Infectious Disease Modelling 7 (2022) 212–230

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Modeling the early transmission of COVID-19 in New York and San Francisco using a pairwise network model

Shanshan Feng [a, 1], Xiao-Feng Luo [a, 1], Xin Pei [b], Zhen Jin [c, d, *], Mark Lewis [e, f], Hao Wang [e, **]

[a] Department of Mathematics, North University of China, Taiyuan, Shanxi, 030 051, China
[b] College of Mathematics, Taiyuan University of Technology, Shanxi, Taiyuan, 030 024, China
[c] Complex System Research Center, Shanxi University, Taiyuan, 030 006, Shanxi, China
[d] Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan, 030 006, Shanxi, China
[e] Department of Mathematics and Statistics Sciences, University of Alberta, Edmonton, Alberta, T6G 2G1, Canada
[f] Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada

ARTICLE INFO

Article history:
Received 31 October 2021
Received in revised form 18 December 2021
Accepted 29 December 2021
Available online 5 January 2022
Handling editor: Dr Lou Yijun

Keywords:
COVID-19
Social network
Quarantine
Social distance
Clustering coefficient

ABSTRACT

Classical epidemiological models assume mass action. However, this assumption is violated when interactions are not random. With the recent COVID-19 pandemic, and resulting shelter in place social distancing directives, mass action models must be modified to account for limited social interactions. In this paper we apply a pairwise network model with moment closure to study the early transmission of COVID-19 in New York and San Francisco and to investigate the factors determining the severity and duration of outbreak in these two cities. In particular, we consider the role of population density, transmission rates and social distancing on the disease dynamics and outcomes. Sensitivity analysis shows that there is a strongly negative correlation between the clustering coefficient in the pairwise model and the basic reproduction number and the effective reproduction number. The shelter in place policy makes the clustering coefficient increase thereby reducing the basic reproduction number and the effective reproduction number. By switching population densities in New York and San Francisco we demonstrate how the outbreak would progress if New York had the same density as San Francisco and vice-versa. The results underscore the crucial role that population density has in the epidemic outcomes. We also show that under the assumption of no further changes in policy or transmission dynamics not lifting the shelter in place policy would have little effect on final outbreak size in New York, but would reduce the final size in San Francisco by 97%.

* Corresponding author. Complex System Research Center, Shanxi University, Taiyuan, 030 006, Shanxi, China.
** Corresponding author.
E-mail addresses: jinzhn@263.net (Z. Jin), hao8@ualberta.ca (H. Wang).
Peer review under responsibility of KeAi Communications Co., Ltd.
[1] The first two authors are contributed equally.

https://doi.org/10.1016/j.idm.2021.12.009

1. Introduction

The number of infected cases and deaths from the coronavirus disease 2019 (COVID-19) continues to increase (https://www.who.int/.; Li et al., 2020a). Since the first case of COVID-19 was confirmed in the United States on January 22th, 2020, the epidemic has spread in North America (https://www.who.int/.). There have been more than 16,000 daily new cases almost every day in the United States since March 23rd, 2020, and the number of total confirmed cases has continued to grow. In this paper, we investigate the early COVID-19 infections in the United States, explore the factors that are important in determining disease outcomes. Our approach is to consider two major US cities with very different infection histories: New York and San Francisco. We compare and contrast the dynamics in these two cities, with a focus on determining the demographic and behavioral factors driving the disease outbreak.

Coronavirus is transmitted mainly through close contact and respiratory droplets, with possible airborne transmission if aerosols are present (Chan et al., 2020; Huang et al., 2020). Available evidence shows that the transmission can arise from symptomatic, pre-symptomatic (exposed) and asymptomatic people infected with COVID-19 (Kimball et al., 2020; Ong et al., 2020; Pan et al., 2020; Yu et al., 2020). Symptoms include cough, fever and difficulty breathing, as with SARS (Severe acute respiratory syndrome coronavirus) and MERS (Middle East respiratory syndrome coronavirus) (Wu et al., 2020). The incubation period for COVID-19, is on average 5–6 days, although it can take up to 14 days. During this "pre-symptomatic" period, some infected people can be contagious. Therefore, transmission from pre-symptomatic cases can occur before symptom onset. It is also possible for infected people to remain asymptomatic while still transmitting the disease.

With the advent of COVID-19 there has been an explosion in the number of mathematical models to study its dynamics. Many of the early models involve analysis of the COVID-19 outbreak in China. A deterministic compartmental model, based on the clinical progression of the disease, epidemiological status of the individuals, and intervention measures to estimate the transmission risk of COVID-19 at the early stage of epidemic was proposed and analyzed by Tang et al. (Tang et al., 2020a, 2020b). Compartmental models have also been augmented with artificial intelligence (AI) approaches to predict the epidemics trend of COVID-19 in China (Yang et al., 2020).

As with other diseases, quarantining plays a major role in the COVID-19 outcomes. Hu et al. modelled quarantine strategies for COVID-19 at different population densities to show how this factor affects the disease outbreak in Guangdong, China (Hu et al., 2020). The effectiveness of quarantine and isolation in China was also discussed in some papers (Hou et al., 2020; Tang et al., 2020c).

One of the main policy strategies for addressing high COVID-19 infection levels is to use social distancing. The effects of such policies have been evaluated by mathematical models. Leung et al. used a susceptible-infectious-susceptible (SIS) model to show the potential effects of relaxing containment measures after the first wave of infection, in anticipation of a possible second wave (Leung et al., 2020). The impact of mixed social distancing and school closure on disease transmission was studied by Zhang et al. and it was found that social distancing alone is sufficient to control COVID-19 in China (Zhang et al., 2020). In other jurisdictions, Anderson et al. estimated the impact of physical distancing on the disease contact rate in British Columbia, Canada (Anderson et al., 2020).

One of the early models that most impacted US policy has come from Imperial College, London (Ferguson et al., 2020). There an individual-based model was used to explore scenarios for COVID-19 in Great Britain and the US, and it was shown that the health care system would be easily overwhelmed without sufficient suppression of the disease via social distancing, home isolation and household quarantine. To fight the COVID-19 pandemic, the United States has adopted a series of the above control measures. These including restricting travel and quarantining close contacts of those with COVID-19, as well as social distancing via advising people to stay at home and by keeping more than 2 m distance from others (https://www.cdc.gov/coron, 2019; https://news.sina.com.cn/, 2020; http://yl.szhk.com/2020/0, 2020).

Among these mathematical models for studying the transmission dynamics of COVID-19, the division of compartments varies. The simplest is to classify individuals as susceptible, infectious or recovered (Roda et al., 2020). It is more realistic to classify individuals as susceptible, exposed, infectious, and recovered. Considering that asymptomatic people infected with COVID-19 can transmit the disease, the infectious are further divided two types: the asymptomatic and the symptomatic (Tang et al., 2020a, 2020b). Given the policy of quarantining the close contacts with the infectious, people in different disease states are further divided into the unquarantined and the quarantined (Tang et al., 2020a, 2020b). To forecast the transmission of COVID-19 in the United States, it is more realistic to classify individuals as susceptible, exposed, asymptomatic infectious, symptomatic infectious, and recovered with considering of quarantine. Except for the compartments, a key factor in the transmission process is the number of contacts, since a susceptible person is infected by an infectious person at the same transmission rate per unit time in the same place. In a well-mixed compartment model, the number of contacts is constant (Roda et al., 2020), which is ideal and can not reflect the clustering effect of population. Although some compartment models considered that the number of contacts was time-varying (Tang et al., 2020a), it is assumed and man-made. The broken links caused by quarantine are not clear in these compartment models. In a social network, nodes represent individuals and links between nodes represent the contacts between two connected individuals. Network models consider the heterogeneity of contact structures, and use the degrees of nodes to reflect the number of contacts. Xue et al. used a degree-based network model to study the transmission of COVID-19 in Wuhan, Toronto and Italy (Xue et al., 2020). However, due to the policies of quarantine and reopening, the distribution of degree should be time-varying. The pairwise model regards the number of links as variables, and simultaneously studies the dynamic behavior of nodes and links between nodes, which effectively reflects human social behavior, and has natural advantages in studying quarantine and social distance. Quarantining the close

contacts and keeping social distance lead to a decrease in the number of links, while resumption of work and school leads to an increase in the number of links. In fact, epidemiological surveys of close contacts of infected persons are based on their social networks. Compared with simple SIR or SEIR models, our pairwise network model is more likely to follow the realistic scenarios during the COVID-19 epidemic and to provide better quantitative results. By explicitly incorporating the effects of family cluster and contact tracing followed by household quarantine and isolation, our model provides a good fit to the trajectory of COVID-19 infections and rational prediction for the epidemic trend.

Here we focus our study on two different cities in the United States: New York and San Francisco. These two cities are chosen based on the following considerations. New York and San Francisco are two large United States cities, one initially hard-hit, the other initially not so hard hit. By looking at the differences between the two large cities we can try to understand factors that are important in determining disease outcomes. New York city includes 5 core counties: New York county (Manhattan), Kings county (Brooklyn), Bronx county (the Bronx), Richmond county (Staten Island), and Queens county (Queens). San Francisco city includes the greater area with 5 core counties: San Francisco, Alameda, Marin, Contra Costa, and San Mateo County.

In this work we build a pairwise network model, based on methods in Ref. (Luo et al., 2021), to investigate the spread of COVID-19 in New York and San Francisco. We use data of COVID-19 infections collected from Ref. (https://covidmapper.ca/ge) to identify the parameters of the proposed model and calculate some common descriptors of epidemic timing and impact, including the basic reproduction number and the effective reproduction number. The sensitivity analysis shows that there is a highly negative correlation between the clustering coefficient and the basic reproduction number. Comparing the transmission in New York and San Francisco, we find that population density has a major effect on disease transmission. When the population density of New York drops to the same as San Francisco, the epidemic duration will be shortened by more than 140 days, and the final size will be reduced by more than 50%. When the population density of San Francisco increases to the same as New York, the final size will be increased 15-fold. Not lifting the shelter in place policy has a little effect on final size in New York, but will reduce the final size in San Francisco by 97%.

2. Materials and methods

2.1. Data collection and description

We collect all the daily reported data on COVID-19 infections, including the laboratory confirmed cases, and death-caused by COVID-19 in two typical cities in the United States: New York city and San Francisco city (https://covidmapper.ca/ge). The data is from March 27th, 2020 to July 20th, 2020. We also collect population sizes and areas of these two cities (https://en.wikipedia.org/a; https://en.wikipedia.org/b; https://en.wikipedia.org/c; https://en.wikipedia.org/d; https://en.wikipedia.org/e; https://en.wikipedia.org/f). The population and area of New York are 8,336,697 and 785 square kilometers, respectively and those of San Francisco are 4,264,934 and 6405.5 square kilometers. This means that the population density of New York is about 16 times as dense as that of San Francisco.

