

Peng, N., & Liu, X. (2024). The Impact of Urban Scaling Structure on the Local-Scale Transmission of COVID-19: A Case Study of the Omicron Wave in Hong Kong Using Agent-Based Modeling. Annals of the American Association of Geographers, 114(5), 1079–1097.

This is an Accepted Manuscript of an article published by Taylor & Francis in Annals of the American Association of Geographers on 03 Apr 2024 (published online), available at: https://doi.org/10.1080/24694452.2024.2313517.

The impact of urban scaling structure on the local-scale transmission of COVID-19: A case study of the Omicron wave in Hong Kong using agent-based modelling

Authors: Ningyezi PENG [a] and Xintao LIU [a, b *]

Authors' affiliations:

[a] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong*

[b] *Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong*

Corresponding authors:

Xintao LIU, PhD

Associate Professor, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong

Email: xintao.liu@polyu.edu.hk

The impact of urban scaling structure on the local-scale transmission of COVID-19: A case study of the Omicron wave in Hong Kong using agent-based modelling

Superspreading events underscore the uneven distribution of COVID-19 transmission among individuals and locations. These heterogenous transmission patterns may stem from human mobility, yet the underlying mechanisms are still not fully understood. Here, we employ an agent-based model incorporating urban scaling structure to simulate fine-grained mobility and human-to-human transmission process. Our results reveal that not only the quantity but also the scaling structure of mobility profoundly influence local transmission risk. Urban scaling structure is characterized by a widely-found power-law scaling distribution of mobility volumes across different locations. By integrating this structure, our model fits reasonably well with empirical Omicron data at various spatial scales in Hong Kong. Further analyses show a positive association between the scaling index, representing the location's importance within the structure, and local transmission risks among urban areas as well as the likelihood of becoming superspreaders among local visitors. This implies that urban scaling structure may offer the first-mover advantage to a minority of places and individuals to infect earlier and thus infect more. This study brings important insights for transmission dynamics of COVID-19 and similar diseases, highlighting the role of urban scaling structure in influencing local transmission risks and superspreading events.

Keywords: COVID-19; scaling structure; complex system; transmission risk; superspreading event

Introduction

The COVID-19 pandemic has caused numerous infections and fatalities since the end of 2019 (Keni et al. 2020; A. T. Levin et al. 2022; Paul, Brown, and Ridde 2020). Extensive evidence reveals that the impact of COVID-19 is unevenly distributed among social groups and urban areas (Ahmed et al. 2020; Blundell et al. 2022; Damme et al. 2020; Huang and Kwan 2021; Xi et al. 2023). For example, disadvantageous groups and individuals residing in densely populated areas often experience a higher COVID-19

incidence (Chang et al. 2021; Chowkwanyun and Reed 2020; Yancy 2020). Moreover, in addition to disparities in incidence rates, local transmission processes exhibit significant disparities as well. Superspreading events demonstrate that a minority of individuals and locations can contribute to the majority of transmission (Adam et al. 2020; Lau et al. 2020; Lewis 2021; Majra et al. 2021). Such uneven transmission patterns place a substantial burden on intervention strategies. To better combat future pandemic, a deep understanding of local transmission patterns within cities is imperative (Koks et al. 2020; Lewis 2021; Rasmussen, Khoury, and del Rio 2020; Yang et al. 2023).

Local-scale transmission of COVID-19 has been closely linked to human mobility patterns (Damme et al. 2020; Franch-Pardo et al. 2020; Kan et al. 2023; Nouvellet et al. 2021; Xu, Jin, and Lee 2023; W. Zhang and Ning 2023). However, the mechanistic connections between the two are still not fully understood (Alessandretti 2022). Many studies identify local mobility volumes as a critical predictor of COVID-19 transmission trends, and they conjectures that the large mobility volume result in more intensive contacts and thus an increased transmission risk (Badr et al. 2020; Hong et al. 2021; Kogan et al. 2021; R. Levin et al. 2021; Nouvellet et al. 2021). Nonetheless, this explanation may not be exhaustive, as it primarily focuses on the magnitude of local mobility while overlooking other important aspects, such as, the underlying structure of mobility.

The underlying structure of human mobility could also exert a substantial influence on the transmission dynamics of COVID-19. Previous research has established that mobility volumes within urban areas follows a power-law scaling distribution (Batty 2008; Bettencourt 2013; Brockmann, Hufnagel, and Geisel 2006; Jiang, Yin, and Zhao 2009). In simple terms, only a minority of urban areas exhibit large mobility volumes, while the majority have lower mobility volumes. This phenomenon is referred as urban

scaling structure in this study. Many studies have confirmed that this structure can be used to simulate fine-grained mobility patterns across urban space (Jiang and Jia 2010; D. Ma et al. 2020; Schläpfer et al. 2021). This structure characterizes how urban areas with heterogenous mobility volumes are spatially distributed and how they interact at the local scale, which can be critical to influencing local-scale transmission of COVID-19.

Despite the importance of urban scaling structure, most existing studies have primarily focused on examining its effects on overall transmission trends of entire cities or countries, with limited exploration of its impact on the local-scale transmission within cities (Aguilar et al. 2022; Lima and Atman 2021; Schläpfer et al. 2014; Tizzoni et al. 2015). For example, one study compared urban scaling structures in three cities and found that more centralized cities, where mobility gathers in a few hotspots, tend to experience larger epidemic size and faster spread rates (Aguilar et al. 2022). Two other studies examined the scaling properties of human interactions across multiple large cities using mobile phone data or Twitter data and found that cities with the higher degree of scaling can facilitate the overall disease transmission (Schläpfer et al. 2014; Tizzoni et al. 2015). While these studies underscore the crucial role of urban scaling structure in disease transmission, they leave the question of how this structure influences local-scale transmission within cities largely unanswered.

To address this gap, this study employed a spatially explicit Agent-Based Model (ABM) that incorporated urban scaling structure to simulate fine-scaled mobility patterns and to reconstruct local transmission processes. The spatial scale employed for simulating local transmission in this study is the subunit of Tertiary Planning Unit (subTPU) in Hong Kong, consisting of 4,863 units. This approach enables us to gain deeper insights into the underlying mechanisms through which urban scaling structure contributes to the heterogeneity in local transmission risks. To validate our model, we compared the

simulation outcomes with empirical case data from the Omicron wave in Hong Kong (HK) across three spatial scales. Further statistical analyses of simulation data examined the relationship between the scaling index, representing a location's importance within the structure, and local transmission risk as well as the likelihood of local visitors becoming superspreader. Additionally, we analysed the unevenly distributed stress on local hospitals resulting from heterogenous transmission patterns and provided qualitative suggestions for hospital emergency preparedness. This study thus enhances our understanding of how human mobility and its scaling structure influence local transmission risks and superspreading events, which may inform precise and effective interventions to combat future pandemic.

Study area and datasets

Our study area is HK that is a worldwide metropolitan city with a population of over 7.4 million. The Omicron wave in HK, from 1 February to 30 March 2022, was chosen as our study event. The Omicron wave infected over 40 percent of population (HKUMed 2022). On contrary, previous waves caused limited number of infections (around 0.2 percent of population), and such a small number might bring large uncertainty to our study. We thus excluded the other waves and chose the Omicron wave. In this study, empirical data regarding four aspects are used to build the model and to validate the model, including case data, demographic data, mobility data, and vaccine data.

Case data

Empirical case data at three spatial scales were used to validate our model (see *Model initialization, calibration and validation* in *Methods*): (1) at the city level, the number of daily confirmed cases was directly used; (2) at the tertiary planning units (TPU) level, we retrieved the number of daily cases confirmed by Rapid Antigen Tests (RAT) from the

RAT reporting platform, launched on 7 March 2022, that records over 460 thousand self-reporting cases and their home building locations (Hong Kong Government 2023a). (3) at the subunit of TPU (subTPU) level, we identified the high-risk subTPUs where is most frequently visited by observed cases based on contact tracing data. The contact tracing data records which buildings the cases had visited during the period 1 to 14 days before their symptoms started, as of 6 February 2022 (Hong Kong Government 2023b), which contains 22,869 visitations of 10,608 cases.

Demographic data

The 2016 census tables were used to generate synthetic population, which contain the number of populations across 1,622 census tracts by sex, age groups, and household sizes. The 2011 household survey data from the Hong Kong Transport Department records the compositions of household members in 35,401 households.

Mobility data

The 2011 travel survey data from the Hong Kong Transport Department was used to generate synthetic mobility behaviors, which provides 121,204 one-day trips of 58,843 residents across 4,863 subTPUs. As shown in Figure S1, even though the 2011 travel survey data do not cover the simulation period, it exhibits a notable representativeness of Hong Kong mobility patterns, when comparing to the mobility structure of 2020 subway data (Zhang et al. 2021). In addition, the Google mobility change index (Google LLC 2023) was used to represent the overall mobility change in response to interventions during the simulation period. Google mobility change index measures variations in mobility volume at the city level. Based on anonymized and aggregated location data from users, Google calculates a percentage change relative to a baseline period (from 3

January to 6 February 2020) to facilitate temporal comparisons. As shown in **Figure S4**, positive values indicate an increase in visits, while negative values suggest a decrease.

Vaccine data

To account for the population immunity, we retrieved the daily counts of vaccination by age group from the Hong Kong Health Bureau Department (DATA.GOV.HK 2023). It records cumulative COVID-19 vaccine uptake of BioNTech or Sinovac, categorized by the vaccine type and order (e.g., BioNTech-BioNTech, or BioNTech-Sinovac) in different age on a daily basis.

Methods

To build the spatially explicit ABM, the census data and household survey data were first used to generate the demographic and household characteristics of synthetic individuals (**Figure 1a**), using iterative proportional fitting (IPF) (Templ et al. 2017; Wong 1992). 727,796 agents, almost 10 percent of the Hong Kong population, with age, sex, and household structure characteristics were generated that have identical joint distributions of the real population across 1,622 census tracts.

We then generated the mobility and social contact behaviours of the synthetic individuals, and disease would start to spread through those behaviours. For mobility behaviours, the traditional mobility model, the departure-diffusion model (Giles and Wesolowski 2023), was first built using the limited number of 2011 travel survey data in Hong Kong, simulating the daily mobility flows across TPUs (**Figure 1b**). To further refine the mobility flows to subTPUs, we built a subTPU complex network that captures the urban scaling structure based on the hierarchical and spatial relationships of subTPUs considering mobility volume (Jiang and Jia 2010; D. Ma et al. 2020; Schläpfer et al. 2021). Based on the complex network, the scaling index for each subTPU was calculated by

using the Google's PageRank algorithm (Brin and Page 1998) to refine the destinations of mobility flows to the subTPU level (Jiang and Jia 2010; Jiang, Yin, and Zhao 2009).

There is a two-layer contact structure in our model (**Figure 1c)**. On a daily basis, agents first had contacts with families at home, and then had contacts with other people in the same places during trips. Each agent had a certain number of daily contacts that was sampled from the age-dependent distributions (**Figure 1d**). For every contact, the susceptible agent had a specific probability to contract the virus from the infectious and enter the latency period and afterward the infectious period.

Most of our parameters were from empirical evidence (**Table** 1), but two constant parameters needed to be calibrated: the infection probability per contact and the initial proportion of latent cases. We calibrated them to the observed city-level daily cases by minimizing the Root Mean Square Error (RMSE). In addition, the in-sample prediction accuracy was evaluated: when calibrated on the number of city-level cases, our best-fit model can predict roughly well the city-level cases and the TPU-level cases.

Simulating synthetic individuals

Demographic and household characteristics

Based on the aggregate census tables and individual-level household survey data, we used IPF to generate and calibrate the 727,796 agents and 263,609 households with age, sex, household structure characteristics across 1,622 census tracts that have identical joint distributions of the real population (**Table 1**) (Templ et al. 2017; Wong 1992). The simulated household structure includes the composition of individuals within a household unit. Our model considers not only the household sizes but also the age and sex composition of household members.

The workflow for simulating the demographic and household characteristics of synthetic individuals follows the following process (Templ et al. 2017): firstly, the individual-level household survey data is calibrated to match the true population numbers by age and sex, using IPF. IPF is a technic to fit an n-dimensional table (household survey data in our context) with unknown entries to match a set of marginal distributions (derived from aggregate census tables). This iterative process involves adjusting the inner cells for each dimension to match the totals for the given dimensions according to related marginal distributions. The iteration continues the n-dimensional table fits all marginal distributions. Then, the calibrated individual-level household survey data is extrapolated to create the realistic household structure of the synthetic population. Finally, synthetic individuals and households are randomly allocated to each census tract, with ongoing adjustments to the allocation process until the convergence is achieved and a good fit to the aggregate census tables is attained.

