scientific **data**

OPEN

DATA DESCRIPTOR

Mobility networks in Greater Mexico City

Marisol Flores-Garrido[1], Guillermo de Anda-Jáuregui[2,3], Plinio Guzmán[4], Amilcar Meneses-Viveros[5], Alfredo Hernández-Álvarez[6], Erika Cruz-Bonilla[7] & Maribel Hernández-Rosales[7]✉

Based on more than 11 billion geolocated cell phone records from 33 million different devices, daily mobility networks were constructed over a 15-month period for Greater Mexico City, one of the largest and most diverse metropolitan areas globally. The time frame considered spans the entire year of 2020 and the first three months of 2021, enabling the analysis of population movement dynamics before, during, and after the COVID-19 health contingency. The nodes within the 456 networks represent the basic statistical geographic areas (AGEBs) established by the National Institute of Statistics, Geography, and Informatics (INEGI) in Mexico. This framework facilitates the integration of mobility data with numerous indicators provided by INEGI. Edges connecting these nodes represent movement between AGEBs, with edge weights indicating the volume of trips from one AGEB to another. This extensive dataset allows researchers to uncover travel patterns, cross-reference data with socio-economic indicators, and conduct segregation studies, among other potential analyses.

Background & Summary

Today's cities and the lifestyles they offer are greatly influenced by mobility. Employing complex networks to depict a city's mobility has emerged as a powerful method for modeling population dynamics across various scales and discerning collective travel trends[1,2]. The daily commutes of urban residents are intricately linked to the city's attributes, including its transportation network, established commercial hubs, areas hosting schools, workplaces, recreational facilities, and medical services, among others. Mobility networks are pivotal in visualizing interactions among various urban areas and establishing spatial patterns. They offer significant utility for decision-making, strategizing for risk scenarios, detecting vulnerabilities, and gaining deeper insights into the unique characteristics of localities within urban landscapes.

Furthermore, these urban mobility networks have a broad spectrum of applications, encompassing the examination of mobility behavior[3–6], optimization of traffic flow[7,8], management of parking demand[9,10], promotion of sustainable transport[11,12], evaluation of environmental impact[13,14], event planning[15–21], infrastructure development[22,23], and public transportation planning[24–30]. These networks contribute to understanding mobility patterns, traffic optimization, sustainability promotion, and environmental impact analysis, and they inform event management, infrastructure planning, and optimization of public transportation systems. Moreover, mobility networks find applications in socio-economic studies, including investigations related to pandemic-related mobility[31–33] and migration analysis[34].

Constructing mobility networks requires records of the location of a group of people over time, so that their travel can be estimated and routes can be established that, when aggregated, describe patterns of population movement. There are studies that use social network data to infer the movements of people in a given area, for example, geolocated Twitter information[35–37]. This has the disadvantage of having data biased towards sectors of the population that use certain platforms and specific apps. Among the data sources for building mobility

[1]Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701, Indeco la Huerta, Ciudad de México, 58190, Michoacan, Mexico. [2]National Institute of Genomics Medicine, Periferico Sur 4809, Arenal Tepepan, Tlalpan, 14610, Mexico City, Mexico. [3]National Council for Science and Technology, Av. Insurgentes Sur 1582, Colonia Crédito Constructor, Benito Juárez, Mexico City, Mexico. [4]Estornuda.me, Mexico City, Mexico. [5]Center for Research and Advanced Studies of IPN, Av Instituto Politécnico Nacional 2508, San Pedro Zacatenco, Gustavo A. Madero, 07360, Mexico City, Mexico. [6]Center for Genomics Sciences, Universidad Nacional Autónoma de México, Avenida Universidad s/n, Universidad Autonoma del Estado de Morelos, 62210, Cuernavaca, Morelos, Mexico. [7]Center for Research and Advanced Studies of IPN, Irapuato Unit, Libramiento Norte Carretera Irapuato León Kilómetro 9.6, 36821, Irapuato, Guanajuato, Mexico. ✉e-mail: maribel. hr@cinvestav.mx
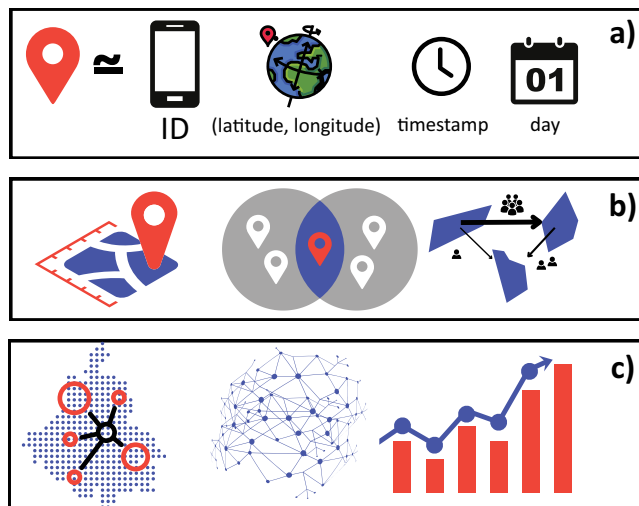
Fig. 1 Schematic overview of the study. (**a**) Input data: anonymized device IDs, locations, and timestamps. (**b**) Device assignment to AGEBs: we geolocate each device to the AGEB where it is observed during the day, identify device counts in all AGEB pairs, and assign them as edge weights in the mobility network. (**c**) Mobility networks: nodes representing regions in Greater Mexico City enable the study of network properties over time.

networks, cell phone records represent the alternative that guarantees the best sampling; cell phone use covers a wide range of population characteristics, especially in urban areas. However, these data are not easy to obtain: they involve costs associated with managing the records with private companies and processing them to achieve a format that allows the analysis of the information.

