^k$ corresponding to the diffusion maps [403], and δ is a

parameter adjusting the kernel width. Based on this kernel, we construct a graph Laplacian $L_{\delta,m}$ through the following procedure. Set

$$K_{\delta,m}(\Omega_i,\Omega_j)=\frac{K_\delta(\Omega_i,\Omega_j)}{\sqrt{\big(\sum_{i=1}^m K_\delta(\Omega_i,\Omega_j)\big)\big(\sum_{j=1}^m K_\delta(\Omega_i,\Omega_j)\big)}},$$
$$Q_{\delta,m}(\Omega_i,\Omega_j)=\frac{K_{\delta,m}(\Omega_i,\Omega_j)}{\sum_{j=1}^m K_{\delta,m}(\Omega_i,\Omega_j)},$$

then the graph Laplacian $L_{\delta,m}=4\delta(Q_{\delta,m}-E_m)$ converges to a Kolmogorov operator $\mathcal{L}$ (Lemma 7.5.1 in Supplementary information). That is

$$\lim_{m\to\infty} L_{\delta,m}=\mathcal{L}=\nabla(\ln p)\cdot\nabla+\Delta, \tag{7.12}$$

where E_m is an identity matrix with dimension $m\times m$. For the Gaussian kernel, the bandwidth $\sqrt{1/\delta}$ is usually set as the median value of the distances between all samples. Similar to the vector $V^{(m)}=[V(t_0,\Omega_1),V(t_0,\Omega_2),...,V(t_0,\Omega_m)]^T$, we let $R^{(m)}$ and $p^{(m)}$ denote the restriction of functions $R(t_0,x,\epsilon)$ and $p(t_0,x)$ to the candidate subregions, and then construct the following inequality

$$diag\big(L_{\delta,m}V^{(m)}+R^{(m)}\big)p^{(m)}\le 0^{(m)}, \tag{7.13}$$

where $diag(\vec{a})$ denotes the diagonal matrix generated by vector $\vec{a}$, and $0^{(m)}$ is a zero vector with dimension m. Comparing inequality (7.13) and inequality (7.8), the convergence of the graph Laplacian $L_{\delta,m}$ (Eq. 7.12) implies that any solution $V^{(m)}$ of inequality (7.13) is a good approximation to the potential V we want to determine, and the error would be very small if the sample size m is large enough.

We have already mentioned that $V=0$ means that there is a complete lockdown in the epidemic area X, and $V=\gamma\ln(p)$ means that there is no restriction (see Eq. 7.6). Our lockdown exit strategy here is to lift travel restrictions as much as possible without worsening the epidemic. Therefore, we give the following optimization to estimate $V^{(m)}$

$$\begin{aligned}\min\quad & H_2=\|V^{(m)}-\gamma\ln(I^{(m)})\|^2,\\ \text{s.t.}\quad & L_{\delta,m}V^{(m)}+R^{(m)}\le 0^{(m)}.\end{aligned} \tag{7.14}$$

Where $I^{(m)}=[I(t_0,\Omega_1),I(t_0,\Omega_2),...,I(t_0,\Omega_m)]^T$, and the constraint of optimization (7.14) is derived from the inequality (7.13) because $p^{(m)}$ is positive in epidemic area.

Note that we use the number I of infectious individuals instead of density p in the cost function H_2 because the latter is difficult to quantify in actual data, and the degree of real lockdown is usually adjusted by the size of the number of regional infectious individuals rather than their density. Without loss of generality, we assume that the calculation result yields $V(t_0, \Omega_i) \geq V(t_0, \Omega_j)$ for $i < j$ because we can always achieve it by adjusting the order of the subregions. The drift process in model (7.6) is the movement of individuals from the region with high potential to nearby regions with low potential, which inspires us to define a net flow matrix $\boldsymbol{J}$ among the distinct subregions with

$$J_{i,j} = max\big\{\theta_{i,j}\big(V(t_0, \Omega_j) - V(t_0, \Omega_i)\big)/D_l(\Omega_i, \Omega_i), 0\big\}, \tag{7.15}$$

where the parameter $\theta_{i,j} \geq 0$ adjusts the net population outflow from subregion Ω_j to other Ω_i. Consequently, this movement strategy yields a patch model of infectious disease transmission.

$$\frac{\mathrm{d}I(t, \Omega_i)}{\mathrm{d}t} = \sum_{j=1}^{m} \big(J_{i,j} I(t, \Omega_j) - J_{j,i} I(t, \Omega_i)\big) + R(t, \Omega_i, \epsilon) I(t, \Omega_i) \tag{7.16}$$

with $i = 1, ..., m$ and $t \in [t_0, t_1)$. Similar to water flow, the net movement of population represented by the matrix $\boldsymbol{J}$ always moves from subregions with higher potential to subregions with lower potential. For the parameter $\theta_{i,j}$ in Eq. (7.15), we develop a source-sink method (Supplementary Text 7.5.3) to determine it. We further extend the one-way movement matrix $\boldsymbol{J}$ to a non-negative matrix $\boldsymbol{B}$ presenting a two-way movements through a population balance equation with random noise

$$B_{i,j} = (B_{j,i} - J_{j,i}) \frac{I(t, \Omega_i)}{I(t, \Omega_j)} + (1 + \xi) J_{i,j}, \tag{7.17}$$

where the random noise ξ breaks the expected balance in movements. Note that the factor $\xi = 0$ means an optimal two-way mobility represented by matrix $\boldsymbol{B}$. In this scenario, population mobility quantified by the movement matrix $\boldsymbol{B}$ is the same as that quantified by matrix $\boldsymbol{J}$.

7.3 Results

As a verification and test of the method, we apply it to the context of the COVID-19 epidemic with strict control in Mainland China from the end of 2019 to the early part

of 2020. More details about calculation and simulation can be found in Supplementary information 7.5. The code is available at: https://github.com/DaipengChen/Travel-fluxes-in-epidemic-areas.

7.3.1 Net growth rate

Based on this novel notion of distance, we model the complex spatiotemporal spread patterns of COVID-19 by a simple reaction-diffusion equation (Eq. 7.4) where the diffusion rate γ and net growth rate R can be estimated from real data of infections. In the early stage of COVID-19 spread, the spatial movements of infected individuals were mainly from the initial outbreak city Ω_0 to others. Assuming that the infectious individuals and net growth rate are homogeneous in each city Ω_i, we get the following ordinary differential equation for the dynamics of the COVID-19 (details in Supplementary information 7.5):

$$\frac{\mathrm{d}I(t,\Omega_i)}{\mathrm{d}t} = \gamma_i(t)I(t,\Omega_0) + R(t,\Omega_i)I(t,\Omega_i), \tag{7.18}$$

where $\gamma_i(t)$ is the degree of movement from the initial outbreak city Ω_0 to other cities Ω_i and assuming that the movement rate $\gamma_i(t)$ was constant γ_i (particularly, $\gamma_0 \propto -\gamma$) before lockdown and was 0 during the period of lockdown. In China, the travel ban and the first level response to major public health emergencies were initiated on 23 January 2020 [380]. These measures definitely prevented the movement of population and reduced the contact rate between people. Therefore, we assume that the incidence rate $\Lambda(t,\Omega_i)$ (Eq. 7.5) contains an activation function $h(t) = 1/(1+e^{t-t_0})$ [410] adjusting the contact rate, where t_0 is the critical time point of 23 January 2020. Furthermore, we can simplify the incidence rate (Eq. 7.5) to $\beta C(t,x)$ since $s(t,x)$ was very close to $n(t,x)$ in the early stages of COVID-19 in China. Consequently, during our research period, the regional net growth rate can be approximated by

$$R(t,\Omega_i) = \Lambda(\Omega_i)h(t) - \Gamma(\Omega_i), \tag{7.19}$$

where $\Lambda(\Omega_i)$ is the maximal incidence, $\Gamma(\Omega_i)$ is the removal rate.

