Communications in Transportation Research 3 (2023) 100110

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Full Length Article

Adapting node–place model to predict and monitor COVID-19 footprints and transmission risks

Jiali Zhou *, Mingzhi Zhou, Jiangping Zhou, Zhan Zhao

Department of Urban Planning and Design, The University of Hong Kong, 999077, Hong Kong, China

ARTICLE INFO

Keywords:
Node–place model
Land use
Transport
Human mobility
COVID-19

ABSTRACT

The node–place model has been widely used to classify and evaluate transit stations, which sheds light on individuals' travel behaviors and supports urban planning through effectively integrating land use and transportation development. This study adapts this model to investigate whether and how node, place, and mobility would be associated with the transmission risks and presences of the local COVID-19 cases in a city. Moreover, the unique metric drawn from detailed visit history of the infected, i.e., the COVID-19 footprints, is proposed and exploited. This study then empirically uses the adapted model to examine the station-level factors affecting the local COVID-19 footprints. The model accounts for traditional measures of the node and place as well as actual human mobility patterns associated with the node and place. It finds that stations with high node, place, and human mobility indices normally have more COVID-19 footprints in proximity. A multivariate regression is fitted to see whether and to what degree different indices and indicators can predict the COVID-19 footprints. The results indicate that many of the place, node, and human mobility indicators significantly impact the concentration of COVID-19 footprints. These are useful for policy-makers to predict and monitor hotspots for COVID-19 and other pandemics' transmission.

1. Introduction

COVID-19, as a global infectious disease, has become an international concern and unprecedentedly hit cities worldwide. Since declared a pandemic by the World Health Organization in March 2020, the disease is still continuing to spread in many parts of the world. Hong Kong, China, for instance, has suffered five surges of locally confirmed COVID-19 caseloads since the first case emerged in the city (Hong Kong SAR Government, 2022a). After initial fear of the unknown health crisis, however, people had gradually considered fighting against or coexisting with the pandemic as a long-lasting task or "new normal". This could be indicated by adjustment in human's perceptions and behaviors, e.g., normality of telecommuting and resuming of some physical social activities rather than universal lockdowns (Awad-Núñez et al., 2021). Hence, the underlying mechanism of how the COVID-19 virus spreads out across space and people's response to the interventions could vary in existing time compared to the earlier pandemic periods. There are emerging calls for more targeted and effective countermeasures to make a balance between mitigating infection risks and maintaining essential human activities and well-being.

Existing studies explored how urban physical environment (especially land use patterns and built environment) are associated with the caseloads and potential virus spreading (Li et al., 2021; Xu et al., 2022). Other physical features like street connectivity and land use mix also potentially contribute to virus transmission (Kan et al., 2021; Nguyen et al., 2020). Nonetheless, little discussion has been given on more spatially granular footprints of individuals infected and interactions among people in different places.

Considering human movement and their activities could impact social interactions in place, human mobility is another essential topic for understanding the COVID-19 virus transmission – in existing scholarship, we know that human mobility patterns can well predict the susceptible-infected-recovered process in cities (Chang et al., 2021). On the basis, public transportation systems, where mobility mostly occurred, also offer fitting indicators explaining the pandemic process (Afrin et al., 2021; Mo et al., 2021). However, little research has integrated concerns on land use patterns, built environment, sociodemographic, transportation systems as well as individuals' mobility and interactions into a certain analytical framework to predict the COVID-19 footprints at a local scale.

Learning that the node–place model has been widely used to describe

* Corresponding author.
E-mail address: jlzhou@hku.hk (J. Zhou).

https://doi.org/10.1016/j.commtr.2023.100110
Received 28 August 2023; Received in revised form 20 October 2023; Accepted 21 October 2023
Available online 27 November 2023
2772-4247/

how characteristics of station-served communities (i.e., services and physical settings) could be related to their attractiveness to human activities (Cao et al., 2020; Dou et al., 2021), we framed this model upon the context of the COVID-19 pandemic. We involved "mobility" as an extended aspect of the normal node–place model to potentially offer knowledge on how human mobility and activities, urban physical settings, and potential health risks are interacted with one another in the city. Accordingly, this study is expected to achieve at least the three following objectives or contributions:

1) An extended node–place–mobility (NPM) model is proposed to measure relative importance of stations with regards to the mobility network. This allows us to better consider the complex links between the built environment, land use, and human mobility network characteristics of stations.
2) We formulate a new metric from detailed visit history of the infected, that is, the COVID-19 footprint. Compared to previous studies that lack detailed epidemiological investigation, the new metric that records individual footprints can help better capture the transmission process and investigate transmission regularity of the COVID-19 virus.
3) Through the proposed node–place–mobility model, we examine how local station characteristics are correlated with COVID-19 footprints. Specifically, using regression analysis, we show how mobility patterns of people, as measured in mobility network centrality, can help us predict spatial distribution of COVID-19 footprints. This could examine the effectiveness of anti-pandemic interventions and indicate potential transmission risks among people in different places.

The remainders of the paper are organized as follows. Section 2 is a literature review of the relationship between land use, built environment, human mobility, and COVID-19 situation, and the node–place model and relevant theory. Sections 3 and 4 elaborate on the methodology used and an empirical study of Hong Kong, China. Section 5 discusses and concludes the empirical results.

2. Literature review

2.1. Impacts of land use and transportation characteristics on COVID-19 situation

There have been many studies discussing how urban physical settings, characterized by land use patterns and built environment features, impact the spreading of COVID-19 virus (Megahed and Ghoneim, 2020; Xu et al., 2022). Types of facilities, for instance, school closure and restrictions on social gathering in public spaces may play a role in containing citywide virus spreading (Litvinova et al., 2019). Notably, Chang et al. (2021) found that full-service restaurants and hotels suffered most. In addition, various land use or built environment characteristics were used to explain the spatial patterns of COVID-19 cases in cities. Nguyen et al. (2020) leveraged Google Street View to detect street features and found that indicators of mixed land use (non-single-family home), walkability (sidewalks), and physical disorder (dilapidated buildings and visible wires) were associated with larger quantity of COVID-19 cases.

Human mobility and its performances in transport systems have also been found to explain the susceptible-infected-recovered process in cities (Chang et al., 2021; Schlosser et al., 2020). Hence, some indicators describing characteristics of transportation systems and peoples' travel behaviors were fitted for prediction. Manzira et al. (2022) examined how different modes of transport, including traffic volume, bus passengers, pedestrians and cyclists, were associated with the reported COVID-19 infections. In particular, public transport has been considered as a potential high-risk environment for virus transmission due to passengers being confined and interacting in limited spaces and difficulty to detect potentially transmission routes among people (Gartland et al., 2022).

Overall, though various indicators extracted from land use patterns, built environment, and mobility have been examined, most of the existing discussions were conducted from a citywide or regional scale. Little research has systematically involved them to explore the impact COVID-19 transmission at the local scale based on footprints of the individual COVID-19 infected.

