Human Mobility based Individual-level Epidemic Simulation Platform

Zipei Fan[1,2], Xuan Song[1,2]*, Yinghao Liu[2], Zhiwen Zhang[2], Chuang Yang[2],
Quanjun Chen[1,2], Renhe Jiang[1,2], Ryosuke Shibasaki[1,2]
[1]SUSTech-UTokyo Joint Research Center on Super Smart City,
Department of Computer Science and Engineering,
Southern University of Science and Technology (SUSTech), Shenzhen, Guangdong, China
[2]Center for Spatial Information Science, University of Tokyo, Kashiwa, Chiba, Japan
fanzipei@iis.u-tokyo.ac.jp; songx@sustech.edu.cn

Abstract

Coronavirus has spread worldwide and about 3.5 million people have been confirmed infected and over 300 thousands people died. Scientists have reached the consensus that human mobility is the principal factor in spreading the virus and mobility should be restricted to control the epidemic. Against this background, we propose a novel coronavirus (COVID-19) fine-grained transmission model based on real-world human mobility data and develop a platform that maps the movement of people before determining transit flows and infection probabilities. Algorithms incorporate a series of incubation period and infection vector analysis. The next step is to work backward to find patients that have not yet been diagnosed by following the chain of transmission. The platform also aims to determine at-risk members of the population based on the travels of infected patients and provide early warning to those members of society. The multi-functional platform improves the opportunities for community leaders and decision-makers to implement different policies at the municipal and local levels, for a safe and healthy society. For example, local decision-makers can set optimal prevention and control policies based on the transmission within their local community.

1   Introduction

Human mobility plays an important role in epidemic control especially for contagious diseases, like COVID-19 [4, 5], which is highly contagious even before symptoms shown up. Thus, to fast identify the potential patients, take effective measures to prevent the spreading of the disease, and simulate the future spreading trends are important to reduce the outbreak risk and get the epidemic under control.

However, sensing human mobility at a citywide/nationwide level and at the individual level is critical to a more fine-grained epidemic control but is difficult before the popularization of Internet of Things and modern localization technologies. Recent years, portable devices equipped with localization functions and internet connection have been popularized and make it possible a nearly real-time sensing and integration of the human mobility at a large-scale.

Moreover, conventional epidemic model (e.g. SIR model [1], SEIR model [2]) the future trends of the epidemic spreading at a coarse level, from the following 3 aspects: 1) the contact ratio is assumed to be constant

---

*Corresponding author: Xuan Song, Email: songx@sustech.edu.cn

during the simulation, however, in practice, the contact ratio is quite different from days to days, especially from weekdays to weekends, 2) the contact ratio is assumed to be uniformed at different locations in the city, while the city/nation is taken as a whole while the downtown area, where people are more crowded and rural area, where people are much less crowded, the contact ratios are quite different, and 3) we can hardly estimate the infected risks of each individual.

Recent advances in epidemic simulations and predictions have taken human mobility into consideration. [3] proposed a proxy for individual mobility and use it to describe the people flow, and predict the spreading of an influenza-like contagious disease. The proxy utilized in the paper is close to the OD information [5] is a recent study on the human mobility effect in the spreading of COVID-19, but only the statistics of OD information is used to estimate the spatial transition of human mobility instead of trajectories of each user. This makes it difficult to find potential contacts and cannot predict the infection risk of each individual.

Moreover, it has been proven that the most effective way to slow down the spreading speed of COVID-19 is to take some public policies such as city lockdown, telework or setting temperature check points. All these public policies are expensive, so it is important for the governments to know how effective a public policy will be before it is taken into actions.

Against this background, in this paper, we develop a novel epidemic simulation platform that provide three principal functions: 1) find out and visualize the primary/secondary contacts on the map, which gives an illustrative view of the spreading risks, 2) simulate the epidemic spreading on the real-world human trajectory dataset, where people contacting behaviors are estimated by GPS trajectories, and 3) simulate the human mobility in response to the public policies, and simulate and compare the effectiveness of each policy by conducting the epidemic spreading simulation on the simulated human mobility.