Due to reporting delays and infection detection durations, we assume that the daily number of reported new cases $\hat{I}(t,\Omega_i)$ is proportional to the numbers $I(t-\tau_i,\Omega_i)$ of infectious individuals at τ_i days ago, and the reporting time-delay τ_i is region specific. Therefore, a dynamic system for the daily number of reported new cases is derived

from model (7.18) and is given by

$$\frac{\mathrm{d}\hat{I}(t,\Omega_i)}{\mathrm{d}t} = \gamma_i(t-\tau_i)\hat{I}(t-\tau_i+\tau_0,\Omega_0) + R(t-\tau_i,\Omega_i)\hat{I}(t,\Omega_i). \tag{7.20}$$

To this end, we initially estimated the unknown parameters in each city Ω_i [411], and then the regional growth rate can be approximated by formula (Eq. 7.19) using the estimated parameters. Specifically, we initially fit the model (Eq. 7.20) to the daily number of reported new cases in city Ω_0 and obtain the estimates for parameters (i.e., γ_0, τ_0, $\Lambda(\Omega_0)$, $\Gamma(\Omega_0)$) associated with this city. Secondly, once we have estimates for all parameters about the initial outbreak city Ω_0, we then apply the same procedure to estimate the relevant parameters (γ_i, τ_i, $\Lambda(\Omega_i)$, $\Gamma(\Omega_i)$) for each city Ω_i by fitting model (7.20) to the daily reported new cases in city Ω_i (see Table 7.2 for results). Fig 7.3A-C displays the fitted results of the daily number of reported new cases and the estimated net growth rate in three cities (Wuhan, Chongqing and Wenzhou) with the largest peak of confirmed cases in all 29 cities.

The estimated net growth rate $R(t,\Omega_i)$ shows a visible decline in these three cities after 20 January 2020 (Fig 7.3A-C), which coincides with fact that on 20 January 2020 confirmation of human-to-human transmission of SARS-CoV-2 was announced and COVID-19 was incorporated as a notifiable disease in the Infectious Disease Law and Health and Quarantine Law in China [412]. The net growth rate of these three cities was reduced to zero on about 23 January 2020 (Fig 7.3A-C), which indicated that the lockdown of Wuhan and the first-level emergency response in other cities greatly reduced the exposure rate and resulted in a balance between the incidence rate and the removal rate of infected individuals (mainly caused by medical isolation and deaths) in the free moving population. After 23 January, the estimated net growth rate in the three cities decreased to negative values (Fig 7.3A-C), indicating that the epidemic curve of new infections in these cities should peak around 23 January 2020, which coincides with the results that the epidemic curve of illness onset peaked around 26 January in China [399] and the median incubation period of Chinese COVID-19 patients was 3 days [413].

It is worth noting that due to the incubation period of COVID-19 [414, 413] and the delays between diagnosis and case reports, there exists a lag between the time when the net growth rate $R(t,\Omega_i)$ reaches zero and the time when the daily number of reported new cases peaks. From Fig 7.3A-C, we find that these delays are region specific. Our estimated reporting time delays in all 29 cities are plotted in Fig 7.3D, which illustrates that the delays varied between different cities. This is in agreement

with a population-level observational study [415] which showed that there were large differences between Wuhan and other cities in China in terms of reporting time delay. On average, except for Wuhan with a reporting lag of 19 days, the reporting time delays in other cities were less than 15 days, and the numbers of infected individuals in most cities were diagnosed within 10 days.

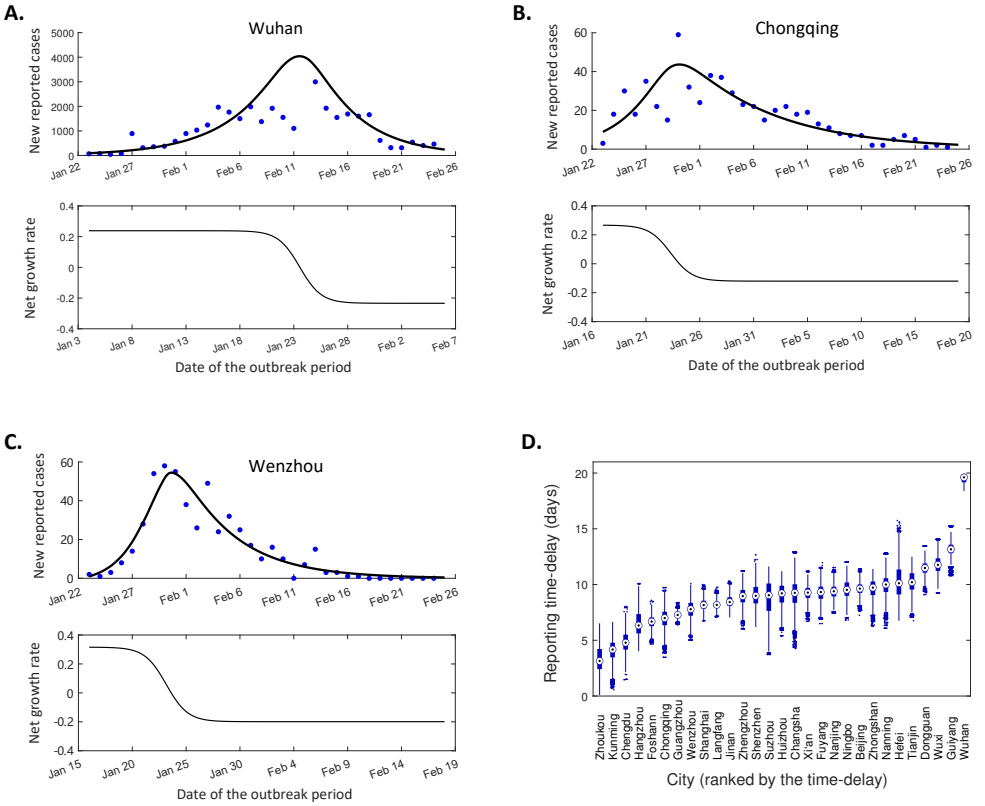


Figure 7.3: The net growth rates and reporting time delays. (A-C) The fitted results and the estimated net growth rates (black curves) in three cities with the highest peaks of prevalence in the selected 29 cities. The solid blue dots are real data and one outlier in Wuhan is not shown in the figure. There were different lags between the estimated net growth rates and the numbers of reported cases in each city. (D) The estimated reporting time delays in all 29 cities. These reporting time delays are region specific.

7.3.2 Where can be unlocked?

To investigate where and when the lockdown exit strategy can be implemented among the selected 29 cities, we solved the optimization problem (Eq. 7.10) with various

adjustment factors ϵ and plotted Fig 7.4 to visualize our calculation results. Note that the adjustment factor ϵ quantifies the changes of contact rate after lifting the travel restrictions (Eq. 7.7), and the larger the value of ϵ the higher the contact rate.

