

Urban dynamics through the lens of human mobility

Received: 6 September 2022

Accepted: 6 June 2023

Published online: 10 July 2023

 Check for updates

Yanyan Xu^{1,2,3}, Luis E. Olmos^{2,4}, David Mateo⁵, Alberto Hernando⁵, Xiaokang Yang¹ & Marta C. González^{2,3,6}✉

The urban spatial structure represents the distribution of public and private spaces in cities and how people move within them. Although it usually evolves slowly, it can change quickly during large-scale emergency events, as well as due to urban renewal in rapidly developing countries. Here we present an approach to delineate such urban dynamics in quasi-real time through a human mobility metric, the mobility centrality index ΔKS . As a case study, we tracked the urban dynamics of eleven Spanish cities during the COVID-19 pandemic. The results revealed that their structures became more monocentric during the lockdown in the first wave, but kept their regular spatial structures during the second wave. To provide a more comprehensive understanding of mobility from home, we also introduce a dimensionless metric, KS_{HBT} , which measures the extent of home-based travel and provides statistical insights into the transmission of COVID-19. By utilizing individual mobility data, our metrics enable the detection of changes in the urban spatial structure.

Urbanization is arguably the most important change in recent human history, as it has greatly transformed how people live, work and travel^{1,2}. However, rapid urban expansion can also endanger socioeconomic and environmental well-being, with consequences for livability^{3–5}. For example, economic growth and the production of new inventions scale with city size^{6,7}, but urban expansion can also introduce negative effects such as the exacerbation of regional air pollution and urban heat island effects^{8–10}. During the process of urbanization, the urban structure affects the environmental and energy costs of settlements^{11–13}, and it confines the development of a city into a given space, including the planning of transportation and expansion of the labor market. In light of this, urban designers often seek to regulate the urban structure with long-term sustainable targets in mind.

The evolution of urban spatial structures—urban dynamics—is only monitored slowly, because the complex interaction between population, resources and urban regulations makes the collection of the required information laborious. In urban planning, urban structure is commonly defined using two complementary components:

(1) the spatial distribution of the population^{14,15} and (2) the spatial distribution of functional facilities or working places¹⁶. In the absence of mobility data, researchers generally utilize census data about employment or commuting flows as a proxy for trip attractors¹⁷. However, national or state-level travel surveys are usually insufficient to provide up-to-date information, failing to take into account the occurrences of large-scale emergency events, such as pandemics, natural disasters or socioeconomic crises. These events might lead to relocations and changes in mobility patterns¹⁸, and can reshape the urban structure. It is thus necessary to identify urban dynamics in a quasi-real-time manner, to facilitate timely policy responses and more effective planning.

In the past decade, the collection of massive amounts of mobility data from dwellers has created a powerful tool for urban planning, as data-driven schemes unravel the dynamic interaction between humans and their complex urban systems^{19–21}. However, the sharing of individual trace data has raised public concern about privacy^{22,23}. Here we argue that there are still important insights to gain from individuals' mobility

¹MoE Key Laboratory of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China. ²Department of City and Regional Planning, University of California, Berkeley, CA, USA. ³Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁴Facultad de Ciencias Básicas, Universidad de Medellín, Medellín, Colombia. ⁵Kido Dynamics SA, Lausanne, Switzerland. ⁶Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA. ✉e-mail: martag@berkeley.edu

data, particularly in generating metrics that can quantify instant urban structures in the face of catastrophic events.

In this Article we study multiple cities and explore the potential of defining urban spatial structure with human mobility. We observe that the location of residence with respect to the central business district (CBD) is a decisive factor in determining the scale of mobility. Motivated by this finding, we propose a mobility centrality index, ΔK_S , as a function of the distance of groups of travelers from the CBD. ΔK_S quantifies how strongly people's home-centered radius of gyration, R_g , changes with distance from the city center. The statistical divergence between the R_g values of different groups of people was measured using the Kolmogorov–Smirnov (KS) test. We analyzed 21 cities from America, Asia and Europe (Supplementary Figs. 1 and 2) and observed that our scale-independent metric is capable of assessing a wide range of urban structures, from monocentric to polycentric forms. More specifically, a large ΔK_S indicates a monocentric urban structure, suggesting that people frequently travel to CBDs even if they live in the periphery. In contrast, a more homogeneous distribution of R_g indicates that a city is more polycentric, with a number of commercial centers having emerged. Our observed values of ΔK_S usually range from 0 to 1, with ΔK_S approaching 0 if trips are uniformly distributed across the city.

We next explored the change in the mobility behavior of people during the COVID-19 pandemic in Spain. At the onset of the COVID-19 outbreak throughout the world in early 2020, to assist policymakers and researchers to propose reasonable prevention and control strategies, several location intelligence companies released their mobility data to the academic community^{22–25}. In this Article, using mobile-phone data from 17 million anonymized users in Spain from 2019 and 2020, we reveal the inequitable impacts of the COVID-19 pandemic by looking at the mobility behavior of different income groups. We observe that the population with lower income took longer trips during the first and second waves in 2020, probably reflecting the fact that they could not afford to stay at home. In contrast, in the holiday seasons of April and August 2019, higher-income groups took longer trips.

We also selected 11 cities in Spain to explore their urban dynamics during the COVID-19 outbreak in 2020, as well as the function of mobility behavior and urban dynamics in spreading the epidemic. Based on the daily reported figures of infections, we divided the study period into three phases, the first wave (before 1 May), the quiet phase (from 1 May to 1 July) and the second wave (after 1 July). We found that all of the studied cities became more monocentric during the first wave of COVID-19, indicating that trips with destinations outside the CBDs decreased. In June 2020, most cities' spatial structures recovered to their usual levels and remained stable during the second wave. We also propose a home-based travel indicator, KS_{HBT} , to measure the extent of mobility restrictions during an emergency. KS_{HBT} measures the statistical distance between the actual R_g and the R_g if the entire population stays at or near home. We then quantified the impacts of time-varying mobility behavior and urban structure on the spread of COVID-19. Our results suggest that R_g and KS_{HBT} are the most important factors for modeling the effective reproduction number R_t , followed by the mobility centrality index ΔK_S . During the first wave in particular, mobility variables show a stronger positive impact on R_t . All the variables and notations introduced in this Article are summarized in Supplementary Section 1.

Results

Urban space through the lens of racial and economic status

Spatial segregation curtails the opportunities for people to access necessary resources, such as education, jobs and public facilities. This enlarges income gaps between different groups and even affects health outcomes among different levels²⁶. By analyzing mobility data, we found that cities display socioeconomic segregation patterns, independent of their specific urban structure. Figure 1a presents the distributions of racial and income groups in Boston and Los Angeles in the

United States, as well as a map of Bogotá, Colombia, showing the socioeconomic strata (SES) there. Details of the datasets are described in the Methods. As expected, residents are segregated by race and income in the subject cities. Boston and Los Angeles exhibit a pattern where higher-income groups reside on the periphery while lower-income groups reside closer to the city center. Bogotá displays a different pattern, with higher-income groups (SES 5 and 6) located in the northern part of the city and lower-income groups (SES 1 and 2) on the periphery. The relationship between income levels and distance from the CBD for the three cities is shown in Supplementary Fig. 3.