Urban scaling structure, the scaling index, and mobility behaviours

To account for the spatial heterogeneity in mobility patterns, we initially simulated the TPU-TPU mobility flows using the departure-diffusion mobility model (Giles and Wesolowski 2023). Considering the trade-off between spatial resolution and prediction accuracy of the traditional mobility models, we subsequently constructed a subTPU complex network to inform mobility at finer spatial scales. The nodes and links of this network serve to characterize the urban scaling structure, specifically delineating the scaling properties of mobility towards different locations within a city. In this case, the scaling indices, which encapsulate the scaling properties of urban scaling structure, were used to distribute the destinations of previously simulated TPU-TPU flows to the subTPU

level. As a result, this approach enabled the simulation of mobility volumes across subTPUs that follows the scaling law.

In the simulation of TPU-TPU mobility flows, the departure-diffusion model (Giles and Wesolowski 2023) was used. This model estimates the travel probability within and outside the origin separately and combines them using conditional probability rules. The model first estimates the travel probability outside the origin location i (the departure process) and then the distribution of travel from the origin by normalizing connectivity values across all j destinations (the diffusion process). These two processes are then combined in the departure-diffusion model as τ_i (the probability of leaving origin i) and π_{ij} (the probability of going from i to j). The probability of travel within the origin i is denoted as **Equation 1**, and the probability of travel outside the origin i is described as **Equation 2**. The expected mean number of trips for route i→j is then as shown in **Equation 3**, where θ is a proportionality constant representing the overall number of trips per person in an origin population of size N_i.

$$Pr\left(\neg depart_i\right) = 1 - \tau_i \quad (1)$$

$$\Pr\left(depart_i, diffuse_{i\to j}\right) = \Pr\left(diffuse_{i\to j} \mid depart_i\right)\Pr\left(depart_i\right) = \tau_i \pi_{ij} \quad (2)$$

$$\lambda_{ij} = \begin{cases} \theta N_i(1-\tau_i) & if\ i = j \\ \theta N_i \tau_i \pi_{ij} & if\ i \neq j \end{cases} \quad (3)$$

As there is a trade-off between spatial resolution and prediction accuracy of traditional mobility models with limited number of travel survey data (**Table S6**), we then characterized urban scaling structure to further refine the mobility patterns across subTPUs.

Urban scaling structure, in this study, particularly refers to the scaling property of human mobility within cities. More precisely, the mobility volumes originating from or

destined for different locations in a city follows the scaling law (Batty 2008; Brockmann, Hufnagel, and Geisel 2006; Jiang, Yin, and Zhao 2009). The scaling property of mobility can be reflected in nodes (origin and destination of movement) and links (between nodes) from a network perspective, which has proven effective in informing mobility at quite fine spatial scales (Jiang and Jia 2010; D. Ma et al. 2020; Schläpfer et al. 2021). Jiang (2018) introduced a topological network representation of taking urban space as a living structure that differentiates from and adapts to each other, as intense local competition for space could be the major reason of the emergent scaling property of cities (Batty 2008). By classifying nodes into different hierarchies and creating links between hierarchies and across hierarchies, the complex network can reflect the scaling property of mobility within cities at fine spatial scales (Jiang 2018; Jiang and Liu 2012; D. Ma et al. 2020). Therefore, we adopted this topological complex network to characterize the urban scaling structure. To build it, several steps were followed (**Figure 1b**):

(1) Firstly, we took subTPUs as distinct nodes, and divided them into 5 hierarchies (see **Table** S7) by mobility volume using head/tail breaks proposed by Jiang (2013). The head/tail breaks classification method is specifically designed for heavy-tailed data (e.g., the power-law scaling distributed data). This method involves iteratively partitioning all data values around the mean into two parts until the head part are no longer heavy-tailed distributed. (2) The nodes in the same hierarchy were then used to create Thiessen polygons, and one hierarchy leads to one way of segmentation of urban space. The hierarchies and the Thiessen polygons we partitioned demonstrate how urban areas differentiate from each other. (3) Based on the polygon-to-polygon relationships, we created edges among subTPUs, which show how urban areas adapt to each other. Specifically, in the same hierarchy, the small sized subTPU points to the adjacent large subTPUs, and across two consecutive hierarchies, the low-level subTPU points to the

high-level subTPU which contains it. As a result, the subTPU complex network was created to characterize the scaling property of human mobility patterns across subTPUs.

Based on the constructed subTPU network, the Weighted PageRank algorithm (Brin and Page 1998) was used to calculated the scaling index for each node, as this algorithm can evaluate the relative importance of nodes in the scaling-characterized networks. The Weighted PageRank algorithm considers not only the mobility volume of the node but also the mobility volume of its neighbours and its neighbours' neighbours through an iterative process. The "neighbours" of a node here is defined as the other nodes that directly point to it. The algorithm first assigns an initial index (the mobility volume) to each node. Then, through an iterative process, it adjusts these indices by considering the indices of linking nodes (**Equation** 4). This iterative nature of this algorithm ensures convergence to stable values, providing a measure of a node's relative importance within the complex network.

$$PR\,(i) = (1 - d) + d\sum_j PR(j)W_{i,j} \quad (4)$$

where d is a damping factor (usually set to 0.85), and j is the nodes that point to i, $W_{i,j}$ is the weight of $link_{i,j}$ which is typically normalized to ensure that the sum of the weights for outgoing links from each node is equal to 1. More mathematics details can be found in Xing and Ghorbani (2004).

As the calculated scaling index is a good indicator to predict the fine-grained aggregate mobility (Jiang and Jia 2010; Jiang, Yin, and Zhao 2009), it was used to prorated TPU-level mobility to subTPUs. In essence, the destinations of original TPU-TPU mobility flows were refined to the subTPU level, but the origins remained at the TPU level. The origins (namely, the home locations of synthetic individuals) could not be simulated at the subTPU level due to the limitations of available census tables. For the technic details, agents were first assigned their TPU destinations by their home TPUs;

and according to the refinement of TPU destinations, agents were then assigned finer subTPU destinations. After the processes, we differentiated agent's visits across 4,863 subTPUs, which would further have important effect on the social contact behaviours in subTPUs.

Mobility changes to Non-pharmacological Interventions (NPI)

We assume that the effect of social distancing, school closure, and work from home can be reflected as mobility decline. The reduced mobility would result in decreased social contacts and thus disease control. Therefore, the Google mobility change index (Google LLC 2023) that represented the mobility change during the simulation period was used to represent the NPI effects on disease spread (**Figure S3**). Notably, we assume that half of infected individuals would undergo 7-day isolation at home (HKUMed 2022). The mobility changes resulting from isolation were integrated into our arrangement of the daily mobility change index. This setting reflects the real-world scenario where individuals who developed symptoms or was confirmed as positive cases are more likely to stay home (Bian et al. 2012; Cai et al. 2022; HKUMed 2022). To apply this setting, we employed probability sampling to determine who would undergo isolations when infected. Subsequently, when they were developing symptoms or being confirmed (whenever is earlier), they would start their 7-day isolations. After this, the remaining reduction in mobility was randomly assigned to other non-infected individuals, using the Monte Carlo method.

Contact behaviours

As shown in the two-layer contact structure (**Figure 1c**), all agents first contacted every other family member on a daily basis. In our study, the contacts occurred within households and among family members were classified as household contacts. Contacts

occurring during trips, excluding those within households, were classified as non-household contacts. Subsequently, after cancelling a certain percentage of agent trips as described in *Mobility changes to Non-pharmacological Interventions (NPI)*, for the agent who still had trips, it visited different subTPUs and had a chance to contact a certain number of contactees in the same subTPU. The contactees were randomly chosen on the first simulation day and fixed afterward, using the Monte Carlo method. A certain number of contacts was sampled from the age-dependent distributions (**Figure 1d**) (Kucharski et al. 2014): we first sampled the total number of contacts from the age-dependent distribution, and then simply assume that the number of non-household contacts was evenly distributed in daily visited subTPUs.

Infection attributes

Susceptible-Latent-Infectious-Removed model was used to represent the course of COVID-19 infection in our model. On the first simulation day, almost all agents are susceptible to the virus, except for a proportion of initial cases. Through daily contacts, susceptible agents could contract the virus and subsequently enter the latency period and the infectious period. As the Gamma distribution is a probability distribution that models right-skewed data and the individual variability in latency period and infectious period tends to be right-skewed, Omicron studies tend to adopt this distribution to characterize individuals' latency period and infectious period (Cai et al. 2022; Manica et al. 2022). Therefore, we used the estimated Gamma distributions to sample the lengths of the latency period and the infectious period for each infected agent, with mean values of 1.20 and 5.64, respectively. Moreover, we further assume that the confirmation delay from being infected to being confirmed follows the Gamma distribution, with a mean delay time of 5.05 days (Manica et al. 2022).

To estimate the number of confirmed cases from the overall infections, we assume that 8 percent of the infections were confirmed before 24 February, 30 percent were confirmed between 24 February and 7 March, and afterward, 35 percent were confirmed, due to the changing reporting standard. On 24 February, the HK government officially acknowledged positive cases tested by commercial laboratories as confirmed cases, reporting them without double confirmation. Additionally, the RAT online reporting system, launched on 7 March, allowing citizens to report positive results tested since 26 February. The values of report rates were from Hong Kong University's study (HKUMed 2022), assuming report rates were 0.08 before and 0.20 after 24 February. However, the post-February 24 value did not account for RAT-reported cases, as these values were estimated on 28 February 2022, before the system's launch. The RAT system-reported cases accounted for 40 percent of the total reported cases during the simulation period, so we adjusted report rates to our current setting as $20\% \div (1 - 40\%) \approx 33\%$. Sensitivity analyses on other values, including increasing from 0.08 to 0.10, and decreasing from 0.30 to 0.25, and 0.35 to 0.30, revealed that our current setting fits better than other scenarios (see *Sensitivity analyses on report rates* in *Supplementary Material*).

Protection behaviours

This work considered vaccine-acquired immunity, due to the limited number of infections (in total 12,258 cases by 20 December, 2021) during previous waves in HK. As there is no geographic information of vaccination data, we randomly assigned the vaccination status to agents by age group, using the Monte Carlo method. The vaccination status includes how many doses and which type of vaccine the agent had been taken. We assume that the vaccination can reduce the per-contact infection probability by certain percentages (**Table S8**) that are determined by the inoculation time, the number of doses,

and the vaccine type (Andrews et al. 2022; HKUMed 2022; Kirsebom et al. 2022). In addition, as almost 96.6 percent Hong Kong residents wear masks, we did not explicitly consider the protective effects of mask-wearing behaviours (Cheng et al. 2020).

To model heterogenous healthcare burden associated with Emergency Department (ED) visits and hospitalizations, we assume that 8.6 percent of confirmed cases visited Emergency Department (ED), and 2.7 percent of confirmed cases required hospitalization (Iuliano et al. 2022). These severe cases would search the medical care from the hospital closest to their residences. To determine the ED capacity for COVID-19 patients in 17 Hong Kong hospitals, we assume that half of the ED capacity (Hong Kong Government 2023c) was used to serve COVID-19 patients (HKUMed 2022). As for the hospital bed capacity, we assume that one in five hospital beds per day were available capacity for new hospitalizations in seven COVID-19-designated hospitals because patients would normally stay in hospital for 5 days (HKUMed 2022).

Model initialization, calibration and validation

Model initialization

The experiments span from 1 February to 30 March 2022. We assume that initially a p_0 proportion of agents was infected and all of them were in the latency period. The latent agents were randomly selected from the whole population, using the Monte Carlo method.

Model calibration

Most model parameters (**Table 1**) can be estimated from empirical data or studies. Apart from this, two parameters need to be calibrated with data: (1) the infection probability per contact p_i; (2) the initial proportion of exposed cases p_0. We calibrated these two

parameters to the city-level number of confirmed cases from 1st February to 30th March 2022 as follows:

We first identified plausible parameter ranges and selected 800 parameter combinations to test which one best fits to the empirical data (see *Selecting Parameter Ranges* in *Supplementary Materials)*. For each combination, 30 replicates were run. After running all combinations, we compared the model prediction to the empirical data by calculating RMSE: for each run,

$$RMSE = \sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(N_{confirmed}^{d} - \hat{N}_{confirmed}^{d}\right)^2} \quad (5)$$

where $\hat{N}_{confirmed}^{d}$ is the observed confirmed cases (per 10,000 people) on day d, and $N_{confirmed}^{d}$ is the corresponding value in the simulation. The average RMSEs over 30 replicates were the final results for all parameter combinations.

The best-fit parameter set is that the one has the lowest value of average RMSE. To consider parameter uncertainty, we selected the parameter sets that have RMSE values within 20 percent of the lowest RMSE (Chang et al. 2021). By gathering the predictions over these selected parameter sets and replicates, we calculated the mean and 2.5th/97.5th percentiles as the model predictions.

Model validation

Three empirical data was used to validate different aspects of the simulation outcomes. Firstly, we used the city-level number of confirmed cases to calculate the city-level RMSE. Secondly, we used the TPU-level number of RAT-confirmed cases from 1st to 30 March 2022 to calculate the RMSEs of 214 TPUs and summed these RMSEs together as the TPU-level RMSE (see *Validation on TPU-level Data* in *Supplementary Materials*). Thirdly, we compared the observed high-risk subTPUs with the simulated superspreading subTPUs that were identified by counting cumulative non-household infections that

occurred in every subTPU. Non-household infections in this context refer to cases that were infected through non-household contacts (refer to *Contact behaviours*). Specifically, these cases do not include household transmission and only consider the community transmission, which is consistent with the criteria of observed data.