2.2. Human mobility and COVID-19

The mobility network has been widely used in air-borne disease studies for detailed transmission modeling and contact-tracing, which examined how human mobility plays an essential role for pandemic prediction. For instance, Mo et al. (2021) proposed a time-varying weighted encounter network to model pandemic spreading in public transportation systems using smart card data and concluded that isolating influential passengers at an early stage can reduce the spreading. Yabe et al. (2020) utilized a contact network based on human mobility trajectories (GPS traces) and web search queries to predict COVID-19 hotspot locations. These studies, however, often employed unipartite networks of people and ignored the interactions between people and the locations where human activities take place.

Bhattacharya et al. (2021) constructed a homogeneous network of locations based on aggregate mobility flows and employed the PageRank (PR) algorithm (Brin and Page, 1998) to identify the high-risk locations for COVID-19 transmission, but did not capture the individual-level human movements. Considering that the transmission processes take place through contact networks of infected individuals, it is important to consider both individual mobility traces across locations and colocation patterns across individuals. This can be done by constructing and analyzing a heterogeneous bipartite network that connects people and their visited locations using large-scale human mobility data. For example, Zhou et al. (2022) constructed a bipartite people–location network with metro smart card data and calculated the Personalized PageRank scores to identify high risk users and locations regarding COVID-19 transmission.

Taking into account the neighborhood's points of interests (POIs), Chang et al. (2021) integrated dynamic mobility networks, using mobile phone data, to simulate the spread of SARS-CoV-2 in US metropolitan areas. They found that a small minority of 'super-spreader' points of interest account for a large majority of the infections. However, as mentioned before, besides land-use, built-environment, human mobility network features, transportation systems also serve as a disseminator of infectious diseases.

In general, land-use, built-environment, sociodemographic, transportation accessibility features, as well as actual human mobility patterns are all needed to capture the dynamics of COVID-19 transmission process accurately. So far, little research has integrated all these aspects well into an appropriate framework for pandemic prediction.

2.3. Node–place model

The node–place model has been widely used to describes stations' realization and balances of good services on the network (the "nodes") with a location that promotes human interactions (the "place") (Bertolini, 1999). It offers a framework to understand and explore the development potential of station-served areas in cities to improve transit-oriented development (TOD). So far, a variety of node–place measures have been used to capture the features of stations, typically including built environment characteristics (Kamruzzaman et al., 2014) and socioeconomic factors (Zemp et al., 2011). To extend the scope of model, some studies add particular aspects to the node–place model, such as street connectivity (Lyu et al., 2016), walkability (Jeffrey et al., 2019), travel network (Dou et al., 2021), accessibility (Cummings and Mahmassani, 2022), design (Zhang et al., 2019), and ridership (Cao et al., 2020).

The node–place model has been used to classify and evaluate transit stations, which provides insights to learn individuals' travel behaviors and support urban planning through effectively integrating physical environment, sociodemographic and transportation development. However, use of the model for COVID-19 studies is still lacking. This study will explore how node, place, and mobility would be associated with the transmission risks and presence of the local COVID-19 footprints in the city.

3. Methodology

The overall research framework is presented in Fig. 1 Specifically, we first collect land-use, sociodemographic, and human mobility data as inputs. Using human mobility data, such as smart card data or GPS data, we construct a bipartite people–location network and calculate the PageRank centrality scores for different stations. Then we introduce the node–place–mobility model, including the index selection and *k*-means station classification processes. After classifying stations into different clusters using their node, place, and mobility indices, we explore the relationship of these indices and their belonging features with COVID-19 infectors' footprints, taking Hong Kong in China as our case study. The linear regression model is employed to further explore and quantify the relationship.

3.1. Node–place–mobility model framework

As mentioned before, the original node–place model ignores the importance of the actual human mobility patterns, that is, how people utilize different locales as a node and a place over time. Only built environment features are considered in the original node–place model. Hence, an extended node–place–mobility model, as shown in Fig. 2, is proposed. The node dimension represents the transportation service around the station areas, while place dimension represents the land use features. The new dimension, mobility dimension measures the importance of stations in a bipartite people–location network.

Regarding people's first-last-mile trips around subway stations, the maximum walking distance for pedestrians is found to be around 1.5 km from a subway stations in (Biba et al., 2010; Park et al., 2021). Therefore, the node and place indicators reflecting the land use and transportation accessibility within 1.5 km from the subway stations are selected.

3.1.1. Node and place indicators

Depending on the form of transportation, node indicators can be classified into two groups: public transit and road networks (Table 1). The place dimension indicators are also shown in Table 1.

3.1.2. Mobility indicators

Most existing network-based disease transmission models are based on a homogenous unipartite network of places or persons. A heterogeneous people–location network is constructed in this study to connect people to their visited locations. This is owing to the nature of disease transmission, in which the infected carry viruses to a location and the viruses can survive (on a surface) for certain time. The SARS-COV-2 virus (the virus that causes COVID-19), for example, can survive on a surface for days, according to Van Doremalen et al. (2020).

To capture the relationship between two subject types, such as people and locations, a ubiquitous data structure, the bipartite graph, has been proposed. In the bipartite graph, all vertices are categorized into two sets, within which vertices in the same set cannot be directly connected to each other; but they can be connected via a vertex from another set. Treating people and their visited locations as two sets of nodes, a bipartite network can be constructed. Each link describes an individual's visit to a location. The input can be human mobility data of any format, such as cell phone data and transit card data.

As shown in Fig. 3, disaggregate-level human mobility data, such as smart card data and cell phone data, is used for bipartite people-location network construction and PR score calculation for transmission risk estimation.

Based on the bipartite network of users and stations, the PageRank centrality scores are calculated for each station. Supposing there are k nodes that link to node p, the PageRank (PR) score of a node p is given as

$$\mathrm{PR}(p) = \frac{1-d}{N} + d\sum_{i=1}^{k} \frac{PR(p_i)}{C(p_i)} \tag{1}$$

where N is the total number of nodes on the web, p_i is the node that links to node p, d is a damping factor, and $C(p_i)$ is the number of out-links of p_i. The damping factor d defines the probability of a web user going to an adjacent link, and thus the probability of a surfer skipping to another page is then given by $(1 - d)$.

The original PR algorithm has been introduced to identify the high-risk locations for COVID-19 transmission (Bhattacharya et al., 2021), but it is only based on aggregate mobility flows and a homogeneous network of locations. The implementation of the PageRank Algorithm can be described below.

