

RESEARCH ARTICLE

A high-frequency mobility big-data reveals how COVID-19 spread across professions, locations and age groups

Chen Zhao^{1,2,3}, Jialu Zhang^{1,2,3}, Xiaoyue Hou^{1,2,3}, Chi Ho Yeung^{4*}, An Zeng^{5*}

1 College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang, P.R. China, **2** Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics and Data Security, Shijiazhuang, P.R. China, **3** Hebei Key Laboratory of Network and Information Security, Shijiazhuang, P.R. China, **4** Department of Science and Environmental Studies, The Education University of Hong Kong, Hong Kong, P.R. China, **5** School of Systems Science, Beijing Normal University, Beijing, P.R. China

* chyeung@eduhk.hk (CHY); anzeng@bnu.edu.cn (AZ)



OPEN ACCESS

Citation: Zhao C, Zhang J, Hou X, Yeung CH, Zeng A (2023) A high-frequency mobility big-data reveals how COVID-19 spread across professions, locations and age groups. *PLoS Comput Biol* 19(4): e1011083. <https://doi.org/10.1371/journal.pcbi.1011083>

Editor: Benjamin Althouse, University of Washington, UNITED STATES

Received: November 25, 2022

Accepted: April 6, 2023

Published: April 27, 2023

Copyright: © 2023 Zhao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw data cannot be shared by the authors due to the Mobile Privacy Policy of China, but might be available upon reasonable request submitted to the mobile network operator (China Mobile) at <http://t.10086.cn/services/#dsjArea>. The derived data supporting the findings of this study to reproduce all figures are stored and annotated on Github: <https://github.com/humlabgroup/spreadmodel1>.

Funding: C.Z. was supported by the National Natural Science Foundation of China

Abstract

As infected and vaccinated population increases, some countries decided not to impose non-pharmaceutical intervention measures anymore and to coexist with COVID-19. However, we do not have a comprehensive understanding of its consequence, especially for China where most population has not been infected and most Omicron transmissions are silent. This paper aims to reveal the complete silent transmission dynamics of COVID-19 by agent-based simulations overlaying a big data of more than 0.7 million real individual mobility tracks without any intervention measures throughout a week in a Chinese city, with an extent of completeness and realism not attained in existing studies. Together with the empirically inferred transmission rate of COVID-19, we find surprisingly that with only 70 citizens to be infected initially, 0.33 million becomes infected silently at last. We also reveal a characteristic daily periodic pattern of the transmission dynamics, with peaks in mornings and afternoons. In addition, by inferring individual professions, visited locations and age group, we found that retailing, catering and hotel staff are more likely to get infected than other professions, and elderly and retirees are more likely to get infected at home than outside home.

Author summary

Since 2019, we have witnessed the worldwide battle against COVID-19. As infected and vaccinated population increases, some countries chose not to impose any non-pharmaceutical measures and to live with COVID-19. In the literature, an extensive effort has been made to understand the effectiveness of intervention measures against COVID-19 and to predict its prevalence, but much less is known about the risk of choosing to coexist with COVID-19. Here, we conduct agent-based simulations overlaying a high-frequency mobility big data which records individual mobility tracks up to every second in a large-scale city in northern China, revealing the comprehensive transmission dynamics from a

(No.61703136) and by the Natural Science Foundation of Hebei (No. F2020205012), and by the Youth Top Talent Project of Hebei Education Department (NO. BJ2020035), and by The Project Supported by Science Foundation of Hebei Normal University (No. L2023K04). The work by C.H.Y. was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Projects No. EdUHK GRF 18301217, and No. GRF 18301119), the Dean's Research Fund of the Faculty of Liberal Arts and Social Sciences, The Education University of Hong Kong (Projects No. FLASS/DRF 04418, No. FLASS/ROP 04396, and No. FLASS/DRF 04624), and the Internal Research Grant, The Education University of Hong Kong (Project No. RG67 2018-2019R R4015 and No. RG31 2020-2021R R4152). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

few initial infected individuals to city-wise infections. Such comprehensive understanding will provide useful insights into our battle against COVID-19.

Introduction

The SARS-CoV-2, aka COVID-19, pandemic has hit the world since early 2020, and all countries were facing a great challenge in suppressing its transmission and saving lives. Several characteristics of COVID-19 has increased such challenges. First, early variants of COVID-19 have a long incubation period. Second, a substantial proportion of infected individuals may only have minimal symptoms. Before the rapid antigen tests were introduced, infections were mainly verified by PCR tests which are inconvenient and time-consuming. These characteristics lead to silent transmission, i.e. infected individuals are prone to transmit the virus to others without being aware of their own infection. In addition, COVID-19 generally has a high transmissivity, for instance, there are evidences of airborne transmission of Omicron [1, 2]. Finally, both hospitalization and fatality by COVID-19 increase with age, which create tremendous pressure on every country's healthcare systems [3].

To battle COVID-19, especially with its high silent transmissivity, other than pharmaceutical means such as vaccines and medicines, non-pharmaceutical measures play an important role [4–6]. The “lockdown” policy which restricts people to stay home is now a common terminology well understood by everyone. Other measures include case isolation and contact tracing, which aim to identify the infected and trace their close contacts and quarantine them [7]; social-distancing to encourage individuals to stay away from each other regardless of being infected or not; travel control to ban travel from local or international origins with infections; closing schools, catering and entertainment premises to avoid gathering [8–10], etc. Many of these measures aim to prevent silent transmissions without identifying the infected individuals. Finally, some countries adopt herd-immunity and do not impose strict intervention measures when a large portion of their population have been infected or vaccinated.

With this large variety of non-pharmaceutical intervention measures, a comparative analysis which reveals their relative effectiveness can contribute significantly to our battle against COVID-19. Nevertheless, such analysis is difficult, since one cannot fair test these measures in reality while potentially risking lives. Conventional approaches involve employing compartmental models, such as the susceptible-infected-recovered (SIR) model, to analyze these measures; this approach often captures the macroscopic trend, but not the details of the transmission dynamics, such as the location of individual infections, which are crucial for evaluating different interventions [11–15]. In comparison, agent-based models reveal both the trend and the detailed transmission dynamics [16], for instance, simulation studies with millions of agents were conducted to identify effective intervention measures for France [17] and Hong Kong [18] respectively. However, their simulation results can be model artefacts, since they depend crucially on agents' mobility patterns which are only generated by models as these studies lack the empirical mobility data [19–22].

As the infected and vaccinated population increases, some countries decided not to impose any non-pharmaceutical measures and to coexist with COVID-19. Despite there is an extensive effort made to understand the effectiveness of intervention measures against COVID-19 and to predict its prevalence, much fewer studies were devoted to reveal the consequence of choosing to live with COVID-19. In particular, some studies show that although vaccination reduces death, hospitalization and symptoms from COVID-19, its effect on reducing transmissions is minimal [23], which makes silent transmissions more prominent. This issue is

particularly important for China as most population has not been infected and most transmissions are silent due to the Omicron variant.

Thus, to reveal the consequence of coexisting with COVID-19, one relies on a full understanding of the dynamics underlying silent COVID-19 transmissions. Empirical mobility data are essential to reveal the mechanism underlying COVID-19 transmissions but are often difficult to obtain. Hence, some studies used aggregated data such as travel statistics instead of individual mobility patterns to study COVID-19 transmissions [5, 24, 25]. Other studies used segmented pieces of individual mobility patterns from a small subset of the population to construct mobility tracks for simulations, but results may be sensitive to the model underlying the construction [26–28]. Other type of data used to study COVID-19 transmissions include self-reported surveys from volunteers [29], cell phone calls [30] and contacts [31], contacts among cruise crews and passengers [32], etc. Nevertheless, a study with a big data of individual mobility tracks absent in existing studies, can reveal the transmission dynamics of COVID-19 to an extent of completeness and realism not attained so far, and lead to a complete understanding of the dynamics and thus useful insights in our battle against the silent transmission of COVID-19.

In this paper, we use a dataset of 4G communication records between mobile phones and base stations to identify the real mobility tracks of 0.7 million citizens in Shijiazhuang, a city in northern China, in a specific week in 2017, the biggest real mobility dataset employed for the study of COVID-19 to date. Since we aim to reveal the consequence of choosing to coexist with COVID-19 without intervention measures as well as prevention awareness from the public, a dataset before the pandemic is necessary and unique as most other studies used only data during the pandemic which are already influenced by intervention measures. There are over 11500 base stations throughout the city, and position of mobile phones and thus individual users was recorded as the location of the nearest base stations at a high frequency up to every second. We then conduct agent-based simulations using these real mobility data and the empirically inferred infection rate of COVID-19 to reveal the comprehensive transmission dynamics from a few initial infected individuals to a city-wise infection within one week. In addition, one can infer (1) the characteristics of individuals such as age and profession based on their mobility patterns, complying with the city's demographical information, as well as (2) the nature of locations they co-visit with others, complying with the city's geographical information. With these two types of information one can reveal how COVID-19 is transmitted within and across age groups, professions, as well as locations of different nature. This study thus reveals a comprehensive dynamics of COVID-19 transmissions with an extent of realism not achieved in existing studies, and provide useful insights into the choice of coexisting with COVID-19.

Results

Data, model and realistic simulations

Our big data of empirical mobility tracks is based on 7-day 4G communication records between base stations and mobile phones served by one of the three major service providers between 22nd and 28th May 2017 in Shijiazhuang, a city in northern China, and a dataset without the influence of the pandemic in terms of both central intervention and public awareness. There are $M = 11594$ base stations throughout the city, and the position of an individual is recorded as the location of the nearest base station as long as his/her mobile phones communicate with the base station in 4G. As some phone applications constantly exchange data with back-end servers, the position of individuals can be recorded up to a high frequency of every second.