Fig 7.4 shows that the travel restrictions could not be lifted during the whole study period in all 29 cities if ϵ was not less than 0.54. This implies that if the incidence rate $\Lambda(t, \Omega_i)$ cannot be maintained at a low level through interventions such as keeping social distancing and maintaining bans on large gatherings, travel restrictions cannot be lifted until the number of infected individuals falls below the local invasion threshold [405]. If the change of contact rate caused by lifting of travel restrictions is relatively small, for example $\epsilon = 0.47$ (Fig 7.4), our calculation results predict that except for densely populated cities such as Chengdu, Chongqing and Guangzhou, the strict travel restrictions in the other 25 cities could be lifted after 26 January. In particular, for an extremely small value of ϵ (for example $\epsilon = 0.1$ in Fig 7.4), travel restrictions in most cities only need to last one day, and it is not necessary to impose strict travel restrictions on the cities such as Langfang, Suzhou, Ningbo, Beijing, Hefei, Wuxi and

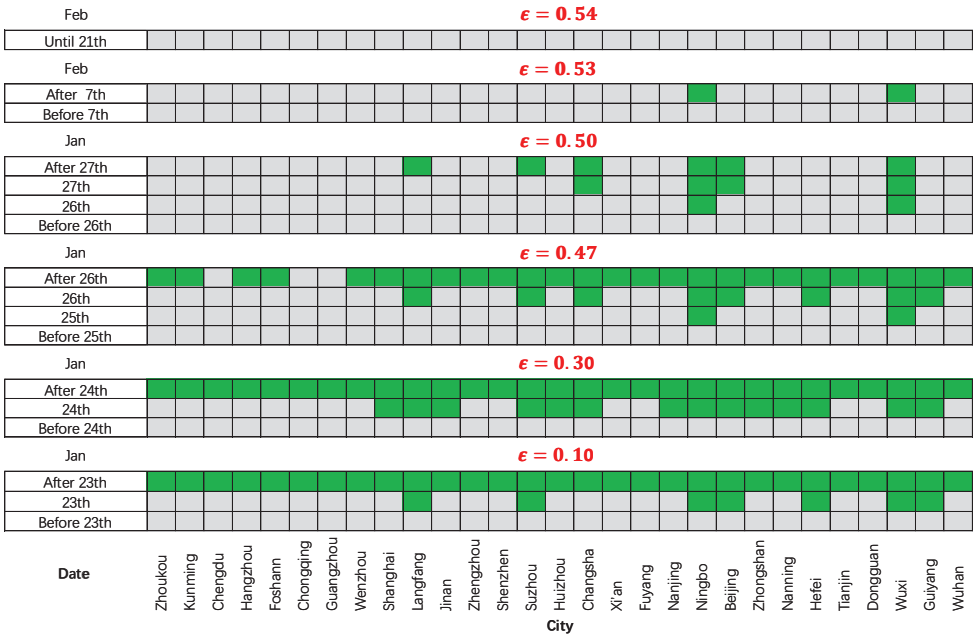


Figure 7.4: Where and when the lockdown exit strategy can be implemented. With different contact rates, a city is marked green if new travel fluxes can be initiated at that time. The values of ϵ quantify the adjustment factor for the contact rate. The larger the value of ϵ the higher the contact rate.

Guiyang. That is to say, the strict travel restrictions can be lifted early or even be unnecessary in more cities if other mild interventions such as keeping social distance, wearing masks and improving hygiene can maintain the contact rate at a lower level. It is worth noting that the lifting of travel restrictions we are discussing here only implies that there is a non-zero movement of the population between the selected cities, not that the individuals can travel freely.

7.3.3 The level of new traffic flow

Although we have determined that where and when the lockdown exit strategy can be implemented by solving the optimization problem (Eq. 7.10), the details on how to start a new travel movement has not been provided yet. In order to obtain the level of optimal traffic flow from one city to another, we first calculate the potential $V(t_0, \Omega_i)$ for each selected city Ω_i at available time t_0 by solving the optimization problem (Eq. 7.14). Here we calculate potentials for the selected cities on 6 February (two weeks after the lockdown of Wuhan) with different adjustment factors. From the optimization (7.14), we know that the potential of each city is related to the growth rates and the numbers of infectious individuals in all selected cities, as well as the diffusion distance among them. In Fig 7.5A, we plot the estimated potential versus the growth rate, the logarithmic number of cases, and the average diffusion distance to other cities. Each of these dots represents a city. We can clearly see that, with an adjustment factor $\epsilon = 0.39$, the cities with larger growth rates and more COVID-19 cases tend to have higher potential, but the estimated potential does not show a clear correlation with diffusion distance. We further investigate the correlation between estimated potential and these three factors under different contact rates. From Fig 7.5B, we find that the estimated potential is positively correlated with the number of COVID-19 cases when the contact rate is relatively low and is positively correlated with the growth rate of COVID-19 cases when the contact rate is high. This result predicts that when the contact rate is kept at a relatively low level, the outward movement of population from regions with a large number of cases can help control the overall epidemic. And when the contact rate level is high, the outward movement of population from regions with high growth rate of COVID-19 cases helps to control the epidemic.

Once we have calculated the potential for each selected city, we define a net flow matrix $\boldsymbol{J}$ (Eq. 7.15) and then approximate the reaction diffusion-drift process (Eq. 7.6) into a patch model (7.16). Fig 7.4 shows that the estimated net movement matrix

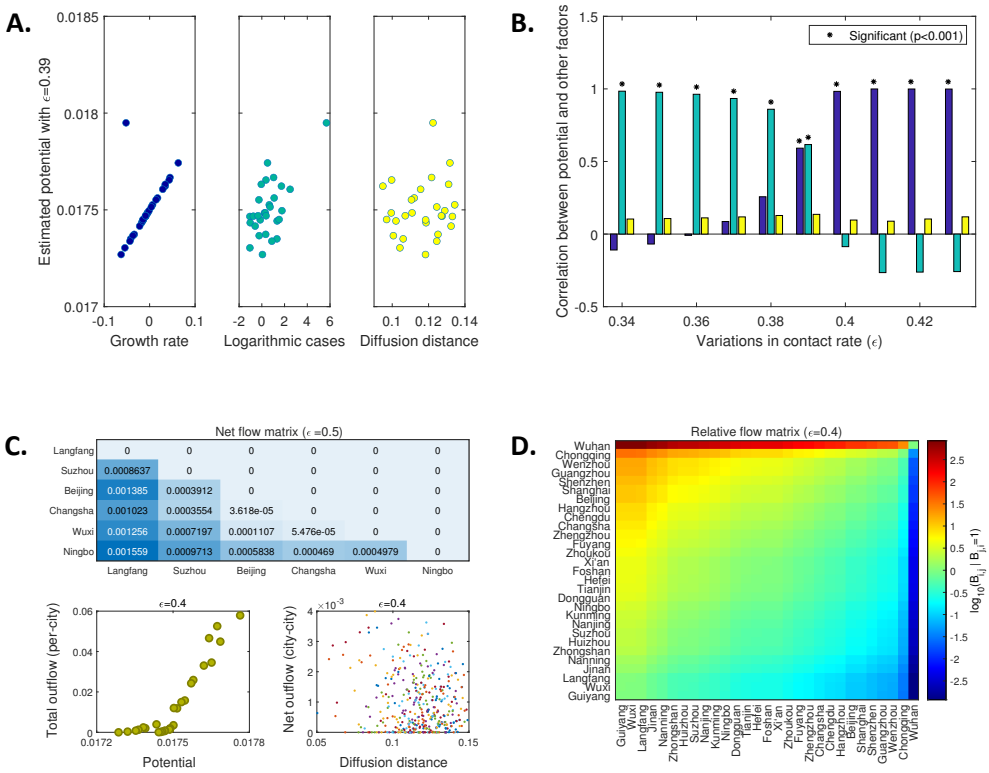


Figure 7.5: Estimated potential and flow matrices for the selected cities on February 6, 2020. (A) The estimated potential versus growth rate, COVID-19 cases and diffusion distance. (B) The correlation between estimated potential and other three factors (i.e., growth rate, COVID-19 cases and diffusion distance) which share the same color used in Fig A. (C) The reconstructed net flow matrix, and the relationship between outflow and potential and diffusion distance, respectively. The value in each cell (i, j) of the net flow matrix represents the movement rate from city j to city i, where i is the row label and j is the column label. (D) The relative flow matrix where the value in cell (i, j) is logarithmic travel flow from city j to city i under the condition that the travel flow from city i to city j is 1. The cities are ranked by numbers of COVID-19 cases.

between Langfang, Suzhou, Beijing, Changsha, Wuxi and Ningbo on 6 February 2020 when $\epsilon = 0.5$. Generally, a city with a higher potential also has more outflows (bottom left corner in Fig 7.5C), for example, Langfang compared with Ningbo. Although the potential does not show a significant correlation with the diffusion distance, it seems that the outflow population from one city is more likely to reach the closer city (bottom right corner in Fig 7.5C). Note that the net flow matrix is a lower triangular matrix which means that it only quantifies a one-way net flow of population from the place

with high potential to another one with low potential, and cannot reflect the two-way flow of population between the two cities in the actual situation. Here we give a two-way travel strategy through the balance equation (Eq. 7.17) without noise, which shows that the relative movement rate $B_{i,j}$ from city Ω_j to city Ω_i can be determined as long as the opposite movement $B_{j,i}$ is given. This provides a quantitative reference for the travel flux, that is, the direct travel flow from one city to another (or vice versa) can be limited to meet the balance equation, which reduces the risk of secondary outbreak of COVID-19 in these cities.