Algorithm 1. PageRank algorithm

1 Input Graph G with N nodes
2 Initialize damping factor d
3 Initialize all nodes in G with original PR score $= 1/N$
4 While not converged
5 For all node v in the graph, do
6 $\mathrm{PR}(p) = \frac{1-d}{N} + d\sum_{i=1}^{k} \frac{PR(p_i)}{C(p_i)}$
7 End for
8 If the error rate for any vertex in the graph falls below a given threshold,
Converged
9 End while

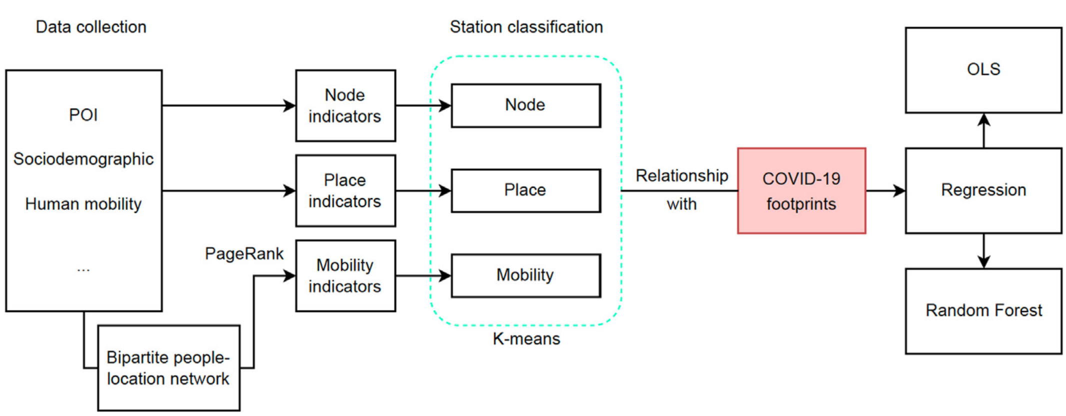


Fig. 1. Research framework.

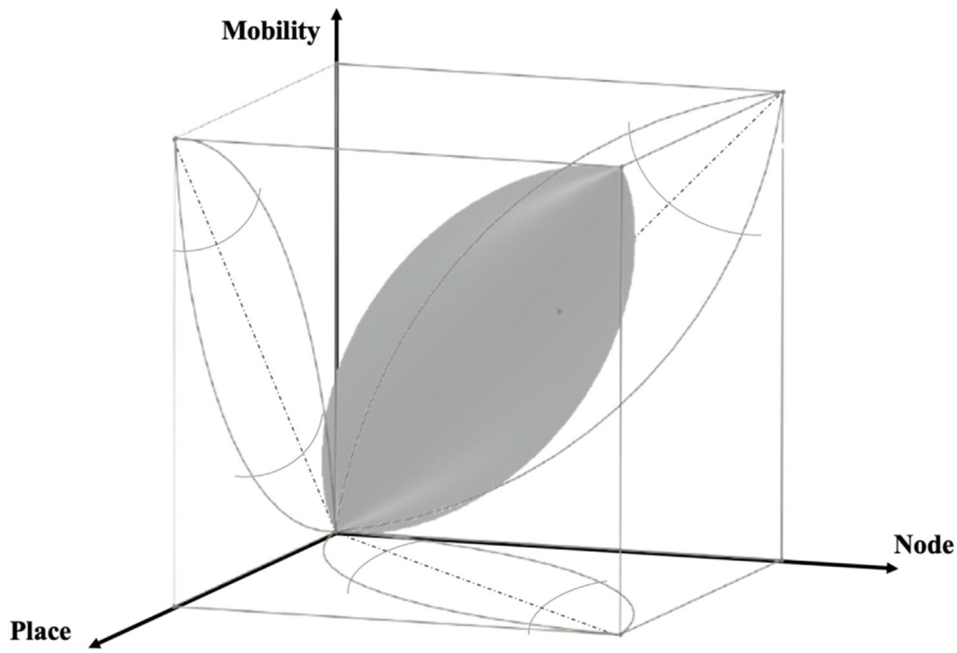


Fig. 2. Illustration of the node–place–mobility network.

Table 1
Indicators of the node dimension.

Indices	Description
Node index	—
Number of bus stations	n1 = Number of bus stations within 1.5 km
Number of intersections	n2 = Number of intersections within 1.5 km
Place index	—
Population density	p1 = Population within 1.5 km
Percentage of unemployed	p2 = Percentage of unemployed in the population within 1.5 km
Percentage of low income	p2 = Percentage of low income (under HK$10,000 per month) in the population within 1.5 km
Number of POIs	p3 = Number of points of interests (POIs) in the population within 1.5 km
Percentage of apartments	p4 = Percentage of apartments in all POIs within 1.5 km
Land use mix	p5 = Land use mix

3.2. Clustering

After data processing, the stations were clustered into different groups using the *k*-means clustering methods and node, place, and mobility index values as input. Cluster analysis is used to create classes of Hong Kong's metro stations with the least variance within each group and the most variance between them. Here, *k*-means cluster analysis is used. The average silhouette approach is used to determine the appropriate number of clusters (Rousseeuw, 1987). The traditional approach would stop here and give management recommendations for the station area. What is not obvious is what role the station plays at the strategic network level. We add a criticality analysis of each station to the cluster analysis to highlight differences between stations within the same clusters and indicate the strategic importance of each station region.

3.3. COVID-19 footprint regression

In this study, we introduce a new COVID-19 metric, the COVID-19 footprint of the infected, which records all locations the infected visited, according to the Hong Kong SAR government's open data on COVID-19 situations in the city. The reason for introducing this new metric is that the place where people get infected remains unknown due to the lack of detailed epidemiological investigation. Only the places visited by the infected are recorded in the open data. As a result, we assume a uniform distribution across all locations the infectors visited before the positive diagnosis.

A uniform distribution across all locations visited by the infectors

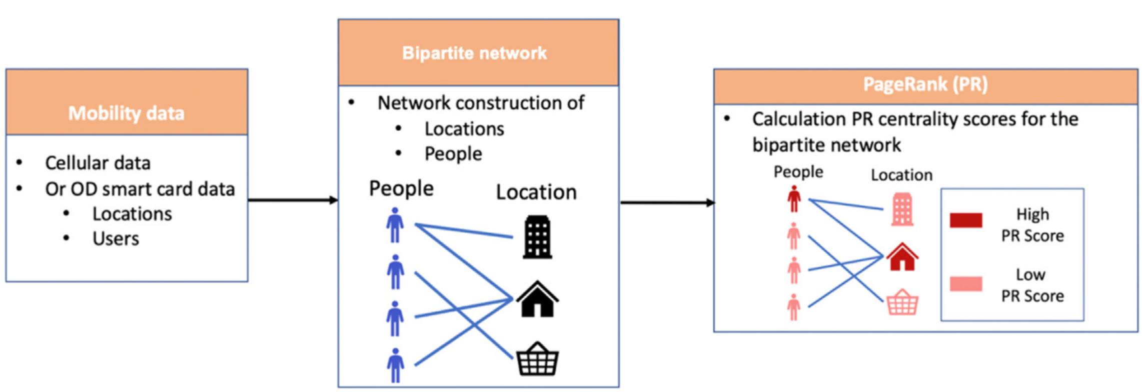


Fig. 3. Construction of bipartite people–location network and calculation of PageRank centrality scores.

before their positive diagnosis is a reflection of the fact that the detailed records about locations visited by them is exhaustive and covers all occurrences of events where the infected visits a place. These records ensure that all instances of the infected individuals visiting a place are captured, forming the basis of our analysis. It provides a baseline understanding of the potential risk associated with various locations visited by infected individuals, regardless of specific work or school-related patterns.