Results

Model fitting

Our model fits empirical case data roughly well at both the city level and the TPU level (**Figure 2**). Noted that the TPU-level prediction is more difficult than the city-level prediction, as the former predicts 214 epidemic curves, and the latter only predicts one curve. At the city level, our model accurately predicts the number of cases during the peak period from 21 February to 10 March 2022, but underestimates the case numbers before 21 February. This underestimation can be attributed to the Chinese New Year celebrations (from 1 to 15 February 2022) that caused intense social contacts and thus large number of infections. The limitations of our model stem from its reliance on Google mobility change data, which may not accurately capture the nuances of social behaviour changes during holidays, despite its proven representativeness during regular periods in other studies (Chang et al. 2021; Cot, Cacciapaglia, and Sannino 2021; Yilmazkuday 2021). To solve this issue, more detailed social contact data during holiday periods are needed in future studies.

At the TPU level, our model predicts well the peak period of the eight most serious TPUs, but underestimates the cases after 7 March. As the empirical case data has an unnatural increase after 7 March, which could be largely resulted from the increased reporting behaviours due to the RAT reporting system launched on 7 March 2022, we infer that our underestimates could be partly originated from this reason. Although our

model considered the increased report rate after 7 March (increasing the report rate from 0.3 to 0.35), other potential factors could still play in a role. For example, the launch of RAT reporting system might change the time gaps between the test date and the report date, and such change may significantly vary among individuals. Despite the underestimation, the analysis reveals a strong positive relationship between the cumulative observed RAT cases and simulated cases, as evidenced by a high Pearson correlation coefficient of 0.97 ($p < 0.01$). Furthermore, both empirical data and simulated results exhibit similar spatial patterns (**Figure 2b**).

The superspreading subTPUs

By calculating the average cumulative non-household infections (see *Contact behaviours* in *Methods*) across 30 runs using the best-fit parameters, we identified the superspreading subTPUs: 20 percent of superspreading subTPUs account for 78 percent of non-household infections (**Figure 3a**). When comparing with the top 20 percent of the empirical high-risk subTPUs that were most frequently visited by cases (see *Study area and datasets*), our results identified 46 percent of them (**Figure 3b**). The top 20 percent of observed subTPUs gathered 83 percent of the visits by cases, which shows a similar nonlinear relationship with our results (**Figure 3c**).

The spatial distribution of simulated superspreading subTPUs is similar to the observed patterns (**Figure 3b, d**). They both confirmed that the financial and commercial centers and the most densely residential areas were highly risky. The differences between the simulated and observed patterns are that the simulated results include more high-risk subTPUs around the populated areas but exclude some high-risk subTPUs in the financial and commercial centers. Notably, the observed patterns include all visits of cases, but only a minority of them would lead to infections. Therefore, they may overestimate the

risk degree of some subTPUs, especially subTPUs that gather a large volume of visits, such as some in the financial and commercial centers.

The effects of urban scaling structure on transmission risks for subTPUs

To understand heterogenous transmission risks across subTPUs, multivariate linear regression was used to estimate the effects of key components, including spatial structure, mobility density, and the structure of social network for each subTPU. The transmission risk for subTPU was measured by the proportion of transmission contacts that successfully transmit the virus among all contacts. The mobility density refers to the number of visiting agents per unit of subTPU area. The social network is generated by the process described in *Contact behaviours* in *Methods* section. As the successful transmission is a chance event, we took 30-run average as the reported result to guarantee a more reliable estimation. After controlling mobility densities and clustering coefficients of social networks for 4,863 subTPUs, the scaling index still have an important effect on the transmission risk: a 1-percent-point increase in the scaling index leads to 2.54 percent increase in transmission risk on average (**Table 2**).

The effects of urban scaling structure on individual's probability of becoming a superspreader

As subTPUs with larger scaling index have higher transmission risks, individuals who visits these subTPUs may have higher chance to infect more people and become superspreaders. The top 10 percent of agents, ranked by the number of secondary cases, were classified as superspreaders, and these superspreaders accounted for on average 77.3 percent (95% CI: 76.5%-78.6%) of infections across 30 simulations. Logistic regressions were used to quantify the average effect of spatial structure on individual's probability of becoming a superspreader. After controlling the number of daily contacts, the average

scaling index has a large influence on the probability of being superspreaders: A 1-percent-point increase in the average scaling index enhances 3.94 percent of the log odds ratio for individual becoming a superspreader, suggesting that the probability of being a superspreader would be increased by 4 percent (**Table 3**).

The unevenly distributed strain on local hospitals

Because the infections are unevenly distributed, the consequent hospital strain also expected to vary across space. To evaluate the local strain, we calculated the ratios of the simulated demands to available local capacities for ED visits and hospital beds (see *Protection behaviours* in *Methods*). During the peak period of the pandemic (from 1 March to 15 March 2022), all hospitals were heavily hit, but the degree of hospital strain varies (**Figure 4**). In 17 hospitals with ED services, three of them experienced the most drastic surge in ED visits that excessed 350 percent of their capacities, and the rest of the hospitals mainly required 150 percent to 300 percent of current capabilities to cope with local ED demands (**Figure 4a**). To show the hospital strain on space, we take the service area (TPUs) of each hospital as a group and show the average hospital strain on each group. TPUs in the central of the city experienced severe ED strain, and remote TPUs in the west and north were slightly better (**Figure 4a**).

The strain on hospital bed capacity was also unevenly distributed across seven COVID-19 designated hospitals (**Figure 4b**). Three in seven hospitals required over 300 percent of their current hospital bed capacities, but one hospital only used 18 percent of hospital beds, during the peak period. Remote TPUs in the west and east of the city experienced severe strain on hospital bed capacity, and the strain for the central TPUs was slightly relieved. The Island District in the southwest corner of the city was an exception, and the demand was much less than the local capacity.

Discussion

This study employed the spatially explicit ABM to reconstruct the transmission processes at the fine spatial scale. After validating by city-level to subTPU-level empirical data, further statistical analyses of model results reveal that not only the quantity of mobility but also the scaling structure of mobility have important impacts on local transmission risk: areas with low scaling indices could be less likely to transmit virus, even when they have a rather large mobility density or a clustering social network; similarly, individuals who tend to visit low-scaling-index areas could be less likely to transmit the virus to others, even when they have a large number of contacts. These findings suggest urban scaling structure may play an important role in the mechanisms driving the heterogeneity in local transmission risks among urban areas and individuals.

How urban scaling structure influences local transmission risks

It is widely acknowledged that high population or mobility density would increase the local transmission risk (Alessandretti 2022; Damme et al. 2020; Hong et al. 2021; Wong and Li 2020). However, our findings suggest areas with a rather high mobility density may not necessarily be more likely to transmit virus, especially when they have low scaling indices. The scaling index measures the relative importance of different areas in the entire structure based on the links pointing to them and the links of those links through an iterative process (see *Urban scaling structure, the scaling index, and mobility behaviours* in *Methods*). Due to the nature of the scaling law (Clauset, Shalizi, and Newman 2009; Jiang 2009; Jiang, Yin, and Zhao 2009), only a minority of areas are of great importance in this structure, and they would be more likely to rapidly import the virus from the initial infections occurred in elsewhere. In this case, the high-scaling-index areas take one step ahead to infect susceptible people first, leaving little chance for most

low-scaling-index areas to transmit the virus, as most of their visitors would have already been infected elsewhere.

Therefore, the scaling index could be a potential risk factor for local transmission. Different from the local population or mobility density, it highlights the urban structure's impact on local risk. Urban scaling structure offers a minority of areas the 'first-mover advantage', allowing them to initiate infections earlier and on a larger scale. In this regard, our study provides a deeper understanding of superspreading places or Points of Interests (POIs): while large population or mobility may be necessary, they are not sufficient conditions for superspreading places. The opportunity for early access to the virus brought by urban scaling structure also matters.

How urban scaling structure contributes to superspreaders

The heterogeneity in local transmission risk would further influence individuals' chance of having more infections and becoming superspreaders. In our model, we assumed that all individuals have an equal probability of successfully transmitting the virus per contact. Therefore, infected people with large number of contacts are more likely to infect more susceptible people and thus become superspreaders. However, our analyses found that where those infected people visited also matters. Individuals tends to visit locations with large scaling indices are more likely to become superspreaders. This suggests that when individuals are embedded in areas with low scaling indices, even those with large number of daily contacts may not necessarily become superspreaders. The reason could be that these individuals fail to infect their contactees when competing with others at high-scaling-index locations who also want to infect the same contacts. The high scaling index offers others an advantage in infecting the same contactee first, as the large value of scaling index increased the local transmission risk. Therefore, individuals at low-scaling-

index locations encounter fewer susceptible people to infect, reducing their likelihood of becoming superspreaders.

In contrast to our findings, some researchers posit that it is the high infectiousness of certain individuals that causes superspreaders (Sidik 2023; Zhou et al. 2023). They think infectiousness is highly unequal distributed among infected cases. However, our results demonstrate that even with the same level of infectiousness, superspreaders are naturally emerged from the interaction of social contacts, in which the location also play an important role. Nevertheless, we conducted a sensitive analyse to consider the heterogeneity in infectiousness and found that the urban scaling structure is a still important risk factor for individuals becoming superspreaders (see *Sensitivity analyses on the effect of heterogeneous infectiousness on superspreader results* in *Supplementary Materials*). Therefore, superspreaders could be originated from a complex interplay of both individual, social, and environmental factors. To comprehensively quantify these intricate interactions, more individual-level viral and behaviour data is needed in future studies.

Theoretical implications from urban science

From the perspective of complexity theory, cities are viewed as complex systems that are constructed from the bottom up in a hierarchical way in which the basic components of cities and their interactions determine the complex networks of cities (Batty 2013). The new perspective emphasises that place is a synthesis of interactions, and it is the networks between places, rather than the intrinsic attributes of places, that influence our understanding of cities. From this point of view, our study constructed the subTPU network to represent interactions between places. This network provides a dynamic and structural view to show how places, mobility, and transmission that depend on the

network that may change and evolve in time by adapting to and differentiating from other places (Jiang 2018). Through this view, we gain a deeper understanding of how local transmission might influence other places' transmission, compared to the traditional static and localized view that emphasizes more on the current attributes of places, namely, the population/mobility densities of places.

By thinking the whole transmission from the bottom up, we found that the scaling properties of superspreading places and superspreaders were emerged from local competition processes in which places and infected cases competed with each other to infect more susceptible people. In a broader sense, many other scaling relationships of urban quantities also reflects the way competition determines the sizes (values) of different places within cities (Batty 2013). For example, interactions between cities in terms of trade or migration, and within cities in terms of shopping and other social movements all exhibit similar scaling properties (Batty 2008). Therefore, the way of thinking and modelling from the bottom up in our study could also be applicable to understand many other urban phenomena.

Practical suggestions on hospital emergency preparedness

The uneven distribution of infections and medical resources gives rise to the heterogeneous strain on the existing healthcare systems (see *Protection behaviours* in *Methods*). After considering the high spatial heterogeneity in hospital strain, we may provide some qualitative suggestions on hospital emergency preparedness to adequately staff local hospitals in advance and wisely share the existing resources. The specific suggestions are as follows:

Firstly, to relieve the strain on local EDs, triage centers are needed in Sha Tin District in the middle of the city and Wan Chai District in Hong Kong Island. During the

fifth wave in Hong Kong, many patients with mild symptoms occupied the ED resources that should have been used to treat those with severe symptoms (A. Ma and Parry 2022). The triage centers in the most severe areas can efficiently filter out a large number of non-urgent patients and identify the critical patients for further ED services. Secondly, more hospital beds should be assigned for local COVID-19 patients in Tin Shui Wai and Tuen Mun New Towns in the west of the city. Thirdly, transferring the patients to the hospital in Island District is another practical suggestion as there would be relatively lower pressure even at the peak of the pandemic. The hospital in Island District (North Lantau Hospital Hong Kong Infection Control Centre) started the services special for COVID-19 patients in 2021 and has over 800 isolation beds, so it is capable to take patients from other hospitals and regions.

Limitations of this study

This study has several limitations from two aspects. The first aspect is the study design. We only chose the fifth wave of the pandemic in Hong Kong to verify the validity of the proposed model. More experiments on different periods and cities could be conducted in future studies to further validate our model if more data can be collected. The first four waves in Hong Kong were excluded because these waves had limited infections (in total 12,258 cases by Dec. 20, 2021), and such a small number of infections might cover up the local-level findings.