We next investigated the mobility patterns of mobile-phone users from diverse racial and economic backgrounds. The mobility data used in this Article are reported in the Methods. By defining an information theory segregation index H , interpreted as the population-weighted difference between the diversity of each spatial unit and the diversity of the whole city, we notice that spatial segregation is notably lower during the day than during the night (Supplementary Section 3 and Supplementary Fig. 4a). This highlights how people's residential location plays a dominant role in the spatial segregation in these cities. Interestingly, when we inspect non-work activities, the results show that the majority of visitors to Boston and Los Angeles are from the same racial group as their destinations (Supplementary Fig. 4b–d). Similar patterns are observed in Bogotá in terms of SES. These observations establish that socioeconomic segregation during leisure time captures homophily in mobility choices.

Urban segregation naturally leads to a divergence in mobility behavior^{27,28}. This divergence comes not only from people's residential location, but also from the low availability of jobs and facilities in peripheral regions. Figure 1b–d presents the distributions of R_g for different socio-demographic groups in Boston, Los Angeles and Bogotá, showing that it captures the spatial coverage of each user's daily mobility, centered at the home location²⁹. Considering a user's mobility behavior during a certain period as a sequence of visited locations, then R_g is calculated as

$$R_g = \sqrt{\sum_{i=1}^n \frac{1}{n} (I_i - I_h)^2} \quad (1)$$

where n is the length of the sequence, I_i and I_h are the geographical coordinates of the i th visited location and the home location, respectively. We then average an individual's daily R_g values over the various data-availability periods. The relation between individuals' R_g and household income level for the aforementioned three cities is presented in Supplementary Fig. 5. In conjunction with Fig. 1b–d, we observe that the inhabitants with the highest income tend to take longer trips in Boston and Los Angeles, but this is not the case in Bogotá. In other words, there is no coincidental relation between the distribution of R_g and socioeconomic characteristics in different cities.

During the COVID-19 pandemic, governments around the world imposed travel-restriction policies to different extents, causing economic activities to stagnate and the unemployment rate to rise^{30,31}. Different socioeconomic groups reacted to COVID-19 and the consequent non-pharmaceutical interventions (NPIs) to different degrees. With data from cities across different countries, researchers observed a greater decrease in the mobility of people with higher socioeconomic status³². We confirm this finding at the national scale in Fig. 1e, where, for 17 million anonymized mobile-phone users in Spain, we associate their annual income with daily R_g during 2019 and 2020. We can identify that long-distance trips taken by high-income residents in April 2019 disappeared in the same period in 2020. In addition, during the early stage of COVID-19, the travel distance of high-income groups dropped more steeply than that of low-income groups, as well as in the second peak season from October to December 2020. High-income groups usually take more long-distance trips during tourist seasons than

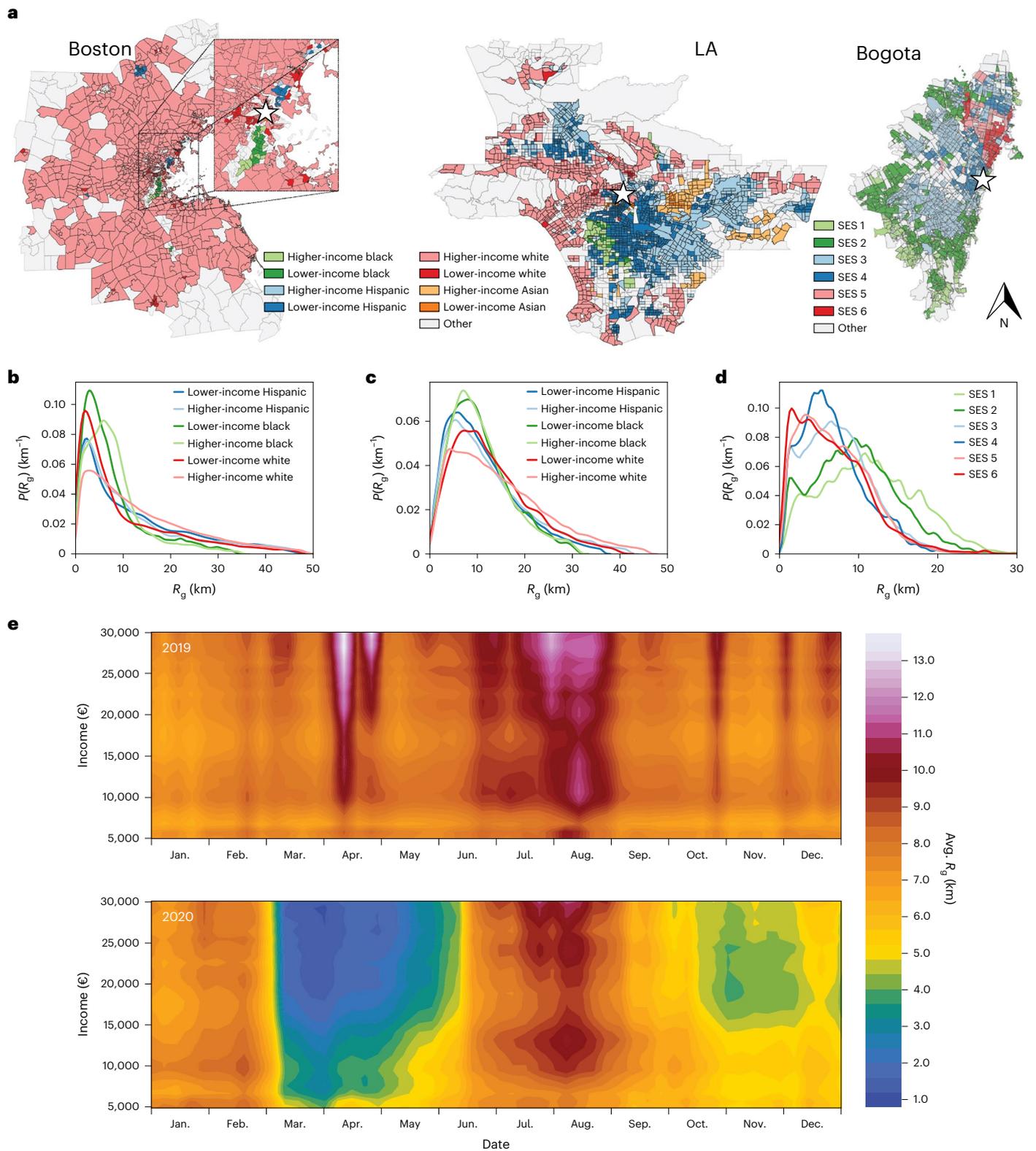


Fig. 1 | Socioeconomic segregation and its relation to the radius of gyration. **a**, Spatial segregation of residents' socioeconomic statuses in Boston, Los Angeles and Bogotá. The white star in each map indicates the location of the CBD. **b–d**, Distribution of R_g for different socioeconomic groups in Boston (**b**), Los Angeles (**c**) and Bogotá (**d**). The sample sizes in Boston, Los Angeles and Bogotá

are 2.58 million, 3.29 million and 4.14 million, respectively. **e**, Relation between annual income and average R_g of the population in Spain during 2019 and 2020, covering the periods before and during the COVID-19 epidemic. The data are from 17 million anonymized mobile-phone users.