3.3.1. COVID-19 footprints

Here we have total number of infected cases n. We define a person p, who has been to m_l distinct locations, the footprint of person p at location j is

$$v_{p,j} = \frac{1}{m_l} \tag{2}$$

And the number of footprints from all persons who visited location j is

$$v_j = \sum v_{p,j} \tag{3}$$

The total number of visited locations by infectors within a specific distance d from a station s can be expressed as a function $l\,(d,s)$, which is a function of d and s.

The COVID-19 footprint around that station is defined as the sum of weighted number of visits by the infected to any locations within a certain distance from a subway station, as shown in Eq. (4).

$$\text{COVID-19 footprints} = \sum_{j=0}^{l(d,s)} v_j \tag{4}$$

3.3.2. Multi-variate regression models

As mentioned in the introduction and literature review sections, socio-economic factors such as low-income population, unemployed population (Chang et al., 2021), types and density of buildings and land use features (Kan et al., 2021), mixed land use (Nguyen et al., 2020), human mobility, and transportation systems (Chang et al., 2021) are found to explain the COVID-19 situations in cities. Based on the previous literature, we conducted multiple linear regression with ordinary least squares to discover individual effects of various factors such as unemployed population, density of apartments, land use mixture, transportation service levels, and human mobility levels on COVID-19 footprints around subway stations. This method allows researchers to investigate the links between a set of indicators and COVID-19 footprints in the context of a larger set of variables. A least squares regression approach was applied.

4. Empirical study

4.1. Study area

In this study, we chose Hong Kong, China as the study area. The Mass Transit Railway Corporation (called "MTR" locally) is Hong Kong's primary public transportation system, with the highest frequency and daily ridership across all modes of public transportation in the territory (Legislative Council Secretariat of Hong Kong SAR, 2016). It carries 41% of the daily passenger trips in Hong Kong. There were 230.9 km of rail track and 98 stations in the system as of February 2022.

The Hong Kong Centre for Health Protection posted daily confirmed COVID-19 cases online (https://data.gov.hk). Each confirmed case's age, gender, type (imported, local, close contact with local cases, perhaps local, close contact with possibly local cases, and close contact with imported cases), and the buildings or venues where they resided or frequented in the 14 days before infected are documented. Case-related sites include confirmed cases' homes and places they frequented (visited locations).

Hong Kong was hit badly by the fifth wave of the COVID-19 outbreak, which started from December 26, 2021 in Hong Kong (Hong Kong SAR Government, 2022b). After February 5, 2022, the Hong Kong SAR government was unable to track the COVID-19 cases in a timely manner. In this study, we used the data between December 27, 2021 and February 5, 2022, and there were 1921 confirmed cases, which covers the beginning and the rising phase of an outbreak, where people's mobility patterns are not impacted significantly due to the mild trend of the transmission.

We present a case study based on built-environment and real-world large-scale mobility data from Hong Kong MTR surrounding all MTR stations to evaluate the relationship between built environment and COVID-19 footprints around subway stations. We chose smart card transaction data from January 21, 2021, a regular day without accidents and holidays. We created a network of 98 location nodes and 1.7 million user nodes. People node's average degree is 3.6 and location node's is 86,163.3.

4.2. Data description

Regarding the exogenous variable, the COVID-19 footprints from infectors, approximately 92% of them are located within 1.5 km from a subway station. The summary of exogenous variable (COVID-19 footprints) and other node, place, and mobility index indicators and the explanatory variables for the linear regression model are included in Table 2.

4.3. Clustering results

Through k-means clustering, 4 clusters are chosen with the best silhouette value 0.31. At the same time, the relationships among node, place, and mobility indexes were also analyzed to reveal the complex links between station-level built environment, human mobility, and COVID-19 footprints around station areas, which are the circled areas within 1.5 km from MTR stations. The four clusters show different trends in the node, place, and mobility indexes and also the number of footprints.

Figure 4 compares node, location, and mobility indices with COVID-19 footprints around stations. The four colors represent 4 clusters, with circle size presenting station footprints. Cluster 1 has high node, low place, and high mobility indices. Cluster 2 has high node, location, and mobility indices. Cluster 3's node, location, and mobility indices are medium. Cluster 4 has low node, location, and mobility indices. Table 3 shows the node, location, and mobility cluster centers, mean and standard deviation for the four clusters. The general trends from the results are each index appears to be positively impacting COVID-19 footprints, even when considered with another index.

Figure 4(a) shows that node, place, and mobility indices have positive impacts on the COVID-19 footprints. Therefore, it appears that the higher any of the node, place, or mobility qualities at a station are, the higher number of COVID-19 footprints are around that station area. However, Figs. 4(a)–4(c) also show the trend of the node index having an outsized impact on COVID-19 footprints relative to place and mobility. This trend of the node index measures shows the impact from transportation service and accessibility. In Figs. 4(b) and 4(c), the typical node–place plot, it is

Table 2
Descriptive statistics of variables.

Variable	Mean (stdev)
COVID-19 footprints	23.11 (87.67)
Population density	36,723 (40,340.79)
Percentage of unemployed (%)	0.0009 (0.00038)
Number of POIs	961 (765.43)
Percentage of apartments (%)	85 (7.65)
Land use mix	0.019 (0.019)
Number of bus stations	58 (29)
PageRank centrality scores	0.0068 (0.025)