The second aspect is the model structure. Firstly, we assume that all contacts have an equal probability of successfully transmitting the virus irrespective of the contact duration and the contact intensity, which is a simplification of reality. If more biological data is available, further study can be conducted to thoroughly examine how the infectiousness per contact may influence the local-scale transmission. Another

simplification is the number of contacts at different places. For an agent, the total number of contacts is evenly distributed across its travel destinations regardless of the activity type. In reality, certain types of activities may lead to more contacts than others. If such kind of data is available in the future, we can further determine how the social contact characteristics may influence local transmission.

Lastly, we did not explicitly consider the spatial heterogeneity in the vaccine-acquired immunity, but our model considered the vaccination by different age groups and the age groups were spatially heterogenous distributed, through which our model may to some extent accounted for the spatial heterogeneity in vaccination. What's more, only a small proportion (12 percent) of population had taken the booster vaccination by Jan. 31st, 2021(DATA.GOV.HK 2023), and most of population that took the primary vaccination had limited protective effect for the Omicron (Andrews et al. 2022). Therefore, we infer that the spatial heterogeneity in vaccine may have limited effects on our results. If we can gather related empirical data, more realistic situations can be considered in the future study.

Conclusion

To gain a deeper understanding of the underlying mechanisms by which urban scaling structure of human mobility contributes to local transmission risks of COVID-19 within cities, this study introduced a spatially explicit ABM which incorporated urban scaling structure to reconstruct fine-grained mobility patterns and transmission processes. After validating by empirical data at various spatial scales, further statistical analyses of simulation outcome reveal that not only the quantity of mobility but also its scaling structure have important impacts on local transmission risks among urban areas and individuals. For urban areas, a large volume of local mobility is only a prerequisite for

high transmission risk, and their significance within the urban scaling structure also plays a crucial regulatory role in it. Consequently, the resulted heterogenous risks among urban areas would further influence the transmission potential of their visitors becoming superspreaders. In summary, urban scaling structure may provide the 'first-mover advantage' to a small group of urban areas and individuals, enabling them to initiate infections earlier and on a more substantial scale. This study thus brings important insights for the transmission dynamics of COVID-19 and similar diseases, highlighting the role of urban scaling structure in driving the heterogeneity in local transmission risks and superspreading events. These insights may serve as valuable guidance for the development of precise and effective interventions to mitigate future pandemics.

Acknowledgments

The authors express sincere appreciation to Prof. Bian Ling and the anonymous reviewers for their invaluable assistance and constructive comments. Our gratitude extends to Prof. Bin Jiang for inspiring this paper with his influential work. We are thankful for Prof. Xun Shi's valuable advice provided during CPGIS 2023.

References

Adam, Dillon C., Peng Wu, Jessica Y. Wong, Eric H. Y. Lau, Tim K. Tsang, Simon Cauchemez, Gabriel M. Leung, and Benjamin J. Cowling. 2020. Clustering and Superspreading Potential of SARS-CoV-2 Infections in Hong Kong. *Nature Medicine* 26 (11):1714–1719. doi:10.1038/s41591-020-1092-0.

Aguilar, Javier, Aleix Bassolas, Gourab Ghoshal, Surendra Hazarie, Alec Kirkley, Mattia Mazzoli, Sandro Meloni, Sayat Mimar, Vincenzo Nicosia, José J. Ramasco, et al. 2022. Impact of Urban Structure on Infectious Disease Spreading. *Scientific Reports* 12 (1). Nature Publishing Group:3816. doi:10.1038/s41598-022-06720-8.

Ahmed, Faheem, Na'eem Ahmed, Christopher Pissarides, and Joseph Stiglitz. 2020. Why Inequality Could Spread COVID-19. *The Lancet Public Health* 5 (5). Elsevier:e240. doi:10.1016/S2468-2667(20)30085-2.

Alessandretti, Laura. 2022. What Human Mobility Data Tell Us about COVID-19 Spread. *Nature Reviews Physics* 4 (1). Nature Publishing Group:12–13. doi:10.1038/s42254-021-00407-1.

Andrews, Nick, Julia Stowe, Freja Kirsebom, Samuel Toffa, Tim Rickeard, Eileen Gallagher, Charlotte Gower, Meaghan Kall, Natalie Groves, Anne-Marie O'Connell, et al. 2022. Covid-19 Vaccine Effectiveness against the Omicron

(B.1.1.529) Variant. *New England Journal of Medicine* 386 (16). Massachusetts Medical Society:1532–1546. doi:10.1056/NEJMoa2119451.

Badr, Hamada S, Hongru Du, Maximilian Marshall, Ensheng Dong, Marietta M Squire, and Lauren M Gardner. 2020. Association between Mobility Patterns and COVID-19 Transmission in the USA: A Mathematical Modelling Study. *The Lancet Infectious Diseases* 20 (11):1247–1254. doi:10.1016/S1473-3099(20)30553-3.

Batty, Michael. 2008. The Size, Scale, and Shape of Cities. *Science* 319 (5864). American Association for the Advancement of Science:769–771. doi:10.1126/science.1151419.

Batty, Michael. 2013. *The New Science of Cities*. MIT Press.

Bettencourt, Luís M. A. 2013. The Origins of Scaling in Cities. *Science* 340 (6139). American Association for the Advancement of Science:1438–1441. doi:10.1126/science.1235823.

Bian, Ling, Yuxia Huang, Liang Mao, Eunjung Lim, Gyoungju Lee, Yan Yang, Murray Cohen, and Deborah Wilson. 2012. Modeling Individual Vulnerability to Communicable Diseases: A Framework and Design. *Annals of the Association of American Geographers* 102 (5). Routledge:1016–1025. doi:10.1080/00045608.2012.674844.

Blundell, Richard, Monica Costa Dias, Jonathan Cribb, Robert Joyce, Tom Waters, Thomas Wernham, and Xiaowei Xu. 2022. Inequality and the COVID-19 Crisis in the United Kingdom. *Annual Review of Economics* 14 (1):607–636. doi:10.1146/annurev-economics-051520-030252.

Brin, Sergey, and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems* 30 (1–7). Elsevier:107–117. doi:10.1016/S0169-7552(98)00110-X.

Brockmann, D., L. Hufnagel, and T. Geisel. 2006. The Scaling Laws of Human Travel. *Nature* 439 (7075). Nature Publishing Group:462–465. doi:10.1038/nature04292.

Cai, Jun, Xiaowei Deng, Juan Yang, Kaiyuan Sun, Hengcong Liu, Zhiyuan Chen, Cheng Peng, Xinhua Chen, Qianhui Wu, Junyi Zou, et al. 2022. Modeling Transmission of SARS-CoV-2 Omicron in China. *Nature Medicine*, May. Nature Publishing Group, 1–8. doi:10.1038/s41591-022-01855-7.

Chang, Serina, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. *Nature* 589 (7840):82–87. doi:10.1038/s41586-020-2923-3.

Cheng, Vincent Chi-Chung, Shuk-Ching Wong, Vivien Wai-Man Chuang, Simon Yung-Chun So, Jonathan Hon-Kwan Chen, Siddharth Sridhar, Kelvin Kai-Wang To, Jasper Fuk-Woo Chan, Ivan Fan-Ngai Hung, Pak-Leung Ho, et al. 2020. The Role of Community-Wide Wearing of Face Mask for Control of Coronavirus Disease 2019 (COVID-19) Epidemic Due to SARS-CoV-2. *Journal of Infection* 81 (1):107–114. doi:10.1016/j.jinf.2020.04.024.

Chowkwanyun, Merlin, and Adolph L. Reed. 2020. Racial Health Disparities and Covid-19 — Caution and Context. *New England Journal of Medicine* 383 (3). Massachusetts Medical Society:201–203. doi:10.1056/NEJMp2012910.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. Power-Law Distributions in Empirical Data. *SIAM Review* 51 (4). Society for Industrial and Applied Mathematics:661–703. doi:10.1137/070710111.

Cot, Corentin, Giacomo Cacciapaglia, and Francesco Sannino. 2021. Mining Google and Apple Mobility Data: Temporal Anatomy for COVID-19 Social Distancing.

Scientific Reports 11 (1). Nature Publishing Group:4150. doi:10.1038/s41598-021-83441-4.

Damme, Wim Van, Ritwik Dahake, Alexandre Delamou, Brecht Ingelbeen, Edwin Wouters, Guido Vanham, Remco van de Pas, Jean-Paul Dossou, Por Ir, Seye Abimbola, et al. 2020. The COVID-19 Pandemic: Diverse Contexts; Different Epidemics—How and Why? *BMJ Global Health* 5 (7). BMJ Specialist Journals:e003098. doi:10.1136/bmjgh-2020-003098.

DATA.GOV.HK. 2023. Daily Count of Vaccination by Age Groups. Accessed July 10. https://data.gov.hk/en-data/dataset/hk-hhb-hhbcovid19-vaccination-rates-over-time-by-age.

Franch-Pardo, Ivan, Brian M. Napoletano, Fernando Rosete-Verges, and Lawal Billa. 2020. Spatial Analysis and GIS in the Study of COVID-19. A Review. *Science of The Total Environment* 739 (October):140033. doi:10.1016/j.scitotenv.2020.140033.

Giles, John, and Amy Wesolowski. 2023. Mobility: An R Package for Modeling Human Mobility Patterns. Accessed September 13. https://covid-19-mobility-data-network.github.io/mobility/index.html.

Google LLC. 2023. Google COVID-19 Community Mobility Reports. Accessed March 11. https://www.google.com/covid19/mobility/data_documentation.html?hl=en.

HKUMed. 2022. HKUMed Proposes Forward Planning after Hong Kong's Fifth Wave of Omicron BA.2. https://sph.hku.hk/en/News-And-Events/Press-Releases/2022/HKUMed-proposes-forward-planning-after-Hong-Kong.

Hong, Boyeong, Bartosz J. Bonczak, Arpit Gupta, Lorna E. Thorpe, and Constantine E. Kontokosta. 2021. Exposure Density and Neighborhood Disparities in COVID-19 Infection Risk. *Proceedings of the National Academy of Sciences* 118 (13). Proceedings of the National Academy of Sciences:e2021258118. doi:10.1073/pnas.2021258118.

Hong Kong Government. 2023a. Declaration System for Individuals Tested Positive for COVID-19 Using Rapid Antigen Test Launched. Accessed March 11. https://www.info.gov.hk/gia/general/202203/07/P2022030700768.htm.

Hong Kong Government. 2023b. Archives of List of Buildings with Confirmed / Probable Cases of COVID-19. Accessed March 11. https://www.chp.gov.hk/en/features/102991.html.

Hong Kong Government. 2023c. Hospital Authority Data Sharing Portal. Accessed March 11. https://www3.ha.org.hk/Data/HAStatistics/MajorReport.

Huang, Jianwei, and Mei-Po Kwan. 2021. Uncertainties in the Assessment of COVID-19 Risk: A Study of People's Exposure to High-Risk Environments Using Individual-Level Activity Data. *Annals of the American Association of Geographers* 0 (0). Taylor & Francis:1–20. doi:10.1080/24694452.2021.1943301.

Iuliano, A. Danielle, Joan M. Brunkard, Tegan K. Boehmer, Elisha Peterson, Stacey Adjei, Alison M. Binder, Stacy Cobb, Philip Graff, Pauline Hidalgo, Mark J. Panaggio, et al. 2022. Trends in Disease Severity and Health Care Utilization During the Early Omicron Variant Period Compared with Previous SARS-CoV-2 High Transmission Periods — United States, December 2020–January 2022. *MMWR. Morbidity and Mortality Weekly Report* 71 (4):146–152. doi:10.15585/mmwr.mm7104e4.

Jiang, Bin. 2009. Street Hierarchies: A Minority of Streets Account for a Majority of Traffic Flow. *International Journal of Geographical Information Science* 23 (8). Taylor & Francis:1033–1048. doi:10.1080/13658810802004648.

Jiang, Bin. 2013. Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *The Professional Geographer* 65 (3). Routledge:482–494. doi:10.1080/00330124.2012.700499.

Jiang, Bin. 2018. A Topological Representation for Taking Cities as a Coherent Whole. *Geographical Analysis* 50 (3):298–313. doi:10.1111/gean.12145.

Jiang, Bin, and Tao Jia. 2010. Agent-Based Simulation of Human Movement Shaped by the Underlying Street Structure. *arXiv:0910.3055 [Physics]*, April. http://arxiv.org/abs/0910.3055.

Jiang, Bin, and Xintao Liu. 2012. Scaling of Geographic Space from the Perspective of City and Field Blocks and Using Volunteered Geographic Information. *International Journal of Geographical Information Science* 26 (2). Taylor & Francis:215–229. doi:10.1080/13658816.2011.575074.

Jiang, Bin, Junjun Yin, and Sijian Zhao. 2009. Characterizing the Human Mobility Pattern in a Large Street Network. *Physical Review E* 80 (2). American Physical Society:021136. doi:10.1103/PhysRevE.80.021136.