In the rest of this paper, we will introduce our human mobility based individual-level epidemic simulation platform focusing on three functions: 1) potential contacts identification, 2) probabilistic contagious disease transmission model, and 3) public policy simulation.
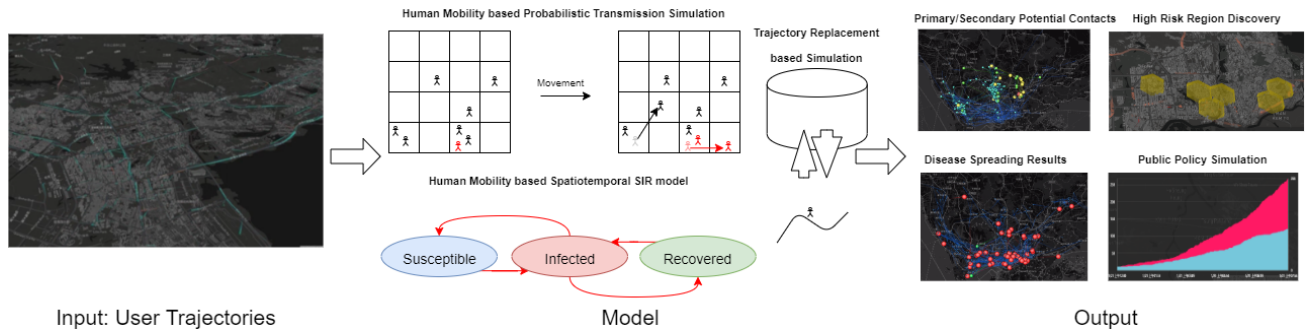
2   Platform Overview

Figure 1: Human mobility based individual-level epidemic simulation platform overview

An overview of our human mobility based individual-level epidemic simulation platform is shown in Figure 1. The input of our platform is user trajectories in a city, like TaxiSZ shown on the left of Figure 1, which is the taxi dataset used by [6], including 14728 raw taxi trajectories of Shenzhen. In this paper, we use the historical Shenzhen Taxi GPS trajectories data to represent the citywide human mobility, and the data can be replaced by other human mobility dataset with minor changes to our platform, especially for the data during the outbreak of the epidemic.

In the middle of Figure 1, we implement mining and simulation models to analyze the trajectory data to support different functionalities of this platform, shown on the right of Figure 1. The mining and simulation
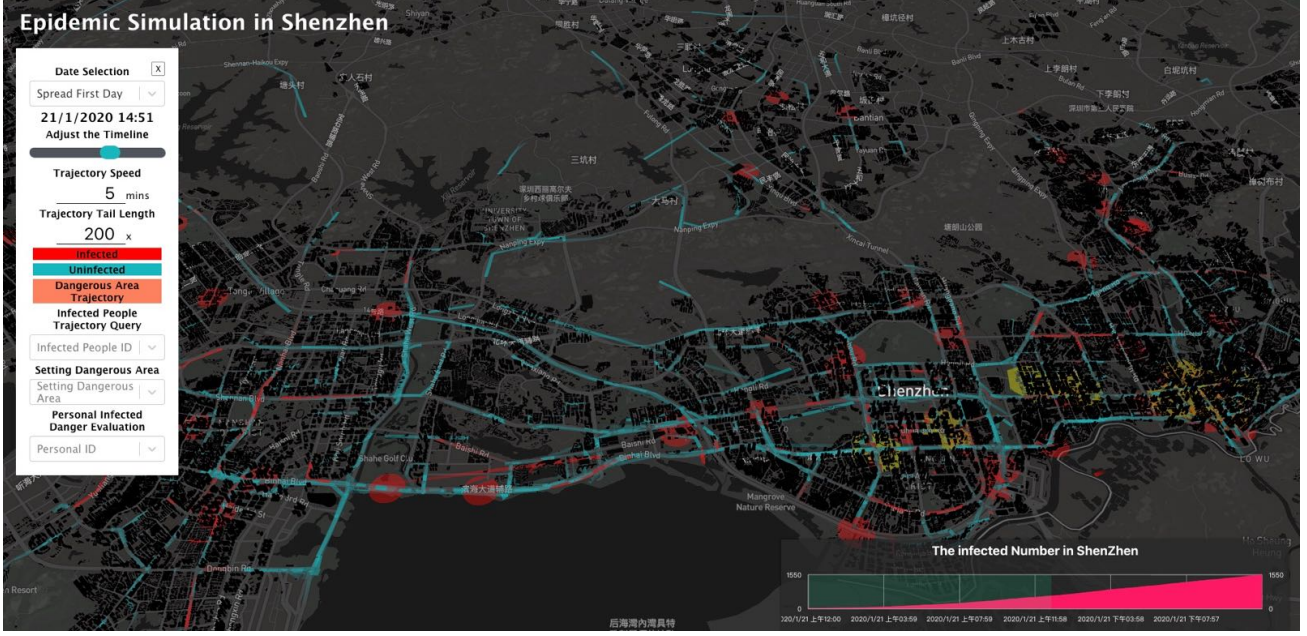
Figure 2: Human mobility based individual-level epidemic simulation platform user interface