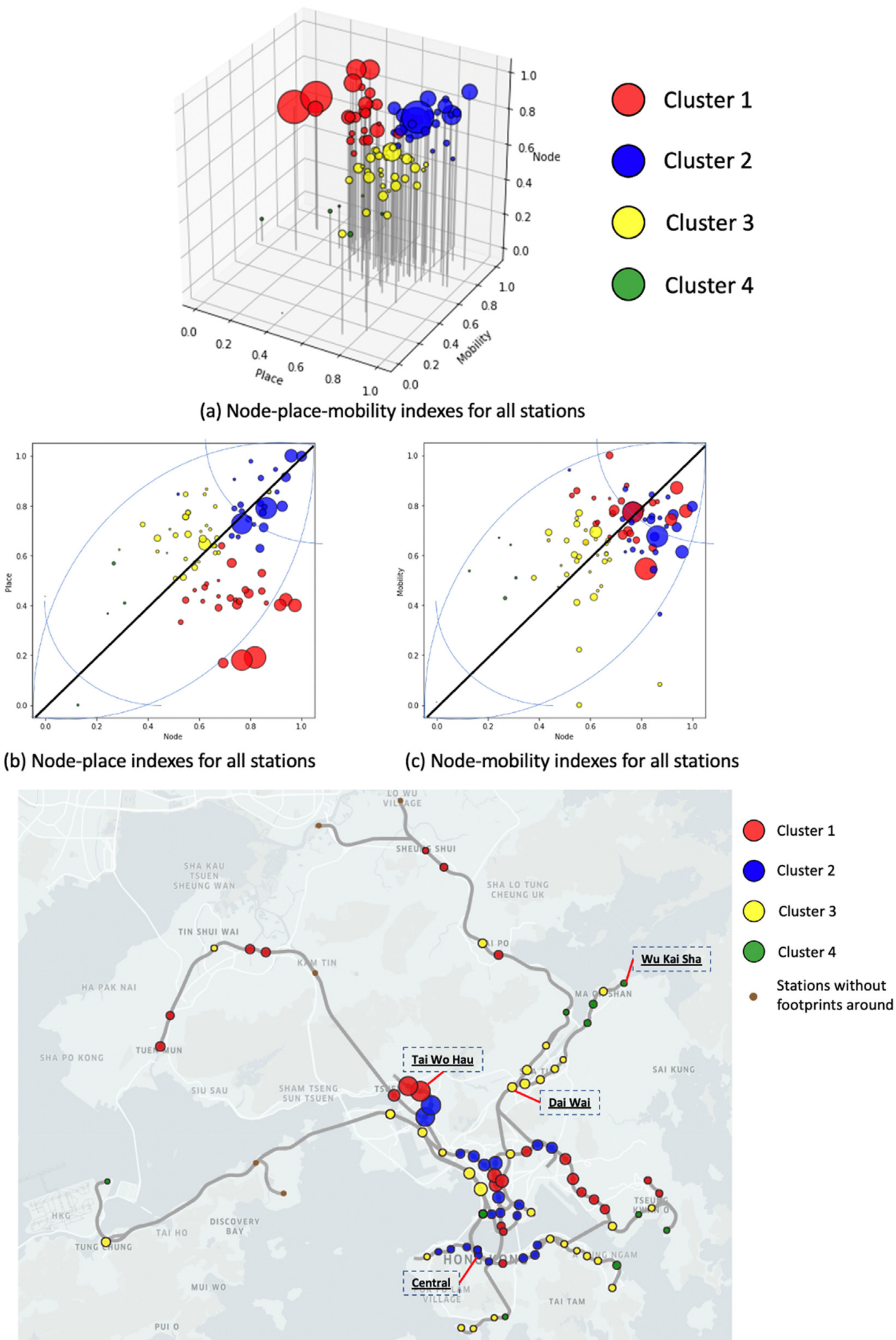


Fig. 4. (a) Node, place, and mobility indexes of the stations (circle size representing footprints). (b) Node and place indexes of the stations (circle size representing footprints). (c) Node and mobility indexes of the stations (circle size representing footprints). (d) Spatial distribution of stations from four clusters (circle size representing COVID-19 footprints).

Table 3
Summary of four clusters.

Cluster	Cluster center			Number of COVID-19 Footprints, mean	Number of COVID-19 Footprints, stdev
	Node	Place	Mobility		
1	0.73	0.41	0.76	81.99	108.88
2	0.82	0.82	0.71	69.86	106
3	0.57	0.69	0.55	22.35	23.37
4	0.19	0.36	0.43	4.66	3.57

shown that the majority of the MTR stations have balanced node–place–mobility values.

Cluster 1, close to the unbalanced node region, has more COVID-19 footprints. This means areas with a higher node index than place index can have more COVID-19 footprints. Fig. 4 shows a typical cluster 1 station, Tai Wo Hau station. Tai Wo Hau in Hong Kong is mainly comprised of public housing apartments, which has lower place index such as land use mix. However, due to its population and the transportation accessibility within 1.5 km, the node and mobility indexes are still high.

Cluster 2, close to the 'stressed' region (high node and place index values), has more COVID-19 footprints. In "stressed" locations, transportation and land use dynamics interact most. Further development due to restricted space leads to incompatibility in transportation and land use systems. The stressed stations also have a high number of footprints. One example, the Central station is at Hong Kong's CBD. The number of jobs and POIs is high around the Central station, and it has large number of bus connections within 1.5 km. Since it is also one of the job centers in Hong Kong, the mobility index for Central station is also one of the highest.

Cluster 3's node, location, and mobility indices are balanced. Balanced stations have low COVID-19 footprints. For example, Tai Wai station, on the outskirts of Kowloon's traditional CBD, has a reasonable amount of bus connections, housing and POI density, and medium mobility level.

Cluster 4 has low node, location, and medium mobility indices. Minimal human interaction, activity intensity, and diversity can be seen in this cluster. As a result, the number of COVID-19 footprints is the lowest for these balanced stations. For instance, Wu Kai Sha is the northeastern terminal station on the Tuen Ma line, and it has relatively low population, POI density, what is more, the transportation accessibility, such as bus stations and road network density there is low, as well as human mobility index.

Another pattern that can be seen in Fig. 4(c) is that the mobility index does not vary a lot for all four clusters, and this is due to the small variance of the PageRank centrality scores for the people–location network.

Figure 4(d) shows the geographic distribution of cluster stations, which demonstrates a line-based pattern. Clusters 1 and 2, which have the most COVID-19 footprints, are in major metropolitan regions with adequate transportation access and diverse land use in Hong Kong Island and Kownloon's CBD. Cluster 3, with a medium number of COVID-19 footprints, is largely along the city's edge and has moderate urban traffic and land use. Cluster 4, having the fewest COVID-19 footprints, lies near the city's edge or close to the end of the rail transit station.

In this study, we applied the *k*-means clustering algorithm to classify stations based on their node, location, and people mobility indices. However, it has been suggested that *k*-means clustering may be more suitable for spherical clusters and may not be as sensitive to the actual situation as other clustering methods. To address this concern, we also applied the Gaussian mixture clustering algorithm (Fig. 5) to our data and compared the results with those obtained using *k*-means clustering. We found that the clustering results obtained using Gaussian mixture clustering were similar to those obtained using *k*-means clustering. This suggests that our conclusions regarding the spatial distribution pattern of clusters are robust and not solely dependent on the choice of clustering algorithm.

Therefore, we continue our discussion from the *k*-means clustering.