Kan, Zihan, Mei-Po Kwan, Jianwei Huang, Jiannan Cai, and Dong Liu. 2023. A Spatial Network-Based Assessment of Individual Exposure to COVID-19. *Annals of the American Association of Geographers* 0 (0). Taylor & Francis:1–11. doi:10.1080/24694452.2023.2266021.

Keni, Raghuvir, Anila Alexander, Pawan Ganesh Nayak, Jayesh Mudgal, and Krishnadas Nandakumar. 2020. COVID-19: Emergence, Spread, Possible Treatments, and Global Burden. *Frontiers in Public Health* 8. doi:10.3389/fpubh.2020.00216.

Kirsebom, Freja C. M., Nick Andrews, Julia Stowe, Samuel Toffa, Ruchira Sachdeva, Eileen Gallagher, Natalie Groves, Anne-Marie O'Connell, Meera Chand, Mary Ramsay, et al. 2022. COVID-19 Vaccine Effectiveness against the Omicron (BA.2) Variant in England. *The Lancet Infectious Diseases* 22 (7). Elsevier:931–933. doi:10.1016/S1473-3099(22)00309-7.

Kogan, Nicole E., Leonardo Clemente, Parker Liautaud, Justin Kaashoek, Nicholas B. Link, Andre T. Nguyen, Fred S. Lu, Peter Huybers, Bernd Resch, Clemens Havas, et al. 2021. An Early Warning Approach to Monitor COVID-19 Activity with Multiple Digital Traces in near Real Time. *Science Advances* 7 (10). American Association for the Advancement of Science:eabd6989. doi:10.1126/sciadv.abd6989.

Koks, Sulev, Robert W Williams, John Quinn, Farzin Farzaneh, Nicola Conran, Shaw-Jeng Tsai, Gordon Awandare, and Steven R Goodman. 2020. COVID-19: Time for Precision Epidemiology. *Experimental Biology and Medicine* 245 (8). SAGE Publications:677–679. doi:10.1177/1535370220919349.

Kucharski, Adam J., Kin O. Kwok, Vivian W. I. Wei, Benjamin J. Cowling, Jonathan M. Read, Justin Lessler, Derek A. Cummings, and Steven Riley. 2014. The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. *PLOS Pathogens* 10 (6). Public Library of Science:e1004206. doi:10.1371/journal.ppat.1004206.

Lau, Max S. Y., Bryan Grenfell, Michael Thomas, Michael Bryan, Kristin Nelson, and Ben Lopman. 2020. Characterizing Superspreading Events and Age-Specific Infectiousness of SARS-CoV-2 Transmission in Georgia, USA. *Proceedings of the National Academy of Sciences* 117 (36). Proceedings of the National Academy of Sciences:22430–22435. doi:10.1073/pnas.2011802117.

Levin, Andrew T., Nana Owusu-Boaitey, Sierra Pugh, Bailey K. Fosdick, Anthony B. Zwi, Anup Malani, Satej Soman, Lonni Besançon, Ilya Kashnitsky, Sachin Ganesh, et al. 2022. Assessing the Burden of COVID-19 in Developing Countries:

Systematic Review, Meta-Analysis and Public Policy Implications. *BMJ Global Health* 7 (5). BMJ Specialist Journals:e008477. doi:10.1136/bmjgh-2022-008477.

Levin, Roman, Dennis L. Chao, Edward A. Wenger, and Joshua L. Proctor. 2021. Insights into Population Behavior during the COVID-19 Pandemic from Cell Phone Mobility Data and Manifold Learning. *Nature Computational Science* 1 (9):588–597. doi:10.1038/s43588-021-00125-9.

Lewis, Dyani. 2021. Superspreading Drives the COVID Pandemic — and Could Help to Tame It. *Nature* 590 (7847):544–546. doi:10.1038/d41586-021-00460-x.

Lima, L. L., and A. P. F. Atman. 2021. Impact of Mobility Restriction in COVID-19 Superspreading Events Using Agent-Based Model. *PLOS ONE* 16 (3). Public Library of Science:e0248708. doi:10.1371/journal.pone.0248708.

Ma, Arisina, and Jane Parry. 2022. When Hong Kong's "Dynamic Zero" Covid-19 Strategy Met Omicron, Low Vaccination Rates Sent Deaths Soaring. *BMJ* 377 (April). British Medical Journal Publishing Group:o980. doi:10.1136/bmj.o980.

Ma, Ding, Toshihiro Osaragi, Takuya Oki, and Bin Jiang. 2020. Exploring the Heterogeneity of Human Urban Movements Using Geo-Tagged Tweets. *International Journal of Geographical Information Science* 34 (12). Taylor & Francis:2475–2496. doi:10.1080/13658816.2020.1718153.

Majra, Dasha, Jayme Benson, Jennifer Pitts, and Justin Stebbing. 2021. SARS-CoV-2 (COVID-19) Superspreader Events. *Journal of Infection* 82 (1):36–40. doi:10.1016/j.jinf.2020.11.021.

Manica, Mattia, Alfredo De Bellis, Giorgio Guzzetta, Pamela Mancuso, Massimo Vicentini, Francesco Venturelli, Alessandro Zerbini, Eufemia Bisaccia, Maria Litvinova, Francesco Menegale, et al. 2022. Intrinsic Generation Time of the SARS-CoV-2 Omicron Variant: An Observational Study of Household Transmission. *The Lancet Regional Health – Europe* 19 (August). Elsevier. doi:10.1016/j.lanepe.2022.100446.

Nouvellet, Pierre, Sangeeta Bhatia, Anne Cori, Kylie E. C. Ainslie, Marc Baguelin, Samir Bhatt, Adhiratha Boonyasiri, Nicholas F. Brazeau, Lorenzo Cattarino, Laura V. Cooper, et al. 2021. Reduction in Mobility and COVID-19 Transmission. *Nature Communications* 12 (1):1090. doi:10.1038/s41467-021-21358-2.

Paul, Elisabeth, Garrett W. Brown, and Valery Ridde. 2020. COVID-19: Time for Paradigm Shift in the Nexus between Local, National and Global Health. *BMJ Global Health* 5 (4). BMJ Specialist Journals:e002622. doi:10.1136/bmjgh-2020-002622.

Rasmussen, Sonja A., Muin J. Khoury, and Carlos del Rio. 2020. Precision Public Health as a Key Tool in the COVID-19 Response. *JAMA* 324 (10):933–934. doi:10.1001/jama.2020.14992.

Schläpfer, Markus, Luís M. A. Bettencourt, Sébastian Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B. West, and Carlo Ratti. 2014. The Scaling of Human Interactions with City Size. *Journal of The Royal Society Interface* 11 (98). Royal Society:20130789. doi:10.1098/rsif.2013.0789.

Schläpfer, Markus, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B. West. 2021. The Universal Visitation Law of Human Mobility. *Nature* 593 (7860):522–527. doi:10.1038/s41586-021-03480-9.

Sidik, Saima. 2023. What Makes a COVID Superspreader? Scientists Learn More after Deliberately Infecting Volunteers. *Nature*, June. doi:10.1038/d41586-023-01961-7.

881 Templ, Matthias, Bernhard Meindl, Alexander Kowarik, and Olivier Dupriez. 2017.
882 Simulation of Synthetic Complex Data: The R Package simPop. *Journal of*
883 *Statistical Software* 79 (August):1–38. doi:10.18637/jss.v079.i10.
884 Tizzoni, Michele, Kaiyuan Sun, Diego Benusiglio, Márton Karsai, and Nicola Perra.
885 2015. The Scaling of Human Contacts and Epidemic Processes in Metapopulation
886 Networks. *Scientific Reports* 5 (1). Nature Publishing Group:15111.
887 doi:10.1038/srep15111.
888 Wong, David W. S. 1992. The Reliability of Using the Iterative Proportional Fitting
889 Procedure*. *The Professional Geographer* 44 (3):340–348. doi:10.1111/j.0033-
890 0124.1992.00340.x.
891 Wong, David W. S., and Yun Li. 2020. Spreading of COVID-19: Density Matters. *PLOS*
892 *ONE* 15 (12). Public Library of Science:e0242398.
893 doi:10.1371/journal.pone.0242398.
894 Xi, Jinglun, Xiaolu Liu, Jianghao Wang, Ling Yao, and Chenghu Zhou. 2023. A
895 Systematic Review of COVID-19 Geographical Research: Machine Learning and
896 Bibliometric Approach. *Annals of the American Association of Geographers* 113
897 (3). Taylor & Francis:581–598. doi:10.1080/24694452.2022.2130143.
898 Xing, W., and A. Ghorbani. 2004. Weighted PageRank Algorithm. In *Proceedings.*
899 *Second Annual Conference on Communication Networks and Services Research,*
900 *2004.*, 305–314. doi:10.1109/DNSR.2004.1344743.
901 Xu, Chen, Libao Jin, and Long Lee. 2023. An Empirical Spatial Network Model Based
902 on Human Mobility for Epidemiological Research: A Case Study. *Annals of the*
903 *American Association of Geographers* 113 (6). Taylor & Francis:1461–1482.
904 doi:10.1080/24694452.2023.2187339.
905 Yancy, Clyde W. 2020. COVID-19 and African Americans. *JAMA* 323 (19):1891–1892.
906 doi:10.1001/jama.2020.6548.
907 Yang, Hao, X. Angela Yao, Ruowei Liu, and Christopher C. Whalen. 2023. Developing
908 a Place–Time-Specific Transmissibility Index to Measure and Examine the
909 Spatiotemporally Varying Transmissibility of COVID-19. *Annals of the American*
910 *Association of Geographers* 113 (6). Taylor & Francis:1419–1443.
911 doi:10.1080/24694452.2023.2182758.
912 Yilmazkuday, Hakan. 2021. Stay-at-Home Works to Fight against COVID-19:
913 International Evidence from Google Mobility Data. *Journal of Human Behavior*
914 *in the Social Environment* 31 (1–4). Routledge:210–220.
915 doi:10.1080/10911359.2020.1845903.
916 Zhang, Nan, Wei Jia, Peihua Wang, Chung-Hin Dung, Pengcheng Zhao, Kathy Leung,
917 Boni Su, Reynold Cheng, and Yuguo Li. 2021. Changes in Local Travel
918 Behaviour before and during the COVID-19 Pandemic in Hong Kong. *Cities* 112
919 (May):103139. doi:10.1016/j.cities.2021.103139.
920 Zhang, Wenjia, and Kexin Ning. 2023. Spatiotemporal Heterogeneities in the Causal
921 Effects of Mobility Intervention Policies during the COVID-19 Outbreak: A
922 Spatially Interrupted Time-Series (SITS) Analysis. *Annals of the American*
923 *Association of Geographers* 113 (5). Taylor & Francis:1112–1134.
924 doi:10.1080/24694452.2022.2161986.
925 Zhou, Jie, Anika Singanayagam, Niluka Goonawardane, Maya Moshe, Fiachra P.
926 Sweeney, Ksenia Sukhova, Ben Killingley, Mariya Kalinova, Alex J. Mann,
927 Andrew P. Catchpole, et al. 2023. Viral Emissions into the Air and Environment
928 after SARS-CoV-2 Human Challenge: A Phase 1, Open Label, First-in-Human
929 Study. *The Lancet Microbe* 0 (0). Elsevier. doi:10.1016/S2666-5247(23)00101-5.

930 **Author biographies**

931 NINGYEZI PENG is a PhD student in the Department of Land Surveying and Geo-
932 Informatics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail:
933 ningyezipeng@hotmail.com. Her research interests include GIScience, complexity
934 science, and environmental health.

935

936 XINTAO LIU is an Associate Professor in the Department of Land Surveying and Geo-
937 Informatics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail:
938 xintao.liu@polyu.edu.hk. His research interest focuses on GIScience, transportation
939 geography, and complex network. He is particularly interested in integrating advances in
940 machine learning into human mobility modeling, human-environment interactions
941 analysis and human behavior analysis across physical and virtual spaces.

942

943 **List of figure captions**

944 Figure 1. Model description of the spatially explicit agent-based model. **a**, synthetic
945 population was generated based on census data and household survey data using iterative
946 proportional fitting, which captures the household structures and age structures among
947 real population across 1,622 census tracts. **b**, synthetic mobility flows were first simulated
948 at the TPU level using the departure-diffusion model, and then were refined the
949 destinations of flows from 214 tertiary planning units (TPUs) to 4,863 subunits of TPUs
950 (subTPUs) using the urban scaling structure, captured by the subTPU network built based
951 on travel survey data. **c**, the two-layer contact structure consists of household contacts
952 and non-household contacts during trips. At visited subTPUs, agents would have contacts
953 with certain number of agents, and the certain number was sampled from the age-
954 dependent distributions for Hong Kong population (**d**).

955

Figure 2. Model fitting results. **a**, epidemic curves. The left plot is the city-level prediction, and the right plot is the TPU-level prediction on the eight worst affected TPUs. The grey bar is the 7-day moving average of observed confirmed cases. Shaded regions denote the 2.5th and 97.5th percentiles across selected parameter sets and stochastic realizations. **b**, the spatial distribution of the total number of observed RAT cases (left) and simulated cases averaged across selected parameter sets and stochastic realizations (right).