In our social network, node represents individual and one link between two individuals describes that such two individuals spend more than 10 minutes at a distance of less than 2 m (https://www.cdc.gov/coron, 2019). The clustering coefficient denoted by φ, one of the network characteristic parameters, can characterize the isolation intensity of "stay at home" order (also called a shelter in place policy) (Luo et al., 2021). New York had a shelter in place policy from March 22nd to May 28th (https://rt.live/us/), and San Francisco had a shelter in place policy from March 19th to May 25th (https://rt.live/us/). The more intensive the household quarantine measure is, the larger the clustering coefficient is. According to Ref. (https://www.census.gov/da), the number of family members varies from 1 person to more than 7 persons in the United States. Since there is no data available on family structure in New York and San Francisco, we use the distribution of household numbers in the United States to describe the distribution in each of the two cities. The ideal case is that every family is completely isolated. In this case, every family is a complete graph.

Mathematically the clustering coefficient φ of a network is defined as the number of triangles over the number of triples in the network (Keeling, 1999). Therefore, in this ideal case, φ behaves as

$$\varphi = \frac{3\sum_{k=3}^{7} C_k^3 F_k}{\sum_{k=3}^{7} A_k^3 F_k} = \frac{\sum_{k=3}^{7} \left(\frac{k}{3}\right) F_k}{2\sum_{k=3}^{7} \left(\frac{k}{3}\right) F_k} = 0.5, \tag{1}$$

where F_k is the number of families each with k family members, C represents combination and A represents permutation. When the number of links with the outside for every family increases by three, ignoring the few addition of triangles, we obtain

$$\varphi = \frac{3\sum_{k=3}^{7} C_k^3 F_k}{\sum_{k=3}^{7} A_k^3 F_k + (6F_k + 6)(k-1)} = 0.297\,5. \tag{2}$$

We consider this case in the following.

2.2. The pairwise model for COVID-19 epidemic

To study the COVID-19 epidemic spread in the United States, based on Luo et al. (Luo et al., 2021), we propose a pairwise epidemic model with family clusters and quarantine of individuals having been contacted with the confirmed cases. As demonstrated in Fig. 1, individuals are grouped into 12 states: unquarantined susceptible (S), quarantined susceptible (S_0), unquarantined incubation (E), quarantined incubation (E_0), unquarantined asymptomatic (A), quarantined asymptomatic (A_0), confirmed symptomatic (I), hospitalized (H), recovered asymptomatic (R^A), quarantined recovered asymptomatic (R_0^A), recovered confirmed symptomatic (R^I) and died while recovering (R^d). The individuals in states of E, A, I and H are infectious, but H individuals do not transmit the disease as they have been isolated in hospital. The transitions between different states are listed as follows. Each S individual moves into the E state when it is exposed to the disease at the transmission rate β_1 by an E individual, at rate β_2 by an A individual and at rate β_3 by an I individual. After an average incubation period of $1/q$ days, each E individual moves to I state with probability p or to A state with probability $1-p$. The individuals in state A enter into the R^A compartment at rate γ_2, while the individuals in state I move into either the H compartment with proportion ϵ after $1/\gamma_4$ days or the R_I compartment with proportion $1-\epsilon$ after $1/\gamma_1$ days. When the I individuals are in the R^I compartment, they will reconnect k links on average with the individuals in states S, E, A, R^A or R^I. After an average period of $1/\gamma_5$ days, the H individuals move into R^I compartment with proportion $1-d$ or R^d compartment with proportion d. When the H individuals move into R^I compartment, they will reconnect k links on average with the individuals in state S, E, A, R^A or R^I. Meanwhile, the individuals in states S, E, A or R^A will be quarantined with probability η and move to S_0, E_0, A_0 or R_0^A, once any one of their neighbors is confirmed to be infected. Accordingly, the individuals in state S_0, E_0, A_0 or R_0^A will go back to the S, E, A or R^A compartment after the quarantine period of average length $1/\gamma_3$ days. At the same time, they will reconnect k links on average with the individuals in states S, E, A, R^A or R^I. Since the individuals in state I are confirmed as symptomatic, we ignore the reconnection links with the individuals in this state.

Social distancing measures, such as keeping more than 2 m distance between individuals in public places, have contributed to the prevention and control of the COVID-19 epidemic. The average degree in our social networks describes the contact number of each individual with other individuals on average within 2 m distance, which is used to characterize the social distance among individuals in this paper. Keeping a social distance means breaking the links. Returning to work and school means reconnecting the links. For the change of social distance, there are two key elements: the intervention time of changing social distance among individuals and the intensity of intervention. We define an intervention function of social distance (IFSD):

$$\theta(t;t_i) = \alpha\delta(t-t_i),$$

where $\delta(\cdot)$ is the Kronecker delta function. Its numerical implementation is given as follows:

$$\theta(t;t_i) = \begin{cases} \alpha, & t_i \le t < t_i+1, \\ 0, & t < t_i \vee t \ge t_i+1, \end{cases}$$

where $t_i > 0$ is the intervention time of changing social distance among individuals, and α is the intensity of intervention. The time unit is day according to the daily data. Solving a system of differential equations generally requires discretization, therefore we consider that social distance changes during the interval $[t_i,\ t_i+1)$ with the intensity of intervention $\alpha \neq 0$ during this interval. After t_i+1, the new social distance remains the same. It is worth noting that this function is defined in cases where social distance measures change quickly. Then the setting of intervention time makes sense. Take the resumption of work and school as an example. Once the time is set, almost all enterprises and schools open at the same time, and social

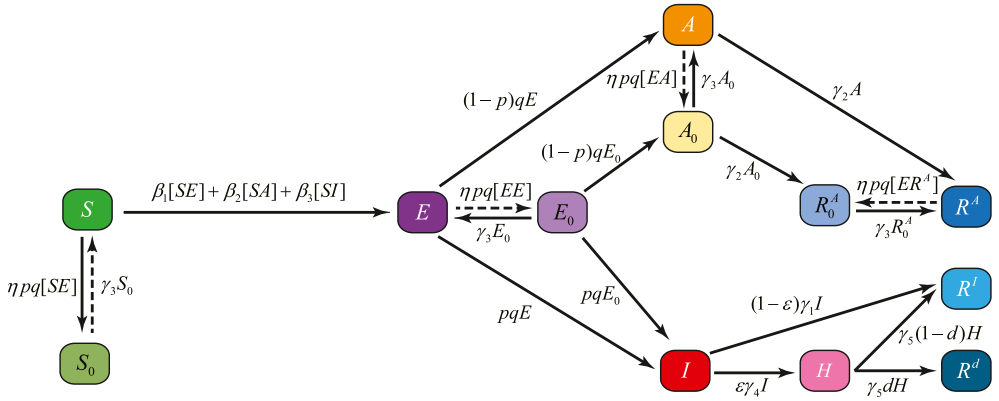

Fig. 1. The flow diagram of pairwise model for COVID-19 epidemic in the United States.

networks have changed a lot at the same time. Even if some enterprises or schools are delayed for a few days, it will have little impact on the overall social network. The function $\theta(t; t_i)$ mirrors the change of social distancing measures. When $\alpha < 0$ (the stricter social distancing measures), more links are broken, and thus the earlier the intervention time t_i the more effective COVID-19 is contained. In contrast, when $\alpha > 0$ (relaxing social distancing measures such as work resumption), the later the intervention time t_i the more beneficial to curb the disease spread. Early resumption of work and school could easily lead to a second outbreak, whereas late resumption would do substantial damage to the economy. Therefore, the appropriate time and intensity of intervention are extremely important.

This pairwise model is applicable to a shelter in place policy. In New York, shelter in place started on March 22nd, and ended on May 28th. In San Francisco, shelter in place started on March 19th and ended on May 25th. When shelter ended, the resumption of work and school began. In this case, the contacts between individuals increase, and links in the social network were correspondingly reconnected, which leads to the larger average degree. In this case, the intensity of intervention α is positive.

Using the notation in Table 1 and parameters in Table 2, and considering the change of social distance measures, the pairwise model for COVID-19 epidemic can be described as follows:

$$\frac{d[S](t)}{dt} = \gamma_3[S_0] - \beta_1[SE] - \beta_2[SA] - \beta_3[SI] - \eta pq[SE], \tag{3}$$

$$\frac{d[E](t)}{dt} = \beta_1[SE] + \beta_2[SA] + \beta_3[SI] + \gamma_3[E_0] - q[E] - \eta pq[EE], \tag{4}$$

$$\frac{d[A](t)}{dt} = (1-p)q[E] + \gamma_3[A_0] - \gamma_2[A] - \eta pq[EA], \tag{5}$$

$$\frac{d[I](t)}{dt} = pq[E] + pq[E_0] - \epsilon\gamma_4[I] - (1-\epsilon)\gamma_1[I], \tag{6}$$

$$\frac{d[H](t)}{dt} = \epsilon\gamma_4[I] - \gamma_5[H], \tag{7}$$

$$\frac{d[R^A](t)}{dt} = \gamma_2[A] + \gamma_3[R_0^A] - \eta pq[ER^A], \tag{8}$$

$$\frac{d[R^I](t)}{dt} = (1-\epsilon)\gamma_1[I] + \gamma_5(1-d)[H], \tag{9}$$

$$\frac{d[R^d](t)}{dt} = \gamma_5 d[H], \tag{10}$$

$$\frac{d[S_0](t)}{dt} = \eta pq[SE] - \gamma_3[S_0], \tag{11}$$

$$\frac{d[E_0](t)}{dt} = \eta pq[EE] - \gamma_3[E_0] - q[E_0], \tag{12}$$

$$\frac{d[A_0](t)}{dt} = \eta pq[EA] + (1-p)q[E_0] - \gamma_2[A_0] - \gamma_3[A_0], \tag{13}$$

$$\frac{d[R_0^A](t)}{dt} = \eta pq[ER^A] + \gamma_2[A_0] - \gamma_3[R_0^A]. \tag{14}$$

