

Understanding the Urban Pandemic Spreading of COVID-19 with Real World Mobility Data

Qianyue Hao*, Lin Chen*, Fengli Xu, Yong Li†

Beijing National Research Center for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing, China
liyong07@tsinghua.edu.cn

ABSTRACT

Facing the worldwide rapid spreading of COVID-19 pandemic, we need to understand its diffusion in the urban environments with heterogeneous population distribution and mobility. However, challenges exist in the choice of proper spatial resolution, integration of mobility data into epidemic modelling, as well as incorporation of unique characteristics of COVID-19.

To address these challenges, we build a data-driven epidemic simulator with COVID-19 specific features, which incorporates real-world mobility data capturing the heterogeneity in urban environments. Based on the simulator, we conduct two series of experiments to: (1) estimate the efficacy of different mobility control policies on intervening the epidemic; and (2) study how the heterogeneity of urban mobility affect the spreading process. Extensive results not only highlight the effectiveness of fine-grained targeted mobility control policies, but also uncover different levels of impact of population density and mobility strength on the spreading process. With such capability and demonstrations, our open simulator contributes to a better understanding of the complex spreading process and smarter policies to prevent another pandemic.

CCS CONCEPTS

• **Information systems** → **Wrappers (data mining)**; • **Applied computing** → **Life and medical sciences**.

KEYWORDS

Urban pandemic spreading; Mobility modeling; Epidemic simulator.

ACM Reference Format:

Qianyue Hao, Lin Chen, Fengli Xu, Yong Li. 2020. Understanding the Urban Pandemic Spreading of COVID-19 with Real World Mobility Data. In *The 26th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'20)*, August 23–27, 2020, Virtual Event, USA. ACM, New York, NY. 8 pages. <https://doi.org/10.1145/3394486.3412860>

* Equal contribution.

† Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '20, August 23–27, 2020, Virtual Event, CA, USA

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7998-4/20/08...\$15.00

<https://doi.org/10.1145/3394486.3412860>

1 INTRODUCTION

With a sudden outbreak, the COVID-19 pandemic has been sweeping the globe with a rapidly growing number of infected cases. Up to May 26, 2020, there have been over 5.4 million people infected, resulting in over 343,000 deaths, according to the situation report from the World Health Organization. At this crucial moment, it is of high significance to understand the COVID-19 spreading patterns in large-scale urban environments, based on which we can design smarter control policies to fight back. Moreover, it is necessary to take human mobility into consideration, as COVID-19 is known to be transmitted mainly via person-to-person contacts, which are highly correlated with human mobility patterns.

In the field of epidemiology, existing works on both the macro- and micro-levels have laid a foundation for understanding the spread of contagious diseases [2, 4, 6, 7, 13]. Facing this COVID-19 pandemic, recent studies focus on evaluating its high-level propagation [5, 14], estimating efficacy of certain containment measures [9, 11], as well as making predictions for future development [12, 18]. However, few is known about how the disease diffuses in a specific urban environment with frequent population flows, namely the intra-city level estimations. It also remains unclear how the heterogeneity in urban population distribution and mobility influences the spreading processes. Such insights are in urgent need for designing scientific containment measures and control policies.

Although modelling the COVID-19 spreading in urban environments with real mobility is important, three major challenges need to be dealt with. First, to balance between the accuracy and simplicity of the modelling, choosing an appropriate spatial-temporal resolution is difficult. Second, how to incorporate the urban-level mobility data with the epidemic model effectively to capture the spreading patterns is also unknown. Finally, the newly emerging COVID-19 has unique features such as a relatively large ratio of asymptomatic infectious cases, a high basic reproduction number, etc., which calls for modification of the existing epidemic models.

To solve the above challenges, we build a data-driven epidemic simulator on the level of city blocks, by incorporating the urban-level real-world mobility data to obtain more realistic simulations of the COVID-19 spreading process. Besides, we add an Asymptomatic Infectious state into the epidemic model, and carefully design the epidemic parameters to be consistent with the real scenario of COVID-19.

Based on the built simulator, we conduct extensive experiments to study the influences of mobility control interventions and heterogeneous population features on the process of disease spreading. First, we design various mobility control policies (dynamic/static,

between-/within- blocks). The implementation strength of the policies can be adjusted by changing the values of the critical parameters in the simulator. We then perform a series of simulations to evaluate their respective efficacy. Second, we take a closer look into the local features of population distribution and mobility strength. After dividing the city blocks into groups with similar levels of these two metrics, we use statistical methods to study the correlations between the population density/mobility levels and the epidemic spreading process. Conclusions are drawn that while both elements make an impact, population mobility is much more influential than population density.

Main contributions of our study can be summarized as follow:

- We build a data-driven epidemic simulator with COVID-19-specific parameters by incorporating urban scale mobility. Our simulator refines the meta-population method, which successfully captures the distinct features of different blocks within the city. Such capability makes it possible to evaluate fine-grained intervention policies in respond to such pandemic crises.
- Based on the simulator, we design and evaluate the efficacy of three types of mobility control policies, under a range of implementation strengths. From the results, we confirm the efficiency of adaptive dynamic control as well as within-block control policies.
- We conduct extensive analysis to explore the influence of population density and mobility on epidemic spreading processes. Experiments uncover different levels of effects caused by population density and mobility on the infection rate, the peak rate as well as the peak height of an epidemic outbreak. Our findings further support the necessity in conducting mobility control policies to intervene the spreading of COVID-19.