models include a movement processing module that estimates the contact behaviors of the users in the dataset, which is the essential part of potential contacts identification, which we will show in details in Section 3, as well as human mobility based probabilistic transmission simulation. A spatiotemporal SIR model is proposed in this paper and utilized to simulate the probability of the infection state (susceptible, infected, recovered) of each individual in the dataset at each timestamp, which we will introduce in details in Section 4, and a trajectory replacement based simulation, which replace the current the trajectory with a historical trajectory from the same user/similar users that satisfies the current public policy.

As is shown in Figure 2, our platform user interface has three view: 1) control panels (left), where we can configure the visualization parameters, select the simulation date, or manually select high dangerous regions. 2) map view, where we show the GPS trajectories of the users, using different colors to distinguish the infected, recovered and susceptible, primary/secondary contacts users. Dangerous regions that are selected manually or estimated by the visit frequency of the infected users. For public policy simulation, we also allows the user's input in the map view such as selecting the lockdown region or set temperature check points; and 3) the trends of the simulated infected numbers, which is simulated by the probabilistic mobility based probabilistic transmission model. If any policy is simulated, the we will also show a comparative view in 3) showing the effectiveness of the current policy.

3 Potential Contacts Identification

An effective identification of contacts is helpful in narrowing the targets of screening or quarantine, especially a timely identification is critical to get the epidemic under control. Note that aside from the confirmed infected user should be treated separately, people who contacted the infected user in a near distance also should also be identified and screened. The identification process is quite costly especially taken the secondary contacts into consideration. As shown in Figure 3, as the time goes by, the potential contacts number grows explosively, one-day contacts (left) and two-day contacts (right), and the loss and risk of spreading will be minimized if we take early measures.
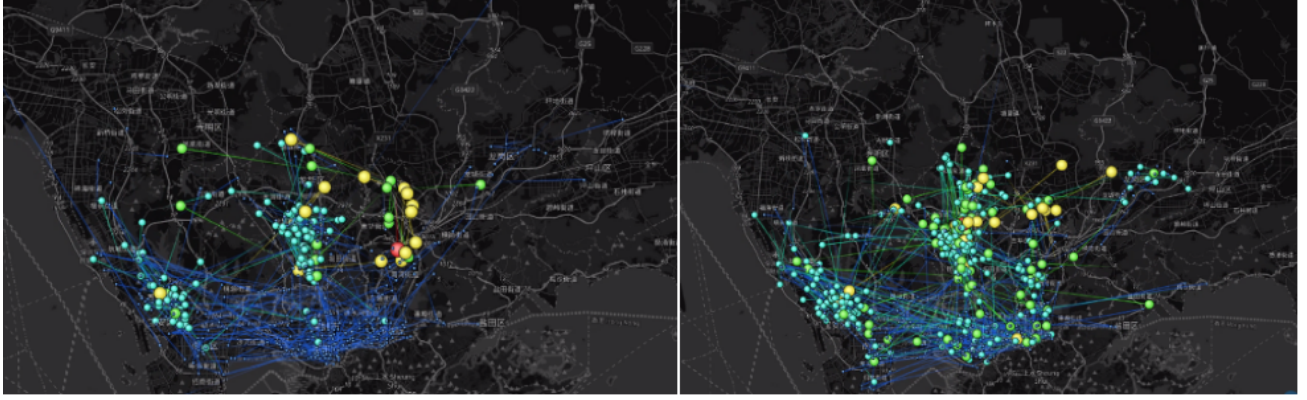
Figure 3: Potential contacts identification of an infected user (red) for a period of one-day (left) and two-day (right). Based on the GPS traces of the users, we can find out the potential contacts between the users. The contacts are classified into primary contacts (yellow) and secondary contacts (green and light blue), which means contacts that have direct contact with the infected user (red) and indirect contact (contact with primary/secondary contacts).