Since the evolution of the number of individuals in each class with time in Eqs.(3-14) depends on the numbers of different types of links, in order to close Eqs.(3-14), we need to derive the evolution equations of the number of links of each type with time. Because individuals in states $S_0, E_0, A_0, R_0^A, H, R^d$ are quarantined, isolated or removed, there are no links among them. Let $\mathbb{V} = \{S, E, A, I, R^A, R^I\}$, the evolution equations of $[XY](X, Y \in \mathbb{V})$ are as follows:

$$\begin{aligned}\frac{d[SS](t)}{dt} = \quad & -2\beta_1[SSE] - 2\beta_2[SSA] - 2\beta_3[SSI] - 2\eta pq[SSE] \\ & +2k\gamma_3[S_0]\frac{[S]}{M} + 2\theta(t; t_i)[SS],\end{aligned} \tag{15}$$

$$\begin{aligned}\frac{d[SE](t)}{dt} = \quad & \beta_1[SSE] + \beta_2[SSA] + \beta_3[SSI] - \beta_1[SE] - q[SE] - \beta_1[ESE] \\ & -\beta_2[ASE] - \beta_3[ISE] - \eta pq[SEE] - \eta pq[ESE] + k\gamma_3[S_0]\frac{[E]}{M} \\ & +k\gamma_3[E_0]\frac{[S]}{M} + \theta(t;t_i)[SE],\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{d[SA](t)}{dt} = \quad & (1-p)q[SE] - \gamma_2[SA] - \beta_2[SA] - \beta_1[ESA] - \beta_2[ASA] \\ & -\beta_3[ISA] - \eta pq[ESA] - \eta pq[SAE] + k\gamma_3[S_0]\frac{[A]}{M} \\ & +k\gamma_3[A_0]\frac{[S]}{M} + \theta(t;t_i)[SA],\end{aligned} \tag{17}$$

$$\begin{aligned}\frac{d[SI](t)}{dt} = \quad & (1-\eta)pq[SE] - \epsilon\gamma_4[SI] - (1-\epsilon)\gamma_1[SI] - \beta_3[SI] - \beta_1[ESI] \\ & -\beta_2[ASI] - \beta_3[ISI] - \eta pq[ESI] - \eta pq[SIE],\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{d[SR^A](t)}{dt} = \quad & \gamma_2[SA] - \beta_1[ESR^A] - \beta_2[ASR^A] - \beta_3[ISR^A] - \eta pq[ESR^A] \\ & -\eta pq[SR^AE] + k\gamma_3[S_0]\frac{[R^A]}{M} + k\gamma_3[R_0^A]\frac{[S]}{M} + \theta(t;t_i)[SR^A],\end{aligned} \tag{19}$$

$$\begin{aligned}\frac{d[SR^I](t)}{dt} = \quad & (1-\epsilon)\gamma_1[SI] - \beta_1[ESR^I] - \beta_2[ASR^I] - \beta_3[ISR^I] \\ & -\eta pq[ESR^I] + k\gamma_3[S_0]\frac{[R^I]}{M} + k((1-\epsilon)\gamma_1[I] \\ & +(1-d)\gamma_5[H])\frac{[S]}{M} + \theta(t;t_i)[SR^I],\end{aligned} \tag{20}$$

$$\begin{aligned}\frac{d[EE](t)}{dt} = \quad & 2\beta_1[SE] + 2\beta_1[ESE] + 2\beta_2[ASE] + 2\beta_3[ISE] - 2q[EE] \\ & -2\eta pq[EEE] + 2k\gamma_3[E_0]\frac{[E]}{M} + 2\theta(t;t_i)[EE],\end{aligned} \tag{21}$$

$$\begin{aligned}\frac{d[EA](t)}{dt} = \quad & (1-p)q[EE] + \beta_2[SA] + \beta_1[ESA] + \beta_2[ASA] + \beta_3[ISA] \\ & -q[EA] - \gamma_2[EA] - \eta pq[EEA] - \eta pq[EAE] + k\gamma_3[E_0]\frac{[A]}{M} \\ & +k\gamma_3[A_0]\frac{[E]}{M} + \theta(t;t_i)[EA],\end{aligned} \tag{22}$$

$$\begin{aligned}\frac{d[EI](t)}{dt} = \quad & (1-\eta)pq[EE] + \beta_3[SI] + \beta_1[ESI] + \beta_2[ASI] + \beta_3[ISI] \\ & -q[EI] - (1-\epsilon)\gamma_1[EI] - \epsilon\gamma_4[EI] - \eta pq[EEI] \\ & -\eta pq[EIE],\end{aligned} \tag{23}$$

$$\begin{aligned}\frac{d[ER^A](t)}{dt} = \quad & \gamma_2[EA] + \beta_1[ESR^A] + \beta_2[ASR^A] + \beta_3[ISR^A] - q[ER^A] \\ & -\eta pq[EER^A] - \eta pq[ER^AE] + k\gamma_3[E_0]\frac{[R^A]}{M} \\ & +k\gamma_3[R_0^A]\frac{[E]}{M} + \theta(t;t_i)[ER^A],\end{aligned} \tag{24}$$

Table 1
Model variables and their definitions.

variable	definition
$[S]$	The number of unquarantined susceptible individuals
$[S_0]$	The number of quarantined susceptible individuals
$[E]$	The number of unquarantined exposed (incubation) individuals
$[E_0]$	The number of quarantined exposed (incubation) individuals
$[A]$	The number of unquarantined asymptomatic individuals
$[A_0]$	The number of quarantined asymptomatic individuals
$[I]$	The number of the confirmed symptomatic individuals
$[H]$	The number of the hospitalized individuals
$[R^I]$	The number of recovered individuals with symptoms
$[R^A]$	The number of recovered individuals without symptoms
$[R_0^A]$	The number of the quarantined recovered individuals without symptoms
$[R^d]$	The number of individuals who die while recovering
$[SS]$	Twice the number of links between nodes with S state and S state
$[SE]$	The number of links between nodes with S state and E state
$[SA]$	The number of links between nodes with S state and A state
$[SI]$	The number of links between nodes with S state and I state
$[SR^A]$	The number of links between nodes with S state and R^A state
$[SR^I]$	The number of links between nodes with S state and R^I state
$[EE]$	Twice the number of links between nodes with E state and E state
$[EA]$	The number of links between nodes with E state and A state
$[EI]$	The number of links between nodes with E state and I state
$[ER^A]$	The number of links between nodes with E state and R^A state
$[ER^I]$	The number of links between nodes with E state and R^I state
$[AA]$	Twice the number of links between nodes with A state and A state
$[AI]$	The number of links between nodes with A state and I state
$[AR^A]$	The number of links between nodes with A state and R^A state
$[AR^I]$	The number of links between nodes with A state and R^I state
$[II]$	Twice the number of links between nodes with I state and I state
$[IR^A]$	The number of links between nodes with I state and R^A state
$[IR^I]$	The number of links between nodes with I state and R^I state
$[R^AR^A]$	Twice the number of links between nodes with R^A state and R^A state
$[R^AR^I]$	The number of links between nodes with R^A state and R^I state
$[R^IR^I]$	Twice the number of links between nodes with R^I state and R^I state
$[SSE]$	The number of triples with the joint structure S–S-E
$[SSA]$	The number of triples with the joint structure S–S-A
$[SSI]$	The number of triples with the joint structure S–S–I
$[SEE]$	The number of triples with the joint structure S-E-E
$[SAE]$	The number of triples with the joint structure S-A-E
$[SIE]$	The number of triples with the joint structure S-I-E
$[SR^AE]$	The number of triples with the joint structure S-R^A-E
$[ESE]$	Twice the number of triples with the joint structure E-S-E
variable	definition
$[ESA]$	The number of triples with the joint structure E-S-A and [ASE] = [ESA]
$[ESI]$	The number of triples with the joint structure E-S-A and [ISE] = [ESI]
$[ESR^A]$	The number of triples with the joint structure E-S-R^A
$[ESR^I]$	The number of triples with the joint structure E-S-R^I
$[EEE]$	Twice the number of triples with joint structures E-E-E
$[EEA]$	The number of triples with the joint structure E-E-A
$[EEI]$	The number of triples with the joint structure E-E-I
$[EER^A]$	The number of triples with joint structures E-E-R^A
$[EER^I]$	The number of triples with joint structures E-E-R^I
$[EAE]$	Twice the number of triples with the joint structure E-A-E
$[EAR^A]$	The number of triples with the joint structure E-A-R^A
$[EAR^I]$	The number of triples with the joint structure E-A-R^I
$[EIE]$	Twice the number of triples with the joint structure E-I-E
$[EIR^A]$	The number of triples with the joint structure E-I-R^A
$[EIR^I]$	The number of triples with the joint structure E-I-R^I
$[ER^AE]$	Twice the number of triples with joint structures E-R^A-E
$[ER^AR^A]$	Twice the number of triples with joint structures E-R^A-R^A
$[ER^AR^I]$	Twice the number of triples with joint structures E-R^A-R^I
$[ASA]$	Twice the number of triples with the joint structure A-S-A
$[ASI]$	The number of triples with the joint structure A-S-I and [ASI] = [ISA]
$[ASR^A]$	The number of triples with the joint structure A-S-R^A
$[ASR^I]$	The number of triples with the joint structure A-S-R^I
$[AAE]$	The number of triples with the joint structure A-A-E
$[AIE]$	The number of triples with the joint structure A-I-E
$[AR^AE]$	The number of triples with joint structures A-R^A-E
$[ISI]$	Twice the number of triples with the joint structure I–S–I
$[ISR^A]$	The number of triples with the joint structure I-S-R^A

Table 1 (*continued*)

variable	definition
$[ISR^I]$	The number of triples with the joint structure I-S-R^I
$[IIE]$	The number of triples with the joint structure I–I-E
$[IR^AE]$	The number of triples with the joint structure I-R^A-E

$$\begin{aligned}\frac{d[ER^I](t)}{dt} = \quad & (1-\epsilon)\gamma_1[EI] + \beta_1[ESR^I] + \beta_2[ASR^I] + \beta_3[ISR^I] \\ & -q[ER^I] - \eta pq[EER^I] + k\gamma_3[E_0]\frac{[R^I]}{M} \\ & +k((1-\epsilon)\gamma_1[I] + (1-d)\gamma_5[H])\frac{[E]}{M} + \theta(t;t_i)[ER^I],\end{aligned} \tag{25}$$

$$\begin{aligned}\frac{d[AA](t)}{dt} = \quad & 2(1-p)q[EA] - 2\gamma_2[AA] - 2\eta pq[AAE] + 2k\gamma_3[A_0]\frac{[A]}{M} \\ & +2\theta(t;t_i)[AA],\end{aligned} \tag{26}$$

$$\begin{aligned}\frac{d[AI](t)}{dt} = \quad & (1-p)q[EI] + (1-\eta)pq[EA] - \gamma_2[AI] - \epsilon\gamma_4[AI] \\ & -(1-\epsilon)\gamma_1[AI] - \eta pq[AIE] - \eta pq[IAE],\end{aligned} \tag{27}$$

Table 2
Definition of parameters and their values (unit time: day).