The original data include records from roughly 3 million users out of a total population of 11 millions in the city. To obtain a dataset of valid mobility tracks from active users, we have implemented strict rules to exclude users who do not move at all and those whose data are largely incomplete. Finally, we single out $N = 702,477$ valid mobility tracks for simulations (see [Method](#) for details, and [S1 Fig](#) for the statistics of the dataset). As co-location visits by users is crucial for transmission of virus [33], we need to identify meaningful stops in a user's mobility track. We first divide the whole period of the dataset into time windows of 15 minutes as in [34], and consider that a user stops in a location if he/she stays there continuously or discontinuously for 10 minutes or more within the time window. Otherwise, if the user did not stay in the same location for more than 10 minutes, we identify the data point as "moving". A **co-location visit** is defined to occur when two individuals stop in the same location and the same time window. Regarding the potential bias of the data, first of all, the data we have is obtained from one of the largest companies among a few major cell phone service providers in China, and cover users from every district in Shijiazhuang city. Although users have different preferences in selecting cell phone companies, bias on the data due to their choice should be small, and hence the original sample of users in the data should be close to a random sample of the whole population in the city. Secondly, as we only take into account users who have continuous mobility records within the week, some users with incomplete mobility records are removed from our data analyses. This criterion may tend to remove senior or young users who do not constantly switch on and connect their phones to the 4G service. However, in order to reduce this bias, we have set the threshold of mobility below which to be defined as elderly, such that the distribution of the inferred age group of users is consistent with that in the census data of the city (see [Materials and methods](#)).

We follow the formula underlying the COVID-19 Essential Supplies Forecasting Tools (ESFT) introduced by the World Health Organization (WHO) [36], such that the *reproduction number* R_0 is given by

$$R_0 = D \times \gamma M \times \beta, \quad (1)$$

where D is the *infectious period*; M is the *average number of co-location visits* with the others per individual per unit time (e.g. day); γ is the *fraction of co-location visits which lead to contacts*, i.e. some co-location visits do not lead to contacts, for instance, users with their mobile phones connected to the same base station do not necessarily imply a contact, and γM is the number of **co-location contacts**; and β is the *probability of infection per co-location contact*. In our simulations, we adopted $R_0 = 7$ for the Omicron variant of COVID-19 [35], an infectious period $D = 7$ days [36] since we have assumed that the incubation period of Omicron lasts for 1.7 days [37] and our dataset spans only for 7 days, and finally an average of $M = 568$ co-location visits with the others by a single individual per day from our empirical data. In this case, one can estimate the quantity $\gamma\beta \approx 0.002$, which is the probability of infection per co-location contact with the others in a single time window.

By adopting the above model, we study the initial stage of the silent transmission of COVID-19 initiated from 0.01% (i.e. 70 people) of the population (see [S2 Fig](#) for the results with fewer initial spreaders). All citizens then move in the city according to their real mobility tracks. We assume that infections start with an incubation period of 1.7 days (see [S4 Fig](#) for the results with a longer incubation period), and afterwards the infected individuals would be able to infect others [37]. If a susceptible individual stays at the same location in the same time window (i.e. a co-location visit) with an infected individual passed his/her incubation period, the susceptible may be infected with a probability $\gamma\beta = 0.002$ (see [S3 Fig](#) for the results of smaller infection probability). We remark that the choice of $\gamma\beta$ would mainly influence transmission

speed instead of the detailed transmission dynamics, hence the main goal of the present study from the present study would be less dependent on $\gamma\beta$.

Since there may be more than one infected individual at a location, if we denote the number of infected individuals who have passed their incubation period at a location α in the time window at time t to be $n_\alpha(t)$, then the probability for a susceptible individual to get infected at α at time t is

$$P_\alpha(t) = 1 - (1 - \gamma\beta)^{n_\alpha(t)}. \quad (2)$$

Even if an individual stays in the same location for multiple consecutive time windows, the number of infected individuals may change as time evolves, and the probability for this individual to get infected in a specific time window is given by Eq (2) if he/she remains un-infected before the time window. Thus, the real mobility track of individuals determines whether they are in contact with the infected, who can be a stranger in malls, a colleague at the workplace, or a family member at home (see [Method](#) and later discussions for the inference of individual professions).

In our work, we analyze the 7-day mobility data of 700,000 cell phone users, finding that the number of people with co-location with a user is 568. Given the R_0 of Omicron is 7, we can obtain the infection probability β^* (i.e. $\gamma\beta$) of the virus (see [Eq \(1\)](#)). According to the census, the size of most family in Shijiazhuang is smaller than 7, so we assume that the number of family members for each family is at most 6. Assuming that the R_0 is also 7 at home, the infection probability β^* would be roughly 100 times larger than that in outdoor as the number of people in contact at home is roughly 100 times smaller than that in outdoor. In fact, considering that people usually wear masks outdoor but do not wear masks at home, R_0 at home may be even larger, and thus in one of our major scenarios of simulations, we made this assumption that the infection probability at home is 100 times larger than outdoors, leading to $\gamma\beta = 0.2$ at home. Nevertheless, we also understand that this is only an assumption, and therefore we also examine in [S5](#) and [S6](#) Figs the case with other $\gamma\beta$ at home, and the results show that our main findings are still valid. As our data span only for one week, which is shorter than the time needed for recovery from COVID-19, we do not include recovery. We also assume that most infected individuals do not show significant symptoms in a week and the virus spreads silently. This assumption may be more valid for Omicron, since it may lead to milder or even no symptoms compared to other COVID-19 variants [38]. Hence, the mobility behavior of citizens remains un-intervened as they are unaware of the silent transmissions. This makes our dataset suitable since it recorded the mobility behavior in 2017 without the influence of the pandemic, and thus our study is unique since most other studies used only macroscopic datasets during the pandemic.

We first illustrate how our big data of high frequency mobility tracks can lead to simulations of COVID-19 transmission with an extent of completeness and realism not attained in existing studies. As an example, we show in [Fig 1](#) two individuals who live far away from each other but co-visit some locations at the same time in a day such that silent transmissions may occur between them. Their spatial mobility tracks with various visited locations are shown in [Fig 1A](#), where both follow a routine schedule like most of us as we can see from their sequence of visited locations in [Fig 1B](#). In this specific day, these two individuals left home in the morning, and then went to work at the same location, so they had two periods of co-location visits i.e. one in the morning and the other in the afternoon at the work place. After work, they went to the same venue for a concert, which led to their third co-location visit in that day. According to our adopted spreading model, if one of them has been infected with COVID-19 before the