In this spirit, the first function of this platform is to identify those users that have direct or indirect contact with infected users. As shown in Figure 3, the big red marker shows an infected user (the user is for simulation only, it does not correspond to the infection of the real-world user), and other small dark blue balls represent the susceptible people. To distinguish the contacts in different risk level, we stratifies the contacts into three levels: level 1, shown by yellow markers, which are the primary contacts, level 1 shown in green markers, and level 3 shown in light blue markers. People in the risk of level 1 (primary contacts) mean that they have a direct contact with infected users, which have the highest risk of being infected. People with indirect contacts, like level 2 and level 3, are those who have contacted with people level 1 and level 2 respectively. They are also at risk to be infected, because people in level 1 or even level 2 may carry the virus. The risk of being infected is decreased with the level number increases, while the number of contacts will increase with the level number. Thus, to identify different level of contacts can be quite helpful in stratifying the strategy of screening and quarantine the potential contacts that can effectively prevent the disease from fast spreading.

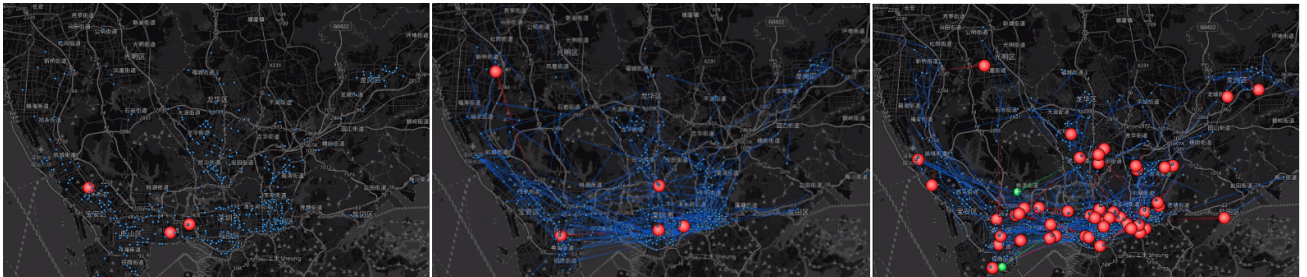## 4 Human Mobility based Probabilistic Transmission Simulation

Figure 4: Infection simulation using human mobility based probabilistic transmission model. The initial infected users are shown on the left (red markers), while the susceptible users are shown in dark blue. As the infected users move (middle), especially moving to the central area of the city, and we can see a simulation of how the contagious disease spreads in the city (right).

Note that not all the contacts will turn into infected users, for the purpose of having a view of future trends of the spreading of the epidemic, we develop a probabilistic transmission model that simulate the spreading process in the city at the individual level with uncertainty.

To model the contagious process by taking human mobility into consideration, we extend the conventional SIR model [1] by:

$$\frac{dS_{l,t}}{dt} = -\beta \frac{S_{l,t} I_{l,t}}{N_{l,t}}$$
$$\frac{dI_{l,t}}{dt} = \beta \frac{S_{l,t} I_{l,t}}{N_{l,t}} - \gamma I_{l,t} \tag{1}$$
$$\frac{dR_{l,t}}{dt} = \gamma I_{l,t}$$

where $S_{l,t}$, $I_{l,t}$ and $R_{l,t}$ are the number of susceptible, infected and recovered users at location l at time t. Location l is represented by the Uber's Hexagonal Hierarchical Spatial Index (H3) (with a hash level of 8, approximately a resolution of 460m). Time is sliced into slots with a constant sampling rate of 5 minutes. β and γ are the contact rates and recovery rates respectively.