Parameter	Definition	New York	Standard deviation	San Francisco	Standard deviation	Data source
β_1	transmission rate by E individuals	0.142 6	0.009 0	0.091 5	0.003 9	MCMC
β_2	transmission rate by A individuals	0.228 4	0.014 3	0.229 9	0.005 7	MCMC
β_3	transmission rate by I individuals	0.456 8	–	0.459 8	–	(Li et al., 2020b; Moghadas et al., 2020)
p	probability of showing symptoms	0.25	–	0.25	–	(https://www.cidrap.umn.ed, 2020; https://www.washingtonpos, 2020)
$1/q$	incubation period for E individuals	5	–	6	–	(Moss et al., 2020; Park et al., 2020)
η	quarantined probability of individuals who have been contacted with the confirmed cases	0.8	–	0.8	–	Moss et al. (2020)
ϵ	proportion of the confirmed symptomatic cases transferring to hospitalized cases	0.2	–	0.2	–	Moss et al. (2020)
$1/\gamma_1$	period of transferring from I individuals directly to R^I individuals	9.68	–	9.68	–	Moss et al. (2020)
γ_2	The recovery rate of A individuals	1/12	–	1/12	–	Luo et al. (2021)
$1/\gamma_3$	quarantine period for S, E, A and R^A individuals	14	–	14	–	Moss et al. (2020)
$1/\gamma_4$	period of transferring from I individuals directly to H individuals	4	–	4	–	Moss et al. (2020)
$1/\gamma_5$	period of transferring from H individuals directly to R^I individuals	5.68	–	5.68	–	Moss et al. (2020)
d	proportion of H individuals who die while recovering	0.075	–	0.075	–	(https://covidmapper.ca/ge)
N	size of total population	8,336,697	–	4,264,934	–	(https://en.wikipedia.org/a; https://en.wikipedia.org/b; https://en.wikipedia.org/c; https://en.wikipedia.org/d; https://en.wikipedia.org/e; https://en.wikipedia.org/f)
φ	clustering coefficient	0.297 5	–	0.297 5	–	(2)
k	the average degree of individuals released from quarantine	1.471 3	0.023 5	1.510 9	0.004 7	MCMC
α	the intensity of intervention	0.069 6	0.010 6	0.122 1	0.005 4	MCMC

$$\begin{aligned}\frac{d[AR^A](t)}{dt} = \quad & \gamma_2[AA] + (1-p)q[ER^A] - \gamma_2[AR^A] - \eta pq[AR^AE] \\ & -\eta pq[EAR^A] + k\gamma_3[A_0]\frac{[R^A]}{M} + k\gamma_3[R_0^A]\frac{[A]}{M} \\ & +\theta(t;t_i)[AR^A],\end{aligned} \tag{28}$$

$$\begin{aligned}\frac{d[AR^I](t)}{dt} = \quad & (1-\epsilon)\gamma_1[AI] + (1-p)q[ER^I] - \gamma_2[AR^I] - \eta pq[EAR^I] \\ & +k\gamma_3[A_0]\frac{[R^I]}{M} + k((1-\epsilon)\gamma_1[I] + (1-d)\gamma_5[H])\frac{[A]}{M} \\ & +\theta(t;t_i)[AR^I],\end{aligned} \tag{29}$$

$$\frac{d[II](t)}{dt} = 2(1-\eta)pq[EI] - 2\epsilon\gamma_4[II] - 2(1-\epsilon)\gamma_1[II] - 2\eta pq[IIE], \tag{30}$$

$$\begin{aligned}\frac{d[IR^A](t)}{dt} = \quad & \gamma_2[AI] + (1-\eta)pq[ER^A] - \epsilon\gamma_4[IR^A] - (1-\epsilon)\gamma_1[IR^A] \\ & -\eta pq[IR^AE] - \eta pq[EIR^A],\end{aligned} \tag{31}$$

$$\begin{aligned}\frac{d[IR^I](t)}{dt} = \quad & (1-\epsilon)\gamma_1[II] + (1-\eta)pq[ER^I] - \epsilon\gamma_4[IR^I] - (1-\epsilon)\gamma_1[IR^I] \\ & -\eta pq[EIR^I],\end{aligned} \tag{32}$$

$$\begin{aligned}\frac{d[R^AR^A](t)}{dt} = \quad & 2\gamma_2[AR^A] - 2\eta pq[ER^AR^A] + 2k\gamma_3[R_0^A]\frac{[R^A]}{M} \\ & +2\theta(t;t_i)[R^AR^A],\end{aligned} \tag{33}$$

$$\begin{aligned}\frac{d[R^AR^I](t)}{dt} = \quad & (1-\epsilon)\gamma_1[IR^A] + \gamma_2[AR^I] - \eta pq[ER^AR^I] + k\gamma_3[R_0^A]\frac{[R^I]}{M} \\ & +k((1-\epsilon)\gamma_1[I] + (1-d)\gamma_5[H])\frac{[R^A]}{M} + \theta(t;t_i)[R^AR^I],\end{aligned} \tag{34}$$

$$\begin{aligned}\frac{d[R^IR^I](t)}{dt} = \quad & 2(1-\epsilon)\gamma_1[IR^I] + 2k((1-\epsilon)\gamma_1[I] + (1-d)\gamma_5[H])\frac{[R^I]}{M} \\ & +2\theta(t;t_i)[R^IR^I].\end{aligned} \tag{35}$$

Here $M = [S] + [E] + [A] + [R^A] + [R^I]$. The terms $\eta pq[EXY]$ $(X, Y \in \mathbb{V})$ represent the case as that individuals in state E turn into I, their neighbors are quarantined, resulting in a decrease in the number of links. The terms $[X][Y]/[M]$ $(X \in \{S_0, E_0, A_0, R_0^A, I, H\}, Y \in \{S, E, A, R^A, R^I\})$ represent a lifting of the quarantine or that recovering results in an increase in the number of links. The change of social distance measures, reflected by the IFSD function, directly leads to the change of links in the social network. The IFSD function reflects both the intervention time t_i of changing social distance among individuals and the intensity of intervention α. The change of links is positively correlated with α, so we consider considered the simplest form $\theta(t; t_i)[XY]$ to represent the change of social distance measures. Since the individuals in state I are confirmed symptomatic, we ignore the change of links with the individuals in this state due to the change of social distance measures.

The numbers of triples could be closed by the following moment closure approximation formula (referred to in (House & Keeling, 2011; Keeling, Rand, & Morris, 1997)),

$$[XYZ] = \frac{n(t)-1}{n(t)}\frac{[XY][YZ]}{[Y]}\left(1 - \varphi + \varphi\frac{N}{n(t)}\frac{[XZ]}{[X][Z]}\right), \qquad X, Y, Z \in \mathbb{V}, \tag{36}$$

where φ is the clustering coefficient, N is the number of nodes in the underlying network and $n(t)$ is the average degree at time t with the initial condition $n_0 \triangleq n(0)$, satisfying

$$n(t) = \frac{1}{N}\left(\sum_{X\in\mathbb{V}}[XX] + 2\sum_{X,Y\in\mathbb{V},X\neq Y}[XY]\right). \tag{37}$$

As a pair approximation model, our model is somewhat different from that of Luo et al. (Luo et al., 2021). The main differences are listed as follows:

- Our model considers the links reconnection between individuals released from quarantine, focusing on studying the effect of social distancing measures and work resumption on COVID-19 spread.
- The quarantined probability of close contacts η is less than one, since close contacts with confirmed cases are quarantined at home rather than centrally.
- Only about 20% are hospitalized (Moss et al., 2020), and most diagnosed patients stay at home and some of them may go out and infect others.
- Our model considers more detailed recovered or removed compartments R^A, R_0^A, R^I, and R^d.

2.3. The basic reproduction number and effective reproduction number

The basic reproduction number $\mathcal{R}_0$, which is defined as an average number of secondary cases that one case produces during the course of its infectious period in a totally susceptible population, determines the growing rate of an emerging infectious disease. Sometimes it can be used to predict whether an outbreak of one infectious disease will occur or not. Adopting the method proposed by the next generation matrix method (van den Driessche & Watmough, 2002), the expression $\mathcal{R}_0$ takes the form of

$$\mathcal{R}_0 = (1-\varphi)\frac{n_0-1}{n_0}\frac{[SS](0)}{[S](0)}\left(\frac{\beta_1}{\beta_1+q} + \frac{(1-p)q\beta_2}{(\beta_1+q)(\beta_2+\gamma_2)} + \frac{pq\beta_3}{(\beta_1+q)[\beta_3+\epsilon\gamma_4+(1-\epsilon)\gamma_1]}\right) = R^{[SE]} + R^{[SA]} + R^{[SI]}, \tag{38}$$

where

$$R^{[SE]} = (1-\varphi)(n_0-1)\frac{\beta_1}{\beta_1+q}, \tag{39}$$

$$R^{[SA]} = (1-\varphi)(n_0-1)\frac{(1-p)q\beta_2}{(\beta_1+q)(\beta_2+\gamma_2)}, \tag{40}$$

and

$$R^{[SI]} = (1-\varphi)(n_0-1)\frac{pq\beta_3}{(\beta_1+q)[\beta_3+\epsilon\gamma_4+(1-\epsilon)\gamma_1]}. \tag{41}$$

As almost all individuals are susceptible when the outbreak starts with the introduction of a single "average" infected individual at time $t = 0$, there is the constraint $[SS](0) = n_0[S](0)$ in networks.

Since the pairwise epidemic model (3−35) is a network model, we are mainly concerned with the variants of the infected links. The biological meaning of $\mathcal{R}_0$ can be regarded as a sum of three quantities. $R^{[SE]}$ is the number of secondary $S-E$ links that one $S-E$ link will generate in an entirely $S-S$ link group $(1-\varphi)(n_0-1)$ during its lifespan $\frac{1}{\beta_1+q}$. Note that $\frac{(1-p)q}{\beta_1+q}$ and $\frac{pq}{\beta_1+q}$ denote the probabilities that one $S-E$ link survives to progress into $S-A$ link and $S-I$ link, respectively. Hence, $R^{[SA]}$ represents the number of secondary $S-A$ links that one $S-A$ link will produce in a totally $S-S$ link group during its lifespan $\frac{1}{\beta_2+\gamma_2}$. $R^{[SI]}$ is the number of secondary $S-I$ links that one $S-I$ link will produce in a totally $S-S$ link group during its lifespan $\frac{1}{\beta_3+\epsilon\gamma_4+(1-\epsilon)\gamma_1}$.