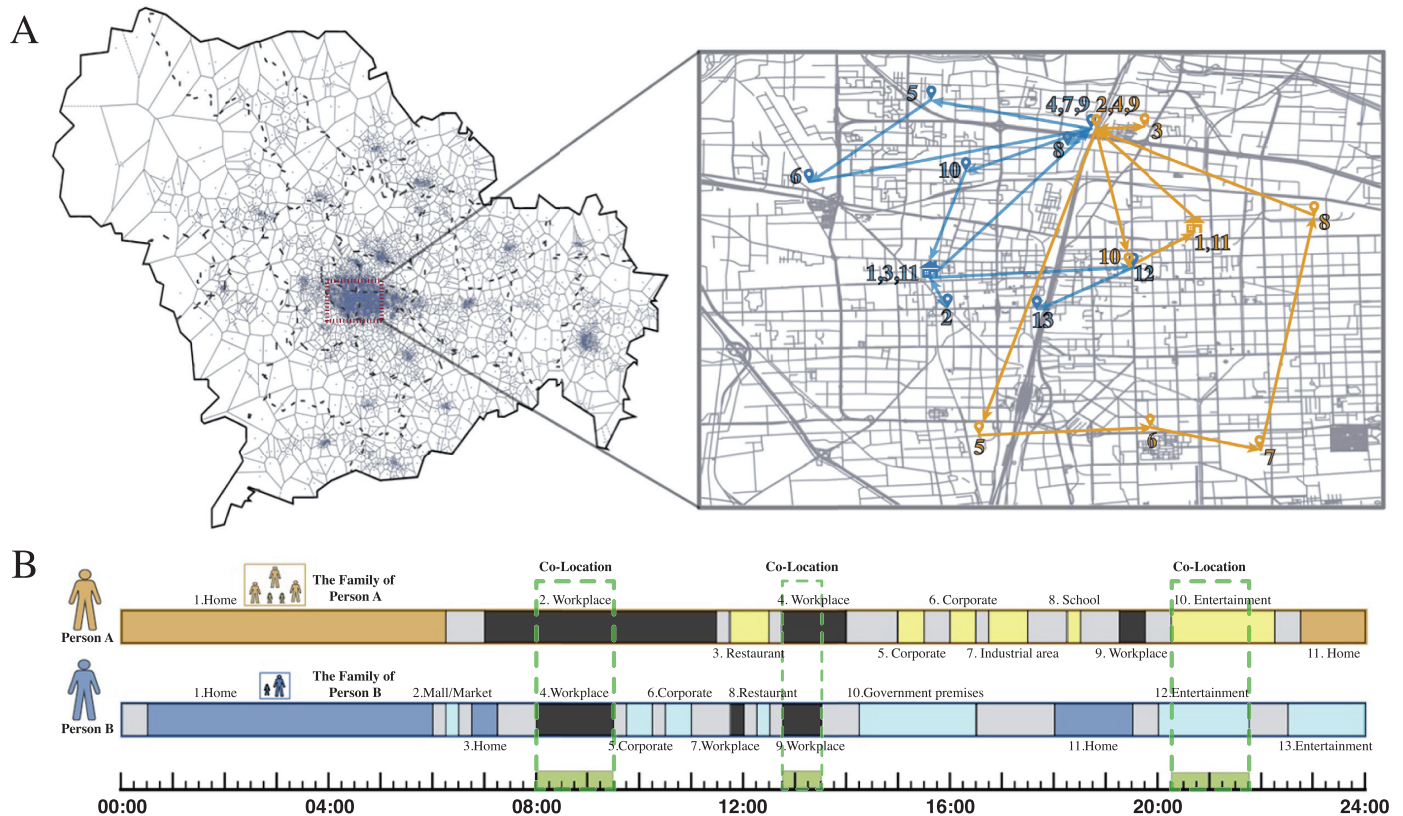


Fig 1. Exemplar high-resolution city-wise human mobility tracks and potential silent transmission of COVID-19 via co-location visits. (A) The map and the distribution of population in Shijiazhuang, and the city center is enlarged and the exemplar mobility tracks from two individuals are shown on the right. (B) Their corresponding sequence of visited locations, numbered according to the location labels in the enlarged map, with the category of each location shown. Blocks in gray correspond to the period when the individuals are “moving” Their co-location visits, i.e. they stop at the same location in the same time window, are marked by dashed squares. The icons used in this figure were obtained or modified from open-source resources in Openclipart (<https://openclipart.org/>). The map was drawn based on open-source shape file with License Creative Commons BY 4.0 (CC BY 4.0) from OpenStreetMap (<https://www.openstreetmap.org/>).

<https://doi.org/10.1371/journal.pcbi.1011083.g001>

day, the other individual would have a probability of $\gamma\beta$ to get infected in each co-location visit in each time window.

Daily periodic transmission dynamics

To reveal details of the transmission dynamics difficult to observe without our high-frequency empirical mobility tracks, we show in Fig 2A the number of new infections $\Delta I(t)$ in the city as a function of time t at a 15-minute interval. According to our analyses, most mobility tracks follow a regular pattern, which are mostly found outside home at different locations in the daytime and stationary at home in the night-time. In general, individuals are more likely to have co-location visits with others and get infected when they visit locations outside home, and less likely to get infected at home, despite the infection rate at home is higher. It is because they came into contact with many others outside home, but only a few family members at home. Therefore, one can observe a significant periodic pattern in $\Delta I(t)$ in Fig 2A, of which peaks and troughs correspond to daytime and night-time respectively. Interestingly, small troughs are found in between mornings and afternoons, which may correspond to individuals going home for lunch or rest at lunch time. All these results suggest that staying home does help suppressing the silent transmissions of COVID-19, even for a short time like lunch time.

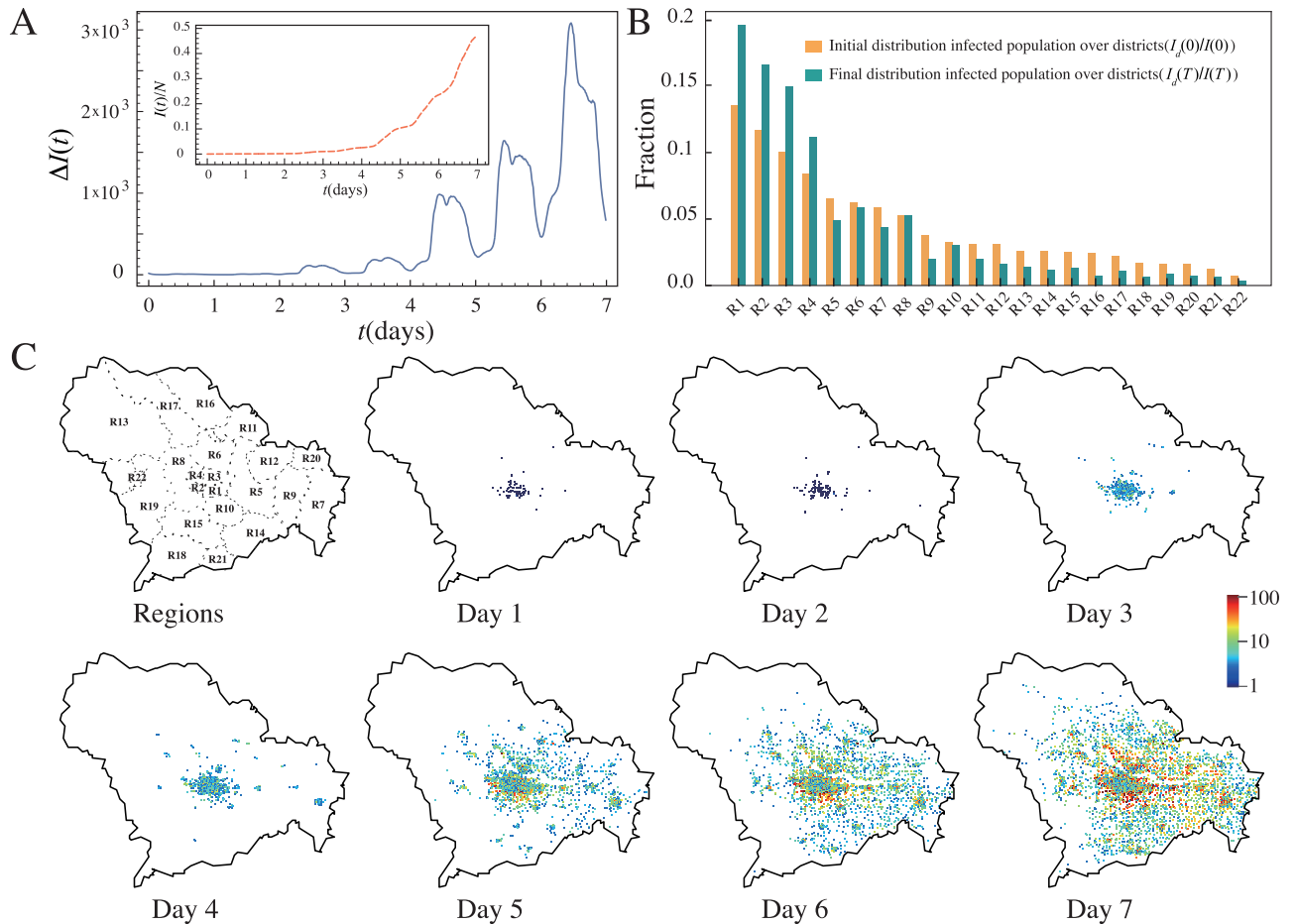


Fig 2. The spatiotemporal patterns of city-wise COVID-19 infection. (A) The number of new infections $\Delta I(t)$ in the city as a function of time t at a 15-minute interval, given 70 randomly selected initial spreaders, averaged over 1000 realizations. A significant periodic pattern is observed, which is caused by the periodic human mobility behavior. Inset: The corresponding fraction of infected population, i.e. $I(t)/N$. (B) The distribution of the initial and the final infected population over the districts, i.e. $I_d(0)/I(0)$ and $I_d(T)/I(T)$ respectively; districts are numbered as shown in the map in C. (C) The evolution of the daily spatial pattern of the infected population in the city; the number of infected population in a location is represented by the color of the dot. The map was drawn based on open-source shape file with License Creative Commons BY 4.0 (CC BY 4.0) from OpenStreetMap (<https://www.openstreetmap.org/>).

<https://doi.org/10.1371/journal.pcbi.1011083.g002>

In the inset of Fig 2A, we show the fraction of infected population as a function of time t , i.e. $I(t)/N$, where $I(t) = \sum_{t'=0}^t \Delta I(t')$ is the total number of infections at time t . Although there are only 70 initial spreaders, a city-wise infection can occur in a week. These transmissions may seem much faster than those in reality; this is because in reality COVID-19 transmissions always come with a high anti-pandemic awareness, preventive and intervention measures, while in our simulations we assume silent transmissions without any interventions. Hence, our results can also serve as a benchmark to demonstrate how COVID-19 spreads given no anti-pandemic awareness nor interventions. However, the fast transmissions observed in our simulations are also dependent on our inferred infection probability $\gamma\beta$, which may be different from that in reality. Nevertheless, our following major results which compare infections across age groups and professions, locations and districts, etc. are relative to each other and less dependent on the pace of transmissions and thus our estimate of $\gamma\beta$.

We then go on to examine how transmissions occur in different districts of the city. Such analysis would be difficult even in reality by contact tracing since it is often hard to determine

how an individual gets infected [7], but would be straightforward in our case by agent-based simulations overlaying a big data of empirical mobility tracks. We first denote $I_d(t)$ as the number of infected individuals at time t who live in district d , with d denoting one of the 22 districts in the city (see [S1 Table](#) for the information of individual districts). We show in [Fig 2B](#) the distribution of the initial and the final infected population over the districts, i.e. $I_d(0)/I(0)$ and $I_d(T)/I(T)$ respectively with T denoting the ending time of the dataset. As the initial infected individuals are randomly selected, their presence should be proportional to the population of a district. As we can see, the final fraction of infected population in a district is not proportional to its initial fraction, i.e. $I_d(T)/I(T) \not\propto I_d(0)/I(0)$, since the infected individuals move across different districts in the city. Interestingly, districts with a large share of initial infected population tend to have an even larger share at last. These districts are mainly regions in the city center with a high population density, and they are also business centers where citizens from other districts come to work or gather in the daytime, causing cross infections. Interestingly, the districts with a small share of initial infected individuals, on the contrary, tend to have an even smaller share of the final infected population. It is because these districts are mostly suburban or rural areas, and their residents are less likely to visit city centers and residents in other areas are also less likely to visit them, which reduces transmissions. In addition, the small population density in these areas may also reduce COVID-19 transmissions since residents have less frequent contacts with others. We show in [Fig 2C](#) the evolution of the spatial pattern of the infected population in the city, which further supports our conjecture that the infections are highly clustered in the city center or regions with a high population density. The results suggest that visiting city centers and meeting strangers from other areas of the city may lead to a high risk of infection.