As a special case, we let the travel flow $B_{j,i}$ from city Ω_i to city Ω_j be 1, and then calculate the relative travel flow $B_{i,j}$ from city Ω_j to city Ω_i according to the balance equation (Eq. 7.17). We plot the calculated results in Fig 7.5D where the order of cities is ranked according to the number of COVID-19 cases. The value in each cell (i, j) of this heatmap is $log_{10}(B_{i,j}|B_{j,i} = 1)$ which is the relative travel flow from city Ω_j to city Ω_i after a logarithmic operation. The values in the triangular area above the antidiagonal of the heatmap are positive, while the values below the antidiagonal are negative, and the cells far away from the diagonal are associated with larger absolute values. This indicates that the travel flow from a city with a severe epidemic to a city with a mild epidemic should be smaller than the travel flow in the opposite direction, and the difference between two opposite travel flows increases as the numbers of COVID-19 cases in two cities have a larger difference. Consequently, the relative travel flow from Wuhan to other cities should be much less than the relative travel flow in the opposite direction during our research period. This conclusion is a result of our estimation and is obtained for a specific date, 6 February 2020. Note that our method can calculate the optimal population flows between selected cities at any specific time t_0 as long as this time meets the restrictions of the zero-one programming (Eq. 7.10).

7.3.4 Testing predictions

To test our lockdown exit strategy, we ran the patch model (7.16) with some movement matrices. As a baseline, we first sample from the estimated movement rates (Table 7.2) to form a basic movement matrix. Once the basic movement matrix is obtained, we replace the matrix $\boldsymbol{J}$ in the patch model (7.16) with it and run model (7.16) to obtain the predictions of the daily number of reported COVID-19 cases. In Fig 7.6A, we plot the simulations of COVID-19 cases in Guangzhou, where the involved movement is the basic movements matrix. With this movements pattern close to the real situation, the

epidemic in Guangzhou begins to rebound when the adjustment factor ϵ for the contact rate exceeds 0.35 (Fig 7.6A). On the one hand, it shows that unrestricted movements of the population are likely to cause a second COVID-19 epidemic in Guangzhou. On the other hand, it also reveals that the growth rate plays a core role in the development of the epidemic, and that prevention of the epidemic can also be achieved by reducing the effective contact rate instead of imposing a strict travel ban.

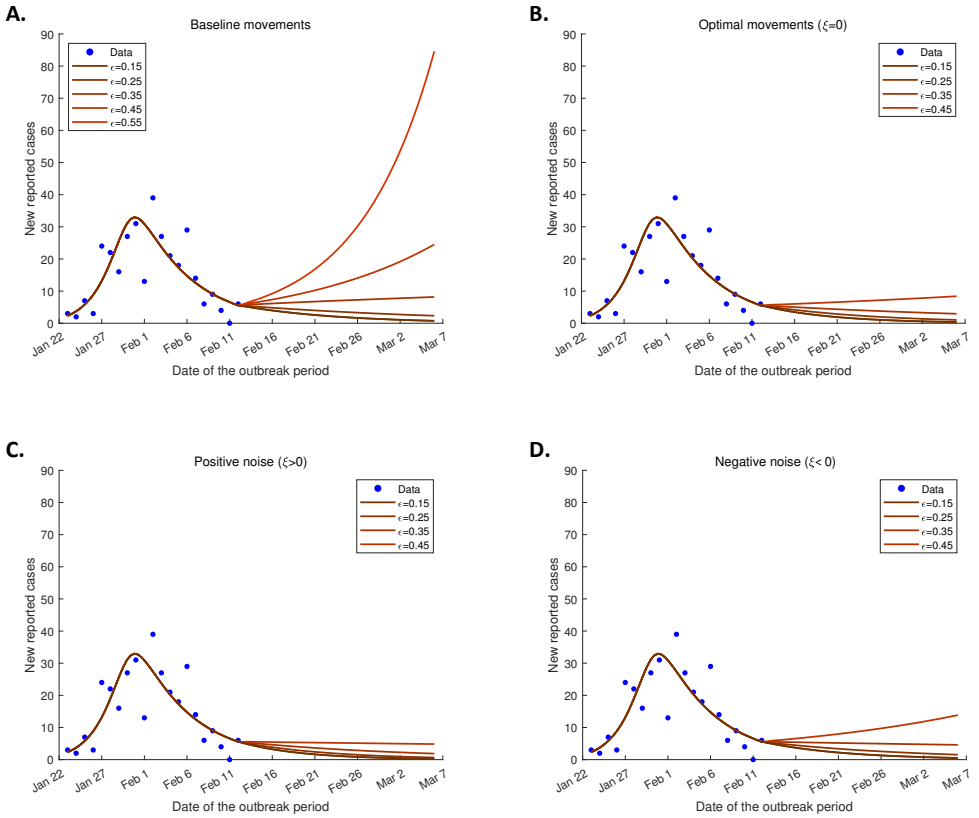


Figure 7.6: Testing predictions of the number of daily cases reported in Guangzhou under our lockdown exit strategy. (A) The baseline movements induce a secondary outbreak of COVID-19 in Guangzhou easily. (B) The travel pattern we present does not cause a secondary outbreak. (C-D) The simulations based on our travel pattern with positive and negative noise, respectively.

To compare our proposed travel pattern with the baseline movements, we simulate the patch model (7.16) with movement matrix $\boldsymbol{B}$ (Eq. 7.17). In Fig 7.6B, we set the interference term ξ in the balance equation (Eq. 7.17) to be 0, and then simulate

the time series of the daily number of reported COVID-19 cases with some acceptable variations in the contact rate. In the first part of our lockdown exit strategy, we already know that this strategy cannot be implemented in Guangzhou if the adjustment factor ϵ is not less than 0.47 (Fig 7.4). We test the acceptable situations and find that the risk of a secondary outbreak of COVID-19 in Guangzhou is greatly reduced according to our calculated movements (Fig 7.6B). The calculated optimal movements always satisfy a noise-free balance equation (Eq. 7.17) which actually requires the difference between the two-way movements of the population between two cities to be equivalent to the net flow. We further explore what will happen if the balance is broken by some random noise. Here we first consider low noise randomly sampled from the intervals $(0, 1)$ or $(-1, 0)$ as the value of ξ. In Fig 7.6C and 7.6D, we show the simulations with $\xi > 0$ (we call this positive noise) and $\xi < 0$ (negative noise), respectively. It seems that the positive noise contributes to the extinction of the COVID-19 epidemic in Guangzhou, while the negative noise increases the risk of the second outbreak. This further supports our idea about the lockdown exit strategy, that the population should move from the cities with higher potential to low-potential cities, because the movement matrix $\boldsymbol{B}$ with positive noise enhances the effect of the net flow matrix $\boldsymbol{J}$ while negative noise is resisting this pattern.