4.4. Linear regression on footprints of COVID-19 infectors

We adopted a two-step regression method as described in Zhou et al. (2021) to fit a series of ordinary least square models using the independent variables mentioned in Section 3.3.1, and the framework is shown in Fig. 6 We first fitted a linear regression model using multiple nodes, place, and mobility indicators. We assume that there exists a linear relationship between COVID-19 footprints and explanatory variables including node, place, and mobility indicators. We also assume that higher node index (higher transportation service levels), higher place value (land use and sociodemographic complexity) and human mobility index can lead to more COVID-19 footprints around stations. Specifically, in the final model after the model adjustments and comparison, we

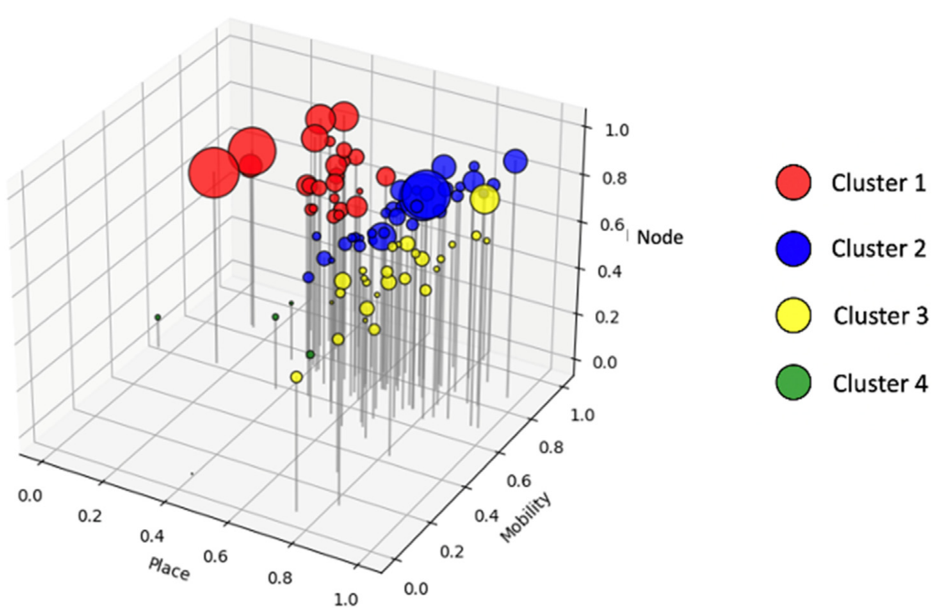


Fig. 5. Gaussian mixture clustering: Node, place, and mobility indexes of the stations (circle size representing footprints).

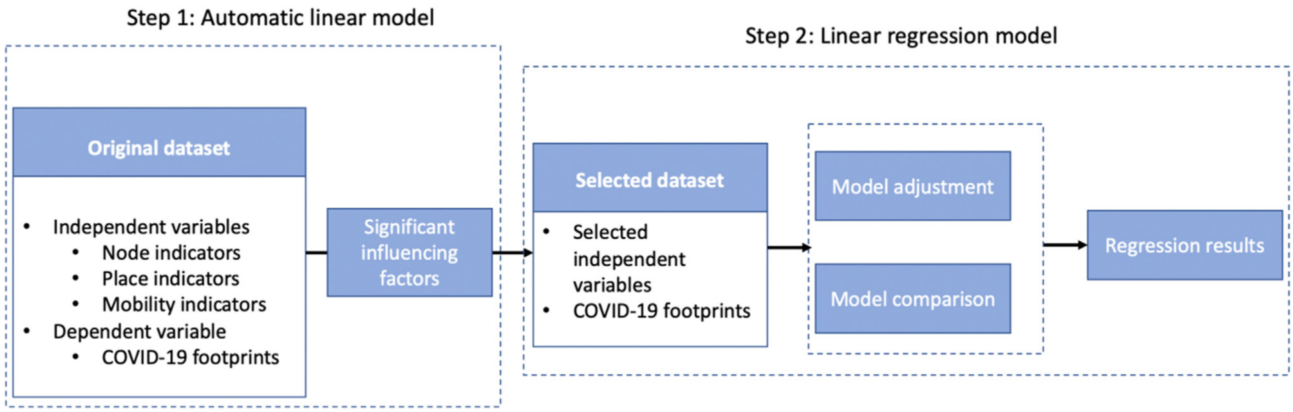


Fig. 6. Regression modeling framework.

included 5 independent variables under the assumptions that unemployed population, more densely distributed apartments, higher land use mixture, higher transportation service levels, and human mobility levels lead to more COVID-19 footprints.

The linear regression model on COVID-19 footprints around stations has five independent variables, including the percentage of apartments among all POIs within 1.5 km of the subway station, the PageRank score of the subway station in the bipartite people-location network, the land use mix, the number of bus stations within 1.5 km of the subway station and the percentage of unemployed in the population within 1.5 km.

Because log transformation can better normalize numerous high-variance variables, the exogenous variable and explanatory variables were in log–log form. For the regression, the logarithm (base 10) of each was calculated. The correlation of the independent variables is shown in Table 4. The results indicate that the variables used in the model have low to moderate correlation with each other.

Table 5 shows model results. Node, place, and mobility indicators are associated to COVID-19 footprints surrounding metro stations. Unsurprisingly, more COVID-19 footprints are seen around MTR stations with a larger percentage of housing apartments in all POIs within 1.5 km. The land use mix around a station is positively associated with more COVID-19 footprints. The increase in percentage of unemployed in the general population also indicates more COVID-19 footprints. Bus station density affects COVID-19 footprints around a station positively and significantly. The higher PageRank scores of mobility network stations, the more COVID-19 footprints could exist.

The results are consistent with the previous findings on the impacts from socio-economic factors (Chang et al., 2021), types of building and density of buildings and land use features (Kan et al., 2021), mixed land use Nguyen et al. (2020), human mobility and transportation (Chang et al., 2021; Schlosser et al., 2020) on the COVID-19 situations in cities. The regression results indicated the positive impact of node, place, and mobility on COVID-19 footprints. In each case, high index scores were correlated with more COVID-19 footprints.

5. Discussion and conclusions

In the existing scholarship, researchers have studied how local land use, transportation, and human mobility affect COVID-19 in cities. Previous research highlighted critical land use features for COVID-19, including population density, building types, socio-economic characteristics, transportation features, and mobility. Most studies have focused on pandemic progress at the local or regional level. Little attention has been paid to more spatially granular footprints of infected individuals and interactions among people in different places, which is critical for more targeted and effective countermeasures to strike a balance between infection risk mitigation and essential human activities and well-being in the post-COVID era. Also, little research has combined land use patterns, the built environment, social demography transportation networks, and individual mobility and interactions into a specific analytical framework to estimate COVID-19 footprints at a local scale. Few studies have focused on transit-reliant cities like Hong Kong. The node–place model helped us define and measure local functions and transit services provided by cities. Functions and services should match, theoretically. This study adds a third dimension to the node–model model: How riders visited different locations. Using empirical data from Hong Kong, we operationalized the revised model to forecast and monitor COVID-19 footprints and transmission risks. In this study, we introduced a third dimension to the model: How riders patronized different locales, that is, we have adapted the existing node-model model. We operationalized the adapted model using empirical data from Hong Kong and illustrated that the model can be used to predict and monitor COVID-19 footprints and transmission risks.