Profession- and location-dependent transmissions

Based on the nature of the location of base stations, one can identify the nature of the locations visited by users in their mobility tracks. According to their nature, we classify locations into 15 *location categories* such as malls, schools, etc. With this information, we can infer the profession of individual users (see [Method](#) for details) and study the relationship between professions and the transmission of COVID-19.

In [Fig 3A](#), we show how the number of newly infected individuals with different inferred professions increases with time t , i.e. $\Delta I_p(t)$ with the subscript denoting inferred profession p . An immediate observation is that the periodic pattern still exists for most professions, but the number of infected individuals is largely different across professions as they have a different population size. The top two infected professions are industrial and corporate workers and freelancers, which are also the two professions with the largest population size. To better compare the infection dynamics across professions, we show in the inset of [Fig 3A](#) the fraction $\Delta I_p(t)/I_p(T)$, i.e. the fraction of infected population of profession p who get infected at time t ; this allows us to compare the periodic pattern of new infections across professions with vastly different population size. One can see that most professions follow the same periodic infection patterns, except retirees who are more likely to be infected in the night-time, possibly caused by their family members who come back home after work.

To compare the likelihood of infection by profession, we show in the *upper panel* of [Fig 3C](#) the distribution of population over professions (orange bars), i.e. N_p/N , which is proportional to the initial distribution of the infected population over professions $I_p(0)/I(0)$ since the initial infected group is randomly selected; we also show the corresponding final infected distribution (green bars), i.e. $I_p(T)/I(T)$. In the *lower panel*, we show the final fraction of infected individuals normalized by the population size in each profession, i.e. $I_p(T)/N_p$. As we can see, despite

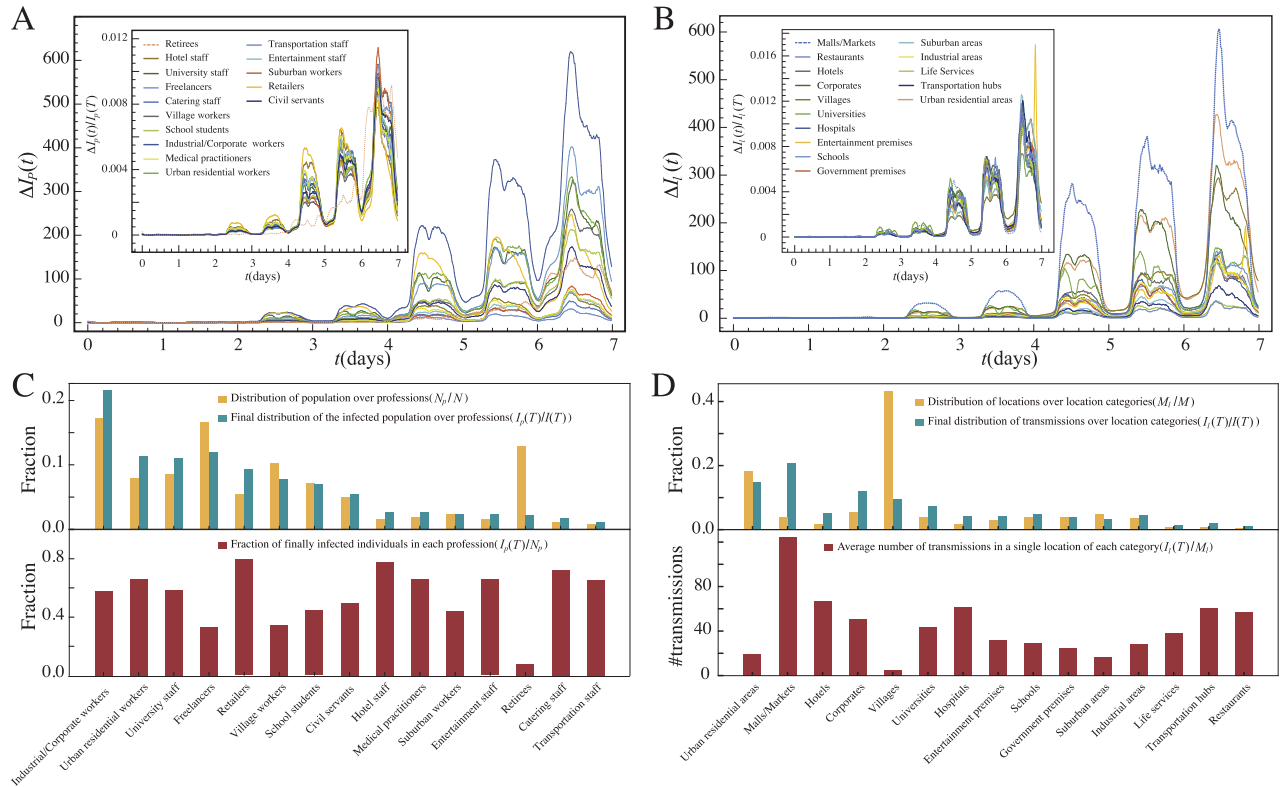


Fig 3. Infected professions and locations. (A) The number of newly infected individuals of different professions p as a function of time t , i.e. $\Delta I_p(t)$. Inset: the fraction of infected population from profession p who get infected at time t , i.e. $\Delta I_p(t)/I_p(T)$, averaged over 1000 realizations. Periodic patterns still exist for different professions, but their infected population is largely different. (B) The number of transmissions in locations of different location category l as a function of time t , i.e. $\Delta I_l(t)$. Inset: the fraction of transmissions in location category l which occur at time t , i.e. $\Delta I_l(t)/I_l(T)$. (C) Upper panel: the distribution of population over professions (orange bars), i.e. N_p/N , which is proportional to the initial distribution of the infected population over professions $I_p(0)/I(0)$, and the corresponding final infected distribution (green bars), i.e. $I_p(T)/I(T)$; lower panel: the fraction of final infected individuals normalized by the population size in each profession, i.e. $I_p(T)/N_p$. (D) Upper panel: the distribution of locations over location categories, i.e. M_l/M and the final distribution of transmissions over location categories, i.e. $I_l(T)/I(T)$; lower panel: the average number of transmissions in a single location of each location category, i.e. $I_l(T)/M_l$.

<https://doi.org/10.1371/journal.pcbi.1011083.g003>

that retirees take up a large fraction of the population as shown by the orange bars, their final share of infection is small, implying a low infection rate for them as also shown by the red bars in the lower panel. This is an interesting characteristic of the pandemic not revealed in previous studies. Nevertheless, here we assume that all the elderly are living in households with at most 6 family members, and thus transmissions in elderly homes are not considered which can greatly increase the infection rate for the elderly. Unlike retirees who are characterized with the smallest infection rate, professions with the highest infection rate include retailing, catering and hotel management. They are all service professions who are in contact with many strangers every day, and some of these strangers may come from other districts leading to long-distance silent transmissions. If these service providers are infected, they may also act like an infection hub to distribute the virus across different areas of the city, again through the strangers they.

To further understand the reason behind profession-dependent infections, we show in Fig 3B the number of transmissions in locations of different category as a function of time t , i.e. $\Delta I_l(t)$ with l denoting the category of location. This is a quantity which can be measured easily in simulations, but not empirically since transmission are difficult to be identified in reality.

As we can see, on top of the periodic patterns, the largest number of transmissions occurs in malls and markets, followed by urban residential areas and corporates. The inset of Fig 3B shows the fraction of transmissions at locations in category l which occur at time t , i.e. $\Delta I_l(t)/I_l(T)$; this again allows us to compare the periodic patterns of new transmissions across location categories with vastly different number of transmissions. One can now see an obvious peak for “entertainment premises” in the last day of the simulations, which came from a massive gathering in an evening concert (see Fig 1 for the mobility tracks of two individuals who attended the concert). This shows that large gathering events do pose a high risk of large-scale transmissions.

In Fig 3D, we show in the *upper panel* the distribution of locations over different categories, i.e. M_l/M , where M_l corresponds to the number locations which belong to category l ; the final distribution of transmissions over location categories, i.e. $I_l(T)/I(T)$, is also shown. In the *lower panel*, we show the average number of transmissions in a single location of each location category, i.e. $I_l(T)/M_l$. The results suggest that the risk to be infected is highest in malls and markets, hotels and transportation hubs, but lowest in villages and urban residential areas, in consistent with our above findings on profession-dependent transmissions as well as our expectation that locations where people gather are likely for transmission of COVID-19. In this case, stricter intervention measures can be imposed in locations of the high-risk location categories.