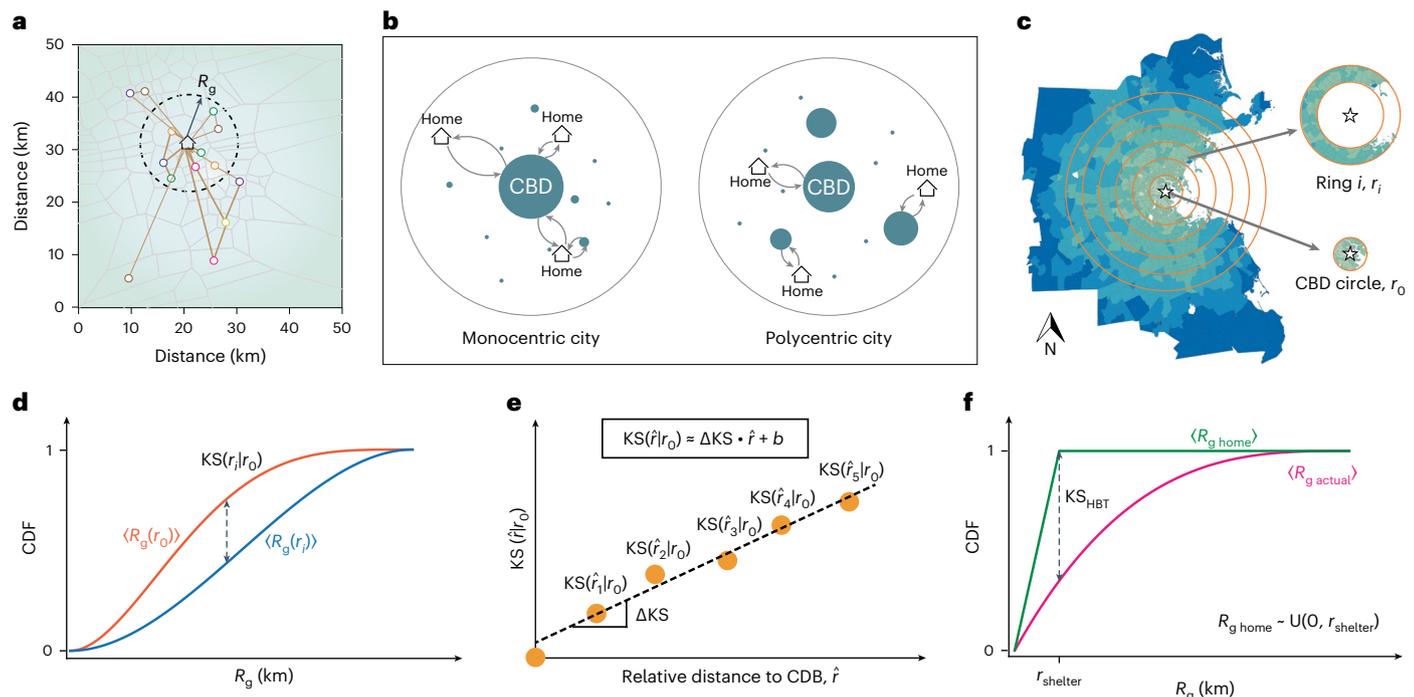


Fig. 2 | Schematic illustration of aggregate metrics relating human mobility to urban structure. **a**, Plot of R_g , showing the spatial coverage of the locations visited by an individual during the period considered. **b**, Illustration of typical mobility behavior in monocentric (left) and polycentric (right) cities. People in monocentric cities tend to travel to the CBD for a job or other resources, whereas in polycentric cities, the CBDs are less attractive and people's travel destinations are distributed in a more dispersed manner in the space. **c**, Illustration of the choropleth map in Boston. The concentric rings represent

incremental radii. The separation of the rings, Δr is set to 3 km for all cities. **d**, Definition of $KS(r_i|r_0)$, the KS statistic, quantifying the distribution disparity of R_g in peripheral ring r_i and CBD circle r_0 . **e**, Definition of the mobility centrality index ΔKS , representing the spatial variation of residents' R_g with respect to their distance to the CBD. **f**, Definition of the home-based-travel indicator KS_{HBT} , referring to the KS statistic between the actual R_g of the population and their R_g under travel restrictions.

low-income groups, but this situation was reversed during the peak seasons of the COVID-19 pandemic. This might be because a higher number of essential trips had to be taken by the low-income population, with fewer of them able to work from home.

Proposed mobility metrics

Humans take trips to work, access resources or attend social activities in complex urban systems. Usually, these trips are centered around residential locations, as shown in Fig. 2a. Meanwhile, cities have organization schemes that can be abstracted as a spectrum ranging from monocentric to polycentric forms. Figure 2b schematically depicts the typical mobility behavior of people residing in monocentric and polycentric cities. In a monocentric city, people with homes closer to the CBD typically have more access to various resources, so they travel over shorter distances than those living in the periphery. However, in a polycentric city, people can also access required resources in sub-centers. We confirmed this empirically with mobility data from multiple cities. Supplementary Fig. 6 depicts the spatial distribution of R_g in Boston, Los Angeles and Bogotá. We note that the southern region of Los Angeles maintains a small R_g , even though it is far away from the CBD. This is consistent with the polycentric urban structure of Los Angeles. As Fig. 2b shows, we can expect monocentric cities to have a more divergent distribution of R_g , whereas the distributions of R_g are expected to be more uniform in polycentric cities due to the dispersed distribution of job opportunities and resources.

Inspired by the work of Bertaud and Malpezzi¹⁵, we defined a sequence of concentric rings of constant width, centered at the CBD, such that a city's population can be divided into multiple groups based

on their residential locations (Fig. 2c). The R_g values of the population in the CBD circle and the i th ring of radius r_i are denoted as $\langle R_g(r_0) \rangle$ and $\langle R_g(r_i) \rangle$, respectively. We observe that, although there is an obvious linear relationship between the average R_g and the radius of the ring in some cities, this is not the case in others (Supplementary Fig. 7). Supplementary Fig. 8 summarizes the average and median values of R_g in the concentric rings for the 21 cities. The results indicate that, regardless of how the cities are organized, individuals residing farther from the CBD tend to take longer trips than those living near the CBD in most cities. From this perspective, R_g , as a function of the distance to the CBD, cannot provide enough information to define the urban spatial structure.

Empirically, as mentioned above, urban structure and residential location are the two major factors in determining mobility radii. Accordingly, we introduce the concept of mobility centrality, an index to quantify the heterogeneity of the mobility scale of the population residing in different areas. To this end, we quantify the divergence between $\langle R_g(r_0) \rangle$ and $\langle R_g(r_i) \rangle$ with the KS statistic, as shown in Fig. 2d:

$$KS(r_i|r_0) = KS(\langle R_g(r_i) \rangle, \langle R_g(r_0) \rangle) = \sup_{R_g} |F_{\langle R_g(r_i) \rangle} - F_{\langle R_g(r_0) \rangle}| \quad (2)$$

where $F_{\langle R_g(r_0) \rangle}$ and $F_{\langle R_g(r_i) \rangle}$ are the cumulative distribution functions of $\langle R_g(r_0) \rangle$ and $\langle R_g(r_i) \rangle$, respectively, and 'sup' is the supremum function. Other metrics to measure the statistical divergence of the R_g values are discussed in Supplementary Section 4.