Specifically, node index measures the number of bus stops and junctions near a metro station. Place index includes population density, unemployment, POIs, apartment percentage, and land use mix. Mobility index is a network's PageRank centrality. Based on the quantification, we divided stations into 4 clusters and fitted regression models to see how node, place, and mobility affect COVID-19 footprints and transmission risks. More COVID-19 footprints occur near a station if its node, location, or mobility attributes are higher. However, the results also show the trend of the node index having an outsized impact on COVID-19 footprints relative to place and mobility. This trend of the node index measures shows the impact from transportation service and accessibility. Node, place, and mobility positively affect COVID-19 footprints. High index values mean more COVID-19 footprints.

Specifically, to capture the impacts from the aforementioned local land use and transportation characteristics, and human mobility patterns features simultaneously, we proposed a novel node–place–mobility model, combining the static land use and transportation accessibility with the human mobility dynamics in cities for subway station area classification. Through the proposed node–place–mobility model, the analysis uses these indices to identify and classify subway stations to different groups, and employs regression to identify important node–-place–mobility variables impacting COVID-19 footprints. The node index characterizing transportation service, the place index representing land use characteristics, and the actual human mobility index, indicating the importance of stations in the bipartite people–location network, showed positive impacts on COVID-19 footprints.

The specific results of the regression models identified a number of characteristics of stations that have significant impact on COVID-19 footprints. They suggest that more COVID-19 footprints are likely to be attained at stations that have (1) more bus stations around, (2) a high PageRank centrality in bipartite people–location network, (3) a higher share of apartments in all POIs around, (4) higher land use mix, and (5) higher share of unemployed population. The results report similar trends

Table 4
Correlation between independent variables.

Independent variable	Percentage of apartments	Percentage of unemployed	Land use mix	Number of bus stations	PageRank centrality
Percentage of apartments	1	0.05	−0.55	−0.16	0.12
Percentage of unemployed	—	1	−0.1	0.43	0.04
Land use mix	—	—	1	0.22	−0.09
Number of bus stations	—	—	—	1	0.01
PageRank centrality	—	—	—	—	1

Table 5
Linear regression results estimating COVID-19 footprints.

Variable	Coefficient	T statistics	P value	Significance level
Intercept	9.299	10.965	$<2\ e^{-16}$	***
Percentage of apartments	7.687	4.815	$6.34\ e^{-6}$	***
Percentage of unemployed	74,010.000	2.525	0.0134	*
Land use mix	0.610	4.240	$5.66\ e^{-5}$	***
Number of bus stations	0.219	2.483	0.015	*
PageRank centrality	0.136	1.784	0.078	—
Number of observations	85.000	—	—	—
Adjusted R^2	0.456	—	—	—

Note: Asterisks indicate significance: "." significant level at 0.1 level, "*" significant at 0.01 level, "**" significant at 0.001 level, and "***" significant at 0.0001 level.

mentioned in previous findings regarding COVID-19 situations in cities.

A better developed understanding of what drives COVID-19 footprints at the local level can help us understand the system, including where to allocate testing, tracking, and medical resources for infectious disease containment. Vulnerable neighborhoods in a transit-reliant city, for example, with high density of apartments and unemployed population, can be the focus of government monitoring with regards to COVID-19 risks. The findings contribute to a better understanding of the factors for COVID-19 footprints; the resulting insights could be used to identify COVID-19 and other infectious disease's hotspot and help with COVID-19 monitoring. Our work also provides some insights for future urban planning and design. For instance, the popular destinations reflected by footprints of the infected infer that some stations and their surroundings might offer better facilities, services and/or opportunities than others to attract a larger quantity of riders/people. The findings indicate inequities in the distribution of and accessibility to essential services and facilities in the city.

Some limitations of the study include that the current node–place–mobility model still represents static subway station situations and human mobility pattern. The dynamic transmission process of infectious disease transmission and its linkages with land use, transportation, and human movement are not thoroughly modeled. Future studies could explore and model time dimension. Also, with more thorough COVID-19 infector epidemiological investigation, multiple regression models can be adopted for different types of locations regarding COVID-19 footprints, such as residential and office regions. Given more detailed mobility data (from GPS or LBS monitoring), future research can focus on building-level models, which gives finer granularity for detecting COVID-19 hotspots in cities and helps with more precise COVID-19 containment and control tactics.

Replication and data sharing

The results and figures in this study can be reproduced with our replication introduction below. Data concerning the number of COVID-19 cases, footprints, number of apartments, bus stations, and land use mix etc. can be obtained freely from Hong Kong government's open data (https://data.gov.hk/en/). Mobility data concerning metro usage is obtained from Hong Kong's MTR, under the MOU between the two organizations and cannot be shared directly. Replication can be made possible with data of similar structures with certain disaggregated OD records, such as smartcard data from Shenzhen Government (https://opendata.sz.gov.cn/interaction/suggest/toSuggestDetails/1485519384033210368). Data processing codes can be found at https://github.com/garychowcmu/nodeplacemobility.

Declaration of competing interest

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

Acknowledgement

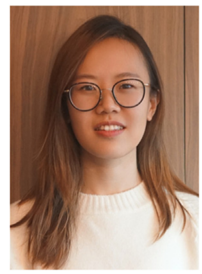
The research underlying this paper is financially supported by the Seed Fund for Basic Research, Platform Technology Funding (URC012530226), Seed Fund for PI Research-Basic Research (URC109000301), and Seed Funding for Strategic Interdisciplinary Research Scheme (URC102010057) at the University of Hong Kong, Hong Kong, China and Guangdong–Hong Kong–Macau Joint Laboratory Program (2020B1212030009).