2 RELATED WORK

According to the different assumptions about population distribution, epidemic modelling approaches can be roughly classified into two categories: homogeneous models and heterogeneous models [1, 6, 15]. Classical homogeneous epidemic models are built upon a strong assumption that individuals of all states are mixed fully and homogeneously, which results in constant contact probabilities [1, 15]. Due to the homogeneity assumption, such models are inadequate to capture important characteristics of the real world. In contrast, our model explicitly incorporates heterogeneity in population density and mobility, to form a realistic picture of the environment.

On the other hand, simulators for heterogeneous models can be further categorized into metapopulation-based and agent-based simulators. (1) Metapopulation-based simulators divide the population into sub-groups connected by mobility flows. Metapopulations are constructed either merely from large-scale transportation networks [6], or supplemented by synthesized commuting flows [2, 3]. In contrast, our simulator directly utilizes real population flow data to connect different blocks, looking into the local features of the city to derive deeper understandings of the spreading mechanism. (2) Agent-based simulators can generate simulations from the micro-level of person-to-person contacts. Important instances include EpiSims [7] and EpiSimdemics [4]. However, due to the huge amount of information, they usually suffer from extremely high computational complexity. Moreover, there are usually a large

quantities of parameters in need of calibration [16]. In contrast, our simulator achieves a balance between accuracy and simplicity, by incorporating individual mobility data and organizing them on the block level.

3 METHOD

There exist three major challenges in modelling the COVID-19 spread in urban environments with real mobility.

- **Choosing an appropriate spatial-temporal resolution to capture the mobility of large-scale urban population.** For one thing, classical models usually focus on a macro level of pandemic spreading, namely the inter-city or even inter-country level [2, 6]. Though successfully revealing certain high-level diffusion trends, they fail to understand the intra-city situations. For the other thing, individual-based models can obtain more accurate descriptions, but they are computationally expensive, making it unrealistic to implement them on a real-world scale. It is also hard to collect sufficient data describing behaviors of every single citizen. Moreover, an excessive amount of detailed information may instead hinder insights of the key mechanisms from the simulation results [15]. Therefore, it is difficult to attain a balance between accuracy and simplicity.
- **Incorporating urban-level mobility data into the epidemic model effectively.** Real world urban mobility data record the mobility of population in the whole city while epidemic models describe the spreading process of the pandemic inside a sub-population. It is necessary to come up with a method that can obtain the descriptions of population flows interconnecting sub-populations and build a model of pandemic spreading process over the whole city.
- **Simulating the special epidemiological characteristics of COVID-19.** Apart from its high R_0 [19], COVID-19 results in a noticeably proportion of asymptomatic infectious cases [17]. This contributes to the unprecedented pandemic, as people may be unawarely infected yet keep moving around, leading to more infections. However, the traditional SIR/SEIR models [8, 10] fail to capture such characteristic, since they mostly do not distinguish different types of infected individuals.

In order to tackle these challenges, we build a simulator on the block level, extract descriptions of population distribution and mobility from real world mobility data and incorporate them into the epidemic model of COVID-19. The detailed methods are shown in the following section.

3.1 Real World Mobility Modelling

As discussed above, a properly chosen spatial-temporal resolution plays a key role in the study. In our scenario, people inside a certain block are regarded as a sub-population. This block-based spatial resolution enables us to simulate on the intra-city level, about various policies intervening the pandemic (see Section 4), and also the features of pandemic spreading processes (see Section 5). Specifically, we divide downtown Beijing into 675 blocks based on major transportation networks, as is shown in Figure 1 (a). For the temporal resolution, we simulate the spreading process of pandemic with a time step of 30 minutes.

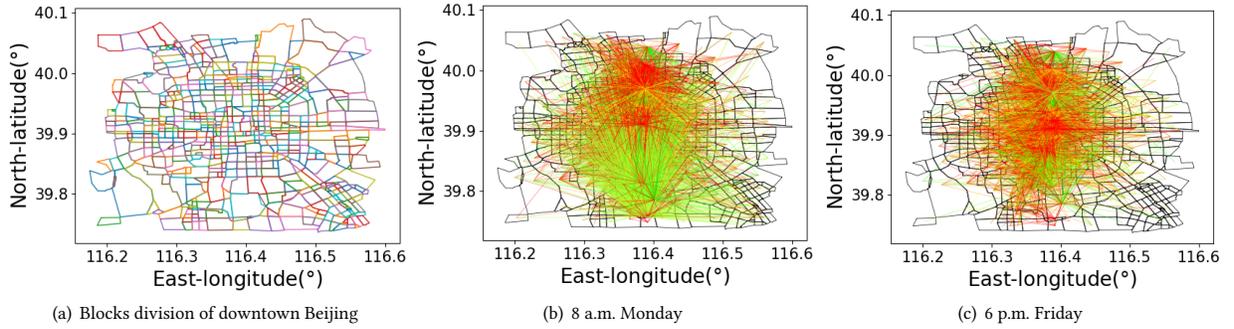


Figure 1: Blocks division of downtown Beijing and the strengths of population mobility related to several selected blocks (red indicates busy population mobility while green indicates the opposite): (a) blocks division of downtown Beijing; (b) strengths of population mobility at 8 a.m. Monday; (c) strengths of population mobility at 6 p.m. Friday.