To represent cities with varying sizes, we scaled each city into a circle of radius 1 with its CBD at the center. Specifically, for each city we first defined r_{max} as the radius of the smallest circle that covers

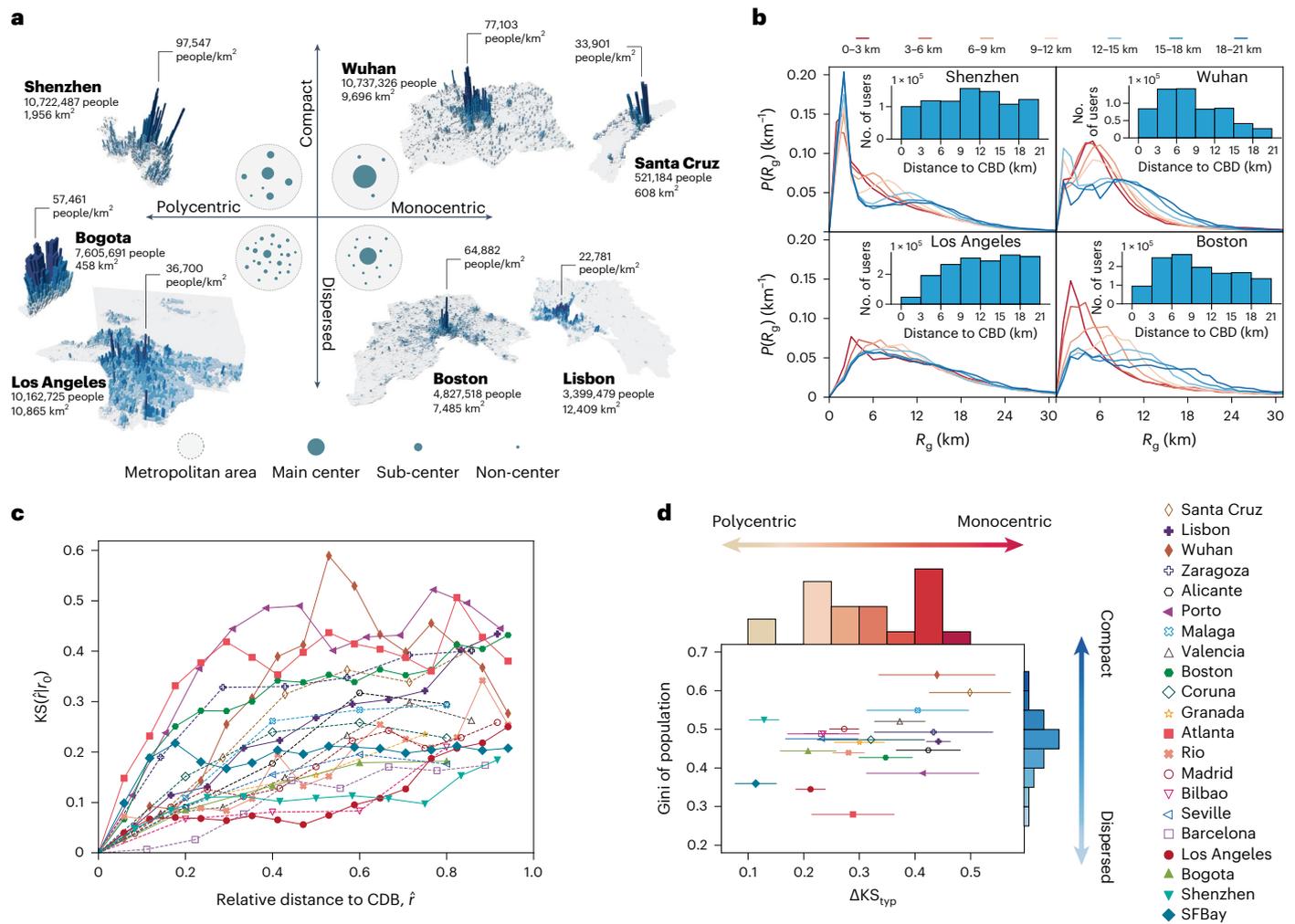


Fig. 3 | Defining urban structure via the mobility behavior of the population.

a, Two dimensions of urban spatial structure. Centrality, ranging from polycentricity to monocentricity, reflects the number of population centers. Concentration, ranging from dispersion to compaction, refers to the degree of heterogeneity or dispersion in the amount of population shared outside the CBD³⁴. The population distribution of some sample cities are visualized with 3D barplots. **b**, Distribution of R_g values in seven rings in Shenzhen, Wuhan,

Los Angeles and Boston. Insets: numbers of users in each ring. The total numbers of users in the seven rings are 868,584, 601,904, 1,747,318 and 1,267,113 for the four cities, respectively. **c**, Relation between mobility distribution changes $\Delta KS(r|r_0)$ in peripheral rings and their relative radii to the CBD. **d**, Distribution of cities in the two-dimensional space defined by ΔKS_{typ} and the Gini index of population. The error bar for each city shows the standard error of linear fitting for the ΔKS calculation. Cities in the legend are shown in descending order of ΔKS .

at least 95% of the total population. The relative radius \hat{r} was then defined as the ratio between the actual radius r and r_{max} . Next, as illustrated in Fig. 2e, we fit the relation between $\langle KS(\hat{r}_i|r_0) \rangle$ and $\langle \hat{r}_i \rangle$ with a linear function, then named the slope ΔKS as the mobility centrality index to assess the urban structure. Note that we can expect ΔKS to approach 0 if all of the required resources are uniformly distributed across the city, whereas it is large if people's distance to the CBD heavily impacts their mobility behavior. In this context, a larger ΔKS means that the scale of people's mobility increases faster in relation to their distance to the CBD, suggesting a more monocentric urban structure.

As a major driving force of the spread of the virus, human mobility received a great deal of attention from researchers and policymakers during the early stage of the COVID-19 pandemic³³. In these works, researchers generally measured the effect of a shelter-in-place order via the aggregated travel flow or the average distance of displacement^{22,25}, while keeping individual trajectories private. However, disparate distributions of travel behavior may share the same flow or average distance. In addition, flow and distance depend greatly on a city's population and scale, hindering the comparison of mobility behavior across regions.

To remedy this, we proposed KS_{HBT} , namely the extent of home-based travel, to assess the effect of shelter-in-place behavior. KS_{HBT} refers to the KS distance between (1) the distribution of observed R_g of all individuals, including $R_g = 0$, and (2) the distribution of R_g if all individuals adhered to the shelter-at-home order. For the latter, we assume that R_g follows a uniform distribution between 0 and a predefined threshold $r_{shelter} \cdot r_{shelter}$ refers to the mobility radius of people following the travel-restriction order. As can be seen in Fig. 2f, KS_{HBT} can be explained as the share of the population with R_g over $r_{shelter}$. Therefore, a lower KS_{HBT} means that more people stay at or near their homes.

Urban dynamics via individual mobility metrics

Following urban study research, in Fig. 3a we present a two-dimensional space of urban spatial structure³⁴, where one dimension quantifies the population distribution, ranging from dispersion to compaction, and the other quantifies the centrality of functional units in space, ranging from polycentricity to monocentricity. The population is shown for blocks with areas of ~1 km² and comes from the LandScan dataset³⁵. A visualization of all studied cities is presented in Supplementary Fig. 2.

Figure 3b presents the distribution of R_g in each ring (3-km width) for four cities (Shenzhen, Wuhan, Los Angeles and Boston). Results for other cities are presented in Supplementary Fig. 9. For Wuhan and Boston, as the distance from the CBD increases, the R_g distribution shifts to a larger average value. Meanwhile, for Shenzhen and Los Angeles, R_g displays a similar distribution, agnostic of the distance from the CBDs. This indicates that, from the perspective of mobility, Wuhan and Boston are both monocentric. Both cities have large proportions of trips attracted to their CBDs, although the population distribution in Wuhan is more disperse than that of Boston. Shenzhen and Los Angeles are both polycentric cities, as people's movement is dispersed among multiple destinations, although the population distributions of the two cities are different. Shenzhen's population distribution is more compact, with the majority of residents living within the CBD circle, whereas Los Angeles has a more uniform distribution of population across the city. We can thus expect that the statistical divergence of R_g may indicate polycentricity and complements well the information drawn from population distributions.