In order to further verify the effectiveness of the travel scheme that we have designed, we simulate the estimated time series of the daily number of reported new cases for all selected cities in our proposed travel pattern and the baseline travel pattern (Supplementary Fig S7.1). Similar to the simulations in Guangzhou, different travel patterns hardly affect the transmission of COVID-19 in all cities if the value of ϵ is relatively small, while with ϵ increasing, our proposed travel scheme has notable advantages in preventing the second outbreak of disease in the selected cities. These findings suggest that it is essential for traffic resumption to coincide with reducing the contact rate by wearing masks and maintaining social distancing. Moreover, our designed travel flux pattern will provide a flow scheme to avoid inducing the second outbreak even if the contact rate cannot reach the relatively low level. The simulations in Guangzhou (Fig 7.6C) show that a moderate increase in the movements flow toward the cities with low potential is beneficial to the prevention and control of the epidemic. However, the Supplementary Fig S7.2 shows that if the positive noise is too strong, the movements still increase the number of COVID-19 cases in many cities.

7.4 Discussion

With the improvement of the epidemic situation, the prevention and control strategies (lockdown, social distancing and others) in epidemic areas have been adjusted accordingly. In order to avoid the second outbreak of the epidemic, when, where and at what level individuals can flow normally are the critical problems that need to be decided. In this study, we redefined a measure of diffusion distance derived from the underlying mobility network instead of geographic distance and embedded it to a reaction diffusion equation to describe the spread of infectious disease in the population without interventions. We further extend the reaction diffusion model to a reaction diffusion-drift model with diffusion process and drift process of the population. On the basis of the extended spatiotemporal spread model, we developed a novel method to explore when and where and to what level travel flux can be triggered during the late stage of an epidemic using the divergence theorem and diffusion map theory, and under this designed movement pattern among various sub-regions the risk of a second outbreak is reduced.

In the case study, the data on a migration index from 1 to 22 January 2020 gave the top 25 cities with the highest emigration or immigration rates in mainland China. The mobility network among them is reconstructed from the population size and per capita GDP of each city using a gravity model [400]. We consequently integrated the multi-source data to our method and retrospectively determined the movement matrix of the population among the selected cities at the "correct" time. Specifically, the estimated growth rate and the number of infected individuals determine where and when the lockdown exit strategy can be implemented. The obtained travel flux of populations among the selected cities is actually the net flow of the population moved from a city with high potential to others with low potential. Finally, we test our travel scheme in the selected cities and find that initiating a new traffic flux among these cities according to our designed travel scheme does not cause a second outbreak of COVID-19 in these cities.

It is worth noting that the case study in this research is geographically specific, since the data for mainland China used here are more complete and regular than those for other countries, but we hope that the approaches we have developed are applicable more generally. Preparing for a responsible lockdown exit strategy [390] has been extremely important not only for the COVID-19 epidemic [391], but also for future pandemic outbreaks. Our method provides a possible reasonable travel flux scheme under which individuals movements cannot induce the second outbreak. There are

still some features and issues about our method to be discussed.

Note that the growth rate of the epidemic and the mobility network of the population are actually input variables in our method, and the estimates of them are part of the preparatory work. The estimation of the growth rate is not limited to a specific method. In the initial outbreak stages where almost all of the population are susceptible (i.e., $S(t)/N(t) \approx 1$), we simplified the incidence (Eq. 7.5) and then fitted the time series of the number of infected individuals using our model to obtain the estimates. For a general community where this simplification is not applicable, the growth rate can also be estimated from epidemic data using the back-tracking method [406].

Lacking available data, we reconstructed the mobility network using a gravity model [400], which takes into account that the links between cities are positively related to their economic activities [198] and population densities [395, 396], but negatively related to the geographic distance [398]. The gravity model provides a feasible approach to estimate the mobility among major cities in mainland China. It is interesting to note that the closer the economic links, the more the high speed trains and flights [416], which has broken the limits of physical distance on mobility. However, it may be better to use real mobility data when the data are available.

There are three free parameters involved in our method, the transfer step size l in the diffusion distance (Eq. 7.3), the dimension k of the embedding space in the Gaussian kernel (Eq. 7.11) and the diffusion rate γ in the optimization (Eq. 7.14). In the case study, we chose $l = 25$, $k = 2$ and $\gamma = 0.03$ which is the estimated $-\gamma_0$ (Table 7.2). We evaluated the variations of these parameters in a relatively large parameter space and found that the changes have little effect on our lockdown exit strategy. Specifically, the change of parameters k and γ almost has no effect on the estimated potential (Supplementary Fig S7.3A-B). Although the variation of l affects the value of the estimated potential in each city, the order of estimates has not changed (Supplementary Fig S7.3C). Cities with higher potential always have higher estimated potential, and the simulations of numbers of COVID-19 cases only rise slightly even under extremely large value of l (Supplementary Fig S7.3D).

7.5 Supporting information

In this section we present some supporting material, which can be accessed through https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009473.

7.5.1 Data preparation for gravity model

The gravity model that quantifies the population flow between any two cities is inspired by Newton's law of gravity. The resident population and GDP per capita in each city can be regarded as the "mass" of each city. The demographic data used in this study are listed in following table.

Table 7.1: The resident population and GDP per capita in selected cities. Note: GDP (CNY), population (ten thousand). Source: Statistical Yearbook of China 2019.

City	Wuhan	Beijing	Chengdu	Dongguan	Foshan	Fuyang
GDP/capita	135136	140211	94782	98939	127691	21589
Population	1108.1	2154.2	1633	839.22	790.57	820.7
City	Hefei	Huizhou	Jinan	Kunming	Langfang	Nanjing
GDP/capita	97470	85418	106302	76387	64906	152886
Population	808.7	483	746.04	685	483.66	843.62
City	Shenzhen	Suzhou	Tianjin	Wenzhou	Wuxi	Xi'an
GDP/capita	189568	173765	120710	72657	174270	85114
Population	1302.66	1072.17	1559.6	925	657.45	1000.37
City	Chongqing	Zhoukou	Hangzhou	Shanghai	Zhongshan	Guiyang
GDP/capita	65933	30817	140180	134982	110585	78449
Population	3101.79	867.78	980.6	2423.78	331	488.19
City	Ningbo	Zhengzhou	Guangzhou	Nanning	Changsha	
GDP/capita	132603	101349	155491	60626	134933	
Population	820.2	1013.6	1490.44	725.41	815.47	

Before applying the gravity model in main text, we first adjust the data to the same order of magnitude. Here we reduce the value of GDP by 100 times, and then use the mean of it and population size as the "mass" of each city. That is

$$M_i = \frac{1}{2}(GDP_i/100 + Population_i), \tag{S.1}$$

where GDP_i and $Population_i$ are the GDP per capita and resident population in city Ω_i, respectively.

7.5.2 The ODE models for parameter estimation

Assume that the infected individuals and net growth rate are homogeneous in each city Ω_i, we integrate the original model (7.4) on the initial outbreak city Ω_0, and get

$$\frac{\mathrm{d}I(t,\Omega_0)}{\mathrm{d}t} = \gamma \int_{\Omega_0} \Delta p(t,x)dx + R(t,\Omega_0)I(t,\Omega_0). \tag{S.2}$$

The divergence theorem derives

$$\int_{\Omega_0} \Delta p(t,x)dx = \int_{\partial\Omega_0} \nabla p(t,x) \cdot d\boldsymbol{S}(x). \tag{S.3}$$

With a Dirac vector function $\boldsymbol{\delta}(x)$ on the boundary $\partial\Omega_0$, the gradient of $p(t,x)$ on the discontinuous boundary $\partial\Omega_0$ is

$$\nabla p(t,x) = \boldsymbol{\delta}(x)(p(t,\Omega_0^{outside}) - p(t,\Omega_0^{inside})). \tag{S.4}$$

In the early stages of the COVID-19 outbreak in China, the epidemic in Wuhan is far worse than other cities. Thus, we approximate the gradient as

$$\nabla p(t,x) = -\boldsymbol{\delta}(x)I(t,\Omega_0)/|\Omega_0|. \tag{S.5}$$

Consequently,

$$\int_{\Omega_0} \Delta p(t,x)dx = -\frac{\int_{\partial\Omega_0} \boldsymbol{\delta}(x) \cdot d\boldsymbol{S}(x)}{|\Omega_0|} I(t,\Omega_0) = -\frac{|\partial\Omega_0|}{|\Omega_0|} I(t,\Omega_0). \tag{S.6}$$

Therefore, we get a ODE model for the epidemic in the city Ω_0

$$\frac{\mathrm{d}I(t,\Omega_0)}{\mathrm{d}t} = \gamma_0 I(t,\Omega_0) + R(t,\Omega_0)I(t,\Omega_0), \tag{S.7}$$

where $\gamma_0 = -\gamma|\partial\Omega_0|/|\Omega_0|$ is the total outflow rate from the initial outbreak city Ω_0.