We denote the list of user IDs at location l at time t by $U_{l,t}$, and the susceptible, infected and recovered subsets to be $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ respectively. By solving the differential equation 1, we get the simulated incremental number $\Delta S_{l,t}$, $\Delta I_{l,t}$ and $\Delta R_{l,t}$ at time t, and increase the lists of users $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ by sampling new infected users and recovered users with corresponding increments. Note that the increments cannot be ignored though it is always very small (the spatial and temporal resolution is high and the incremental number of a small area during a short time period is always very small). However, for simulating the infection process, the infection is discrete, with the minimal change of 1. Thus, to get the discrete values of the increments $\Delta \hat{S}_{l,t}$, $\Delta \hat{I}_{l,t}$ and $\Delta \hat{R}_{l,t}$, we quantify the incremental numbers by implementing a sampling strategy to use randomness to represent the small values. For example, if our simulated incremental number of infection is 0.001 at location l at time t, then we sample 1 new infected user from susceptible list $U_{l,t}^S$ with the probability of 0.001.

Thereafter, we update the $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ by 1) sampling the incremental IDs of $U_{l,t}^R$ from $U_{l,t}^I$ with a number of $\Delta R_{l,t}$; 2) sampling the incremental IDs of $U_{l,t}^I$ from $U_{l,t}^S$ with a number of $\Delta I_{l,t}$; and 3) updating $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ by the above two draws.

As the time proceeds, the movements of the users will make another update to the user lists $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$. For example, for one infected user u moves from l_0 to l_1 between time t and $t+1$, we update the user lists by 1) remove u from $U_{l_0,t}^I$, and 2) add u to $U_{l_1,t}^I$.

Combing these two above updates to $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$, we get the initial state for the next time step simulation $U_{l,t+1}^S$, $U_{l,t+1}^I$ and $U_{l,t+1}^R$. This process is summarized in the Algorithm 1 iterates over the target time range, and we collect the $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ at each time step as our simulation results and visualize these users trajectories with their infection state in out platform, as shown in Figure 4. The red markers show the infected users, dark blue markers show the susceptible users and green markers show the recovered users. The initial infected users shown on the left of Figure 4, and move to the central area of Shenzhen (in the middle of Figure 4, where the contagious disease can spread all city very fast on the right of Figure 4.

5   Public Policy Simulation

For public policies, we allow the users to input the public policies in multiple ways. For example, as shown in Figure 5 (a), the user can select a region of city by clicking on the map to select a polygon region that should be locked down. Then the users within the locked down region are not allowed to go outside the region and also the outside users are not allowed to come into this region. Another example is shown in Figure 5 (b) to set

Result: $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ for all location l and time t

Randomly or manually initialize $U_{l,0}^S$, $U_{l,0}^I$ and $U_{l,0}^R$;

for $t = 1, 2, 3, \ldots, T$ **do**

    for $l \in L$ **do**

        Get simulated incremental numbers $\Delta S_{l,t}$, $\Delta I_{l,t}$ and $\Delta R_{l,t}$ from Equation 1;

        Quantified by $\Delta \hat{I}_{l,t} = \begin{cases} \lfloor \Delta I_{l,t} \rfloor & \text{with probability of } 1 - (I_{l,t} - \lfloor \Delta I_{l,t} \rfloor) \\ \lfloor \Delta I_{l,t} \rfloor + 1 & \text{with probability of } I_{l,t} - \lfloor \Delta I_{l,t} \rfloor \end{cases}$;

        $\Delta \hat{R}_{l,t} = \begin{cases} \lfloor \Delta R_{l,t} \rfloor & \text{with probability of } 1 - (R_{l,t} - \lfloor \Delta R_{l,t} \rfloor) \\ \lfloor \Delta R_{l,t} \rfloor + 1 & \text{with probability of } R_{l,t} - \lfloor \Delta R_{l,t} \rfloor \end{cases}$;

        Sample the incremental IDs of $U_{l,t}^R$ from $U_{l,t}^I$ with a number of $\Delta R_{l,t}$;