Figure 3c presents the $KS(\hat{r}|r_0)$ of 21 cities. Note that, as we are measuring the disparity of mobility behavior, we exclude users who did not move during one day ($R_g = 0$). A constant value as a function of \hat{r} suggests that the daily mobility of the population residing in different regions appears to have similar R_g distributions. We then calculate the mobility centrality index ΔKS to capture the change of $KS(\hat{r}|r_0)$ in space (shown in Fig. 2e), and use ΔKS_{typ} to represent its typical value over long-term periods of observations, varying from one to six months (details in the Methods). The results of the fit are presented in Supplementary Fig. 12, and the goodness-of-fit is presented in Supplementary Table 1. We observe a relatively high r^2 value in most cities, except for the San Francisco Bay Area (SFBay), which has an irregular spatial topology in comparison with the others.

Regarding the population distribution, we use the Gini index to measure the uniformity across urban blocks with population over 500 (ref. 5). A larger Gini value indicates that a city is more compact in its population distribution, and a lower value means a city's population is more dispersed. In Fig. 3d we present the Gini index versus ΔKS_{typ} and find that ΔKS increases mildly with the index, in line with the intuition that more compact cities tend to be more monocentric (Supplementary Fig. 13). For example, Wuhan, where the COVID-19 pandemic began in December 2019, is the most compact in terms of Gini index. The ΔKS_{typ} of Wuhan is also the second largest, after Santa Cruz. In comparison with Wuhan, cities like Porto, Alicante, Lisbon and Zaragoza have similar ΔKS_{typ} values but are more dispersed in their population distributions, suggesting that they may be undergoing processes of urban sprawl^{36,37}. As a crowded city, Shenzhen's population peaks (compactly) in several districts, and the uniform distributions of job opportunities and resources shape it into a polycentric form. Other polycentric regions, such as SFBay and Los Angeles, have dispersed population distributions.

In the analysis above, note that we set the widths of the rings to be constant at 3 km. We tested our definition of urban structures with different ring widths (Supplementary Fig. 14). The results show that, although the 3-km width performs slightly better than 1- and 2-km widths in terms of the r^2 of the linear fit, the different ring widths did not affect the conclusions drawn from the results (Supplementary Fig. 14c–g).

Sensing urban dynamics in times of crisis

During February 2020, cities in Spain were successively hit by the COVID-19 pandemic, and the government imposed a nationwide mandatory lockdown on 14 March 2020. Figure 4a shows the daily infection incidence in each province per 1,000 people before 30 September 2020. The numbers of newly confirmed infections are presented in Supplementary Fig. 15. The study period can be broken down into three phases: phase I covers the first wave of COVID-19 before 30 April, phase

II covers the quiet period from 1 May to 30 June, and phase III covers the second wave after 1 July. We use individual mobile-phone data collected from February to September 2020 to dissect the changes in mobility behavior and urban dynamics during this period in 11 major cities in Spain. In Supplementary Fig. 16 we show the median R_g for each city and find that during the lockdown this is nearly 0.5 km. We thus set $r_{shelter}$ as 0.5 km in the definition of KS_{HBT} .

With the observed individual R_g values, we calculated the daily metrics $\overline{R_g^d}$, ΔKS^d and KS_{HBT}^d per city, as depicted in Supplementary Fig. 17. All three mobility metrics display a weekly periodicity. To curb the behavioral fluctuation caused by weekdays, we also averaged each metric over seven consecutive days, namely $\overline{R_g^{7d}}$, ΔKS^{7d} and KS_{HBT}^{7d} , as presented in Supplementary Fig. 18. We present the changes of the three variables in Fig. 4b–d, in relation to their values in the first week of February 2020. The relative change in ΔKS is formulated as $(\Delta KS^{7d} - \Delta KS_{Feb.1-7}^{7d}) / \Delta KS_{Feb.1-7}^{7d} \times 100\%$. As can be seen, the three metrics changed after the mandatory lockdown and started to recover to normal levels once the shelter-in-place order was relaxed in June.

Figure 4b presents the dynamic evolution of the urban structure in each city. In the initial stage of COVID-19, people kept commuting as usual and amenities were open. Once the lockdown began, ΔKS increased in all cities to varying degrees. This meant that all cities became monocentric during lockdown, particularly Barcelona, Seville and Madrid, which were polycentric before the pandemic. More importantly, their shift to monocentricity implies that trips to destinations outside the CBDs were reduced. After late June, travel restrictions were lifted and economic activities gradually resumed. The country began to transition into a 'new normality', accompanied by other NPIs. In this context, urban structures recovered to their normal levels in most cities, with Alicante being a notable exception. Alicante's polycentricity became stronger than before COVID-19. This might have been caused by the migration of second-home owners. In the early stage of the pandemic, second-home owners tended to migrate from crowded cities to low-density areas. Therefore, in Alicante, an increase in the number of second-home owners might have occurred after travel restrictions were lifted³⁸.

Figure 4c shows that, when the number of infections gradually increased at the end of February, there were notable fluctuations in $\overline{R_g}$ for several cities (for example, Seville, Malaga, La Coruna and Granada). This is mainly caused by the emerging long-distance mobility as a response to future lockdowns. In Fig. 4d, we can see that our shelter-in-place indicator KS_{HBT} is very similar for all cities. At the end of phase II, all mobility metrics resumed their normal pre-lockdown levels. The resumption of KS_{HBT} occurred much earlier than that of $\overline{R_g}$. Moreover, as the spatial range of mobility depends not only on the urban structure but also on the spatial scale, the studied cities display varying levels of $\overline{R_g}$ in phase III. In contrast, KS_{HBT} displays a more stable behavior during the entire period, so it might be a more reliable measure to capture mobility restrictions, agnostic of city scale.

Next, we explored the role of urban dynamics and mobility behavior during the spread of COVID-19 across the studied cities. To this end, we estimated the effective reproduction number R_t using the EpiEstim R package³⁹ (Supplementary Section 5). Generally, R_t started to be high at the beginning of COVID-19 and then decreased below 1.0 about one week after mandatory lockdowns. Afterward, R_t fluctuated in June, indicating that the second wave was coming. Next, we used gradient boosting machine (GBM)⁴⁰ to model R_t with the mobility and population variables described before, and assessed each predictor's impact using the SHAP package⁴¹. Figure 4e–g presents each feature's SHAP values for each of the COVID-19 phases. The conjunction of high feature value (in red) and positive SHAP value implies that the impact is salient and positive. The mixture of red and blue dots implies an indeterminate impact on R_t . The insets in these figures present the relative importance of each factor, measured by the ratio between the means of its absolute SHAP values and the average R_t in the phase (Methods).