Figure 3. The simulated superspreading subTPUs and the observed high-risk subTPUs. **a**, the top 20 percent of simulated subTPUs (or superspreading subTPUs), ranked by the average number of infections occurred in subTPUs across 30 runs, accounts for 78 percent of the average number of simulated infections. **b**, the spatial distribution of simulated superspreading subTPUs. **c**, the top 20 percent of observed subTPUs, ranked by the total number of visits by infected cases occurred in subTPUs, account for 83 percent of visits. **d**, the spatial distribution of the observed high-risk subTPUs.

Figure 4. Simulated hospital strain on the Emergency Department (ED) services (**a**) and hospital beds (**b**). The strain on each hospital was estimated by the ratio of simulated local demand to available hospital capacity. In **a** and **b**, Left, the hospital strain varied by hospital and time. Right, the average hospital strain during the peak period (from 1 March to 15 March 2022) varied across TPUs.

The impact of urban scaling structure on the local-scale transmission of COVID-19: A case study of the Omicron wave in Hong Kong using agent-based modelling

Table 1. Individual attributes, values, and sources

Attributes	Value	Source
Demographics and households		
Identify of individual	Agent ID	Simulated
Age	Twelve age groups ([0,15), [15,20), [20,25), [25,30), [30,35), [35,40), [40,45), [45,50), [50,55), [55,60), [60,65), 65+)	(2016 HK Census)
Sex	Male, female	(2016 HK Census)
Identify of household	Household ID	Simulated
Household location	Census tract ID	(2016 HK Census)
Travel behaviors		
Trip destinations	subTPU IDs	Derived from HK travel survey data
Probability of cancelling all trips (stay at home) when being infected	50%	(HKUMed 2022)
Contact behaviors		
The total number of total contacts	17.5 (the mean value)	Derived from HK contact survey data (Kucharski et al. 2014)
Number of contacts at home	Decided by household size	Simulated
Number of contacts outside home	Decided by the number of total contacts and household size	Simulated
Infection attributes		
Infection probability per contact	0.05	Estimated
Probability of being initial latent cases	6×10^{-5}	Estimated
Latency period	1.20 (the mean value)	(Cai et al. 2022)
Infectious Period	5.64 (the mean value)	(Cai et al. 2022; Manica et al. 2022)
Period from infectious to confirmed	5.05 (the mean value)	(Manica et al. 2022)
Probability of being confirmed when being infected	8% before 24 Feb, 30% between 24 Feb. and 7 Mar., 35% after 7 Mar.	Estimated based on HKUMed (2022)
Protection behaviors		
Vaccination information	Vaccine dose and type, inoculation date	Vaccine data (DATA.GOV.HK 2023)

Vaccination effectiveness	Decided by vaccine information (see **Table S8**)	(HKUMed 2022)
Probability of visiting Emergency Department when being infected	8.60%	(Iuliano et al. 2022)
Probability of requiring hospitalization when being infected	2.70%	(Iuliano et al. 2022)

Table 2. The effect of the scaling index on the average transmission risk (the average proportion of transmission contacts across 30 runs with the best-fit parameters) for subTPUs through multivariate linear regression.

Variable	Coefficient
The scaling index	**2.54*****
Average mobility density	0.16***
Average clustering coefficient of social network	0.09***

*** Significant result with p-value < 0.001.

Table 3. The average effects of the scaling index on individual's probability of becoming a superspreader (the top 10% of individuals ranked by the number of secondary cases) through logistic regressions across 30 runs with the best-fit parameters.

Variable	The mean of coefficients	Exponential of (coefficients*1 unit)
The average scaling index	**3.94** $^{*\wedge}$	**104% (a 1-percent-point increase)**
The number of contacts	0.03 ***	103% (a 1-point increase)

*** Significant result with p-value < 0.001; $^{*\wedge}$ 60% of results are significant with p-value < 0.05.

Supplementary Materials

The impact of urban scaling structure on the local-scale transmission of COVID-19: A case study of the Omicron wave in Hong Kong using the agent-based model

Supplementary Methods

Selecting Parameter Ranges

We assume that R_0 of Omicron ranges from 6 to 10. This is a relatively wide range, as a previous study estimated the R_0 of 8.2 (Liu and Rocklöv 2022). To determine the values of p_i that make R_0 in the assumed range, we created a well-mixed agent-based model. This is due to the definition of R_0 that requires a completely susceptible and fully mixed population. The well-mixed model does not have the immunity and mobility settings, so all agents are fully susceptible, and they randomly choose their contacts in the whole population.

Under this scenario, we initialized a proportion of index cases ($p_0 = 10^{-4}$) and recorded the number of secondary cases. Let N_0 be the number of index cases, and N_s be the number of secondary cases, which gives $R_0 = \frac{N_s}{N_0}$. We averaged these R_0 values over 30 replicates for each p_i value, and results shows that R_0 is linear in p_i (Chang et al. 2021). To allow R_0 in the plausible range, p_i should range from 0.04 to 0.07, and this range is then used in the model fitting.

It is still not clear that how many infections existed at the beginning of the simulation (1st February 2022). We set p_0 with a large range between 10^{-5} to 10^{-3} to account for the uncertainty of the real situation.

Validation on TPU-level Data

The number of cases confirmed by rapid antigen tests (RAT) at every TPU can be used to evaluate the TPU-level model accuracy. Different from the city-level data, this data only records partial confirmed cases (not including cases confirmed by the polymerase chain reaction test) and does not cover the whole period of simulation (starting from 26th February 2022). Due to its incompleteness, we do not use it to calibrate our model, but we directly use it to validate the model accuracy on the local scale. We sum the RMSEs of TPUs together to evaluate the TPU-level prediction,

$$RMSE_{TPU} = \sum_{tpu}^{TPU} \sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(N_{RAT}^{d,tpu} - \hat{N}_{RAT}^{d,\mathrm{tpu}}\right)^2} \quad (1)$$

where $\hat{N}_{RAT}^{d,tpu}$ is the observed RAT-confirmed cases in a tpu on a day d, and $N_{RAT}^{d,tpu}$ is the corresponding value in our simulation. As our model predict on all confirmed cases, we assume that a $p_{RAT} = 0.4$ proportion of cases are confirmed by RAT, $N_{RAT}^{d,tpu} = p_{RAT} \times N_{confirmed}^{d,tpu}$. p_{RAT} is determined by the proportion of the total number of RAT-confirmed cases in the total number of observed confirmed cases. The time series of this validation also spans from 1st March to 30th March 2022.

Supplementary Results

Comparison of 2011 travel survey data with 2020 mobility data

To verify the representativeness of 2011 travel survey data, we compared the mobility structure of 2011 travel survey data with that of 2020 subway data in Hong Kong as presented in Zhang et al. (2021), the most recent figure available. Using the Louvain

heuristics algorithm, as employed in Zhang et al. (2021), to detect communities in complex networks, we found consistent community structures in both 2011 travel survey and the latest 2020 data (**Figure S1**). This indicates a high degree of temporal and spatial regularity in human mobility patterns (Gonzalez, Hidalgo, and Barabasi 2008; Song et al. 2010). Therefore, even though the 2011 data do not cover the simulation period, it is still acceptable to apply the 2011 data for modelling human travel behaviours.

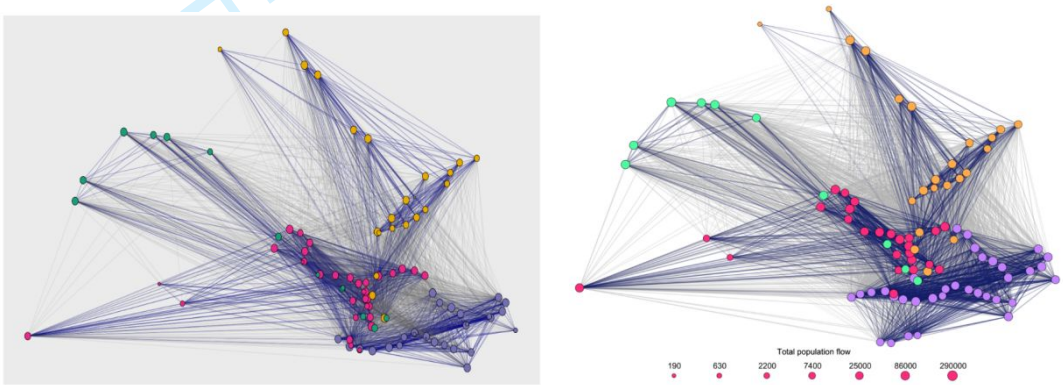

Figure S1. The comparison of the mobility structure of 2011 travel survey data (**Left**) and 2020 MTR (Mass Transit Railway) card data (**Right**) from Zhang et al. (2021), both using Louvain heuristics. **Left**, 45,960 trips involved MTR stations in 2011 travel survey data were used, and four groups of communities were found. **Right**, MTR card data on a typical day between 1 January to 31 March 2020 were used, and also four groups of communities were found. In **Left** and **Right**, the edges in grey and dark blue represent population flows between and within communities, respectively. The vertex size is proportional to the total population flow after log transformation.

Sensitivity analyses on report rates

To account for the uncertainty of report rate values, sensitivity analyses were conducted on three additional parameter sets (**Table S2**). As mentioned in *Infection attributes* in ***Method*** section, the report rates referenced from Hong Kong University was 0.08 and 0.20 before and after 24 February (HKUMed 2022). These values did not consider the RAT online reporting platform, which accepts RAT positive results tested since 26

February. Therefore, the actual values are more likely larger than those values. To address this, we increased the report rate between 24 February and 7 March from 0.20 to 0.25 or 0.30, and the report rate after 7 March from 0.20 to 0.30 or 0.35. What's more, report rate before 24 February was increased from 0.08 to 0.10 to account for the potential bias in referenced value.

After replicating the model calibration process for three scenarios (see *Model calibration* in *Methods*), **Table S2** and **Figure S2** show that compared to our model, scenario S2 yielded comparable model fits, and scenario S1 and S3 generated slightly worse results. In other words, larger values in post-February 24 report rates fit better. Given that the report rate is a global index, and changes in report rates equally affect local case numbers, the relative differences in local-scale transmission are likely to remain consistent across scenarios. Thus, although there are uncertainties in report rates in our model, these uncertainties are expected to exert limited influence on our conclusions.

Table S1. Scenarios of report rates

Scenarios	Report rate before 24 February	Report rate between 24 February and 7 March	Report rate after 7 March
Our model	*0.08*	*0.30*	*0.35*
S1	0.08	0.25	0.30
S2	0.10	0.30	0.35
S3	0.10	0.25	0.30

Table S2. The comparison of the city-level and TPU-level RMSE of the best-fit models

Scenarios	City-level RMSE	TPU-level RMSE
Our model	**7.486**	**9.257**
S1	10.976	9.287
S2	7.668	9.558
S3	10.509	9.472