Our real world mobility data comes from one of China’s most popular Internet service provider, by collecting GPS coordinates of users accessing location-based services. It offers fine-grained descriptions of intra-city population mobility. The raw data contains records of users’ positions every 30 minutes during 3 weeks, with the following information in detail:

- The number of Internet service users who travel from the i -th block to the j -th during the t -th time step, which is denoted as: $m_{ij}(t)$, where $i, j \in \{0, 1, 2, \dots, 674\}$, $t \in \{0, 1, 2, \dots, 1007\}$.
- The number of Internet service users in the i -th block at the start of t -th time step, which is denoted as: $n_i(t)$, where $i \in \{0, 1, 2, \dots, 674\}$, $t \in \{0, 1, 2, \dots, 1007\}$.

Considering the population mobility in a city has an obvious periodicity of 7 days (a week, 336 time steps), due to the alteration of weekdays and weekends, we calculate an average number of $m_{ij}(t)$ and $n_i(t)$ over the three weeks and get the average number of Internet service users who travel from the i -th block to the j -th during the t -th time step, denoted as $M_{ij}(t)$; and the average number of Internet service users in the i -th block at the start of t -th time step, denoted as $N_i(t)$, using the following method:

$$M_{ij}(t) = \frac{m_{ij}(t) + m_{ij}(t + 336) + m_{ij}(t + 672)}{3}, \quad (1)$$

and

$$N_i(t) = \frac{n_i(t) + n_i(t + 336) + n_i(t + 672)}{3}, \quad (2)$$

where $i, j \in \{0, 1, 2, \dots, 674\}$, $t \in \{0, 1, 2, \dots, 335\}$.

The total population size in the i -th block is denoted as $N[i]$, $i \in \{0, 1, 2, \dots, 674\}$, and the strength of population flows interconnecting the blocks is measured by the possibility for people in the i -th block to travel to the j -th at the t -th time step, denoted as $P_{ij}(t)$, $i, j \in \{0, 1, 2, \dots, 674\}$, $t \in \{0, 1, 2, \dots, 335\}$. All $N[i]$ and $P_{ij}(t)$ are estimated through the real world mobility data. The strengths of population flow related to several selected blocks are shown in Figure 1 (b) (c). We can observe the difference between Monday morning and Friday afternoon, confirming the periodicity mentioned above. Since we have obtained the population distribution as well as mobility strength from the real world mobility data, we can further incorporate them into the epidemic model to build a reliable simulator.

3.2 Simulator Building

In this section, we introduce the construction of our epidemic simulator, which refines a GLEaM structured framework [2] to incorporate the real world mobility data. Inside the framework, we design a compartment model with 6 states to describe the development of the disease. Especially, a state of Asymptomatic Infectious is included for a better description of the influence of the notably relatively high occurrence rate of asymptomatic infection cases [17], which is one of the outstanding features of COVID-19. Therefore, possible states include Susceptible (S), Latent (L), Infectious (I), Asymptomatic Infectious ($I^{[a]}$), Dead (D) and Recovered (R).

Specifically, all possible state transition processes are described in Figure 2, with the respective transition probabilities. The parameter β is the average number of new latent cases caused by one existing infectious individual per day, through Infectious-Susceptible ($I - S$) contacts. Thus it measures the infectiousness of the pandemic. For a Latent-Susceptible ($L - S$) contact and an Asymptomatic Infectious-Susceptible ($I^a - S$) contact, it is weighted by r_L and r_a . The parameter ϵ describes the probability for a Latent individual to become Infectious or Asymptomatic Infectious per day. Further, P_a determines the proportion of Asymptomatic cases among all infections. The parameters α , μ and μ_a respectively characterize the $I \rightarrow D$, $I \rightarrow R$, $I^a \rightarrow R$ transition probabilities per day. The number of cases in state m in the j -th block at t -th time step is denoted as $X_j^{[m]}$, $m \in \{S, L, I, I^{[a]}, D, R\}$, $j \in \{0, 1, 2, \dots, 674\}$.

At initialization, each block is assigned an initial infection number $X_j^{[I]}$, where the subscript j stands for the j -th region. By subtracting $X_j^{[I]}$ from the total population of the j -th region, we obtain $X_j^{[S]}$, the initial number of susceptible individuals in each region.

In each time step, the simulator updates the number of individuals in different epidemic states in the following two steps:

- **Step 1: update $X_j^{[m]}$, according to individual disease development (state transitions):**

The transition number from state m to n in the j -th block is denoted as $F_j[m][n]$, given by the following polynomial distribution. (Polynomial distribution is a discrete probability distribution form, which is an extension of binomial distribution):

$$\{F_j[m][S], F_j[m][L], \dots\} \sim PN(X_j^{[m]}; P_{m \rightarrow S}, P_{m \rightarrow L}, \dots), \quad (3)$$

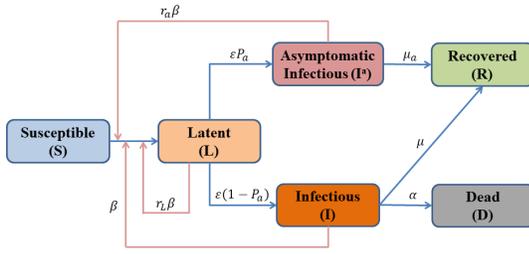


Figure 2: All possible state transition processes in our epidemiological model, along with the respective transition probabilities.

where $P_{m \rightarrow n}$ denotes the possibility for people in state m to transfer to n in one time step.