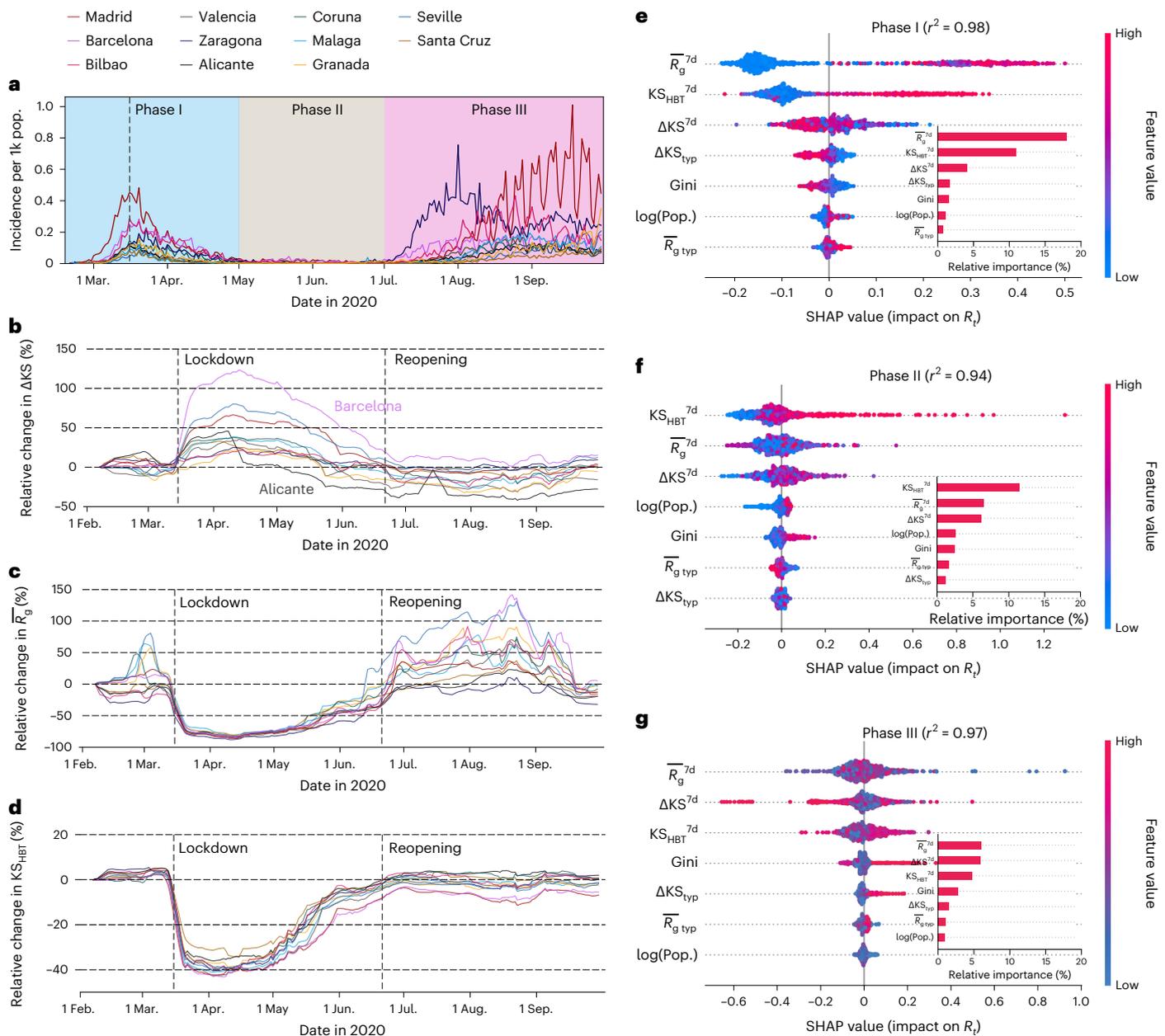


Fig. 4 | Changes in urban spatial structure and mobility behavior during the COVID-19 pandemic in the 11 Spanish cities. a, Daily reported infections per 1,000 population in each province before 30 September 2020. Nationwide lockdown was imposed on 14 March, marked with the dashed line. We divide the entire period into three phases: the first wave, a quiet period and the second wave. **b**, Relative change of mobility centrality index ΔKS , with respect to the first week

in February. **c**, Relative change of mobility scale measure by \bar{R}_g . **d**, Relative change of the home-based-travel indicator KS_{HBT} . **e–g**, Impacts of mobility behavior on the effective reproduction number R_t during phases I (**e**), II (**f**) and III (**g**). The scatter plots display the impact of each factor on R_t on a given day and in one city. Insets: the relative importance of each factor, represented by the ratio between means of their absolute SHAP values and the average R_t during each phase.

As illustrated in Fig. 4e, the mobility variables (\bar{R}_g and KS_{HBT}) have stronger and positive impacts on R_t , because phase I covers the early stage of COVID-19 as well as the mandatory lockdown period. In this phase, \bar{R}_g shows the highest impact, followed by KS_{HBT} , while the impact of urban dynamics is relatively weak. These observations confirm that in phase I when people traveled more frequently and over longer distances, the epidemic spread faster. Figure 4f shows the results in the quiet phase II. In conjunction with the insets, we notice that KS_{HBT} still shows a considerable and positive impact on R_t but the impact of \bar{R}_g is ambiguous and weakened. Interestingly, we observe that the weight of the urban structure indicator ΔKS becomes more important in comparison with phase I. In the results of phase III (the second COVID-19

wave) shown in Fig. 4g, all variables became less significant because non-mobility NPIs were the major protection measures.

Discussion

This Article seeks to understand urban dynamics by studying how people interact with their resources in space. We used the radius of gyration of individual traces in the calculation of the proposed mobility metrics. However, to address privacy concerns, most smartphone platforms are moving to anonymized aggregate data collection (k-anonymity and so on). In this context, although R_g only presents the mobility scale of one user, without any private information, we needed the individual trajectory of each user to calculate their R_g . It was thus not possible

to directly calculate R_g (Fig. 2a) with k-anonymity. To solve this issue, a specially designed k-anonymity strategy may be required, or the mobile operator could collect R_g from the smartphone, rather than people's mobility traces.

This study offers several promising avenues for further investigation. Although we have observed that low-income individuals traveled longer distances than their high-income counterparts during the initial COVID-19 outbreak, the purposes of these trips remain unclear. This knowledge gap limits our understanding of the unequal impact of lockdown measures. To address this, future research could leverage long-term mobility traces to identify essential workers and explore their behavior during emergency events, with a focus on socioeconomic equity.

Overall, our work highlights the value of fine-grained individual mobility data for quantifying urban dynamics. Near-real-time sensing of urban dynamics is more important in large-scale emergency events like pandemics or natural disasters to better plan for the reopening or reconstruction of cities.

Methods

Data description

Selection of city boundaries. The selection of city boundaries impacts the spatial centrality of the population distribution and also the values of urban spatial structure metrics in a city. In this study we used the metropolitan region of each city. For big cities in the United States and Portugal we used their metropolix boundaries from census data, including multiple connected cities and towns. For Spanish cities we chose the urban areas defined by the AUDES initiative (Areas Urbanas De España)⁴². For cities in China and Latin America we chose areas enclosed by their administrative boundaries. As for the CBDs, we selected the CBD in each city, by convention, by their locations in Google Maps. Note that, as we are defining a CBD circle of radius 3 km, the geographical coordinates of the CBD do not have to be very precise. The selected boundaries, the areas and the location of the CBDs of the 21 cities are illustrated in Supplementary Fig. 1 at the same scale, in descending order of area. The shapefiles of the boundaries are available at <https://github.com/humnetlab/UrbanForm/tree/main/data/Geo/Cities/>.

Demographic and socioeconomic data. We used block-level population data (2015) from LandScan for all 21 cities³⁵. LandScan provides global population data at ~1-km resolution (30" × 30"). We selected blocks with a population over 500 to calculate the Gini index. Three-dimensional barplots of the population distributions of the 21 cities studied are presented in Supplementary Fig. 2, as well as their total populations and Gini indices.

To compare the residents of different socio-demographic groups in US urban areas, we classified tracts into lower income and higher income based on a household's median income being less than or more than the national middle-class threshold of US\$45,200. Similarly, we classified tracts as majority non-Hispanic white, non-Hispanic black, Hispanic or Asian using a threshold of 50%. Racial and economic data were taken from the 2015 American Community Survey (ACS 2015) aggregated at the census tract level⁴³. Because, in Boston, there were too few Asian tracts to permit reliable analyses for that group, we decided not to include it in this analysis. In Bogotá, Colombia, we used the official socioeconomic stratification as a proxy for income level. The department of city planning (DCP) has assigned socioeconomic strata (SES, ranging from 1 to 6, representing income levels from lowest to highest) at the census block level in the metropolitan area. The socioeconomic level of each census block is determined by a DCP official, who relies on direct observation of the block and its surroundings and must take into account several factors in their assessment²⁷.