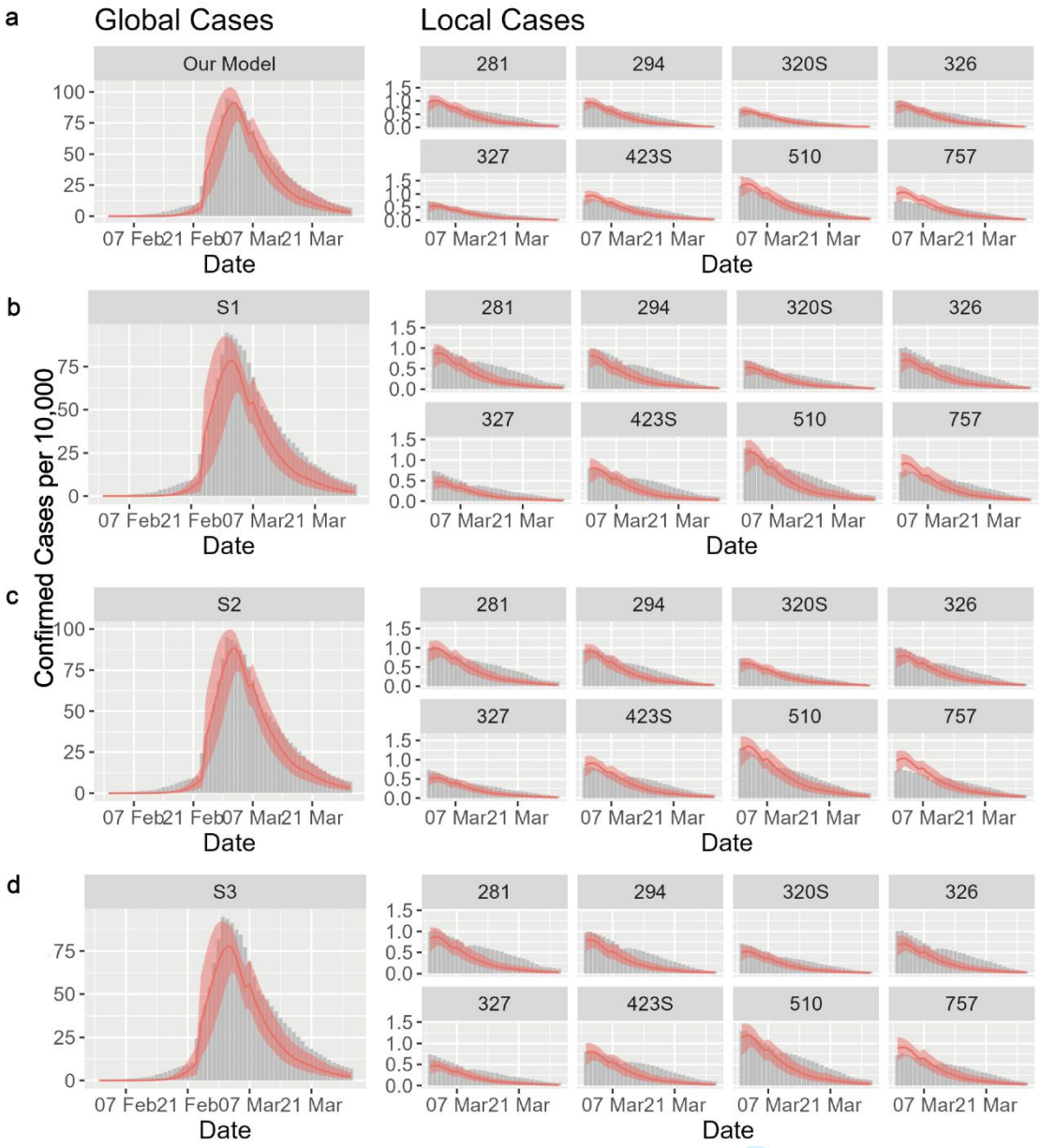

Figure S2. Model fittings under different report rate settings. The report rates for our model (**a**) were 0.08 before 24 February, 0.30 between 24 February and 7 March, and 0.35 after 7 March. Similarly, the report rates for S1 (**b**) are 0.08, 0.25, and 0.30; for S2 (**c**) were 0.10, 0.30, and 0.35; for S3 (**d**) were 0.10, 0.25, and 0.30.

Sensitivity analyses on partially-fixed contactees

Instead of assuming that all contacts remain the same, we assume that 70% of contacts are fixed, and 30% of contacts are updated daily. On the first simulation day, agents were randomly connected to a certain number of contactees in the same subTPU. After the initial setup, the planned links in each subTPU were partitioned into two parts: 70% of

them were classified as fixed contacts, and 30% of them were classified as random contacts. The random contacts were then rearranged, meaning 30% of links in each subTPU were rewired on a daily basis. All other dynamics remains the same with the original model. After 30 runs using the best-fit parameters, we repeated the multivariate linear regression (see **Table 2**) and logistic regression (see **Table 3**) used in the main manuscript.

The regression results demonstrate that our conclusion holds true for the partially-fixed contactee scenario. As shown in **Table S3** and **Table S4**, the scaling index still play an important role in influencing the local transmission risk and individual's probability of becoming a superspreader. These results echo with the main findings of Smieszek, Fiebig, and Scholz (2009), who compared two types of epidemic models— one assuming no repetition of contacts, the other assuming the same contacts repeat day-by-day. The study found that when the transmission probability is very high (e.g., for Measles and Chickenpox), both models exhibit similar outcomes. Given Omicron's notably high transmission probability (R0 is comparable to that of Measles and Chickenpox), it is acceptable to assume fully fixed contactees in our study.

Table S3. For partially-fixed contactee scenario, the effect of the scaling index on the average transmission risk (the average proportion of transmission contacts across 30 runs with best-fit parameters) for subTPUs through multivariate linear regression.

Variable	Coefficient
The scaling index	**2.61*****
Average mobility density	0.17***
Average clustering coefficient of social network	0.09***

*** Significant result with p-value < 0.001.

Table S4. For partially-fixed contactee scenario, the average effects of the scaling index on individual's probability of becoming a superspreader (the top 10% of individuals

ranked by the number of secondary cases) through logistic regressions across 30 runs with best-fit parameters.

Variable	The mean of coefficients	Exponential of (coefficients*1 unit)
The average scaling index	**4.80** *^	**105%** **(a 1-percent-point increase)**
The number of contacts	0.03 ***	103% (a 1-point increase)

*** Significant result with p-value < 0.001; *^ 73% of results are significant with p-value < 0.05.

Sensitivity analyses on the effect of heterogeneous infectiousness on superspreader results

In this scenario, instead of assuming equal infection probabilities for all agents, we assume that the infection probabilities of agents follow the lognormal distribution with $\mu = \log(0.05)$ and $\sigma = 0.5$. As shown in **Figure S3**, the lognormal distribution can guarantee that most of samples would be low values (around 0.05) and a small proportion of samples would be large values (the maximum can reach 0.5 that is 10 times of the original infection probabilities 0.05 (see **Table 1**)). What's more, the mean value of sampled infection probabilities would be comparable to the original infection probabilities 0.05 (see **Table 1**). We randomly sampled the infection probabilities from the lognormal distribution to account for the heterogeneous infectiousness across individuals, and all other dynamics remains the same with the original model. After 30 runs, we repeated the multivariate linear regression used in the main manuscript (see *The effects of urban scaling structure on individual's risk of becoming superspreaders* in *Results).*

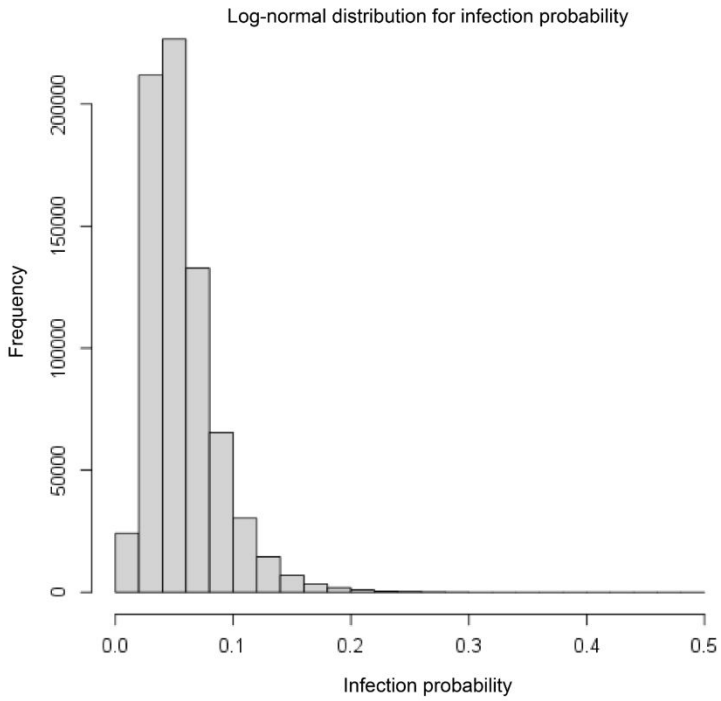

Figure S3. The log-normal distribution for sampling the infection probability.

The results of this scenario did not show much difference. Firstly, superspreaders (the top 10% agents with the large number of secondary cases) still accounted for a large proportion (76.6%, 95% CI: 76.4%-76.7%) of infections. Secondly, conditioning on the infection probability does not substantially change the magnitude and the direction of the coefficients of the average scaling index (**Table S5**). The results suggest the infection probability could be one of the important reasons that influence the individual's risk of becoming superspreaders, but similar to the number of contacts, individuals with high infection probability need to be embedded in areas with high scaling indices, so that they can reach full potential in infecting more people and becoming superspreaders.

Table S5. When the infection probabilities are heterogeneously distributed among agents, the average effects of the scaling index on individual's risk of becoming superspreaders (the top 10% of individuals ranked by the number of secondary cases) through logistic regressions across 30 simulations.

Variable	The mean of coefficients	Exponential of (coefficients*1 unit)
The average scaling index	**3.42** * ^	**104%** **(a 1-percent-point increase)**
The infection probability	6.82 ***	107% (a 1-percent-point increase)
The number of contacts	0.03 ***	103% (a 1-point increase)

*** Significant result with p-value < 0.001; * ^ 46% of results are significant with p-value < 0.05.

Supplementary Tables

Table S6. Model fit statistics for the mobility models at the Tertiary Planning Unit (TPU) and sub-TPU scales

The mobility model	DIC	RMSE	R^2	Mean trips (per person)
Spatial scale: 214 TPUs				
The gravity model	10071869	1812.46	0.93	-
The radiation model	NA	4550.83	0.73	-
The departure-diffusion model	**8652189**	**1936.98**	**0.90**	**2.17**
*Spatial scale: 1622 LSBGs**				
The gravity model	13369799	742.45	0.71	-
The radiation model	NA	595.97	0.79	-
The departure-diffusion model	13259806	732.64	0.70	20.2

* LSBG refers to Large Street Block Groups, which is a set of subunits of TPUs used by the Census and Statistics Department in Hong Kong.

Table S7. Head/tail breaks of 4,863 subTPUs (subunits of Tertiary Planning Units) by mobility volume

Hierarchy	Range	subTPU count
1	(0, 15.0)	3,780
2	(15.0, 56.2)	772
3	(56.2, 122.2)	222
4	(122.2, 227.8)	58
5	(227.8, 703)	31

Table S8. Vaccine effectiveness in reducing susceptibility to Omicron by time since the second or third dose, the estimates from HKU Li Ka Shing Faculty of Medicine (HKUMed 2022)

Vaccine effectiveness	Time since 2nd and 3rd dose		
Vaccine	*14 days*	*90 days*	*180 days*
BioNTech × 1	0	0	0
Sinovac × 1	0	0	0
BioNTech × 2	0.20	0.05	0.01
Sinovac × 2	0.03	0.01	0.01
BioNTech × 3	0.89	0.86	0.77
BioNTech × 2 + Sinovac	0.81	0.67	0.44
Sinovac × 2 + BioNTech	0.64	0.47	0.29
Sinovac × 3	0.36	0.19	0.08

Supplementary Figures

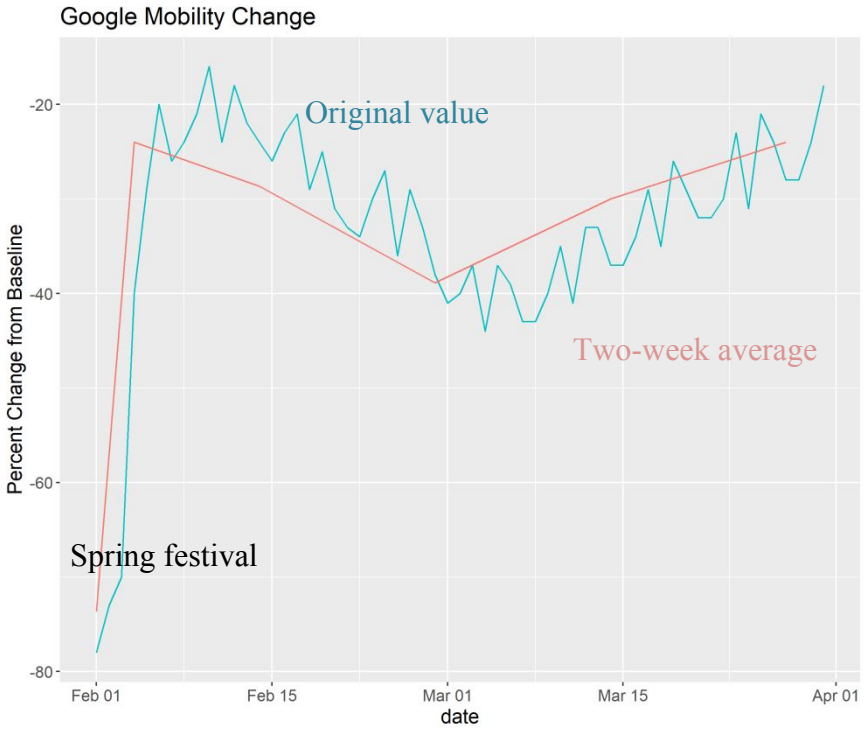

Figure S4. Google mobility change index in workplaces during Feb.1st to Apr. 1st, 2022. We took the two-week average of the original data to avoid weekly fluctuation due to noise or other bias. To better account for the large decline during Spring festival (Feb. 1st

to 3rd), three-day average was calculated and taken as the mobility change index for the three special days.

References

Chang, Serina, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. 2021. Mobility Network Models of COVID-19 Explain Inequities and Inform Reopening. *Nature* 589 (7840):82–87. doi:10.1038/s41586-020-2923-3.

Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabasi. 2008. Understanding Individual Human Mobility Patterns. *Nature* 453 (7196):779–782. doi:10.1038/nature06958.

HKUMed. 2022. HKUMed Proposes Forward Planning after Hong Kong's Fifth Wave of Omicron BA.2. https://sph.hku.hk/en/News-And-Events/Press-Releases/2022/HKUMed-proposes-forward-planning-after-Hong-Kong.