Then we update $X_j^{[m]}$ as:

$$X_j^{[m]} \leftarrow X_j^{[m]} + \Delta X_j^{[m]}, \quad (4)$$

where

$$\Delta X_j^{[m]} = \sum_{[n_1]} F_j[n_1][m] - \sum_{[n_2]} F_j[m][n_2], \quad (5)$$

$$m, n, n_1, n_2 \in \{S, L, I, I^{[a]}, D, R\}, j \in \{0, 1, 2, \dots, 674\}.$$

- **Step 2: update $X_j^{[m]}$, according to human mobility among blocks:**

The number of people in state m , traveling from the i -th block to the j -th is denoted as $T_{ij}[m]$, given by the following polynomial distribution:

$$\{T_{i0}[m], T_{i1}[m], \dots\} \sim PN(X_i^{[m]}, P_{i0}(t \bmod 336), P_{i1}(t \bmod 336), \dots). \quad (6)$$

Then we update $X_j^{[m]}$ as:

$$X_j^{[m]} \leftarrow X_j^{[m]} + \Delta X_j^{[m]}, \quad (7)$$

where

$$\Delta X_j^{[m]} = \sum_{i=0}^{674} (T_{ij}[m] - T_{ji}[m]), \quad (8)$$

$$m \in \{S, L, I, I^{[a]}, D, R\}, i, j \in \{0, 1, 2, \dots, 674\}.$$

We obtain $P_{m \rightarrow n}$ $m, n \in \{S, L, I, I^{[a]}, D, R\}$ according to medical researches on COVID-19, thus we can achieve an accurate description of such pandemic.

The simulator is programmed in Python, using efficient matrix computing libraries and programming skills. It is worth mentioning that due to the rational spatial-temporal resolution choice and proper programming technique, we can simulate the spreading of pandemic very efficiently, only consuming no longer than 8 seconds to simulate the spreading process for 1 day (48 time steps) in the whole downtown Beijing, on a simple laptop.

Our built simulator is open for the community and available at <https://github.com/KYHKL-H/Epidemic-Simulator>.

4 EFFECT OF MOBILITY CONTROL POLICY

Population mobility is a critical element governing peoples' contact patterns in the city. Therefore, mobility control is a common policy to intervene the spread of pandemic. The positive effect of mobility control between cities on the global level has been proven [5]. However, it is still unclear how mobility control on block levels within a single city affects the pandemic spread. To better understand the process, we design various mobility control policies both on the level of inter-blocks and within-blocks. We perform the spreading simulations using the built simulator mentioned above to estimate the effectiveness of these policies.

4.1 Experimental Settings

We simulate a pandemic outbreak with 50 initial cases caused by zoonotic exposure, a typical order of magnitude [5]. The initially outbreak spot is set to be block #0, and it stays the same during this section, eliminating the influence of different initially outbreak spots. The simulation lasts 500 days, covering whole process of the pandemic from outbreak to fading away. The following three different mobility control policies are performed:

- **Simple-ratio mobility control between blocks:** The traffic flows between different blocks (any flow crosses the boundary of blocks, i.e., public transport taking, private car driving or walking) in the whole city are restricted by a given restriction rate γ ($0 < \gamma \leq 1$).
- **Dynamic mobility control between blocks:** To find a practical dynamic policy, we flexibilize the mobility control policy with two tunable parameters, control-threshold C_t and adjusting-frequency C_f (both positive integers), rather than using a simple restriction ratio. We perform a judgment after every period of time lasting C_f (h). If the number of currently infectious cases (I) in a block is greater than C_t (person), all traffic flows from or to this block are banned until the next judgment.
- **Mobility control within blocks:** The traffic flows within every block, i.e., without crossing any block boundaries, in the whole city are restricted. In lack of data describing the actual traffic flows within blocks, it is difficult to directly set a restriction rate as above. Considering restriction within blocks uniformly reduces the chance for people to contact with each other, we solve this problem by setting a reduction factor γ_0 ($0 < \gamma_0 \leq 1$). γ_0 acts on β , leading to a decreased $\beta^* = \gamma_0 * \beta$ and this new β^* reflects the restriction within blocks.

The result of simulations are analysed by viewing the curves that show the spreading process of the pandemic, and a further analyse through applying statistical techniques on the critical three indexes:

- **Infect-rate:** the percentage of infected cases at a certain time, which is defined as:

$$r_I(t) = \frac{N_t[I] + N_t[I^{[a]}]}{N_0}. \quad (9)$$

where $N_t[m]$, $m \in \{S, L, I, I^{[a]}, D, R\}$ stands for the number of cases in state- m at the t -th time step, N_0 denotes the total number of population in downtown Beijing.

- **Peak-date:** time when infect-rate reach the top during the whole process of pandemic, which is defined as:

$$t_p = \arg \max_t r_I(t). \quad (10)$$

- **Peak-height:** the highest level of infect-rate during the whole process of pandemic, i.e., infect-rate at peak-date, which is defined as:

$$h_p = \max_t r_I(t). \quad (11)$$

4.2 Results and Analysis

Results of simulation under simple-ratio mobility control between blocks are shown in Figure 3, where "baseline" refers to situation without any traffic restriction.