Mobility data. We leveraged travel survey data, including home locations and daily visited locations, for 567,301 and 444,127 users in Atlanta

and Rio de Janeiro, respectively. For the other cities we used mobile-phone data to measure the mobility behavior of the population. Mobile-phone data were passively collected by mobile operator carriers when anonymized users' cell phones interacted with mobile base towers (for example, antennas). Each record provided the anonymized user ID, the time and the geographical location of cell phone use. The distributions of the mobile base towers are presented in Supplementary Fig. 1. The spatial resolution of the localization of mobile-phone users ranged from 50 m in densely populated areas to 300 m in sparse areas. These data are usually referred to as call detail records (CDRs). The recorded activities of mobile-phone users in the CDRs include making or receiving phone calls and sending or receiving text messages. For the Chinese cities and 11 Spanish cities, besides calling and messaging, data-using activities were also recorded in the mobile-phone data (termed 'extended detail records', XDRs). XDRs are more frequent than CDRs and can be directly used to capture human mobility.

For cities using CDRs, for example, Boston, SFBay, Los Angeles, Bogotá, Lisbon and Porto, we apply the TimeGeo framework to model individuals' travel behavior at fine granularity (for example, every 10 min) with CDR data⁴⁴. TimeGeo extracts stay points from each individual's sequence of records, and separates commuters from non-commuters by checking users' work locations. For each census tract (or equivalent unit area) in the metropolitan area, expansion factors are calculated for commuters and non-commuters using census data. Based on the distributions of the empirical mobility parameters extracted from the active user data, a simulation of how the entire urban population moves is achieved. The simulation results of TimeGeo for some of the cities have been validated with travel survey data in previous work^{21,44,45}. As the home locations of users are not available in either CDRs or XDRs, we identified users' most frequently visited location during weekends and weekday nights as their home locations.

Periods of mobile-phone data. The observation periods for the mobile-phone data ranged from one month to six months in the studied cities. The period of the data collection was six weeks in Los Angeles and SFBay in October and November 2012, two months in Boston in February and March 2010, six months in Bogotá across 2013 and 2014, one month in Shenzhen in October 2013, two months in Wuhan in August and September 2012, and six months in Lisbon and Porto across 2006 and 2007. Note that, for the Spanish cities, we had access to the mobile-phone data during two periods, the month of October 2019 and from 1 February to 30 September 2020. The data in October 2019 were used to identify the mobility behavior during a typical period before the outbreak of the COVID-19 pandemic, and the data from February to September 2020 were used to identify the impact of urban form and mobility on the spreading of COVID-19, covering the first wave and rising stage of the second wave. Note that, although the periods of the mobile-phone data are not exactly matched with the population data, this does not affect our analysis, because the urban structure for one city remains almost unchanged over a period of a few years.

Estimate of R_t from reported cases

The EpiEstim R package was developed by Cori et al.³⁹ and has been adopted to estimate the transmission intensity of SARS-CoV-2 in various countries^{46,47}. EpiEstim provides a way to measure the reproduction number R_t of an epidemic based on the daily number of new infections. Due to delayed reports and the limited accuracy of epidemiological data, several hypotheses are involved. First, EpiEstim uses a Bayesian inference framework to estimate the posterior probability of R_t , assuming its prior probability as a Gamma distribution. Based on the results reported by Imai et al.⁴⁸, we used a prior Gamma distribution for $R_{t,\tau}$ with a mean of 2.6 and standard deviation of 2.0. Second, EpiEstim requires the distribution of the serial interval, which means the time difference between the onset of the symptoms of a primary case and the corresponding secondary cases. This work follows observations from

ref. 49 and assumes the serial interval to be a discrete Gamma distribution, with a mean of 3.6 days and a standard deviation of 4.9 days. Third, to achieve stable estimation, we used a seven-day ($\tau = 7$ days) time window before day t to calculate R_t for each province in Spain. The estimated R_t and the 95% confidence interval in each city are illustrated in Supplementary Fig. 19. More details about the implementation of EpiEstim are provided in Supplementary Section 5 and ref. 39.

Assessing feature importance with SHAP

For the selected Spanish cities, we adopted the SHAP package to assess the impacts of the mobility and urban dynamics variables on the effective reproduction number R_t , which serves as the proxy for how COVID-19 spreads. SHAP is a game-theoretic approach to explain the output of a given machine-learning model⁴¹. To build a regression model for R_t , we utilized the GBM implemented by the LightGBM package⁴⁰. More specifically, we collected the time-varying variables (\bar{R}_g^{7d} , KS_{HBT}^{7d} , ΔKS^{7d}) and constant variables ($\bar{R}_{g_{typ}}$, ΔKS_{typ} , logarithmic value of the total population, and Gini value of population distribution) in the cities, using them as the input to the GBM model, and set the corresponding R_t as the output. With the GBM model, R_t can be well modeled in each phase of COVID-19 with high r^2 values (0.98, 0.94 and 0.85 for the three phases).

Given this well-trained GBM model, we can use the SHAP package to quantify how the factors in the GBM model contribute to the prediction of R_t for each city on a daily basis. The quantification was done with a selected SHAP value, as shown in Fig. 4e–g. A large absolute SHAP value implies that the prediction of R_t would be more sensitive to the given factor in the GBM regression model. In these charts, the x axis shows the SHAP values of the features in the modeling of R_t , and the y axis shows all the mobility and urban variables we input into GBM. Each colored point on the chart indicates the SHAP value of the feature in the prediction of R_t . Red and blue colors imply higher and lower values of a feature, respectively. Thus, we can read features' directionality impact on R_t in the well-trained GBM model based on the distribution of the red and blue dots. Taking the SHAP values of \bar{R}_g^{7d} in Fig. 4e as an example, we can see that a higher value of \bar{R}_g^{7d} leads to higher R_t in the GBM model, and a lower value of \bar{R}_g^{7d} leads to a lower value of R_t . In other words, \bar{R}_g^{7d} shows a clearly positive relation with the spread of COVID-19 in phase I. Next, for each phase of COVID-19, we aggregated the SHAP values of all samples to assess the importance of each factor holistically, as shown in the insets of Fig. 4e–g. To obtain the relative importance of each factor, we calculated the mean of the absolute SHAP values and then divided it by the average R_t for that phase.

Data availability

All data needed to evaluate the conclusions in the paper are described in the paper and the Supplementary Information. For contractual and privacy reasons, we cannot make the raw mobile-phone data available. One can contract Kido Dynamics SA to try to get access to the raw mobile-phone data. A sample of the data is available in ref. 50. Source data are provided with this paper.

Code availability

The implementation of this work is available at GitHub (https://github.com/humnetlab/Urban_Dynamics) and Zenodo⁵¹.