Other than profession-related locations, everyone goes home and stay probably a long time there in a day, it is therefore important to examine how often transmissions occur at home. We remark that we do not associate home to any of the location categories, since we define home as the location an individual stayed the longest time over-night instead of based on the nature of the location (see Method for details). In Fig 4A, we compare how the number of transmissions inside and outside home increases with time t . As we can see, transmissions occur more frequently outside than at home. The inset of Fig 4A shows the fraction of transmissions inside and outside home which occur at time t , i.e. $\Delta I(t|\text{home})/I(T|\text{home})$ and $\Delta I(t|\text{outside})/I(T|\text{outside})$ respectively. This again enables us to compare the periodic patterns in spite of the difference in their total number of transmissions. As we can see, both fractions show a periodic pattern but are roughly out of phase, i.e. the peaks of transmissions at home are found roughly in the night-time which are also the minima for transmissions outside.

The dependence of infection on age groups

Based on the locations individuals visited, their inferred professions and the demographic data from the 7th official Census of Shijiazhuang, we can further infer and classify individual users into three age groups, namely (1) under 15, (2) 15 to 60, and (3) above 60 (see Method for details). We then go on to investigate how transmissions occur within and across age groups. In particular, both hospitalization and fatality by COVID-19 increase with age, leading to immense pressure on the public healthcare systems, and it is worthwhile to examine how transmissions to the senior population can be suppressed.

In Fig 4B, we show how the number of newly infected individuals increases with time t in different age groups, i.e. $\Delta I_a(t)$ with a denoting the age group. As the largest fraction of the population falls in the second group with age between 15 and 60, it has the largest share of infected population. We further show in the inset of Fig 4B the fraction of new infections in different age groups which occur at time t , i.e. $\Delta I_a(t)/I_a(T)$. As we can see, senior population with age above 60 evolves differently from those in the other two age groups. Specifically, the senior age group tends to have a higher probability to be infected during the weekend, i.e. the 6th and the 7th day in our dataset, possibly because their younger infected family members stay

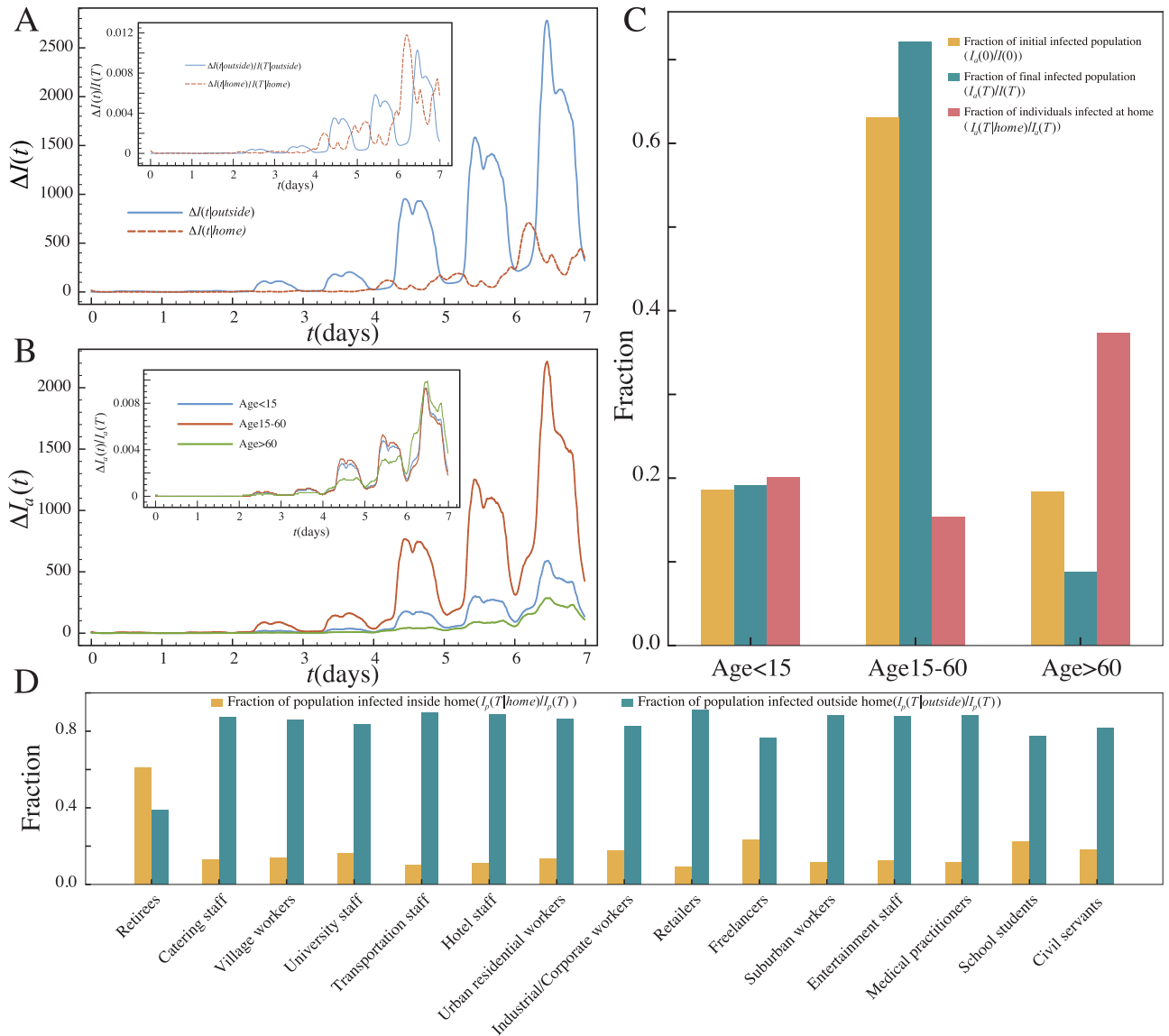


Fig 4. Transmissions inside and outside home and their dependence on age. (A) The number of new transmissions inside and outside home as a function of time t , averaged over 1000 realizations. Inset: the fraction of transmissions inside and outside home normalized by the total number of transmissions as a function of time t , i.e. $\Delta I(t|_{home})/I(T|_{home})$ and $\Delta I(t|_{outside})/I(T|_{outside})$ respectively. (B) The number of new infected individuals as a function of time t in different age groups, i.e. $\Delta I_a(t)$. Inset: the fraction of infections in different age groups which occur at time t , i.e. $\Delta I_a(t)/I_a(T)$. (C) The distribution of initial and final infected population across different age groups, i.e. $I_a(0)/I(0)$ and $I_a(T)/I(T)$ as orange and green bars respectively, and the fraction of individuals infected at home in different age groups, i.e. $I_a(T|_{home})/I_a(T)$ (red bars). (D) The fraction of population infected inside and outside their home according to their professions, i.e. $I_p(T|_{home})/I_p(T)$ and $I_p(T|_{outside})/I_p(T)$ respectively.

<https://doi.org/10.1371/journal.pcbi.1011083.g004>

at home for a long time or they go out with their family members during the weekend. Fig 4C further shows the distribution of initial and final infected population across different age groups, i.e. $I_a(0)/I(0)$ and $I_a(T)/I(T)$ respectively, which suggests that in general the senior population is less likely to be infected due to their smaller mobility. The red bars of Fig 4C also show the fraction of individuals infected at home in different age groups, i.e. $I_a(T|_{home})/I_a(T)$. One can see that the senior population are much more likely to be infected at home than younger people.

The above results suggest some specific ways for the senior population to get infected, and are further supported by our profession-dependent analyses. We show in Fig 4D the fraction of population infected inside and outside their home according to their professions, i.e. $I_p(T| \text{home})/I_p(T)$ and $I_p(T|\text{outside})/I_p(T)$ respectively. Since retirees are mainly composed of the senior population, they are the only group who are more likely to be infected at home than outside. As COVID-19 impacts the senior population more severely, to reduce death toll, our results suggest that it is important to avoid the senior population to get infected at home, for instance, to impose prevention measures at home such as wearing masks or to reduce the frequency of high-risk family members visiting or staying with them during the pandemic [8].

Profession-specific and location-specific source of infection

Finally we study how the final state of infection in the city depends on the initial group of infected [39]. Intuitively, the final state depends significantly on the professions and locations of the initial infected group. The radar maps in Fig 5A and 5B show the total number of the infected population over professions until the 3rd and the 7th day of the simulations respectively, i.e. $I_p(3^{\text{rd}} \text{ day})$ and $I_p(7^{\text{th}} \text{ day})$, given that each of the initial infected group falls