The effective reproduction number $\mathcal{R}_e(t)$ is defined as an average number of secondary cases that one case produces during the course of its infectious period. Unlike the basic reproduction number, the effective reproduction number is given in real-time and depends upon the current state of the epidemic. Using the same method as with the basic reproduction number, the effective reproduction number $\mathcal{R}_e(t)$ takes the form of

$$\mathcal{R}_e(t) = (1-\varphi)\frac{n(t)-1}{n(t)}\frac{[SS](t)}{[S](t)}\left(\frac{\beta_1}{\beta_1+q} + \frac{(1-p)q\beta_2}{(\beta_1+q)(\beta_2+\gamma_2)} + \frac{pq\beta_3}{(\beta_1+q)[\beta_3+\epsilon\gamma_4+(1-\epsilon)\gamma_1]}\right). \tag{42}$$

If there is a single "average" infected individual at the onset of COVID-19 outbreak at time $t = 0$ then we have $\mathcal{R}_e(0) = \mathcal{R}_0$, as given in Eqs. (38) and (42).

2.4. Parameters and initial conditions

For the parameters β_1, β_2, k and α that cannot be easily obtained, we use Markov Chain Monte Carlo (MCMC) simulations (Gamerman & Lopes, 2006) to estimate their values based on COVID-19 cases (https://covidmapper.ca/ge). The random-walk Metropolis-Hastings (MH) algorithm is used to adapt and readjust the MCMC procedure. We assume that the observations over time for data is Gaussian distributed (Gamerman & Lopes, 2006; Roda, 2020). For the parameter β_3, it is about twice the parameter β_2 (Li et al., 2020b; Moghadas et al., 2020). Note that transmission rates in this paper are real transmission probability at which a susceptible individual is infected by an infectious individual in a unit time. For the network topological parameters, the clustering coefficient φ and the initial value of average degree n_0, we apply the method in Ref. (Luo et al., 2021). to calculate their values according to the distribution of household structure in the studied two cities (https://www.census.gov/da). The clustering coefficient φ is calculated by (2). The average degree is equal to the sum of the degrees of all nodes divided by the total number of nodes. At the initial time, the sum of the degrees of all nodes equals to $3\sum_{k=1}^{7}F_k + \sum_{k=1}^{7}k(k-1)F_k$, and the total number of nodes is $\sum_{k=1}^{7}kF_k$. Thus, the initial value of average degree equals

$$n_0 = \frac{3\sum_{k=1}^{7}F_k + \sum_{k=1}^{7}k(k-1)F_k}{\sum_{k=1}^{7}kF_k} = 3.465\,0. \tag{43}$$

In New York it appears that 88% of infected pregnant patients may be asymptomatic (https://www.washingtonpos, 2020). In Seattle more than half of the infected patients in a nursing home were asymptomatic (https://www.cidrap.umn.ed, 2020). Therefore, we consider that the probability of showing symptoms p equals to 0.25 for these two cities. For the remaining parameters, we obtain their values based on references (Luo et al., 2021; Moss et al., 2020; Park et al., 2020). All parameters are summarized in Table 2.

The initial values of variables in the pairwise model (3)–(35) are given as follows. The population sizes of New York and San Francisco are about 8, 336, 697 and 4, 264, 934, respectively (https://en.wikipedia.org/a; https://en.wikipedia.org/b; https://en.wikipedia.org/c; https://en.wikipedia.org/d; https://en.wikipedia.org/e; https://en.wikipedia.org/f). Let t_0 stand for the starting date, which is March 27th, 2020. Values $[E](t_0)$, $[A](t_0)$, $[SS](t_0)$, $[SE](t_0)$, $[SA](t_0)$, and $[SI](t_0)$ are estimated by MCMC method. According to the collected data, the initial values of cumulative confirmed cases (denoted by $C(t_0)$) and deaths caused by COVID-19 $[R_d](t_0)$ are obtained directly. The remaining initial values are obtained by data processing and calculations. Since the H individuals move into R^I compartment with proportion $1-d$ or R^d compartment with proportion d, the initial value of recovered individuals with symptoms $[R^I](t_0)$ is calculated by the following approximation formula:

$$[R^I](t_0) = \frac{1-d}{d}[R^d](t_0). \tag{44}$$

Since each E individual moves to I state with probability p or to A state with probability $1-p$, and $p = 0.25$, the ratio of the initial value of recovered individuals without symptoms $[R^A](t_0)$ and the initial value of recovered individuals with symptoms $[R^I](t_0)$ is less than $(1-p)/p = 3$. We assume that the initial value of recovered individuals without symptoms $[R^A](t_0)$ is about 2.5 times of that of those with symptoms, that is,

$$[R^A](t_0) = 2.5[R^I](t_0). \tag{45}$$

For any time t, the number of cumulative confirmed cases is the sum of the confirmed symptomatic individuals $[I](t)$, the number of recovered cases $[R^I](t)$ and the number of died cases $[R^d](t)$. Hence the initial value of the confirmed symptomatic individuals $[I](t_0)$ approximately equals to the difference between the number of cumulative confirmed cases and the sum of the number of recovered cases $[R^I](t_0)$ and the number of died cases $[R^d](t_0)$, that is,

$$[I](t_0) = C(t_0) - [R^I](t_0) - [R^d](t_0). \tag{46}$$

Since the individuals in state I move into the H compartment with proportion ϵ, the number of the hospitalized individuals $[H](t_0)$ is calculated by the following approximation formula:

$$[H](t_0) = \epsilon[I](t_0). \tag{47}$$

Quarantined individuals are caused by the diagnosis of infectious ones. The initial value of quarantined individuals approximately equals to the total number of close contacts of new confirmed cases in a incubation period: $n_0([I](t_0 + 1/q) - [I](t_0))$. Among the number of quarantined individuals, $[S_0](t_0)$ accounts for about 2/3, and the sum of $[E_0](t_0)$, $[A_0](t_0)$ and

$[R_0^A](t_0)$ approximately accounts for 1/3 (https://www.soundhealthan). We assume that $[R_0^A](t_0) = [A_0](t_0)$, then the approximation equations are

$$[S_0](t_0) = \frac{2}{3} n_0([I](t_0 + 1/q) - [I](t_0)), \tag{48}$$

$$[E_0](t_0) = \frac{1}{3((1-p)q+1)} n_0([I](t_0 + 1/q) - [I](t_0)), \tag{49}$$

$$[A_0](t_0) = \frac{(1-p)q}{6((1-p)q+1)} n_0([I](t_0 + 1/q) - [I](t_0)), \tag{50}$$

$$[R_0^A](t_0) = \frac{(1-p)q}{6((1-p)q+1)} n_0([I](t_0 + 1/q) - [I](t_0)). \tag{51}$$

Since the number of all nodes is N which remains the same at any time, the initial value of unquarantined susceptible individuals $[S](t_0)$ is calculated by the following approximation formula:

$$\begin{aligned}[S](t_0) = \quad & N - [I](t_0) - [E](t_0) - [A](t_0) - [H](t_0) - [R^A](t_0) - [R^I](t_0) \\ & - [R^d](t_0) - [S_0](t_0) - [E_0](t_0) - [A_0](t_0) - [R_0^A](t_0).\end{aligned} \tag{52}$$

For the initial values of links except estimated ones and $[SR^A](t_0)$, we determine their values based on the following approximation formula:

Table 3
The initial values of variables.

Variable	New York	Standard deviation	San Francisco	Standard deviation	Data source
$[S](t_0)$	8,158,859	–	4,259,357	–	(https://en.wikipedia.org/a; https://en.wikipedia.org/b; https://en.wikipedia.org/c; https://en.wikipedia.org/d; https://en.wikipedia.org/e; https://en.wikipedia.org/f)
$[E](t_0)$	26,809	157.208 0	763.455 8	4.845 2	MCMC
$[A](t_0)$	54,872	420.920 6	1,357	7.322 7	MCMC
$[I](t_0)$	20,693	–	763	–	(46)
$[H](t_0)$	4,139	–	152.7	–	(47)
$[R^A](t_0)$	11,285	–	431.7	–	(45)
$[R^I](t_0)$	4,514	–	172.7	–	(44)
$[R^d](t_0)$	366	–	14	–	(https://covidmapper.ca/ge)
$[S_0](t_0)$	36,773	–	1,282	–	(48)
$[E_0](t_0)$	15,988	–	569.800 0	–	(49)
$[A_0](t_0)$	1,199	–	35.612 5	–	(50)
$[R_0^A](t_0)$	1,199	–	35.612 5	–	(51)
$[SS](t_0)$	14,545,774	129,689	9,559,037	114,793	MCMC
$[SE](t_0)$	66,189	266.492 9	1,764	27.537 7	MCMC
$[SA](t_0)$	128,521	822.550 8	2,231	14.231 3	MCMC
$[SI](t_0)$	2,063	21.625 0	1,102	12.209 7	MCMC
$[SR^A](t_0)$	13,910,916	–	5,207,559	–	(37)
$[SR^I](t_0)$	15,307	–	597.507 6	–	(53)
$[EE](t_0)$	298.724 8	–	0.473 5	–	(53)
$[EA](t_0)$	611.427 9	–	0.841 4	–	(53)
$[EI](t_0)$	230.575 8	–	0.473 5	–	(53)
$[ER^A](t_0)$	125.745 3	–	0.267 7	–	(53)
$[ER^I](t_0)$	50.298 1	–	0.107 1	–	(53)
$[AA](t_0)$	1,251	–	1.494 9	–	(53)
$[AI](t_0)$	471.941 0	–	0.841 2	–	(53)
$[AR^A](t_0)$	257.374 7	–	0.475 7	–	(53)
$[AR^I](t_0)$	102.949 9	–	0.190 3	–	(53)
$[II](t_0)$	177.973 8	–	0.473 4	–	(53)
$[IR^A](t_0)$	97.058 6	–	0.267 7	–	(53)
$[IR^I](t_0)$	38.823 5	–	0.107 1	–	(53)
$[R^AR^A](t_0)$	52.931 3	–	0.151 4	–	(53)
$[R^AR^I](t_0)$	21.172 5	–	0.060 6	–	(53)
$[R^IR^I](t_0)$	8.469 0	–	0.024 2	–	(53)
n_0	3.465 0	–	3.465 0	–	(43)

$$[XY](t_0) = n_0 \times [X](t_0) \times \frac{[Y](t_0)}{N}, \quad X,\ Y \in \mathbb{V}, \tag{53}$$

where $[X]$ or $[Y]$ represents the number of nodes in state X or Y and $[XY]$ represents the number of XY-links. According to formula (37), letting $t = t_0$, we obtain the initial value of $[SR^A](t_0)$. All the initial values are summarized in Table 3.

3. Results

3.1. Fitting results

Based on the diagnosed cases of COVID-19 infections (https://covidmapper.ca/ge), we adopt the MCMC method (Gamerman & Lopes, 2006) for 20 000 iterations with a burn-in of 10 000 iterations to fit the model (3−37) to estimate the parameters and the initial conditions of variables (see Tables 2 and 3). The transmission rates are subject to the condition that $\beta_1 \leq \beta_2$. According to the shelter in place policy, $t_i = 62$ (May 28th) for New York, and $t_i = 59$ (May 25th) for San Francisco. Notice that there is no overfitting because the ratio of the number of data points (=115) and the number of estimated parameters or the initial values of variables (=10) is far over 5:1. The details have been presented in Section 2.4.