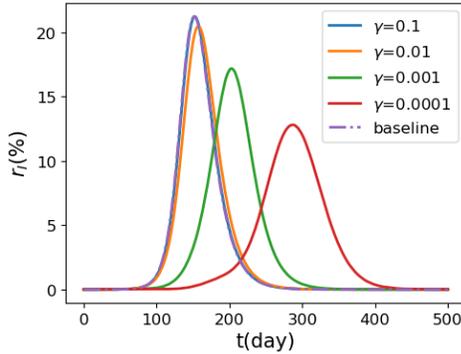


Figure 3: Infect-rate under simple-ratio mobility control between blocks with different γ , where the baseline-cure overlaps with the blue one.

Five different values (1.0,0.1,0.01,0.001,0.0001) of γ are chosen, and their effect on are shown in Table 1. It can be noticed that when γ is set to 1.0, 0.1 or 0.01, there is no obvious effect of delaying the arrival of the peak of this pandemic (no more than 3.56%) and neither an obvious effect of decreasing the height of it (no more than 3.62%). Only when γ is set to value 0.001 or 0.0001 (i.e., 99.9% or 99.99% of the traffic flow is banned), will it present an obvious effect on delaying and decreasing the peak. For such strict mobility control is almost unattainable in the real world, it has very limited positive effect on relieving the pressure of the medical system and intervening this pandemic. It suggests that we need more effective policies than this simple-ratio mobility control between blocks.

Table 1: Effect of different γ on t_p and h_p under simple-ratio mobility control between blocks.

$-\log(\gamma)$	$t_p/day(\text{delay}/\%)$	$h_p/\%(decrease/\%)$
baseline	151.7(-)	21.25(-)
1	152.1(0.26)	21.23(0.09)
2	157.1(3.56)	20.48(3.62)
3	202.8(33.68)	17.19(19.11)
4	287.1(89.26)	12.81(39.72)

As an improvement of the simple-ratio mobility control between blocks, we put forward a dynamic mobility control between blocks, and the values of the two parameters are $C_t \in \{10, 5, 2\}$ (person), $C_f \in \{24, 48, 96\}$ (h), respectively, resulting in 9 simulations in total. The results are shown in Figure 4, where the "baseline" also refers

to the situation without any mobility control. The effects of this policy with different C_f and C_t on the spread of the pandemic are shown in Table 2.

Table 2: Effect of different C_f, C_t on t_p and h_p under dynamic mobility control between blocks.

$C_f - C_t$	$t_p/day(\text{delay}/\%)$	$h_p/\%(decrease/\%)$
baseline	151.7(-)	21.25(-)
24 - 10	150.2(-0.99)	20.55(3.29)
24 - 5	157.1(3.56)	18.91(11.01)
24 - 2	163.1(7.51)	14.19(33.22)
48 - 10	152.6(0.59)	20.59(3.11)
48 - 5	156.5(3.16)	19.12(10.02)
48 - 2	160.3(5.67)	16.05(24.47)
96 - 10	151.8(0.07)	20.62(2.96)
96 - 5	152.0(0.20)	19.48(8.33)
96 - 2	157.8(4.02)	15.96(24.89)

It can be noticed that whatever C_f is, when the C_t is set to 2, an obvious decrease on peak-height happens (24.47% or more). Meanwhile, when C_t is a fixed small number, even if C_f is not very small, it still has an obvious positive effect on intervening the pandemic. This suggests that there is no need to adjust the policy very frequently (i.e., one time per day), which adds to the implementability of this policy in the real world. Although this dynamic mobility control does not lead to an obvious delay in peak-date, it does lead to a decrease in peak-height, which may serve to relieve the pressure of the medical systems in the real world. And thus this dynamic mobility control policy is of great meaning.

The influence of this dynamic mobility control policy on the daily operation of the city is also evaluated. Only the situation of $C_t = 2$ is evaluated, for larger control-thresholds do not have an obvious effect on the intervention of this pandemic. Figure 5 shows the influence on traffic flows under dynamic mobility control policies with $C_t = 2$ (restrict-ratio refers to the percentage of banned traffic). As showing above, only when γ is set to 0.001 (red line in Figure 5) or smaller, will it make an obvious difference. In comparison, this dynamic one has a lower influence on daily traffic flows, and the influence goes down with the pandemic fading away. In other words, there is no need to keep a strict restriction until the number of infected cases turns to zero.

We extend the mobility control policy further into inner-block level. The method is by setting the reduction factor γ_0 as defined above to infect rate β , 4 values of γ_0 are chosen, i.e., $\gamma_0 \in \{0.9, 0.8, 0.7, 0.6\}$ and the result is shown in Figure 6 ("baseline" refers to situation without any mobility control). The effect of different values of γ_0 is shown in Table 3.

Table 3: Effect of different γ_0 on t_p and h_p under mobility control within blocks.

γ_0	$t_p/day(\text{delay}/\%)$	$h_p/\%(decrease/\%)$
baseline	151.7(-)	21.25(-)
0.9	170.5(12.39)	17.97(15.44)
0.8	208.2(37.24)	14.31(32.66)
0.7	253.4(67.04)	10.30(51.53)
0.6	379.7(150.30)	6.10(71.29)