Before the lockdown in Wuhan, the infected individuals mainly spread from Wuhan to other cities. We generalize the ODE model for city Ω_0 to a general form for each city Ω_i, yields

$$\frac{\mathrm{d}I(t,\Omega_i)}{\mathrm{d}t} = \gamma_i I(t,\Omega_0) + R(t,\Omega_i)I(t,\Omega_i), \tag{S.8}$$

where γ_i is the movement rate of population form the initial outbreak city Ω_0 to

another city Ω_i. The parameters of model (S.8) are summarized in Table 7.2.

Table 7.2: The Estimated epidemiological parameters of each city are listed in this table.

City(Ω_i)	Parameter			
	$\Lambda(\Omega_i)$	$\Gamma(\Omega_i)$	$\tau(\Omega_i)$	$\gamma(\Omega_i)$
Wuhan	0.482	0.238	19.50	$-3.0 * 10^{-2}$
Beijing	0.370	0.182	9.732	$7.868 * 10^{-5}$
Chengdu	0.492	0.142	4.693	$2.028 * 10^{-4}$
Dongguan	0.455	0.180	11.461	$6.241 * 10^{-5}$
Foshan	0.516	0.154	5.718	$6.664 * 10^{-5}$
Fuyang	0.507	0.189	9.423	$4.226 * 10^{-5}$
Guangzhou	0.510	0.161	7.045	$2.581 * 10^{-4}$
Guiyang	0.513	0.205	12.762	$1.860 * 10^{-5}$
Hangzhou	0.477	0.167	5.483	$2.012 * 10^{-5}$
Hefei	0.459	0.208	10.59	$3.997 * 10^{-5}$
Huizhou	0.517	0.181	8.546	$2.434 * 10^{-5}$
Jinan	0.522	0.208	7.756	$3.538 * 10^{-5}$
Kunming	0.500	0.155	4.556	$1.079 * 10^{-4}$
Langfang	0.471	0.197	7.650	$4.388 * 10^{-4}$
Nanjing	0.500	0.204	9.636	$3.443 * 10^{-4}$
Nanning	0.478	0.206	10.219	$3.551 * 10^{-5}$
Ningbo	0.434	0.228	9.725	$2.103 * 10^{-4}$
Shanghai	0.407	0.170	7.561	$9.118 * 10^{-5}$
Shenzhen	0.481	0.200	8.264	$6.091 * 10^{-4}$
Suzhou	0.451	0.207	8.510	$9.565 * 10^{-5}$
Tianjin	0.503	0.196	10.768	$2.081 * 10^{-5}$
Wenzhou	0.515	0.199	7.156	$1.00 * 10^{-3}$
Wuxi	0.470	0.234	11.847	$2.750 * 10^{-5}$
Xi'an	0.475	0.191	9.407	$5.005 * 10^{-5}$
Changsha	0.395	0.195	9.079	$2.469 * 10^{-4}$
Zhengzhou	0.469	0.176	7.868	$6.269 * 10^{-5}$
Zhongshan	0.494	0.198	9.983	$3.465 * 10^{-5}$
Chongqing	0.387	0.120	6.304	$1.725 * 10^{-5}$
Zhoukou	0.535	0.143	4.327	$2.036 * 10^{-4}$

7.5.3 Source-sink method

Based on the behavior-related growth rate $R(t_0, \Omega_i, \epsilon)$, we distinguish the subregions into source subregion with $R(t_0, \Omega_i, \epsilon) > 0$ and sink subregion with $R(t_0, \Omega_i, \epsilon) < 0$. After getting the estimated potential $V(\Omega_1) > V(\Omega_2) > ... > V(\Omega_m)$, we first give a basic matrix $\widetilde{J}$ with the element

$$\widetilde{J}_{i,j} = max\big\{\big(V(\Omega_j) - V(\Omega_i)\big)/D(\Omega_i, \Omega_j), 0\big\}, \tag{S.9}$$

where $D(\Omega_i, \Omega_j)$ is the diffusion distance between Ω_i and Ω_j. And then, we test the patch model (7.16) using this matrix $\widetilde{J}$ as a mobility matrix. For every $i = 1, ..., m$,

we calculate the right term of patch model (7.16). For

$$\sum_{j=1}^{m}\left(\widetilde{J}_{i,j}I(t_0,\Omega_j)-\widetilde{J}_{j,i}I(t_0,\Omega_i)\right)+R(t_0,\Omega_i,\epsilon)I(t_0,\Omega_i)>threshold, \tag{S.10}$$

we orderly set $\bar{J}_{i,j}=0$ $(j=1,...,m)$ if Ω_j is a sink subregion until the left term of the inequality (S.10) is non-positive or all sink subregions have been traversed, and $\bar{J}_{i,j}=\widetilde{J}_{i,j}$ otherwise. In this way, we reduce the risk of a short-term outbreak of the epidemic in subregion Ω_i due to the population flow from the sink subregion Ω_j. To further reduce the rebound probability of infectious diseases, we traverse source subregions and repeat the above process. Larger *threshold* in (S.10) means higher population flow following the gradient of potential V. We set $threshold=0.1$ in our study.

With the updated matrix $\bar{J}=\{\bar{J}_{i,j}\}$, we calculate the right term of patch model (7.16) again. For every $i=1,...,m$, if

$$\sum_{j=1}^{m}\left(\bar{J}_{i,j}I(t_0,\Omega_j)-\bar{J}_{j,i}I(t_0,\Omega_i)\right)+R(t_0,\Omega_i,\epsilon)I(t_0,\Omega_i)<0, \tag{S.11}$$

we denote

$$\vartheta_i=\frac{\sum_{j=1}^{m}\bar{J}_{j,i}I(t_0,\Omega_i)-R(t_0,\Omega_i,\epsilon)I(t_0,\Omega_i)}{\sum_{j=1}^{m}\bar{J}_{i,j}I(t_0,\Omega_j)+eps}, \tag{S.12}$$

where $eps>0$ is small enough to ensure that the denominator is not zero. And then, we update

$$J_{i,j}=\min\{\vartheta_i\bar{J}_{i,j},\widetilde{J}_{i,j}\}. \tag{S.13}$$

In this process, we promote the movement of population to subregions with low potential, while ensuring that the elements of the mobility matrix do not exceed the basic matrix to avoid population gathering in specific areas.

7.5.4 Theory Supplement

Here, we list the theoretical results used to determine the available subregions for reopening and travel fluxes in these epidemic areas. Lemma 7.5.2 refers to the work of Coifman and Lafon [402].

Lemma 7.5.1. *For any continuous (piecewise continuous) function $g(x)$ in the measure space $(X,\mathcal{A},\mu)$ with $\int_\Omega g(x)d\mu(x)\le 0$, there is a integrable function $h(x)$ with $\int_\Omega h(x)d\mu(x)=0$, such that $g(x)+h(x)\le 0$ for any $x\in\Omega\subset X$.*

Proof. We decompose the function $g(x)$ by

$$g(x) = g_+(x) + g_-(x)$$

with $g_+(x) \geq 0$ and $g_-(x) \leq 0$.