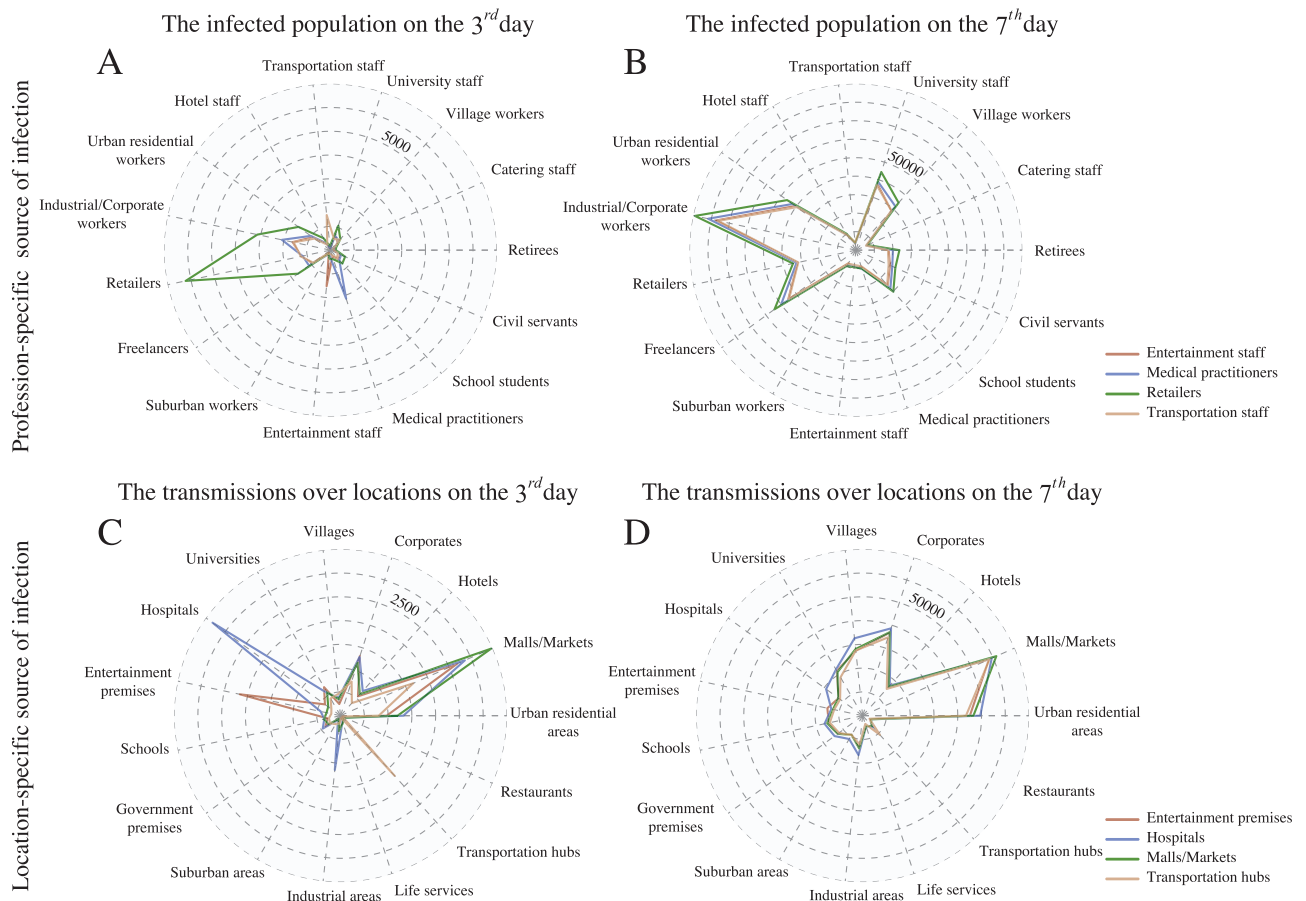


Fig 5. Effect of initial spreaders. The number of the infected population over professions until (A) the 3rd and (B) the 7th day of the simulations respectively, i.e. $I_p(3^{\text{rd}} \text{ day})$ and $I_p(7^{\text{th}} \text{ day})$, given that each of the initial infected group falls completely in four different professions. Similar investigation on (C) $I_l(3^{\text{rd}} \text{ day})$ and (D) $I_l(7^{\text{th}} \text{ day})$ for location categories, given that the initial infection starts at locations in four different location categories.

<https://doi.org/10.1371/journal.pcbi.1011083.g005>

completely in four different professions (see [S7 Fig](#) for the results with the initial infected group in other professions). These results suggest that the distribution of infected professions on the 3rd day can be substantially different, depending on the initially infected profession. For instance, retailing is the most infected profession given that retailers are initially infected, but less affected if the initial infected group comes from other professions. On the other hand, one can see from [Fig 5B](#) that the final distributions of the infected population initiated by different infected groups look similar, suggesting that the distributions of the infected individuals over professions become more independent of the initial infected groups as time evolves. These results imply that the professions of the initial infected population may have a short-term impact on the distribution of the infected population, but this impact reduces as the pandemic evolves, and ultimately the final state of infection may become independent of the sources.

Similar results can be seen in [Fig 5C and 5D](#) when we investigate the total number of transmissions over location categories until the 3rd and the 7th day of the simulations respectively, i.e. $I_i(3^{rd} \text{ day})$ and $I_i(7^{th} \text{ day})$, given that initial infection starts at locations in four different location categories (see [S8 Fig](#) for the results where initial infection starts at other location categories). The results suggest that in the first few days, tracing the sources of infection is important as it affects the professions and locations which are shortly infected. However, in the later stage, source tracing is no longer important as the final infection state becomes independent of the sources.

Finally, we show in the [S9 Fig](#) the simulation results where transmissions through both contact and environment are considered. In this case, we first randomly select 70 individuals as the initial infected group and they spread COVID-19 via co-location contacts as in our previous simulations. We then consider four location categories for environmental transmissions, i.e. an individual who visits locations in these four categories would have a probability 0.002 to be infected on top of the probability via co-location contacts with infected individuals. In other words, visits to these locations may result in infections even if no infected individual is present in these locations in the same time window. As we can see from [S9 Fig](#), the results of transmissions are qualitatively similar to those in [Fig 5C and 5D](#), where one can see a high heterogeneity of infections at the early stage when the environmental spreading occurs at different location categories, but then the states of infection become similar at the later stage.

Discussion

Since more countries decided to coexist with COVID-19, it becomes essential to understand the consequence of such coexistence strategy. From the perspective of China where most population has not been infected and most infections are silent due to the Omicron variant, a comprehensive understanding of the risk of coexisting with COVID-19 is particularly important. In particular, vaccination rate increases in every country but it may make silent transmissions more prominent [23]. Our study served as a first example to simulate the silent transmission of COVID-19 across a community using a big data of more than 0.7 million empirical human mobility tracks without any intervention measures for a week. There were previous studies on COVID-19 transmission using similar agent-based simulations, but they lack the empirical mobility data and have to either generate the mobility patterns of individual agents by models [17, 18], or construct virtual mobility tracks using segmented pieces of real data [27]. Details including the daily routine of agents, how they co-visit various locations with other agents, how much time they spend in each location, how their mobility patterns depend on their age and professions, etc., are all crucial factors affecting the transmission of COVID-19, but are only modeled or constructed in previous studies, which may lead to a large discrepancy from real mobility patterns and thus a large discrepancy in the results and insights generated. Our

study thus represents a call for the use of empirical big data for revealing the realistic transmission dynamics of COVID-19.

Among the existing works about COVID-19, the majority of them aim to predict the number of infected cases and to understand the influence of the pandemic on human mobility patterns. In this paper, we mainly study the spreading of COVID-19 within a city when all mobility restrictions are cancelled. The mobility data between 2020 and 2022 is thus not appropriate for this purpose, yet the mobility data before the pandemic characterizes better the travel patterns of citizens after their lives return to normal. In order to support the above claim that the mobility patterns before the pandemic is closer to the pattern after pandemic restrictive measures are lifted, we compute the similarity of inter-city mobility patterns between different years in 2019–2022 with that of 2023. We find that compared to the years during the pandemic, the time before the pandemic is more similar in human mobility to the time after opening up (see [S10 Fig](#)). Meanwhile, we remark that there are some limitations for using the data before the pandemic. First of all, some travel habit may be formed during the pandemic and be kept even after opening up. This may cause some difference between the mobility patterns before and after the pandemic. Secondly, the income of some citizens is reduced during the pandemic, which may change their travel patterns such as less leisure travels and more traveling due to works. Finally, the pandemic may cause some long-lasting sickness for some senior people, which largely reduces their travel frequency and range.

With the big data of empirical mobility tracks, we can reveal details for silent transmissions of COVID-19 not observed in previous studies. For instance, our high frequency data record the location of individual users up to every second, such that we can show the number of new infections at a 15-minute interval; this leads us to a daily periodic transmission pattern, which peaks in the mornings and the afternoons as expected since individuals are most active at these times, but an interesting minima in between when some of them go home for lunch or rest. In addition, we also observe less transmissions overnight, suggesting staying home does mitigate COVID-19 transmissions. If individual mobility patterns are generated by models instead of obtained from empirical data, the results would be largely sensitive to model formulation and we are not sure whether such interesting results are simply model artefacts or real phenomenon. This again suggests that agent-based simulation using empirical big data can reveal realistic transmission dynamics and thus useful insights that help in our battle against COVID-19.

Indeed, the largest advantage brought by our big data of mobility tracks is not limited to the high frequency macroscopic trends, but the very details such as how transmissions depend on age, profession, and location. They are obviously crucial factors influencing transmissions, but to reveal their relationship with transmissions one need to model how mobility depends on them which are difficult and are yet to be examined in previous studies. In our case, with the empirical data of mobility tracks, we can reveal the dependence of transmission mechanism on age, profession and location relatively easily by inferring the characteristics of individuals and locations. This again gives us useful insights such as the professions and the nature of locations at which transmissions are more likely to occur, or the common ways senior citizens get infected, which are hidden mechanisms difficult to be revealed even in reality. In particular, we found that retailing, catering and hotel management are the professions prone to COVID-19 infection, while retirees and elderly are less likely to be infected due to their limited mobility. As for the nature of locations, the largest number of transmissions occurs in malls and markets, followed by urban residential areas and corporates. Moreover, staying home does help suppressing COVID-19, even for a short time and with a larger infection rate at home.