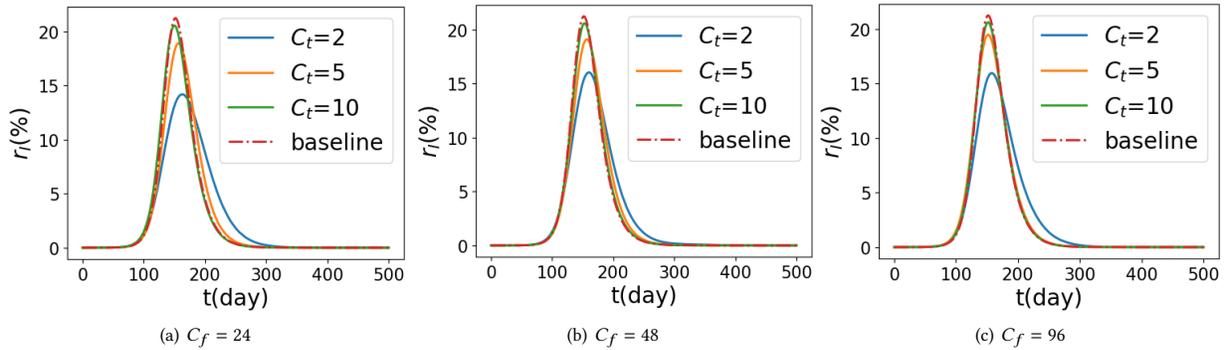


Figure 4: Infect-rate under dynamic mobility control between blocks with different $freq$ and C_t .

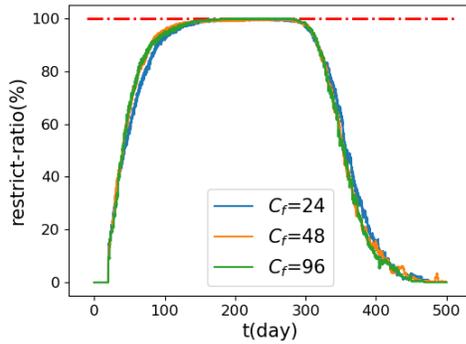


Figure 5: Influence on traffic flow under dynamic mobility control between blocks with $C_t = 2$ (the red line marks 99.99%).

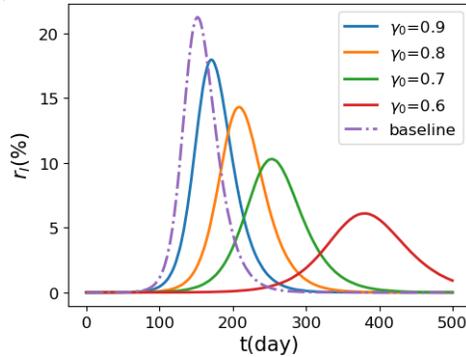


Figure 6: Infect-rate under mobility control within blocks with different γ_0 .

We can tell from the result that the mobility control on the inner-block level, which reduces the chance for people to contact with each other and finally decrease the chance for the pandemic to spread, make a much more obvious difference on intervening the pandemic. Even when the reduction factor γ_0 is close to 1.0, both an obvious delay in peak-date ($\gamma_0 = 0.8$, 37.24% delay on peak-date) and an obvious decrease in peak-height ($\gamma_0 = 0.8$, 32.66% decrease on peak-height) happen.

4.3 Summary

From the result of three different mobility control policies shown above, we can conclude that simple-ratio mobility control between blocks makes a difference only when the restriction is extremely

strict, while dynamic mobility control between blocks shows a more obvious effect and has a lower influence on daily traffic flows when the control-threshold is small enough. In comparison, mobility control within blocks has a much more obvious effect on delaying the arriving time and decreasing the height of the peak, even if the restriction is not that strict. This does inspire us that in order to intervene the pandemic, implementing a multi-level mobility control, especially one with restrictions on small scales (i.e., within the blocks), is of vital importance.

5 EFFECT OF POPULATION AND MOBILITY

Above study shows how different traffic restriction policies affect the spreading of pandemic. However, it is still unclear how the pandemic spreading processes inside the city are affected by population distribution and mobility of blocks. In order to solve this confusion, we conduct the following experiments and analyses.

5.1 Population Distribution and Mobility

Among the 675 blocks divided originally, we wipe out the ones with too few data related to them, and 627 remain. In order to study the effect of population distribution and mobility on pandemic spreading features quantitatively, we define the following two metrics of a given block ($i = 0, 1, \dots, 626$):

- **Population density:** population size on unit area, which is defined as:

$$\rho_0[i] = \frac{N[i]}{area[i]}, \quad (12)$$

where $area[i]$ refers to the area of the i -th block.

- **Population mobility:** the sum of possibility for people in the i -th block to travel to other blocks and possibility for people in other blocks to travel to the i -th, which is defined as:

$$f_0[i] = \sum_{t=0}^{335} \sum_j (P_{ij}(t) + P_{ji}(t)). \quad (13)$$

For convenience in mathematical operations, we perform normalization on ρ_0 and f_0 , and obtain the relative population density $\rho[i] = \frac{\rho_0[i]}{\max \rho_0[i]}$ and relative population mobility $f[i] = \frac{f_0[i]}{\max f_0[i]}$. The relative population density and flow of the 627 regions are shown in Figure 7 (a). We discover that the blocks are naturally associated with various combinations of features, i.e., high/medium/low population density and high/medium/low population flow. In other

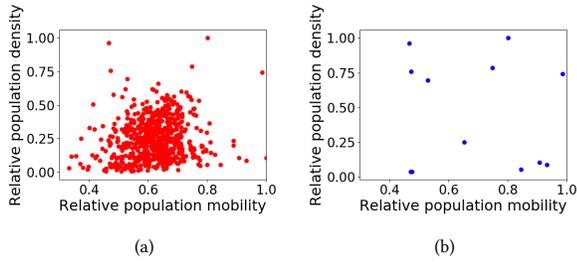


Figure 7: Relative population density and flow: (a) all 627 blocks; (b) blocks selected as outbreak spots.

words, the blocks are diversified and representative, which offers great convenience to our study.