Liu, Ying, and Joacim Rocklöv. 2022. The Effective Reproductive Number of the Omicron Variant of SARS-CoV-2 Is Several Times Relative to Delta. *Journal of Travel Medicine* 29 (3):taac037. doi:10.1093/jtm/taac037.

Smieszek, Timo, Lena Fiebig, and Roland W. Scholz. 2009. Models of Epidemics: When Contact Repetition and Clustering Should Be Included. *Theoretical Biology and Medical Modelling* 6 (1):11. doi:10.1186/1742-4682-6-11.

Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of Predictability in Human Mobility. *Science*, February. American Association for the Advancement of Science. doi:10.1126/science.1177170.

Zhang, Nan, Wei Jia, Peihua Wang, Chung-Hin Dung, Pengcheng Zhao, Kathy Leung, Boni Su, Reynold Cheng, and Yuguo Li. 2021. Changes in Local Travel Behaviour before and during the COVID-19 Pandemic in Hong Kong. *Cities* 112 (May):103139. doi:10.1016/j.cities.2021.103139.

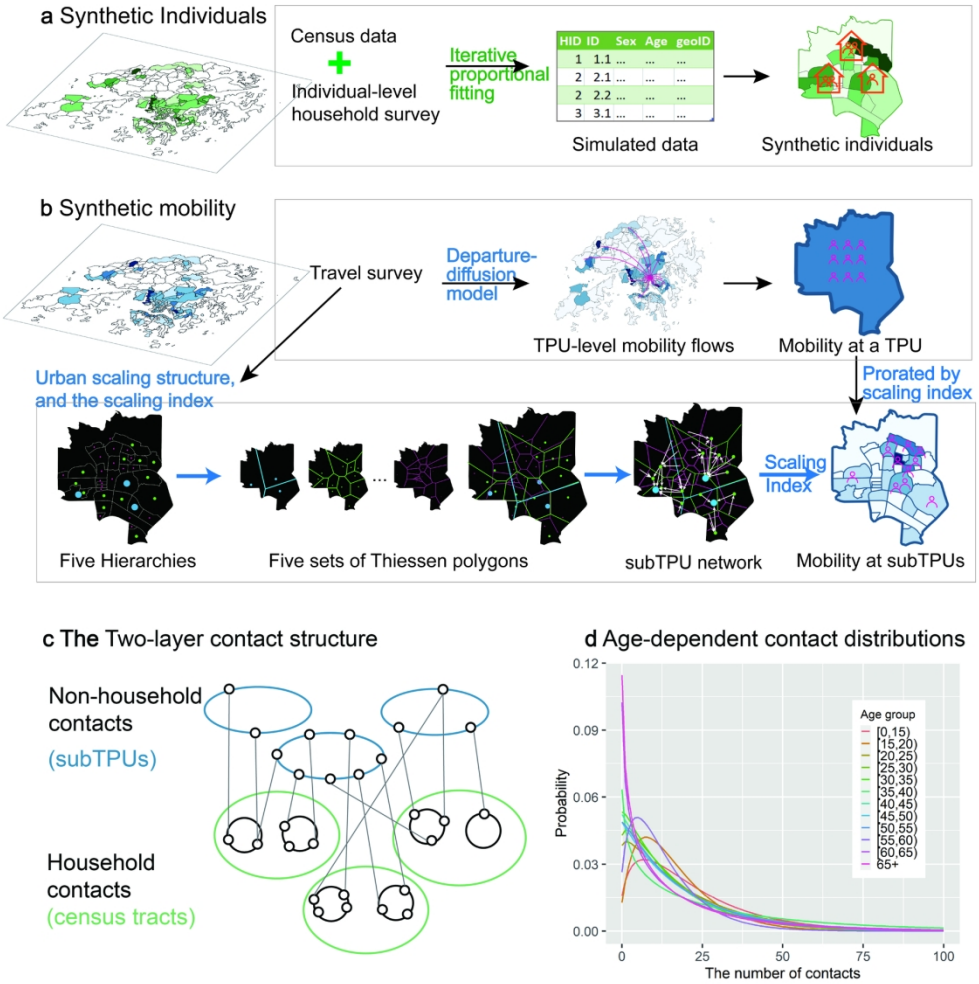

Figure 1. Model description of the spatially explicit agent-based model. a, synthetic population was generated based on census data and household survey data using iterative proportional fitting, which captures the household structures and age structures among real population across 1,622 census tracts. b, synthetic mobility flows were first simulated at the TPU level using the departure-diffusion model, and then were refined the destinations of flows from 214 tertiary planning units (TPUs) to 4,863 subunits of TPUs (subTPUs) using the urban scaling structure, captured by the subTPU network built based on travel survey data. c, the two-layer contact structure consists of household contacts and non-household contacts during trips. At visited subTPUs, agents would have contacts with certain number of agents, and the certain number was sampled from the age-dependent distributions for Hong Kong population (d).

167x169mm (300 x 300 DPI)

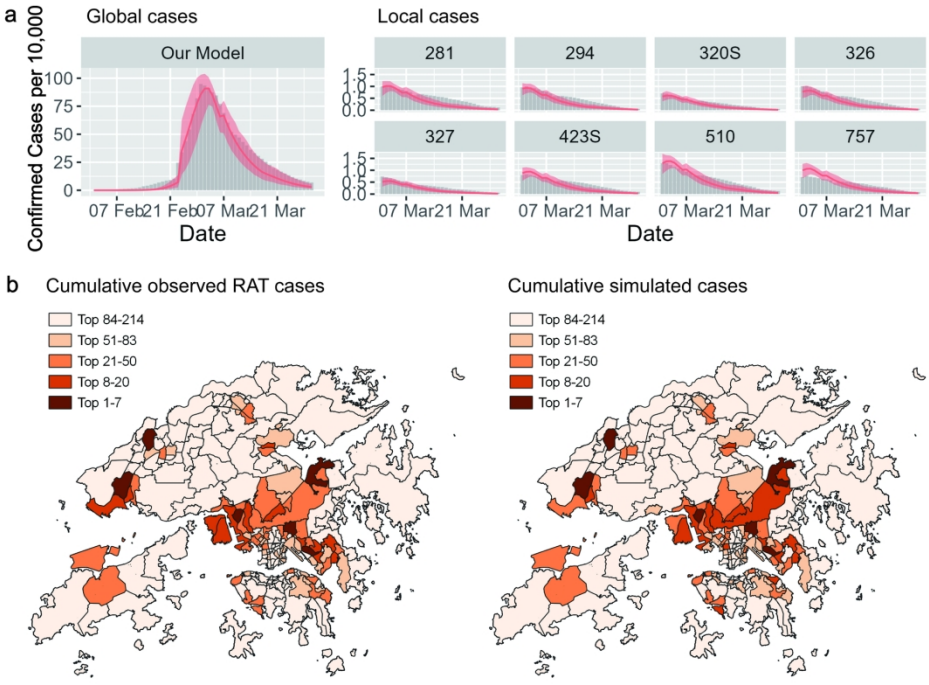

Figure 2. Model fitting results. a, epidemic curves. The left plot is the city-level prediction, and the right plot is the TPU-level prediction on the eight worst affected TPUs. The grey bar is the 7-day moving average of observed confirmed cases. Shaded regions denote the 2.5th and 97.5th percentiles across selected parameter sets and stochastic realizations. b, the spatial distribution of the total number of observed RAT cases (left) and simulated cases averaged across selected parameter sets and stochastic realizations (right).

177x124mm (300 x 300 DPI)

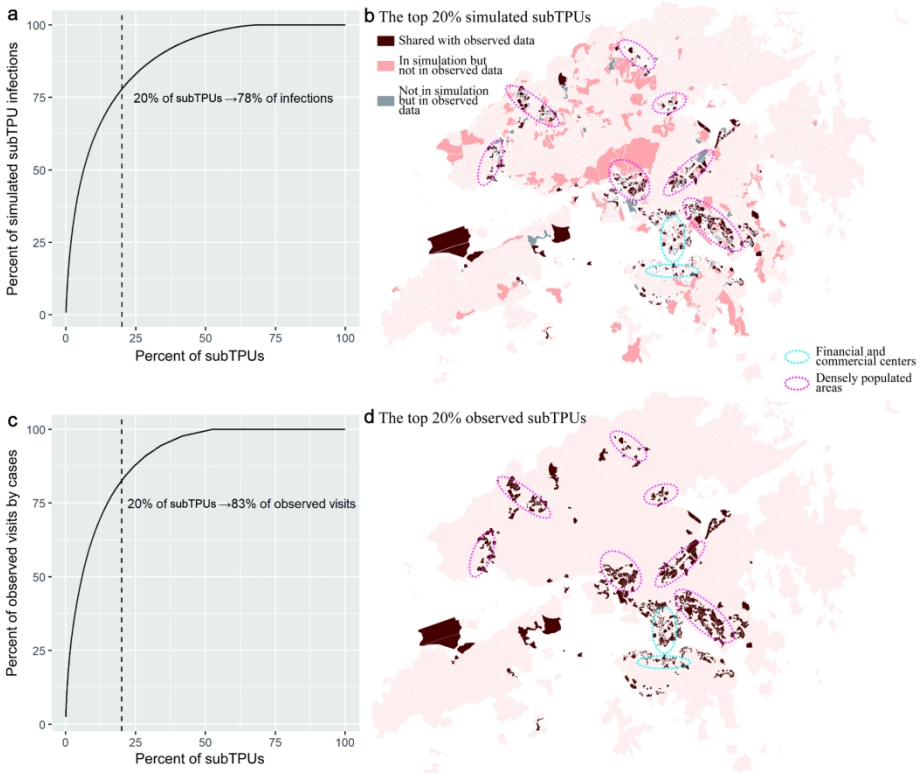

Figure 3. The simulated superspreading subTPUs and the observed high-risk subTPUs. a, the top 20 percent of simulated subTPUs (or superspreading subTPUs), ranked by the average number of infections occurred in subTPUs across 30 runs, accounts for 78 percent of the average number of simulated infections. b, the spatial distribution of simulated superspreading subTPUs. c, the top 20 percent of observed subTPUs, ranked by the total number of visits by infected cases occurred in subTPUs, account for 83 percent of visits. d, the spatial distribution of the observed high-risk subTPUs.

177x137mm (300 x 300 DPI)

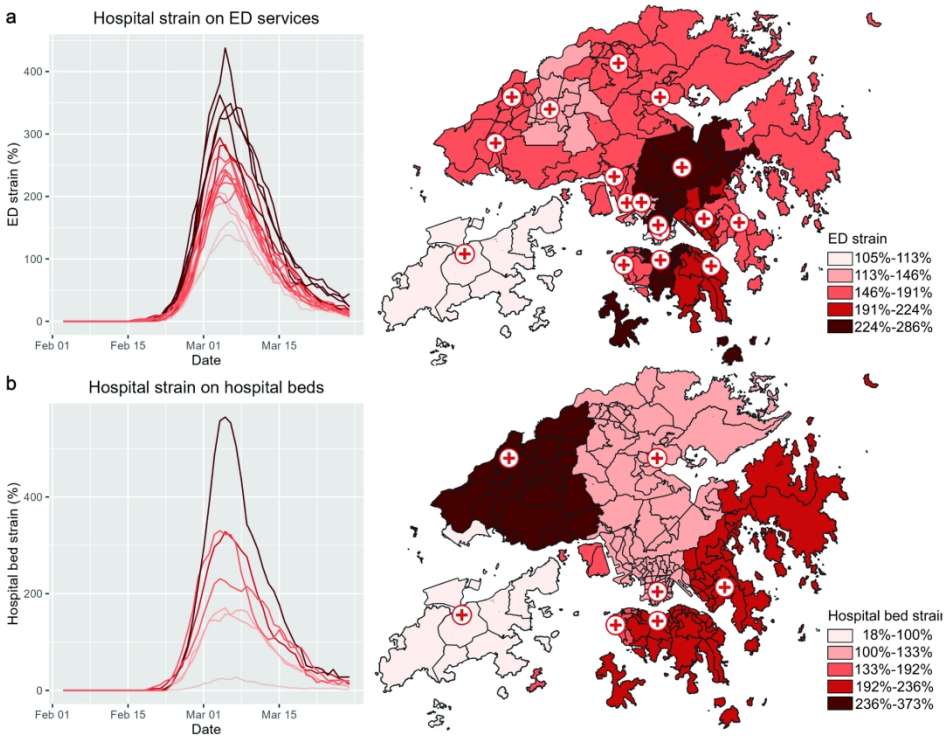

Figure 4. Simulated hospital strain on the Emergency Department (ED) services (a) and hospital beds (b). The strain on each hospital was estimated by the ratio of simulated local demand to available hospital capacity. In a and b, Left, the hospital strain varied by hospital and time. Right, the average hospital strain during the peak period (from 1 March to 15 March 2022) varied across TPUs.

177x135mm (300 x 300 DPI)