        Sample the incremental IDs of $U_{l,t}^I$ from $U_{l,t}^S$ with a number of $\Delta I_{l,t}$;

        Update $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ by the above two draws;

    end

    Update $U_{l,t}^S$, $U_{l,t}^I$ and $U_{l,t}^R$ to $U_{l,t+1}^S$, $U_{l,t+1}^I$ and $U_{l,t+1}^R$ by user trajectories;

end

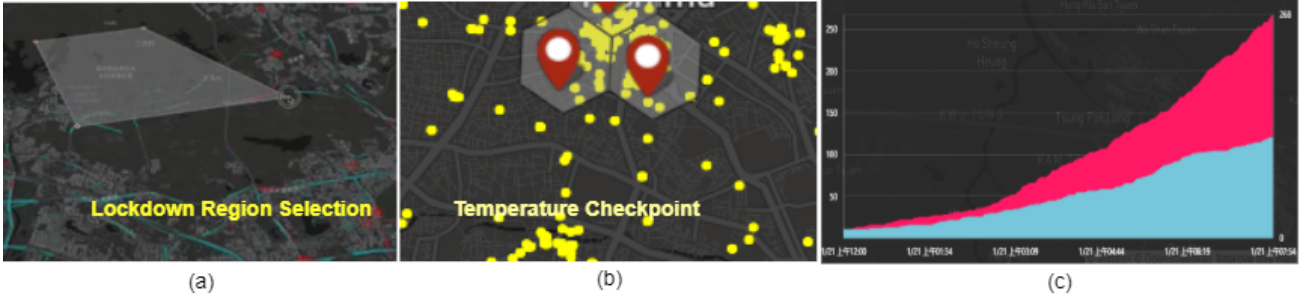**Algorithm 1:** Human Mobility based probabilistic transmission model

Figure 5: Public policy simulation. (a) lockdown region selection, (b) setting temperature checkpoints, (c) a comparative view of epidemic spreading trends for measuring the effectiveness of policy

temperature checkpoints, where the passengers need to be checked whether their body temperature is within the normal range. This is another effective measure to screen the infected users and give them an early quarantine.

The next question is how to simulate the human mobility in response to the public policy. Here we proposed a simple but effective approach named trajectory replacement based simulation. The basic idea of trajectory replacement based simulation is to replace the trajectory that does not satisfies the current policy with a history trajectory that satisfies the policy. For example, we replace the trajectory of going from home to work in the morning and come back home in the night with a trajectory that almost stay at home whole day long if the city is locked down.

Thus, we can simulate the epidemic spreading on the simulated human mobility and compared with previous simulation in Section 4. As is shown in Figure 5(c), we can see a difference between the trends of epidemic spreading with (shown in light blue) the policy of lockdown the selected region and without (shown in red) the policy. This provides a quantitative view of how effective a public policy will be.

6 Conclusion

In this paper, we develop a novel human mobility based individual-level epidemic simulation platform, which can effectively identify the potential contacts with different risks (primary/secondary contacts). This function can help the government take effective stratified measures to screening and quarantining the potential contacts at an early stage. This platform also implements a novel human mobility based probability transmission model that simulates the citywide contagious disease spreading process at the individual level. This function is helpful in having an illustrative view of the spatiotemporal trends of the spreading of the epidemic. Last but not the least, this platform can also simulate the human mobility as well as the epidemic spreading trends in response to different public policies. This provides a quantitative way of measuring the effectiveness of a policy before taking the policy into action.

References

[1] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[2] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet Public Health*, 2020.

[3] M. Tizzoni, P. Bajardi, A. Decuyper, G. K. K. King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza. On the use of human mobility proxies for modeling epidemics. *PLoS computational biology*, 10(7), 2014.

[4] Y. Wang, Y. Wang, Y. Chen, and Q. Qin. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (covid-19) implicate special control measures. *Journal of medical virology*, 92(6):568–576, 2020.

[5] J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.

[6] D. Zhang, J. Zhao, F. Zhang, and T. He. Urbancps: A cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, ICCPS '15, page 238–247, New York, NY, USA, 2015. Association for Computing Machinery.