5.2 Experimental Settings

Now we research into the pandemic spreading process on a much smaller scale, i.e., in each block. With a substantial increase in the spatial resolution, it is necessary to consider the randomness in the place of the initial outbreak. Thus, we choose 13 blocks with different population densities and mobility as the outbreak spots, shown in Figure 7 (b). In the set of selected blocks, both the population density and the mobility strength vary from low to high, making them representative of blocks with different features.

We take the selected blocks as the outbreak spot one by one for simulations, assuming 50 cases caused by zoonotic exposures each time and each simulation lasts 1 year (365days). To get more credible results, we run every simulation for 5 times (i.e., 13 outbreak spots and 5 simulations for each outbreak spots, 65 simulations in total). In seek of a deeper understanding, besides the three indexes (infect-rate r_I , peak-date t_P and peak-height h_P) defined above, we further study on another index, namely the accumulated infect-rate. It is the percentage of people previously or currently infected at time t , which is defined as follows:

$$r_{al}(t) = \frac{N_t[L] + N_t[I] + N_t[I^{[a]}] + N_t[R] + N_t[D]}{N_0}. \quad (14)$$

5.3 Results and Analysis

The result in one of the blocks (block NO.101) is shown in Figure 8, and other blocks has a similar variation trend.

5.3.1 Curve Fitting. In order to quantitatively describe the epidemic curves, we perform the method of curve fitting, expressing the spreading process using a brief function as follows:

- **Fitting infect-rate:**

$$r_I(t) \sim m * e^{-\frac{(t-\mu)^2}{\sigma}}. \quad (15)$$

Considering the attributions of Gaussian-like function, we have

$$h_P \Leftrightarrow m(\%), \quad t_P \Leftrightarrow \mu(\text{day}). \quad (16)$$

The value of σ measures the relative lasting time of the pandemic, the larger σ is, the longer this pandemic lasts.

- **Fitting accumulated infect-rate:**

$$r_{al}(t) \sim k * e^{at}, \quad (17)$$

where the value of a measures the relative spreading speed of the pandemic: the larger a is, the faster this pandemic diffuses.

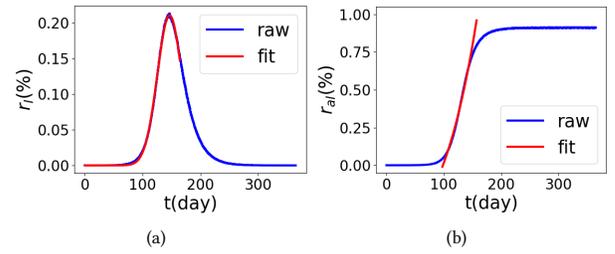


Figure 8: Result of simulation in block NO.101: (a) infect-rate; (b) accumulated infect-rate.

The fitting result of block NO.101 is also in Figure 8. For the infect-rate, $R^2 \in [0.9983, 0.9998]$; while for the accumulated infect-rate, $R^2 \in [0.9947, 0.9980]$, both indicating a very good fitting result.

As the outbreak period of the pandemic rather than the fading-away period is of more concern, curve fitting is only applied to the first half of the infect-rate curve and the exponentially growing part on the accumulated infect-rate curve, as is shown in Figure 8. Moreover, we average the results of all the simulations, eliminating the randomness brought by different outbreak spots and other factors. The mean and median value of a , σ , μ , m of 627 blocks are shown in Table 4. According to the median of every index, we divide them into the lower part and the higher part.

Table 4: Average and median value of indexes in 627 blocks.

index	average	median
a	0.01499	0.01467
σ	895.84	897.29
$\mu(\text{day})$	146.56	146.60
$m(\%)$	20.98	21.00

5.3.2 Effect of Population Density on the Spreading of Pandemic. According to the quantile of relative density of the blocks, we divide 627 blocks into 5 groups, with very low, low, medium, high, very high population density, respectively. The medium group is chosen as the reference. From the effects of population density on the pandemic spreading, we can draw the following conclusions.

First, very high population density tends to cause a decrease on a [OR=0.328 (95%CI: 0.227-0.642), p-value=0.0004], in other words, decrease the spreading speed. Note that the spreading is measured by the infect-rate rather than the absolute number of infected cases. Thus blocks with higher population density may witness a faster increase in the number of infected cases, but a slower increase on the infect-rate. However, when the population density varies among other groups (very low, low or high), no obvious effect on a happens, indicated by a p-value>0.05.

Secondly, very low population density tends to cause a decrease on σ [OR=0.446 (95%CI: 0.268-0.742), p-value=0.0027], μ [OR=0.504 (95%CI: 0.302-0.840), p-value=0.0119] and m [OR=0.475 (95%CI: 0.285-0.791), p-value=0.0058], in other words, causes a shorter lasting time, an earlier peak-date and a lower peak-height. However, when the population density varies among other groups (low, high or very high), no obvious effect on σ , μ and m is observed, indicated by a p-value>0.05.

5.3.3 Effect of Population Mobility on the Spreading of Pandemic. Using a similar statistical technique as above, 627 blocks are divided into 5 groups, and the medium group is chosen as the reference.