A follow-up study with a great potential to fully utilize our big data of mobility tracks is to reveal the effectiveness of different non-pharmaceutical measures in suppressing COVID-19. Early studies which investigate aggregated data such as travel statistics have already shown

correlations between the implementation of non-pharmaceutical measures, such as travel control, with the reported number of COVID-19 infections, but comprehensive individual mobility patterns are not studied [5, 24]. Large-scale agent-based simulations are also used for this purpose, but as we have mentioned before, agents' mobility tracks in these studies are generated by models [17, 18]. We expect that with the big data of mobility tracks, one can investigate intervention measures such as lockdown, isolation, contact tracing, quarantine, travel control, etc. Nevertheless, there are also shortcomings of studying intervention measures with real mobility tracks as they are real data at a time without the pandemic and models are required if we would like to simulate individual mobility subject to these measures. However, the extent of modeling in this case would be less than those without empirical data.

In this paper, we discuss the spreading of COVID-19 within a city when all mobility restrictions are cancelled. The results suggest that even starting with a few initial spreaders, within one week, almost half of the population in a city might get infected. In reality, there are indeed a large number of citizens in China got infected in a short time after opening up. Such fast infection is predicted in our work based on microscopic human mobility behavior. In addition, our works predict that such fast outbreak of virus may cause the shortage of medical materials such as hospital beds and antipyretic. From early December 2022 to early January 2023, people in many cities of China indeed experienced a roughly one-month difficulty in buying medicine for fever.

COVID-19 has impacted every aspect of our daily life since early 2020 [40], and its characteristics make it prone to silent transmissions and thus difficult to be identified in the community before some infected individuals show up with obvious symptoms. There is no easy way to reveal how the virus is transmitted throughout the community, and we believe that our approach of large-scale agent-based simulations overlaying a big data of individual mobility tracks is a promising one with a sufficient extent of completeness and realism not attained in existing studies. We hope that the insights generated in our study would contribute to our battle against COVID-19.

Methods

Dataset of mobility tracks

Our original data is composed of 7-day 4G communication records between base stations and mobile phones served by one of the three major service providers between 22nd and 28th May 2017, from close to 3 million users out of a total population of 11 millions in Shijiazhuang, a city in northern China. There are over 11500 base stations throughout the city, and the position of an individual is recorded as the location of the nearest base station as long as his/her mobile phones communicate with the base stations in 4G. As some phone applications constantly exchange data with back-end servers, the position of individuals can be recorded up to a high frequency of every second. We then divide each single day into 96 time windows, each with a duration of 15 minutes, and consider a user stops in a location if he/she stays there continuously or discontinuously for 10 minutes or more within the 15-minute time window. Otherwise, if the user does not stay in the same location for more than 10 minutes within a time window, we identify the data point as “moving”.

To identify those active users with a valid mobility track, we require the following characteristics to be present in each mobility track: (1) a location of home (see below for the inference of home position); (2) records of mobility in all 7 days; (3) a location outside home recorded in at least one time window per day; (4) at most 30 “moving” time windows among all the 96 time windows per day; (5) less than 24 time windows at home among the 60 time windows per

day during the daytime which we define as the period between 6am to 9pm. Finally, 702,477 mobility tracks satisfy the above requirements and were used in our simulations.

Inference of users' home, workplace and profession

To analyze how the virus transmits inside and outside home, we define the home location of each individual to be the location he/she stayed the longest time between 8pm to 7am the next morning during weekdays, and with at least 6 hours of stay. Since 4G communications from mobile phones sometimes switch between nearby base stations even if the phones are stationary, we substitute the overnight locations of a user which were within 250m from his/her inferred home location by his/her inferred home location, following the definition in [41]. Similarly, we define the workplace for each individual to be the location he/she stayed the longest time between 7am to 8pm on the same day during weekdays, and with at least 3 hours of stay.

Based on the workplace, we can first group users into 13 profession categories which include catering staff, village workers, university staff, transportation staff, hotel staff, urban residential workers, industrial and corporate workers, retailers, suburban workers, entertainment staff, medical practitioners, school students and civil servants. We remark that urban residential workers, suburban workers and village workers include individuals whose workplace locations are identified in urban residential areas, suburban areas and villages respectively, but the exact nature of the location is not known. On the other hand, entertainment staff include providers of entertainment and accessory services such as karaoke, hair salon, public bathrooms, car maintenance, etc. Users whose workplace cannot be identified using the above criteria are first considered as “freelancers”. Some of the freelancers are then considered as retirees as we will describe below. We also remark that although our inference of users' home, workplace, profession and age group may not be perfect, such inference from real data would still lead to much more realistic simulations compared to agent-based simulation studies of which the properties of individual agents are all modeled.

Inference of users' age

Based on the 7th official Census of Shijiazhuang in 2022, we group users into three age groups, namely (1) under 15, (2) 15 to 60, and (3) above 60. To infer the age group of individual users based on their mobility tracks, we first classify users with profession “school students” as under 15. On the other hand, to identify elderly users, we calculate the daily total traveling distance and the radius of gyration for all recorded users as shown in Table 1, and for those “freelancers” with both of these measures ranked in the bottom 15% percentile, they are reclassified as “retirees” and are included in the age group with age above 60. Finally, the population in the three age groups in each district is consistent to the statistics given by the 7th Census of Shijiazhuang.

Modeling household members

Since our dataset only includes the mobility patterns of individual users without other information, to analyze transmissions at home, we have to select the household members of

Table 1. The average, median, 25th and 15th percentile of daily total traveling distance and the radius of gyration estimated from the mobility tracks of all users.

	Average	Median	25 th percentile	15 th percentile
Daily total traveling distance (meters)	17485	10331	2829	791
Radius of gyration (meters)	2435	1423	497	213

<https://doi.org/10.1371/journal.pcbi.1011083.t001>

Table 2. The statistics of the number of members per household in Shijiazhuang based on the 7th Census in 2022.

Number of household members	Percentage
1	0.1243
2	0.2716
3	0.2316
4	0.1902
5	0.097
6	0.0943

<https://doi.org/10.1371/journal.pcbi.1011083.t002>

individual users. From the 7th Census of Shijiazhuang in 2022, we obtain the statistics on the number of members in a household as shown in Table 2. We then randomly group individual users with the same inferred home location in households according to these statistics, with only one requirement that individuals under 15 cannot form a household by themselves.

Supporting information

S1 Fig. The basic statistics of the data.

(PDF)

S2 Fig. The effect of the number of initial spreaders on the prevalence of the virus.

(PDF)

S3 Fig. The effect of a smaller infection rate on the prevalence of the virus.

(PDF)

S4 Fig. The effect of a longer incubation period on the prevalence of the virus (an infected individual cannot infect others during the incubation period).

(PDF)

S5 Fig. The role of home infection rate on the prevalence of the virus.

(PDF)

S6 Fig. The role of home infection rate on the prevalence of virus.

(PDF)

S7 Fig. We simulate the spreading results given that the infection starts at individuals of different professions, respectively.

(PDF)

S8 Fig. We simulate the spreading results given that the infection starts at locations in different location categories, respectively.

(PDF)

S9 Fig. The simulation results where transmissions through both contact and environment are considered.

(PDF)

S10 Fig. We first calculate the Pearson correlation between the mobility patterns of each city in year 2019, 2020, 2021, 2022 and the patterns after opening up in year 2023. The figure shows the distribution of highest Pearson correlation in each year.

(PDF)

S1 Table. The population in different districts of Shijiazhuang city, together with the percentage of population in different age groups in each district.

(PDF)

Author Contributions**Conceptualization:** Chen Zhao, Chi Ho Yeung, An Zeng.**Formal analysis:** Chen Zhao, Chi Ho Yeung, An Zeng.**Investigation:** Chen Zhao, Chi Ho Yeung, An Zeng.**Methodology:** Chen Zhao, Chi Ho Yeung, An Zeng.**Resources:** Chen Zhao, Jialu Zhang, Xiaoyue Hou.**Software:** Chen Zhao, Jialu Zhang, Xiaoyue Hou.**Supervision:** Chi Ho Yeung, An Zeng.**Validation:** Chi Ho Yeung, An Zeng.**Visualization:** Chen Zhao, Jialu Zhang, Xiaoyue Hou.**Writing – original draft:** Chi Ho Yeung, An Zeng.**Writing – review & editing:** Chen Zhao, Chi Ho Yeung, An Zeng.**References**