References

Afrin, S., Chowdhury, F.J., Rahman, M.M., 2021. COVID-19 pandemic: rethinking strategies for resilient urban design, perceptions, and planning. Front. Sustain. Cities 3, 668263.
Awad-Núñez, S., Julio, R., Gomez, J., Moya-Gómez, B., González, J.S., 2021. Post-COVID-19 travel behaviour patterns: impact on the willingness to pay of users of public transport and shared mobility services in Spain. Eur. Transp. Res. Rev. 13, 1–18.
Bertolini, L., 1999. Spatial development patterns and public transport: the application of an analytical model in The Netherlands. Plann. Pract. Res. 14, 199–210.
Bhattacharya, S., Siva Sathya, S., Sharmiladevi, S., 2021. Analyzing the spread of COVID-19 in India through PageRank and diffusion techniques. In: Emerging Technologies in Data Mining and Information Security, pp. 723–733.
Biba, S., Curtin, K.M., Manca, G., 2010. A new method for determining the population with walking access to transit. Int. J. Geogr. Inf. Sci. 24, 347–364.
Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. Comput. Netw. ISDN Syst. 30, 107–117.
Cao, Z., Asakura, Y., Tan, Z., 2020. Coordination between node, place, and ridership: comparing three transit operators in Tokyo. Transport Res D-TR E. 87, 102518.
Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., et al., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 589, 82–87.
Cummings, C., Mahmassani, H., 2022. Does intercity rail station placement matter? Expansion of the node–place model to identify station location impacts on Amtrak ridership. J. Transport Geogr. 99, 103278.
Dou, M., Wang, Y., Dong, S., 2021. Integrating network centrality and node–place model to evaluate and classify station areas in Shanghai. ISPRS Int. J. Geo-Inf. 10, 414.
Gartland, N., Fishwick, D., Coleman, A., Davies, K., Hartwig, A., Johnson, S., et al., 2022. Transmission and control of SARS-CoV-2 on ground public transport: a rapid review of the literature up to May 2021. J. Transport Health 26, 101356.
Hong Kong SAR Government, 2022a. Data in coronavirus disease (COVID-19) in HK (geodatabase). https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent.
Hong Kong SAR Government, 2022b. Archive of statistics on 5th wave of COVID-19. https://www.coronavirus.gov.hk/eng/5th-wave-statistics.html.
Jeffrey, D., Boulangé, C., Giles-Corti, B., Washington, S., Gunn, L., 2019. Using walkability measures to identify train stations with the potential to become transit

oriented developments located in walkable neighbourhoods. J. Transport Geogr. 76, 221–231.
Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2014. Advance transit oriented development typology: case study in Brisbane, Australia. J. Transport Geogr. 34, 54–70.
Kan, Z., Kwan, M.P., Wong, M.S., Huang, J., Liu, D., 2021. Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong. Sci. Total Environ. 772, 145379.
Legislative Council Secretariat of Hong Kong SAR, 2016. Transport Statistical Highlights. https://www.legco.gov.hk/research-publications/english/1617issh06-public-transport-20161028-e.pdf.
Li, B., Peng, Y., He, H., Wang, M., Feng, T., 2021. Built environment and early infection of COVID-19 in urban districts: a case study of Huangzhou. Sustain. Cities Soc. 66, 102685.
Litvinova, M., Liu, Q.H., Kulikov, E.S., Ajelli, M., 2019. Reactive school closure weakens the network of social interactions and reduces the spread of influenza. Proc. Natl. Acad. Sci. U.S.A. 116, 13174–13181.
Lyu, G., Bertolini, L., Pfeffer, K., 2016. Developing a TOD typology for Beijing metro station areas. J. Transport Geogr. 55, 40–50.
Manzira, C.K., Charly, A., Caulfield, B., 2022. Assessing the impact of mobility on the incidence of COVID-19 in Dublin City. Sustain. Cities Soc. 80, 103770.
Megahed, N.A., Ghoneim, E.M., 2020. Antivirus-built environment: lessons learned from Covid-19 pandemic. Sustain. Cities Soc. 61, 102350.
Mo, B., Feng, K., Shen, Y., Tam, C., Li, D., Yin, Y., et al., 2021. Modeling epidemic spreading through public transit using time-varying encounter network. Transport. Res. C Emerg. Technol. 122, 102893.
Nguyen, Q.C., Huang, Y., Kumar, A., Duan, H., Keralis, J.M., Dwivedi, P., et al., 2020. Using 164 million google street view images to derive built environment predictors of COVID-19 cases. Int. J. Environ. Res. Publ. Health 17, 6359.
Park, K., Farb, A., Chen, S., 2021. First-/last-mile experience matters: the influence of the built environment on satisfaction and loyalty among public transit riders. Transport Pol. 112, 32–42.
Rousseeuw, P., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Schlosser, F., Maier, B.F., Jack, O., Hinrichs, D., Zachariae, A., Brockmann, D., 2020. COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. Proc. Natl. Acad. Sci. USA 117, 32883–32890.
Van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N., et al., 2020. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N. Engl. J. Med. 382, 1564–1567.
Xu, G., Jiang, Y., Wang, S., Qin, K., Ding, J., Liu, Y., et al., 2022. Spatial disparities of self-reported COVID-19 cases and influencing factors in Wuhan, China. Sustain. Cities Soc. 76, 103485.
Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., Ukkusuri, S.V., 2020. Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Sci. Rev. 10, 18053.
Zemp, S., Stauffacher, M., Lang, D.J., Scholz, R.W., 2011. Classifying railway stations for strategic transport and land use planning: context matters. J. Transport Geogr. 19, 670–679.
Zhang, Y., Marshall, S., Manley, E., 2019. Network criticality and the node–place–design model: classifying metro station areas in Greater London. J. Transport Geogr. 79, 102485.
Zhou, J., Zhao, Z., Zhou, J., 2022. Quantifying COVID-19 Transmission Risks Based on Human Mobility Data: A Personalized PageRank Approach for Efficient Contact-Tracing. https://doi.org/10.48550/arXiv.2210.01005.
Zhou, J., Wu, J., Ma, H., 2021. Abrupt changes, institutional reactions, and adaptive behaviors: an exploratory study of COVID-19 and related events' impacts on Hong Kong's metro riders. Appl. Geogr. 134, 102504.

Jiali Zhou is a research assistant professor in the Department of Urban Planning and Design at the University of Hong Kong. He holds a B.E. degree from South China University of Technology, a M.S. degree from Carnegie Mellon University (CMU), and a Ph.D. degree from Northeastern University in Boston, Massachusetts. His research interest is in public transportation, transport system resilience, simulation-based decision-making, machine learning, and multi-modal mobility.



Mingzhi Zhou is a Ph.D. candidate from the Department of Urban Planning and Design at the University of Hong Kong. She obtained a M.S. degree from the University of Southern California (USC) and a B.S. degree from the Fudan University (China). Her research focuses on urban resilience, human mobility and travel behaviors based on big data analysis, and accessibility to public services.

Jiangping Zhou is an associate professor in Department of Urban Planning and Design, The University of Hong Kong (HKU). He holds a Ph.D. degree from the University of Southern California, a M.S. degree from the Nanjing University and University of Illinois, Chicago, and a B.S. degree from Tianjin University. His research interests include transport policy and planning, big-data applications and analytics for transport/urban studies, comparative urban and regional studies, and sustainability.

Zhan Zhao is an assistant professor in Department of Urban Planning and Design, The University of Hong Kong (HKU), and also affiliated with HKU Urban Systems Institute and Musketeers Foundation Institute of Data Science. He holds a Ph.D. degree from the Massachusetts Institute of Technology, a M.S. degree from the University of British Columbia, and a B.S. degree from Tongji University. His research interests include human mobility modeling, public transportation systems, and urban data science.