The results show that the value of a decreases obviously with the population mobility going up (OR>1 when population mobility is lower than the reference group and OR<1 when population mobility is higher than reference, with p-value no larger than 0.0241), indicating blocks with a high population mobility show a slower spreading inside themselves. In contrast, the value of m increases obviously when population mobility goes up (OR<1 when population mobility is lower than the reference group and OR>1 when population mobility is higher than the reference group, with p-value no larger than 0.0290), indicating blocks with high population mobility meet a higher infect-rate peak. This especially alarms us that the medical systems in high-mobility blocks are highly risky to suffer from heavier pressure. And thus more effective control policies and more adequate preparations are called for. Moreover, very low population mobility tends to cause a decrease on σ [OR=0.228 (95%CI: 0.131-0.399), p-value<0.0001] and μ [OR=0.146 (95%CI: 0.080-0.267), p-value<0.0001], in other words, causing a shorter lasting time and a earlier peak-date. However, when the population mobility varies among other groups (low, high or very high), there is no obvious effect on σ and μ , indicated by a p-value>0.05.

5.3.4 Summary. We found in the above experiments that the spreading process of COVID-19 is shaped by different population distribution and mobility patterns. Very high population density leads to a decrease on the spreading speed, while very low population density results in a shorter lasting time, an earlier peak-date and a lower peak-height. The spreading speed inside the block decreases and peak-height increases obviously as population mobility goes up, and very low population density leads to a shorter lasting time and a earlier peak-date. It can also be found that population mobility has a much more obvious effect on the spreading process than population density. This sheds light on the key role that the population mobility plays in the process of pandemic spreading, providing an evidence for the necessity of intervening pandemic through mobility control.

6 CONCLUSION

In this paper, we incorporate real world mobility data into the epidemic model of COVID-19 and build an efficient simulator on the block level, which is open to the community. By studying the influences brought by various mobility control policies under different implementation strength, we conclude that mobility control within blocks is the most effective one, which offers a meaningful suggestion for the real world pandemic intervention. We further study the differences in pandemic spreading due to varies population density and population mobility levels. Our study not only sheds light on the key role of mobility control policies, but can also be very meaningful for making targeted policies to fight the pandemic in different regions. Based on this open simulator, more intelligent policies and fundamental spreading phenomena can be investigated, which calls for collaborative future work of the community.

ACKNOWLEDGMENTS

This work was supported in part by The National Key Research and Development Program of China under grant 2018YFB1800804, the National Nature Science Foundation of China under U1936217,

61971267, 61972223, 61941117, 61861136003, Beijing Natural Science Foundation under L182038, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

REFERENCES

- [1] Linda J S Allen. 2017. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling* (2017).
- [2] Duygu Balcan, Vittoria Colizza, Bruno Goncalves, Hao Hu, Jose J Ramasco, and Alessandro Vespignani. 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* (2009).
- [3] Duygu Balcan, Hu Hao, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, and Van Den Wouter Broeck. 2009. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. (2009).
- [4] Christopher L. Barrett, Keith R. Bisset, Stephen G. Eubank, Xizhou Feng, and Madhav V. Marathe. 2008. EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *International Conference for High Performance Computing, Networking, Storage Analysis*.
- [5] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore Y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020).
- [6] Vittoria Colizza, Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America* (2006).
- [7] S Eubank, H Guclu, Kumar Vsa, M. V. Marathe, A Srinivasan, and Z Toroczkai. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* (2004).
- [8] P. Giles. 1977. The Mathematical Theory of Infectious Diseases and Its Applications. *Journal of the Operational Research Society* (1977).
- [9] Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Fiona Sun, Stefan Flasche, Billy J Quilty, Nicholas Davies, Yang Liu, Samuel Clifford, Petra Klepac, Mark Jit, Charlie Diamond, Hamish Gibbs, Kevin van Zandvoort, Sebastian Funk, and Rosalind M Eggo. 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* (2020).
- [10] W O Kermack and A G Mckendrick. 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences* (1927).
- [11] Moritz U G Kraemer, Chiahung Yang, Bernardo Gutierrez, Chiehhsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno Rodrigues Faria, Ruoran Li, William P Hanage, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *medRxiv* (2020).
- [12] Fotios Petropoulos and Spyros Makridakis. 2020. Forecasting the novel coronavirus COVID-19. *PLoS ONE* (2020).
- [13] Filippo Simini, Marta C Gonzalez, Amos Maritan, and Albert-laszlo Barabasi. 2012. A universal model for mobility and migration patterns. *Nature* (2012).
- [14] Ashleigh R Tuite, Victoria Ng, Erin Rees, and David Fisman. 2020. Estimation of COVID-19 outbreak size in Italy. *The Lancet Infectious Diseases* (2020).
- [15] Van, Mieghem, Piet, Vespignani, Alessandro, Castellano, Claudio, Pastor-Satorras, and Romualdo. 2015. Epidemic processes in complex networks. *Reviews of Modern Physics* (2015).
- [16] Srinivasan Venkatramanan, Bryan Lewis, Jiangzhuo Chen, Dave Higdon, Anil Vullikanti, and Madhav Marathe. 2018. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics* (2018).
- [17] Yixuan Wang, Yuyi Wang, Yan Chen, and Qingsong Qin. 2020. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *Journal of Medical Virology* (2020).
- [18] Joseph T Wu, Kathy Leung, and Gabriel M Leung. 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* (2020).
- [19] Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S. Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, and Maggie H. Wang. 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases* (2020).