Let

$$A = \int_\Omega g_+(x)d\mu(x), \quad B = \int_\Omega g_-(x)d\mu(x).$$

If $B = 0$ (i.e., $g_-(x) = 0$), then $g(x) = 0$. We take $h(x) = 0$.

If $B < 0$, then $\frac{A}{B} + 1 \geq 0$. We take $h(x) = \frac{A}{B}g_-(x) - g_+(x)$. □

Lemma 7.5.2. *Assume that the data set $\{x_1, x_2, ..., x_m\}$ are obtained from a probability density $f(x)$ supported on manifold $\Omega \subset \mathbb{R}^k$. Given a rotation-invariant kernel $k_\delta(x, y) = \left(\frac{\delta}{\pi}\right)^{k/2} e^{-\delta\|x-y\|^2}$ on the manifold $\Omega \subset \mathbb{R}^k$. Set $b_\delta(x_i) = \sum_{j=1}^m k_\delta(x_i, x_j)$, $k_{\alpha,\delta}(x_i, x_j) = \frac{k_\delta(x_i,x_j)}{b_\delta^\alpha(x_i)b_\delta^\alpha(x_j)}$, $d_{\alpha,\delta}(x_i) = \sum_{j=1}^m k_{\alpha,\delta}(x_i, x_j)$, and $P_{\alpha,\delta}(x_i, y_j) = \frac{k_{\alpha,\delta}(x_i,x_j)}{d_{\alpha,\delta}(x_i)}$ is the transition kernel of matrix P_m, then the matrix $L_{\delta,m} = 4\delta(P_m - E)$ converges to a Kolmogorov operator $\mathcal{L}_\alpha$, i.e., $\lim_{\delta,m\to\infty} L_{\delta,m}\phi = \mathcal{L}_\alpha\phi = 2(1-\alpha)\nabla(\ln f)\cdot\nabla\phi + \Delta\phi$ for any smooth function $\phi \in C^2(\Omega)$.*

Proof. When the sample size m are large enough, the law of large numbers yeilds

$$b_\delta(x) = \lim_{m\to\infty} \sum_{j=1}^m k_\delta(x, x_j) = \int_\Omega k_\delta(x, y)f(y)dy.$$

Using the Taylor expansion of $f(y)$, we have

$$b_\delta(x) - f(x) = \int_\Omega k_\delta(x, y)\Big((y-x)\nabla f + \frac{1}{2}(y-x)H(y-x)^T + O(\|y-x\|^3)\Big)dy,$$

where $\nabla f = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_k})^T$ is the gradient of f and H is the corresponding Hessian.

First, for $i \neq j$,

$$\int_\Omega k_\delta(x, y)(y-x)_i dy = 0, \quad \int_\Omega k_\delta(x, y)(y-x)_i(y-x)_j dy = 0. \tag{S.14}$$

Second, for $i = j$,

$$\int_\Omega k_\delta(x, y)(y-x)_i^2 dy = \Big(\frac{\delta}{\pi}\Big)^{k/2} \int_\Omega (y-x)_i^2 e^{-\delta\|x-y\|^2} dy = \frac{1}{2\delta}. \tag{S.15}$$

Further, we have

$$\int_{\Omega} k_{\delta}(x,y)\|y-x\|^3 dy = \Big(\frac{\delta}{\pi}\Big)^{k/2} \int_{\Omega} \|y-x\|^3 e^{-\delta\|x-y\|^2} dy = O(\delta^{-\frac{3}{2}}). \tag{S.16}$$

These equations (S.14-S.16) immediately imply that

$$\begin{aligned} b_{\delta}(x) &= \int_{\Omega} k_{\delta}(x,y) f(y) dy = f(x) + \frac{1}{4\delta} \sum_{i=1}^{k} \frac{\partial^2 f(x)}{\partial^2 x_i} + o(\frac{1}{\delta}) \\ &= f(x) + \frac{1}{4\delta} \Delta f(x) + o(\frac{1}{\delta}). \end{aligned} \tag{S.17}$$

Similarly, we can calculate $d_{\alpha,\delta}(x)$ by

$$\begin{aligned} d_{\alpha,\delta}(x) &= \lim_{m\to\infty} \sum_{j=1}^{m} k_{\alpha,\delta}(x_i, x_j) = \int_{\Omega} k_{\alpha,\delta}(x,y) f(y) dy \\ &= \frac{1}{b_{\delta}^{\alpha}(x) f^{1-\alpha}(x)} \Big[1 + \frac{1}{4\delta} \Big(\frac{\Delta(f^{1-\alpha}(x))}{f^{1-\alpha}(x)} - \alpha \frac{\Delta f(x)}{f(x)} \Big) + o(\frac{1}{\delta}) \Big]. \end{aligned} \tag{S.18}$$

Consequently, the shift operator $P_{\alpha,\delta}$, which is the limitation of the transition matrix P_m for infinite samples, has the asymptotic expansion as

$$\begin{aligned} P_{\alpha,\delta}\phi(x) &= \frac{1}{d_{\alpha,\delta}(x)} \int_{\Omega} k_{\alpha,\delta}(x,y)\phi(y) f(y) dy \\ &= \phi(x) + \frac{1}{4\delta} \Big[\frac{\Delta(\phi(x) f^{1-\alpha}(x))}{f^{1-\alpha}(x)} - \phi(x) \frac{\Delta f^{1-\alpha}(x)}{f^{1-\alpha}(x)} \Big] + o(\frac{1}{\delta}) \end{aligned}$$