1. Riediker M, Briceno-Ayala L, Ichihara G, Albani D, Poffet D, Tsai DH, et al. Higher viral load and infectivity increase risk of aerosol transmission for Delta and Omicron variants of SARS-CoV-2. *Swiss Medical Weekly*. 2022;(1). PMID: [35019196](https://pubmed.ncbi.nlm.nih.gov/35019196/)
2. Wong SC, Au AKW, Chen H, Yuen LLH, Li X, Lung DC, et al. Transmission of Omicron (B. 1.1. 529)-SARS-CoV-2 Variant of Concern in a designated quarantine hotel for travelers: a challenge of elimination strategy of COVID-19. *The Lancet Regional Health—Western Pacific*. 2022; 18. <https://doi.org/10.1016/j.lanwpc.2021.100360> PMID: [34961854](https://pubmed.ncbi.nlm.nih.gov/34961854/)
3. Paireau J, Andronico A, Hozé N, Layan M, Crepey P, Roumagnac A, et al. An ensemble model based on early predictors to forecast COVID-19 health care demand in France. *Proceedings of the National Academy of Sciences*. 2022; 119(18):e21103302119. <https://doi.org/10.1073/pnas.21103302119> PMID: [35476520](https://pubmed.ncbi.nlm.nih.gov/35476520/)
4. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Physics Reports*. 2021; 913:1–52. <https://doi.org/10.1016/j.physrep.2021.02.001> PMID: [33612922](https://pubmed.ncbi.nlm.nih.gov/33612922/)
5. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MU, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*. 2020; 368(6491):638–642. <https://doi.org/10.1126/science.abb6105> PMID: [32234804](https://pubmed.ncbi.nlm.nih.gov/32234804/)
6. Willem L, Abrams S, Libin PJ, Coletti P, Kuylén E, Petrof O, et al. The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. *Nature Communications*. 2021; 12(1):1–9. <https://doi.org/10.1038/s41467-021-21747-7> PMID: [33750778](https://pubmed.ncbi.nlm.nih.gov/33750778/)
7. Davis EL, Lucas TC, Borlase A, Pollington TM, Abbott S, Ayabina D, et al. Contact tracing is an imperfect tool for controlling COVID-19 transmission and relies on population adherence. *Nature Communications*. 2021; 12(1):1–8. <https://doi.org/10.1038/s41467-021-25531-5> PMID: [34518525](https://pubmed.ncbi.nlm.nih.gov/34518525/)
8. Lessler J, Grabowski MK, Grantz KH, Badillo-Goicoechea E, Metcalf CJE, Lupton-Smith C, et al. Household COVID-19 risk and in-person schooling. *Science*. 2021; 372(6546):1092–1097. <https://doi.org/10.1126/science.abh2939> PMID: [33927057](https://pubmed.ncbi.nlm.nih.gov/33927057/)
9. Liu QH, Zhang J, Peng C, Litvinova M, Huang S, Poletti P, et al. Model-based evaluation of alternative reactive class closure strategies against COVID-19. *Nature Communications*. 2022; 13(1):1–10. <https://doi.org/10.1038/s41467-021-27939-5> PMID: [35031600](https://pubmed.ncbi.nlm.nih.gov/35031600/)
10. García Bulle B, Shen D, Shah D, Hosoi AE. Public health implications of opening National Football League stadiums during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*. 2022; 119(14):e2114226119. <https://doi.org/10.1073/pnas.2114226119> PMID: [35316127](https://pubmed.ncbi.nlm.nih.gov/35316127/)

11. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*. 2020; 395(10225):689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
12. Ye Y, Zhang Q, Ruan Z, Cao Z, Xuan Q, Zeng DD. Effect of heterogeneous risk perception on information diffusion, behavior change, and disease transmission. *Physical Review E*. 2020; 102(4):042314. <https://doi.org/10.1103/PhysRevE.102.042314> PMID: 33212602
13. Yang J, Zhang Q, Cao Z, Gao J, Pfeiffer D, Zhong L, et al. The impact of non-pharmaceutical interventions on the prevention and control of COVID-19 in New York City. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2021; 31(2):021101. <https://doi.org/10.1063/5.0040560> PMID: 33653072
14. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. 2021; 589(7840):82–87. <https://doi.org/10.1038/s41586-020-2923-3> PMID: 33171481
15. Cai J, Deng X, Yang J, Sun K, Liu H, Chen Z, et al. Modeling transmission of SARS-CoV-2 omicron in China. *Nature Medicine*. 2022; p. 1–8. <https://doi.org/10.1038/s41591-022-01855-7> PMID: 35537471
16. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*. 2021; 17(7):e1009149. <https://doi.org/10.1371/journal.pcbi.1009149> PMID: 34310589
17. Thomine O, Alizon S, Boennec C, Barthelemy M, Sofonea M. Emerging dynamics from high-resolution spatial numerical epidemics. *Elife*. 2021; 10:e71417. <https://doi.org/10.7554/eLife.71417> PMID: 34652271
18. Zhou H, Zhang Q, Cao Z, Huang H, Dajun Zeng D. Sustainable targeted interventions to mitigate the COVID-19 pandemic: A big data-driven modeling study in Hong Kong. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2021; 31(10):101104. <https://doi.org/10.1063/5.0066086> PMID: 34717342
19. Vespignani A, Tian H, Dye C, Lloyd-Smith JO, Eggo RM, Shrestha M, et al. Modelling COVID-19. *Nature Reviews Physics*. 2020; 2(6):279–281. <https://doi.org/10.1038/s42254-020-0178-4> PMID: 33728401
20. Li R, Dong L, Zhang J, Wang X, Wang WX, Di Z, et al. Simple spatial scaling rules behind complex cities. *Nature Communications*. 2017; 8(1):1841. <https://doi.org/10.1038/s41467-017-01882-w> PMID: 29184073
21. Li R, Richmond P, Roehner BM. Effect of population density on epidemics. *Physica A: Statistical Mechanics and its Applications*. 2018; 510:713–724. <https://doi.org/10.1016/j.physa.2018.07.025>
22. Li R, Wang W, Di Z. Effects of human dynamics on epidemic spreading in Côte d'Ivoire. *Physica A: Statistical Mechanics and its Applications*. 2017; 467:30–40. <https://doi.org/10.1016/j.physa.2016.09.059>
23. Wilder-Smith A. What is the vaccine effect on reducing transmission in the context of the SARS-CoV-2 delta variant? *The Lancet Infectious Diseases*. 2022; 22(2):152–153. [https://doi.org/10.1016/S1473-3099\(21\)00690-3](https://doi.org/10.1016/S1473-3099(21)00690-3) PMID: 34756187
24. Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. 2020; 368(6490):493–497. <https://doi.org/10.1126/science.abb4218> PMID: 32213647
25. Hu S, Xiong C, Yang M, Younes H, Luo W, Zhang L. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. *Transportation Research Part C: Emerging Technologies*. 2021; 124:102955. <https://doi.org/10.1016/j.trc.2020.102955> PMID: 33456212
26. Aleta A, Martín-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*. 2020; 4(9):964–971. <https://doi.org/10.1038/s41562-020-0931-9> PMID: 32759985
27. Aleta A, Martín-Corral D, Bakker MA, Pastore y Piontti A, Ajelli M, Litvinova M, et al. Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *Proceedings of the National Academy of Sciences*. 2022; 119(26):e2112182119. <https://doi.org/10.1073/pnas.2112182119> PMID: 35696558
28. Hou X, Gao S, Li Q, Kang Y, Chen N, Chen K, et al. Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*. 2021; 118(24):e2020524118. <https://doi.org/10.1073/pnas.2020524118> PMID: 34049993
29. Allen WE, Altae-Tran H, Briggs J, Jin X, McGee G, Shi A, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nature Human Behaviour*. 2020; 4(9):972–982. <https://doi.org/10.1038/s41562-020-00944-2> PMID: 32848231
30. Vigfusson Y, Karlsson TA, Onken D, Song C, Einarsson AF, Kishore N, et al. Cell-phone traces reveal infection-associated behavioral change. *Proceedings of the National Academy of Sciences*. 2021; 118(6):e2005241118. <https://doi.org/10.1073/pnas.2005241118> PMID: 33495359

31. Rüdiger S, Konigorski S, Rakowski A, Edelman JA, Zernick D, Thieme A, et al. Predicting the SARS-CoV-2 effective reproduction number using bulk contact data from mobile phones. *Proceedings of the National Academy of Sciences*. 2021; 118(31):e2026731118. <https://doi.org/10.1073/pnas.2026731118> PMID: 34261775
32. Pung R, Firth JA, Spurgin LG, Lee VJ, Kucharski AJ. Using high-resolution contact networks to evaluate SARS-CoV-2 transmission and control in large-scale multi-day events. *Nature Communications*. 2022; 13(1):1–11. <https://doi.org/10.1038/s41467-022-29522-y> PMID: 35414056
33. Liu C, Yang Y, Chen B, Cui T, Shang F, Fan J, et al. Revealing spatiotemporal interaction patterns behind complex cities. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2022; 32(8):081105. <https://doi.org/10.1063/5.0098132> PMID: 36049958
34. Lucchini L, Centellegher S, Pappalardo L, Gallotti R, Privitera F, Lepri B, et al. Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures. *Scientific Reports*. 2021; 11(1):1–12. <https://doi.org/10.1038/s41598-021-04139-1>
35. Burki TK. Omicron variant and booster COVID-19 vaccines. *The Lancet Respiratory Medicine*. 2022; 10(2):e17. [https://doi.org/10.1016/S2213-2600\(21\)00559-2](https://doi.org/10.1016/S2213-2600(21)00559-2) PMID: 34929158
36. Organization WH, et al. COVID-19 essential supplies forecasting tool overview of the structure, methodology, and assumptions used: interim guidance, 7 April 2021. World Health Organization; 2020.
37. Killingley B, Mann AJ, Kalinova M, Boyers A, Goonawardane N, Zhou J, et al. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Medicine*. 2022; 28(5):1031–1041. <https://doi.org/10.1038/s41591-022-01780-9> PMID: 35361992
38. Rajpal VR, Sharma S, Kumar A, Chand S, Joshi L, Chandra A, et al. “Is Omicron mild”? Testing this narrative with the mutational landscape of its three lineages and response to existing vaccines and therapeutic antibodies. *Journal of Medical Virology*. 2022; 94(8): 3521–3539. <https://doi.org/10.1002/jmv.27749> PMID: 35355267
39. Ódor G, Czifra D, Komjáthy J, Lovász L, Karsai M. Switchover phenomenon induced by epidemic seeding on geometric networks. *Proceedings of the National Academy of Sciences*. 2021; 118(41): e2112607118. <https://doi.org/10.1073/pnas.2112607118> PMID: 34620714
40. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*. 2020; 4(5):460–471. <https://doi.org/10.1038/s41562-020-0884-z> PMID: 32355299
41. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, González MC. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*. 2015; 58:162–177. <https://doi.org/10.1016/j.trc.2015.04.022>