References

- Bettencourt, L. M. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
- Batty, M. A theory of city size. *Science* **340**, 1418–1419 (2013).
- Batty, M. The size, scale and shape of cities. *Science* **319**, 769–771 (2008).
- Keuschnigg, M., Mutgan, S. & Hedström, P. Urban scaling and the regional divide. *Sci. Adv.* **5**, eaav0042 (2019).
- Xu, Y., Olmos, L. E., Abbar, S. & González, M. C. Deconstructing laws of accessibility and facility distribution in cities. *Sci. Adv.* **6**, eabb4112 (2020).
- Bertaud, A. *The Spatial Organization of Cities: Deliberate Outcome or Unforeseen Consequence?* IURD Working Paper 2004-01 (Univ. California, 2004).
- Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306 (2007).
- Ewing, R. & Rong, F. The impact of urban form on U.S. residential energy use. *Housing Policy Debate* **19**, 1–30 (2008).
- Lamsal, L., Martin, R., Parrish, D. & Krotkov, N. Scaling relationship for NO₂ pollution and urban population size: a satellite perspective. *Environ. Sci. Technol.* **47**, 7855–7861 (2013).
- Li, D. et al. Urban heat island: aerodynamics or imperviousness? *Sci. Adv.* **5**, eaau4299 (2019).
- Anderson, W. P., Kanaroglou, P. S. & Miller, E. J. Urban form, energy and the environment: a review of issues, evidence and policy. *Urban Studies* **33**, 7–35 (1996).
- Tsekeris, T. & Geroliminis, N. City size, network structure and traffic congestion. *J. Urban Econ.* **76**, 1–14 (2013).
- Kaza, N. Urban form and transportation energy consumption. *Energy Policy* **136**, 111049 (2020).
- Clark, C. Urban population densities. *J. R. Stat. Soc. A (Gen.)* **114**, 490–496 (1951).
- Bertaud, A. & Malpezzi, S. *The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition*, 54–55 (The Center for Urban Land Economics Research, Univ. Wisconsin, 2003).
- Pereira, R. H. M., Nadalin, V., Monasterio, L. & Albuquerque, P. H. Urban centrality: a simple index. *Geogr. Anal.* **45**, 77–89 (2013).
- Sohn, J. Are commuting patterns a good indicator of urban spatial structure? *J. Transport Geogr.* **13**, 306–317 (2005).
- Acosta, R. J., Kishore, N., Irizarry, R. A. & Buckee, C. O. Quantifying the dynamics of migration after Hurricane Maria in Puerto Rico. *Proc. Natl Acad. Sci. USA* **117**, 32772–32778 (2020).
- Calabrese, F., Ferrari, L. & Blondel, V. D. Urban sensing using mobile phone network data: a survey of research. *ACM Comput. Surveys* **47**, 1–20 (2014).
- Olmos, L. E., Çolak, S., Shafiei, S., Saberi, M. & González, M. C. Macroscopic dynamics and the collapse of urban traffic. *Proc. Natl Acad. Sci. USA* **115**, 12654–12661 (2018).
- Xu, Y., Çolak, S., Kara, E. C., Moura, S. J. & González, M. C. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Energy* **3**, 484–493 (2018).
- Grantz, K. H. et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat. Commun.* **11**, 4961 (2020).
- Oliver, N. et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 (2020).
- Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2020).
- Alessandretti, L. What human mobility data tell us about COVID-19 spread. *Nat. Rev. Phys.* **4**, 12–13 (2022).
- Bor, J., Cohen, G. H. & Galea, S. Population health in an era of rising income inequality: USA, 1980–2015. *Lancet* **389**, 1475–1490 (2017).
- Florez, M. A. et al. Measuring the impact of economic well being in commuting networks—a case study of Bogota, Colombia. *Proc. Transportation Research Board 96th Annual Meeting Paper no. 17-03745* (Transportation Research Board, 2017).
- Wang, Q., Phillips, N. E., Small, M. L. & Sampson, R. J. Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc. Natl Acad. Sci. USA* **115**, 7735–7740 (2018).

29. González, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
30. Bonaccorsi, G. et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 15530–15535 (2020).
31. Verschuur, J., Koks, E. E. & Hall, J. W. Observed impacts of the COVID-19 pandemic on global trade. *Nat. Hum. Behav.* **5**, 305–307 (2021).
32. Brough, R., Freedman, M. & Phillips, D. C. Understanding socioeconomic disparities in travel behavior during the COVID-19 pandemic. *J. Reg. Sci.* **61**, 753–774 (2021).
33. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
34. Meijers, E. J. & Burger, M. J. Spatial structure and productivity in U.S. metropolitan areas. *Environ. Planning A Econ. Space* **42**, 1383–1402 (2010).
35. LandScan Global Population Database (ORNL, 2017); <http://web.ornl.gov/sci/landscan/>
36. Martins, H. Urban compaction or dispersion? An air quality modelling study. *Atmos. Environ.* **54**, 60–72 (2012).
37. Rubiera Morollón, F., González Marroquín, V. M. & Pérez Rivero, J. L. Urban sprawl in Spain: differences among cities and causes. *Eur. Planning Studies* **24**, 207–226 (2016).
38. Zoğal, V., Domènech, A. & Emekli, G. Stay at (which) home: second homes during and after the COVID-19 pandemic. *J. Tourism Futures* **8**, 125–133 (2020).
39. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
40. Ke, G. et al. LightGBM: a highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (eds Guyon, I. et al.) (2017).
41. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
42. AUDES project (ESI, 2022); <http://alarcos.esi.uclm.es/per/fruiz/audes/>
43. Census data (United States Census Bureau, 2016); <https://www.census.gov/data.html>
44. Jiang, S. et al. The TimeGeo modeling framework for urban mobility without travel surveys. *Proc. Natl Acad. Sci. USA* **113**, E5370–E5378 (2016).
45. De Nadai, M., Xu, Y., Letouzé, E., González, M. C. & Lepri, B. Socio-economic, built environment and mobility conditions associated with crime: a study of multiple cities. *Sci. Rep.* **10**, 13871 (2020).
46. Abbott, S. et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res.* **5**, 112 (2020).
47. Nouvellet, P. et al. Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1090 (2021).
48. Imai, N. et al. *Report 3: Transmissibility of 2019-nCoV*, 625 (Imperial College London, 2020).
49. Ryu, S., Kim, D., Lim, J.-S., Ali, S. T. & Cowling, B. J. Serial interval and transmission dynamics during SARS-CoV-2 delta variant predominance, South Korea. *Emerg. Infect. Dis.* **28**, 407–410 (2022).
50. Xu, Y. et al. Sample data for the paper titled urban dynamics through the lens of human mobility (Zenodo, 2023); <https://doi.org/10.5281/zenodo.8001784>
51. Xu, Y. et al. Source code for the paper titled urban dynamics through the lens of human mobility (Zenodo, 2023); <https://doi.org/10.5281/zenodo.8001855>

Acknowledgements

We thank P. Wang for the data provided. This work was supported by Berkeley DeepDrive (BDD) and the ITS-SB1 Berkeley Statewide Transportation Research Program. Y.X. and X.Y. were also supported by the National Natural Science Foundation of China (62102258), the Shanghai Pujiang Program (21PJ1407300), the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102) and the Fundamental Research Funds for the Central Universities.

Author contributions

All contributors who fulfill the authorship criteria are listed as co-authors in this paper. Other contributors who do not meet all criteria for authorship are listed in the Acknowledgements. Y.X., L.E.O., X.Y. and M.C.G. conceived the research and designed the analyses. D.M. and A.H. processed the Spanish data. Y.X. and L.E.O. performed the analyses. M.C.G. and Y.X. wrote the paper. M.C.G. supervised the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43588-023-00484-5>.

Correspondence and requests for materials should be addressed to Marta C. González.

Peer review information *Nature Computational Science* thanks Elsa Arcaute, Benjamin F. Maier, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: Fernando Chirigati, in collaboration with the *Nature Computational Science* team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023