A good fitting between the model solution and real data in two cities is shown in Fig. 2. From the inset of Fig. 2(a), the real data point on April 23rd is a singularity with a negative value. This is due to a drop in the number of cumulative confirmed cases that day. In fact, ignoring this point has no effect on the overall fitting result. According to the estimated parameter values and initial conditions as given in Tables 2 and 3, we obtained the mean value of the reproduction number $\mathcal{R}_0 = 1.472$ (95% CI of [1.444, 1.498]) in New York, $\mathcal{R}_0 = 1.445$ (95% CI of [1.432, 1.459]) in San Francisco. Here CI indicates credible interval, the range containing a particular percentage of probable values. In other words, the 95% credible interval is simply the central portion of the posterior distribution that contains 95% of the values. In what follows, we mainly employ our model with estimated parameters to explore the epidemic behavior of COVID-19 infections in two cities.

Fig. 3 shows the effective reproduction number. For these two cities, the effective reproduction number decreases quickly due to the social distancing. When shelter ended, the effective reproduction number increases quickly. For New York, the effective reproduction number is always smaller than one. But for San Francisco, the effective reproduction number is always larger than one, and the effective reproduction number is larger than 1.3 when shelter in place ended, which leads to a rapid increase of cumulative confirmed cases. Overall, the effective reproduction number is consistent with that shown on a popular website calculator for $\mathcal{R}_e(t)$ (https://rt.live/us/; https://rt.live/us/). The major difference is that when policy shifts the increase in our results is faster. The reason is that the intensity of intervention α is not equal to zero only at the intervention time. At the intervention time t_i, shelter ended, and the resumption of work and school began. This leads to an sudden increase in the total number of links in the network and hence an sudden increase in the average degree.

3.2. Differences in COVID-19 infections

Perhaps surprisingly, under the assumption of no further changes in policy or transmission dynamics, the final size of the San Francisco epidemic is about 3.7 times that of New York, and that the epidemic duration of San Francisco is more than 140 days longer than that of New York. From Tables 2 and 3, transmission rates, population size and initial infections are different in New York and San Francisco. The transmission rate β_1 is higher in New York than in San Francisco. The transmission rate β_2 is slightly higher in San Francisco than in New York. According to Ref. (https://en.wikipedia.org/a; https://en.wikipedia.org/b; https://en.wikipedia.org/c; https://en.wikipedia.org/d; https://en.wikipedia.org/e; https://en.wikipedia.org/f), the population

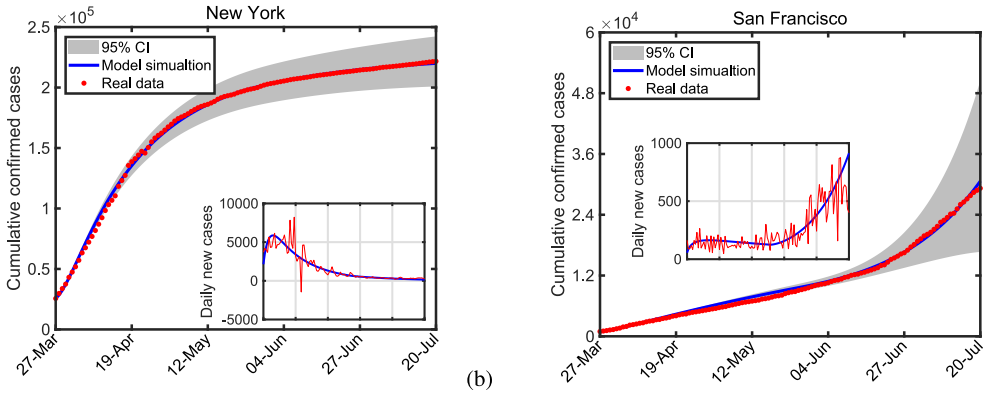

Fig. 2. The fitting results of our model to real data of the cumulative and daily new cases (see inset) of COVID-19 infections in New York city and San Francisco city. The blue solid lines represent model simulation. Red dots and red lines are the real data. The grey area marks the 95% CI of MCMC estimations.

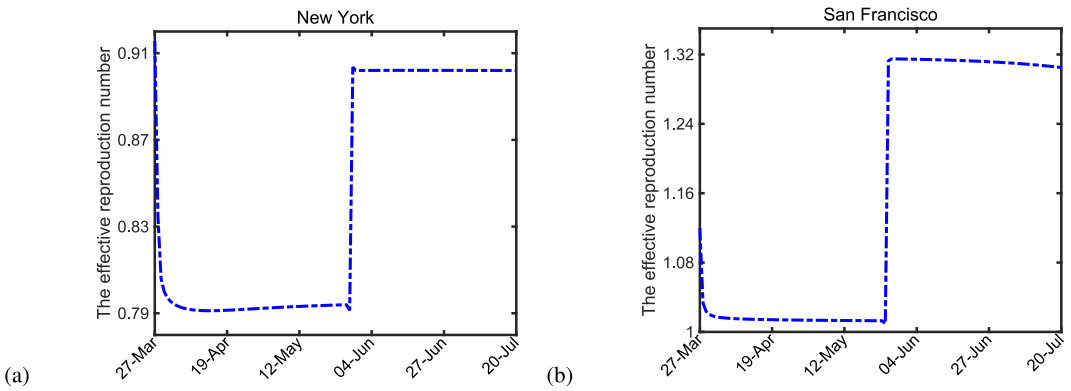

Fig. 3. The effective reproduction number.

and area of New York are 8,336,697 and 785 square kilometers, respectively and those of San Francisco are 4,264,934 and 6405.5 square kilometers. So the population density of New York is about 16 times as dense as that of San Francisco. The starting date is not the true initial time of the disease, so it is of little significance to study the effect of the initial infections.

Here we only compare the effect of transmission rates and population density on disease dynamics by asking how the disease dynamics would progress if transmission rates or population densities were switched between the two cities. In detail, we set the transmission rates of New York being the estimated values of San Francisco to study the effect of transmission rates. According to the population sizes and areas in San Francisco and in New York, we calculate the population density of New York and modify the number of links $[SS]$, $[SE]$, $[SA]$, $[SR^A]$ accordingly. In a similar way, we investigate their effects in San Francisco. As shown in Fig. 4, population density has a major effect on disease transmission, and transmission rates have a smaller impact. For New York, when the population density drops to the same as San Francisco, the epidemic duration will be shortened by more than 140 days, and the final size will be reduced by more than 50%. When transmission rates in New York are made the same as those of San Francisco, the epidemic duration will be prolonged by more than 2 months, but the final size will be reduced by more than 11%. For San Francisco, when the population density increases to the same as New York, the epidemic duration will be extended by 19 days, and the final size will be increased 15-fold. When San Francisco transmission rates are made the same with those of New York, the epidemic duration will be shortened by 135 days, but the final size will be increased by more than 11%. Overall, the final size increases and the epidemic duration lengthens as population density increases. Also, as the rate of infection increases, so does the final size of the epidemic.

3.3. *Sensitivity analysis*

To identify which parameters are important for the basic reproduction number R_0, the final size and the effective reproduction number $R_e(t)$, we conduct a sensitive analysis by evaluating the partial rank correction coefficients (PRCCs) for key parameters based on Latin hypercube sampling (Marino et al., 2008). All parameters are sampled 1000 times using a Latin hypercube sampling. For the probability density function (PDF) of these parameters, a normal distribution is selected for these

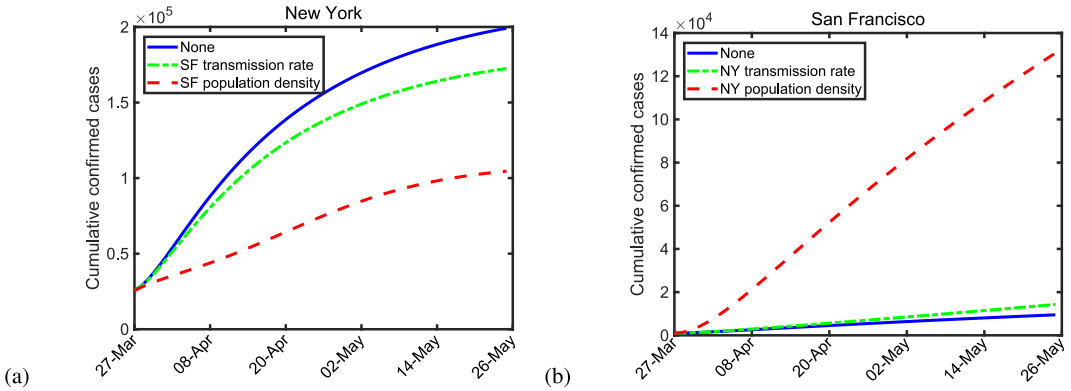

Fig. 4. The effect of transmission rates and population density on COVID-19 infections in New York city and San Francisco city. In panel (a), the green dot-dashed line is obtained by replacing transmission rates of New York with those of San Francisco, and the red dashed line is under the case of replacing population density of New York with that of San Francisco. In panel (b), the green dot-dashed line is obtained by replacing transmission rates of San Francisco with those of New York, and the red dashed line is under the case of replacing population density of San Francisco with that of New York.