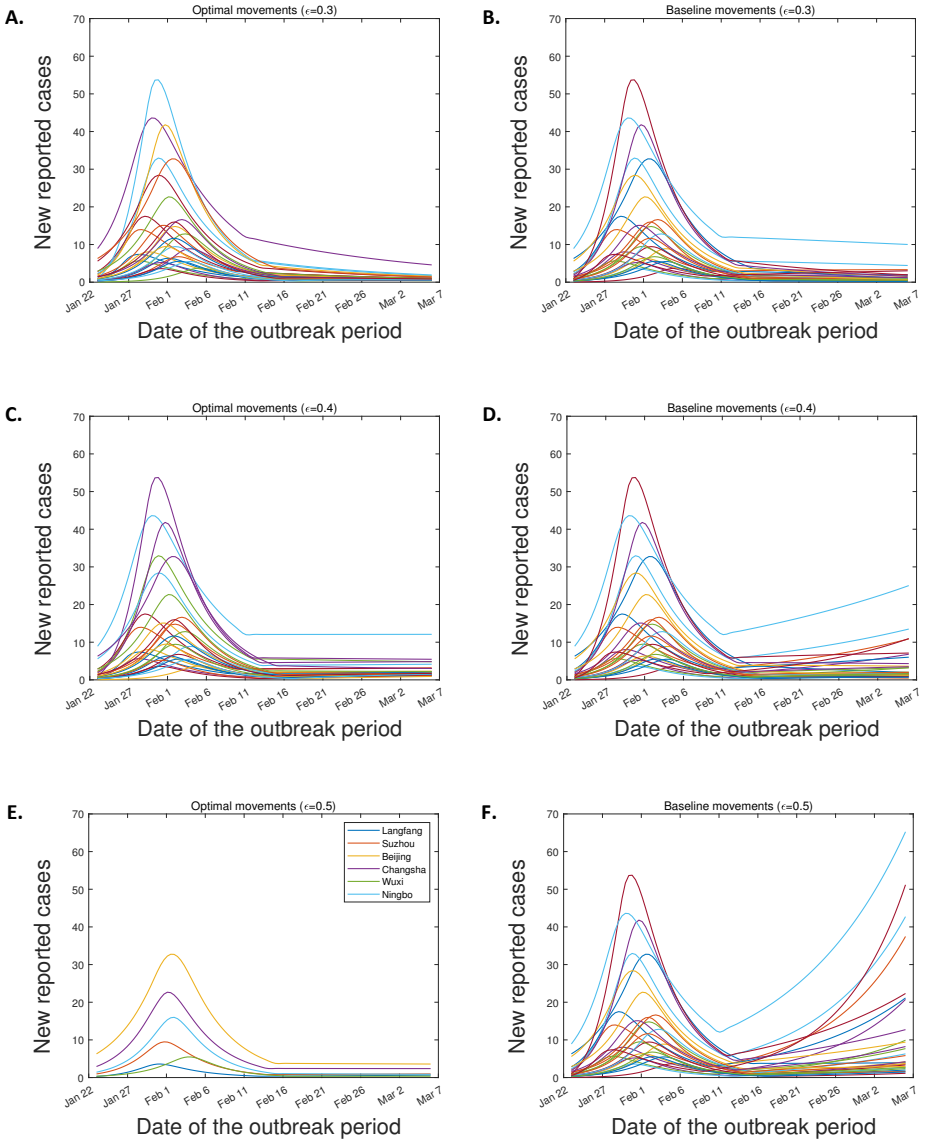
since

$$\begin{aligned} & k_{\alpha,\delta}(x,y)\phi(y) f(y) dy \\ =& \frac{1}{b_{\delta}^{\alpha}(x) f^{1-\alpha}(x)} \Big[\phi(x) + \frac{1}{4\delta} \Big(\frac{\Delta(\phi(x) f^{1-\alpha}(x))}{f^{1-\alpha}(x)} - \alpha\phi(x) \frac{\Delta f(x)}{f(x)} \Big) + o(\frac{1}{\delta}) \Big]. \end{aligned}$$

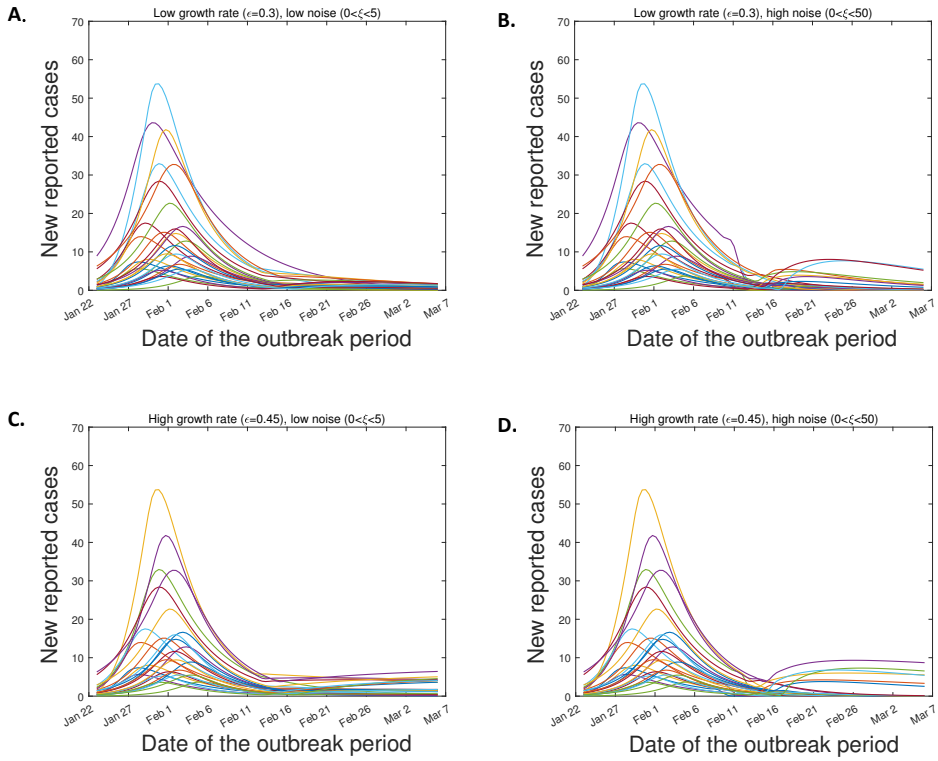
Therefore, we have the infinitesimal generator $\mathcal{L}_{\alpha}$ of the diffusion as

$$\mathcal{L}_{\alpha}\phi = \lim_{\delta\to\infty} 4\delta(P_{\alpha,\delta} - E)\phi = \frac{\Delta(\phi f^{1-\alpha})}{f^{1-\alpha}} - \frac{\Delta(f^{1-\alpha})}{f^{1-\alpha}}\phi = 2(1-\alpha)\nabla(\ln f)\cdot\nabla\phi + \Delta\phi.$$

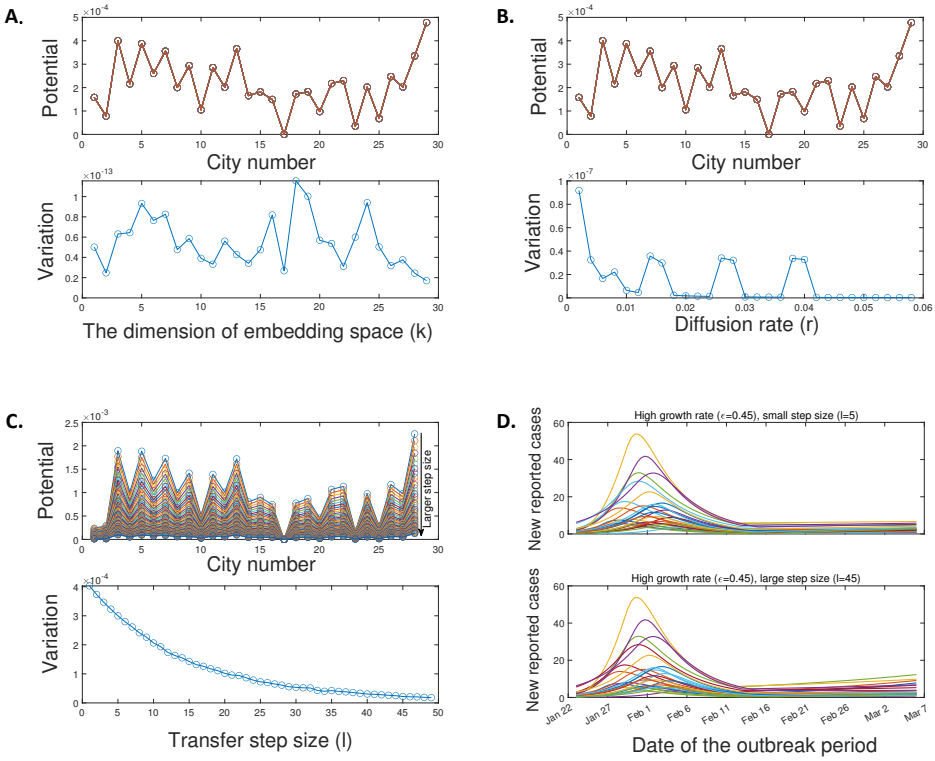
□



Supplementary Fig 7.1: Predictions of reported numbers of daily cases in all cities. The sub-figures in the left column are the simulations in our movement pattern. The sub-pictures in the right column are the simulations in the baseline movement pattern. The simulations in Wuhan are not shown in the graphics window.



Supplementary Fig 7.2: Predictions of reported daily cases in all cities. The impact of variations in noise and contact rates on our lockdown exit strategy. The simulations in Wuhan are not shown in the graphics window.



Supplementary Fig 7.3: The sensitivity of the estimated potential to three free parameters. A-C: The estimated potential function and the variation of potential with changes in the parameters. The variation is the distance between two potential functions with adjacent parameter values. For instance, the variation at $k = 5$ is the distance between the potential function with $k = 5$ and the potential function with $k = 6$. Little variation means that the parameter change has little effect on the calculation of the potential function. **D:** Simulations of reported numbers of daily cases in all cities. The adjustment factor $\epsilon = 0.45$ and the potential are minimum-centered.