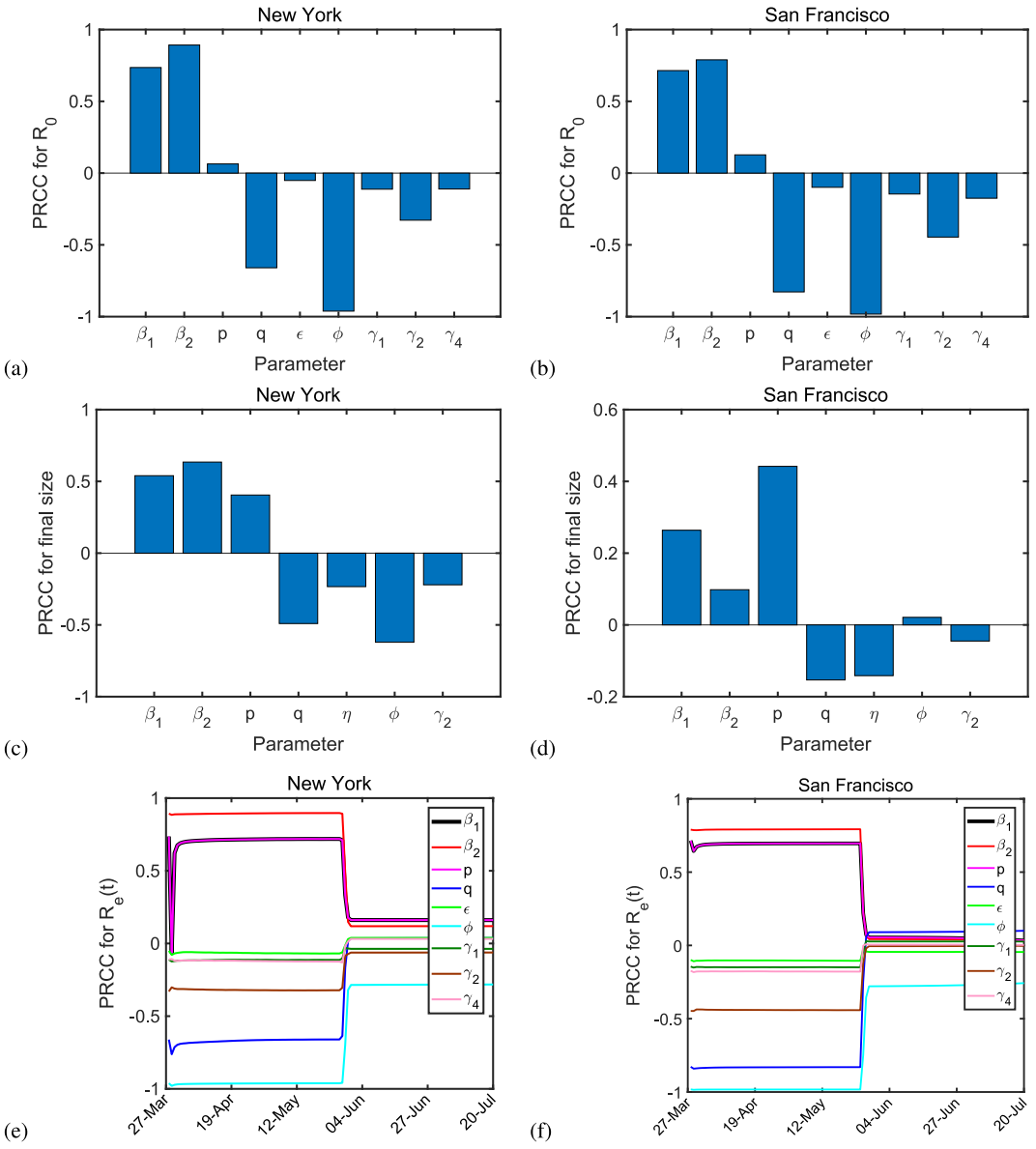

Fig. 5. Partial rank correlation coefficients (PRCCs) for R_0, final size and $R_e(t)$ in New York and San Francisco. The panels (a) and (b) show PRCCs for R_0 and each parameter, (c) and (d) show PRCCs for final size and each parameter, while (e) and (f) show temporal variation of the sensitivity of the effective reproduction number $R_e(t)$ to each parameter. All parameters are sampled 1000 times using a Latin hypercube sampling. For the probability density function (PDF) of these parameters, a normal distribution is selected for these parameters.

parameters. Fig. 5 shows PRCCs which illustrate the dependence of R_0, the final size and $R_e(t)$ on all parameters, respectively. The panels (a) and (b) show PRCCs for R_0. For New York, the four parameters β_1, β_2, q, φ have high correlations ($|PRCCs| > 0.4$) with R_0. The most influential parameter on R_0 is found to be the clustering coefficient φ ($|PRCC| = 0.961\ 7$), and then in descending order are transmission rate by A individuals β_2 ($|PRCC| = 0.882\ 3$), transmission rate by E individuals β_1 ($|PRCC| = 0.735\ 6$), incubation period for E individuals $1/q$ ($|PRCC| = 0.660\ 5$). The parameter γ_2 has moderate correlations ($0.2 \leq |PRCCs| < 0.4$) with R_0. The other four parameters p, ε, γ_1, γ_4 have low correlations ($0 \leq |PRCCs| < 0.2$) with R_0. The PRCCs associated with these parameters except the parameter ε are statistically significant. For San Francisco, the five parameters β_1, β_2, q, φ, γ_2 have high correlations with R_0. The most influential parameter on R_0 is found to be the clustering coefficient φ ($|PRCC| = 0.982\ 2$), and then in descending order are incubation period for E individuals $1/q$ ($|PRCC| = 0.827\ 5$), transmission rate by A individuals β_2 ($|PRCC| = 0.789\ 5$), transmission rate by E individuals β_1 ($|PRCC| = 0.714\ 9$), the recovery rate of A individuals γ_2 ($|PRCC| = 0.446\ 4$). The other four parameters p, ε, γ_1, γ_4 have low correlations with R_0. The PRCCs associated with

these nine parameters are also statistically significant. From the panels (a) and (b), it is obvious to conclude that the higher the transmission rates are, the higher the basic reproduction number R_0 is, and that the longer the incubation period is, the higher the basic reproduction number R_0 is. In addition, there is a highly negative correlation between the clustering coefficient φ and the basic reproduction number R_0. The shelter in place policy makes the clustering coefficient increase and then reduces the basic reproduction number.

The panels (c) and (d) show PRCCs for the final size. For these two cities, The PRCCs associated with the two parameters γ_1 and γ_4 are not statistically significant, thus we do not plot their PRCCs. For New York, this sensitivity analysis shows the first five parameters with most impact on the final size are transmission rate by A individuals β_2 ($|PRCC| = 0.615\ 8$), the clustering coefficient φ ($|PRCC| = 0.617\ 2$), transmission rate by E individuals β_1 ($|PRCC| = 0.531\ 9$), incubation period for E individuals $1/q$ ($|PRCC| = 0.483\ 6$), and probability of showing symptoms p ($|PRCC| = 0.381\ 7$). The PRCCs associated with the seven parameters (β_2, φ, β_1, q, p, γ_2, ε) are statistically significant. For San Francisco, the first five parameters with most impact on the final size are probability of showing symptoms p ($|PRCC| = 0.441\ 9$), transmission rate by E individuals β_1 ($|PRCC| = 0.263\ 9$), incubation period for E individuals $1/q$ ($|PRCC| = 0.153\ 4$), proportion of the confirmed symptomatic cases transferring to hospitalized cases ε ($|PRCC| = 0.141\ 5$), and transmission rate by A individuals β_2 ($|PRCC| = 0.098\ 1$). The PRCCs associated with these five parameters are statistically significant. The PRCCs associated with the other two parameters φ and γ_2 are not statistically significant.

The panels (e) and (f) show the temporal variation of the sensitivity of the effective reproduction number $R_e(t)$ to each parameter. For these two cities, it is shown that variations in the clustering coefficient φ, the transmission rate by A individuals β_2, transmission rate by E individuals β_1, probability of showing symptoms p, and incubation period for E individuals $1/q$ dominate the PRCCs during the shelter in place policy. When the shelter in place policy ends, the correlations between these nine parameters (β_2, β_1, q, p, ε, φ, γ_1, γ_2, γ_4) and the effective regeneration number $R_e(t)$ decrease. During the shelter in place policy, the clustering coefficient φ is the most influential parameter on the effective reproduction number. There is a highly negative correlation between the clustering coefficient φ and $R_e(t)$. The shelter in place policy makes the clustering coefficient increase and then reduces the effective regeneration number.

3.4. The effect of social distancing

Fig. 6 illustrates the effect of social distancing in New York and San Francisco. The panels (a) and (b) show the effect of social distancing on cumulative confirmed cases, while (c) and (d) show the effect of social distancing on the effective

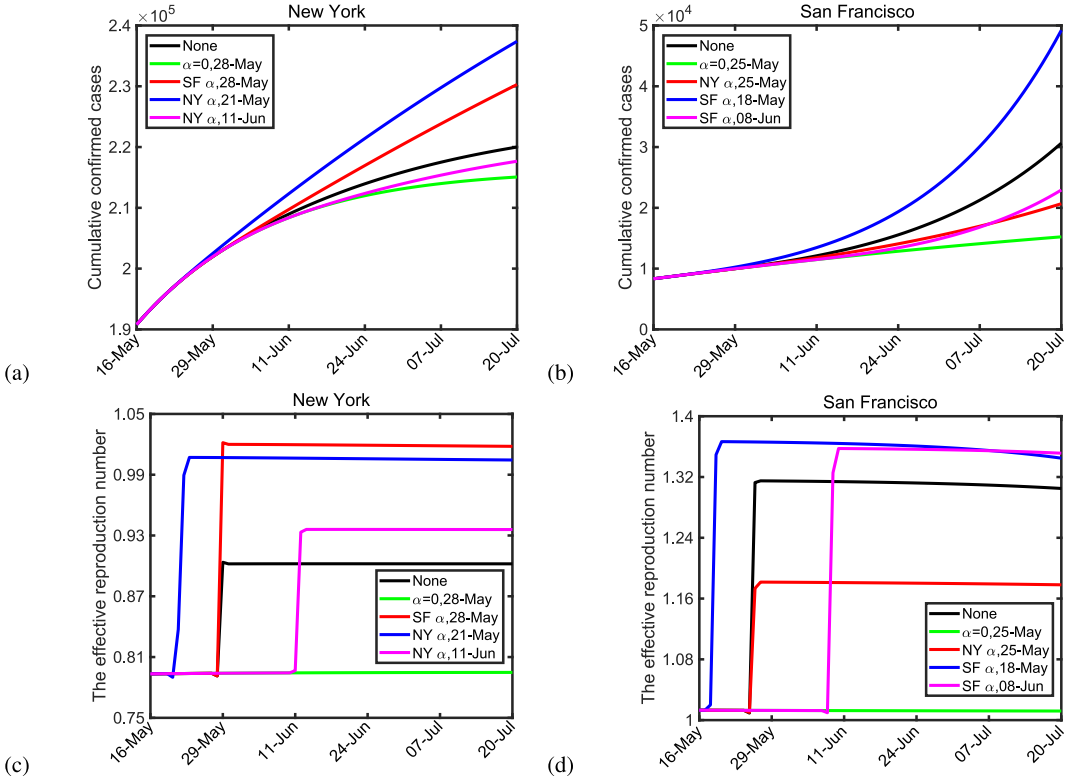

Fig. 6. The effect of social distancing in New York and San Francisco. The panels (a) and (b) show the effect on cumulative confirmed cases, while (c) and (d) show the effect on the effective reproduction number.

reproduction number. In all panels, black lines denote the actual case; green lines correspond to the case of keeping the shelter in place policy; red lines represent the case of changing the intensity of intervention α when shelter in place policy ended; blue lines represent bringing the shelter in place end time forward by one week; and purple lines correspond to delaying the shelter in place end time by two weeks. For New York, not lifting the shelter in place policy has a little effect on final size, but the epidemic duration will be shortened by more than 80 days. When the New York α keeps the same with that of San Francisco, that is, nearly doubling the intensity of intervention α, the final size will increase nearly 35%, and the epidemic duration will extend by one and a half more years. Bringing the New York shelter end time forward by one week will increase the final size by nearly 33%, and extend the epidemic duration by more than one year, whereas delaying the end time by two weeks has little effect on final size, but delays the epidemic duration by about 40 days. For San Francisco, not lifting the shelter policy will reduce the final size by 97%. When the San Francisco α keeps the same with that of New York, that is, reducing the intensity of intervention α by about half, the epidemic duration will be delayed by about 300 days and the final size will be reduced by about 21%; moving the shelter end time forward or later will shorten the epidemic duration and increase the final size. This means that New York will be less affected by reopening, but San Francisco will be more affected. Moreover, keeping a social distance helps reduce the final size of these two cities. The effective reproduction number in San Francisco is always more than one, which may lead to the results of San Francisco different from those of New York. It is noticed that these results are obtained under the assumption of no further changes in policy or transmission dynamics.

4. Discussion

Our comparative mathematical study of COVID-19 outbreaks in New York and San Francisco have allowed us to assess the impacts of population density and social distancing in these cities. Our approach was to use a pairwise network model to investigate COVID-19 epidemic in these two cities under a series of control measures. Using MCMC fitting to confirmed cases we were able to estimate the parameters and initial values of variables.

We explored the reasons for major differences between New York and San Francisco. Comparing the transmission in New York and San Francisco, we found that, under the current dynamics, the final size of San Francisco is predicted to be about 3.7 times as much as that of New York, and that the epidemic duration of San Francisco is more than 140 days longer than that of New York. This is because the effective reproduction number for New York was initially very high, but declined quickly due to social distancing. By way of contrast, San Francisco started with a lower effective reproduction number, but this did not decline as quickly. Unless there are major changes, the outcome for San Francisco may have an extended epidemic that ends off being larger than New York's.

We conducted a sensitive analysis by evaluating the PRCCs, and found that there is a highly negative correlation between the clustering coefficient and the basic reproduction number and the effective reproduction number. The shelter in place policy makes the clustering coefficient increase and then reduces the basic reproduction number and the effective reproduction number.

By mathematically switching population densities between San Francisco and New York, we were able to assess the impact of high-density interactions on COVID-19 dynamics. As expected, we found that population density has a major effect on disease transmission. The greater the population density is, the greater the final size of the epidemic is. Alternatively, by switching the transmission rates between the two cities we found that the final size increases as the rate of infection increases. As expected, social distance measures have a crucial and highly sensitive role to play in determining disease outcomes. Small changes such as increasing or decreasing the average number of contacts have the effect of allowing or preventing the outbreak in both New York and San Francisco. Thus, keeping strict social distance is conductive to disease prevention and control, and when social distance measures are relaxed, the epidemic may become more serious. Importantly, the timing of social distancing was also key, which is consistent with the conclusions in (Morris et al., 2020) on robust epidemic control.

Other approaches have been used to study the impact of physical distancing. In British Columbia, Canada a Bayesian model was used to analyze the population according to those willing and able to participate in distancing measures. The population was divided into two types of individuals: the non-physical-distancing and the physical-distancing (Anderson et al., 2020). By way of contrast, our model is based on the requirement of “shelter in place”. The quarantined individuals have close contacts with the confirmed individuals but do not participate in the COVID-19 infection.

As with many modelling problems, it is not always the case that a more complicated model is better. For example, using the Akaike Information Criterion (AIC) for model selection, Roda et al. showed that a susceptible-infectious-removed (SIR) model performs much better than a more realistic SEIR model in representing the information contained in the confirmed-case data (Roda et al., 2020). Of course, with the AIC model selection approach, the appropriate choice of model complexity depends on the type and amount of data available, and could change as more data become available. In the context of the models in this paper, assessing the appropriate level of complexity for pairwise interaction models, given the available data, is a possible venue for future work in this field, although it is likely that the optimal complexity from an information theoretic perspective will change as more data become available.

Funding

This work is partially supported by the National Natural Science Foundation of China grants 61 873 154 and 12 101 573, Health Commission of Shanxi Province grants 2020XM18, Shanxi Provincial Department of Science and Technology COVID-19 Emergency Special Fund grants 202003D31011/GZ, and Fundamental Research Program of Shanxi Province grants 20 210 302 124 608 and 20 210 302 124 381, and partially supported by a Canada Research Chair (MAL), NSERC Discovery Grants (HW and MAL), NSERC Discovery Accelerator Supplement Award (HW), and an Alberta Innovates grant 202 100 502.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Anderson, S. C., Edwards, A. M., Yerlanov, M., et al. (2020). *Estimating the impact of COVID-19 control measures using a Bayesian model of physical distancing*. medRxiv.
Chan, J., Yuan, S., Kok, K., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet, 395*(10223), 514–523.
van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences, 180*, 29–48.
Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., et al. (16-03-2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Imperial College London. https://doi.org/10.25561/77482
Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.
Hou, C., Chen, J., Zhou, Y., et al. (2020). The effectiveness of the quarantine of Wuhan city against the Corona Virus disease 2019 (COVID-19): Well-mixed SEIR model analysis. *Journal of Medical Virology, 92*(7), 841–848.
House, T., & Keeling, M. J. (2011). Insights from unifying modern approximations to infections on networks. *Journal of The Royal Society Interface, 8*, 67–73.
Huang, C., Wang, Y., Li, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet, 395*, 497–506.
Hu, Z., Cui, Q., Han, J., et al. (2020). Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province, China. *International Journal of Infectious Diseases*, (95), 231–240.
Keeling, M. J. (1999). The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London B, 266*, 859–867.
Keeling, M. J., Rand, D. A., & Morris, A. J. (1997). Correlation models for childhood epidemics. *Proceedings of the Royal Society of London B, 264*, 1149–1156.
Kimball, A., Hatfield, K. M., Arons, M., et al. (2020). Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility ł king county, Washington, March 2020. *MMWR Morb Mortal Wkly Rep, 69*(13), 377–381.
Leung, K., Wu, J. T., Liu, D., et al. (2020). First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: A modelling impact assessment. *Lancet, 395*, 1382–1393.
Li, Q., Med, M., Guan, X., et al. (2020a). Early transmission dynamics in Wuhan, China, of noval coronavirus-infected pneumonia. *New England Journal of Medicine, 382*, 1199–1207.
Li, R., Pei, S., Chen, B., Song, Y., Yang, W., & Shaman, J. (2020b). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science, 368*(6490), 489–493.
Luo, X., Feng, S., Yang, J., et al. (2021). Nonpharmaceutical interventions contribute to the control of COVID-19 in China based on a pairwise model. *Infectious Disease Modelling, 6*, 643–663.
Marino, S., Hogue, I. B., Ray, C. J., et al. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology, 254*(1), 178–196.
Moghadas, S. M., Shoukat, A., Fitzpatrick, M. C., Wells, C. R., Sah, P., Pandey, A., et al. (2020). Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proceedings of the National Academy of Sciences of the United States of America, 117*(16), 9122–9126.
Morris, D. H., Rossine, F. W., Plotkin, J. B., et al. (2020). *Optimal, near-optimal, and robust epidemic control*. arXiv preprint, Article 02209. arXiv:2004.
Moss, R., Wood, J., Brown, D., et al. (2020). *Modelling the impact of COVID-19 in Australia to inform transmission reducing measures and health system preparedness*. Preprint.
Ong, S. W. X., Tan, Y. K., Chia, P. Y., et al. (2020). Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a symptomatic patient. *JAMA, 323*(16), 1610–1612.
Pan, X., Chen, D., Xia, Y., et al. (2020). Asymptomatic cases in a family cluster with SARS-CoV-2 infection. *The Lancet Infectious Diseases, 20*(4), 410–411.
Park, M., Cook, A. R., Lim, J. T., et al. (2020). A systematic review of COVID-19 epidemiology based on current evidence. *Journal of Clinical Medicine, 9*(4), 967.
Roda, W. C. (2020). Bayesian inference for dynamical systems. *Infectious Disease Modelling, 5*, 221–232.
Roda, W. C., Varughese, M. B., Han, D., et al. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling, 5*, 271–281.
Tang, B., Bragazzi, N. L., Li, Q., et al. (2020a). An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infectious Disease Modelling, 5*, 248–255.
Tang, B., Wang, X., Li, Q., et al. (2020b). Estimation of the transmission risk of 2019-nCov and its implication for public health interventions. *Journal of Clinical Medicine, 9*, 462.
Tang, B., Xia, F., Tang, S., et al. (2020c). The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemics in the final phase of the current outbreak in China. *International Journal of Infectious Diseases, 95*, 288–293.
Wu, P., Hao, X., Lau, E. H. Y., et al. (2020). Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Euro Surveill, 25*(3), 2000044.
Xue, L., Jing, S., Miller, J. C., et al. (2020). A data-driven network model for the emerging COVID-19 epidemics in Wuhan, Toronto and Italy. *Mathematical Biosciences, 326*, 108391.
Yang, Z., Zeng, Z., Wang, K., et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease, 12*(3), 165–174.
Yu, P., Zhu, J., Zhang, Z., et al. (2020). A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. *The Journal of Infectious Diseases, 221*(11), 1757–1761.
Zhang, J., Litvinova, M., Liang, Y., et al. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science, 368*(6498), 1481–1486.
http://yl.szhk.com/2020/03/17/283059062276979.html.
https://covidmapper.ca/geomap/.
https://en.wikipedia.org/wiki/Alameda County, (California).
https://en.wikipedia.org/wiki/Contra Costa County, (California).

https://en.wikipedia.org/wiki/Marin County, (California).
https://en.wikipedia.org/wiki/New York City.
https://en.wikipedia.org/wiki/San Francisco.
https://en.wikipedia.org/wiki/San Mateo County, (California).
https://news.sina.com.cn/w/2020-04-11/doc-iirczymi5615931.shtml.
https://rt.live/us/CA.
https://rt.live/us/NY.
https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html.
https://www.census.gov/data/tables/time-series/demo/families/households.html.
https://www.cidrap.umn.edu/news-perspective/2020/04/study-many-asymptomatic-covid-19-cases-undetected.
https://www.soundhealthandlastingwealth.com/health-news/one-third-of-people-in-massachusetts-study-tested-positive-for-covid-19-antibodies.
https://www.washingtonpost.com/outlook/2020/04/20/we-tested-all-our-patients-covid-19-found-lots-asymptomatic-cases/.
https://www.who.int/.