Greater Mexico City, with approximately 22 million inhabitants, is among the largest metropolitan areas in the world. However, in contrast to the availability of data on other large cities, we could locate only one work that analyzes mobility patterns in Mexico with a network approach based on Twitter data[38].
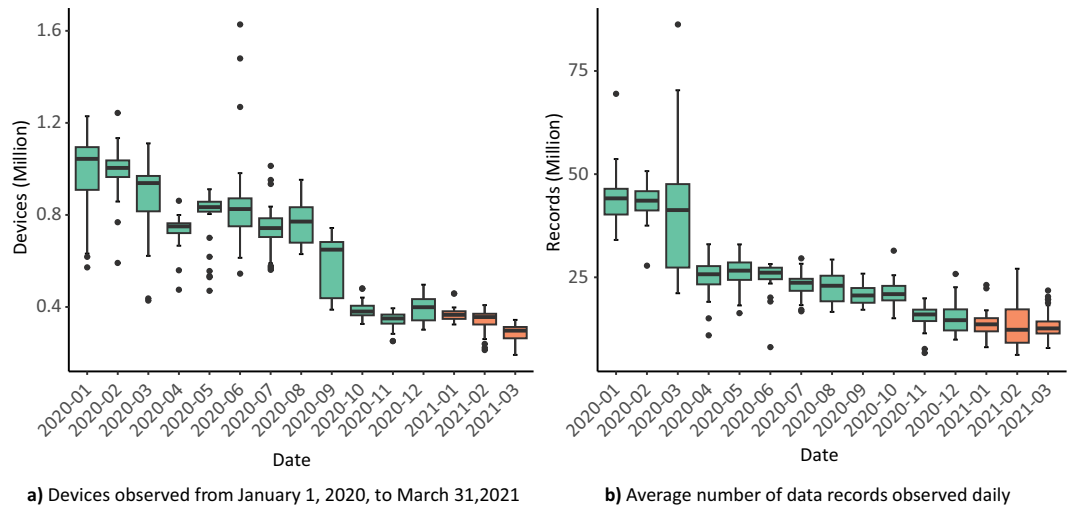
In this work we managed to obtain and process an average of 25 million records per day, during a period of fifteen months, corresponding to cell phone-based location data of inhabitants of Greater Mexico City. Then, we aggregated the information to build a set of 456 mobility networks between zones in Greater Mexico City. To establish the zones we were guided by the basic geostatistical areas (AGEBs) defined by the National Institute of Statistics, Geography and Informatics[39]. Each of these areas corresponds to a census block group and can be easily linked to studies that determine its socioeconomic profile. Figure 1 presents a schematic overview detailing the methodology employed in this study.

The dataset represents an important contribution due to the amount of data processed and the work invested in structuring them in mobility networks, following a reconstruction based on mobile device location data or *pings*. In addition, the data (1) is organized with a time axis that allows the study of mobility over fifteen months, and (2) covers the year 2020, so that it is possible to analyze the disruption in the behavior of the population derived from the health contingency of SARS-Cov-2.

In fact, the recent COVID-19 crisis has underscored the significance of human flow datasets across diverse geographical scales. These datasets support evidence-based public health policies and various applications related to social sensing and transportation. To illustrate, the work by Guidotti *et al*.[40] presents time-series data encompassing various parameters associated with the health crisis arising from Covid-19 in 230 countries. Similar datasets explore specific regions; for example, Liu *et al*.[41] provide a detailed account of cases in China, integrating demographic, clinical, and mobility information. This intersection of Covid-19 cases and mobility within a region is also evident in datasets released by Ponce de Leon *et al*.[42] for Spain, Pepe *et al*.[43] for Italy, and Kang *et al*.[44] for the United States. These research groups aggregate information and present it through mobility networks, spatial networks, or descriptive sets of edges, respectively, recognizing the importance of studying the unique characteristics of each country, including territorial organization, population size, and public policies. As highlighted by Kang *et al*.[44], examining human flow at various geographic scales enhances our understanding of human dynamics during a health crisis.

Mexico represents one of the countries that did not have a strict lockdown. With an economy heavily based on services and informal trade, the country faced the crisis with a dilemma between emergency measures to manage the health sector crisis and keep the economy functioning. Far from strict measures, the government implemented different recommendations that had a very dissimilar effect among different sectors of the population, mainly between those who could work from home and isolate themselves and those who were forced to provide services to keep the city running. The difference in the situation of different sectors of the population translates into different patterns of displacement in AGEBS, depending, in part, on the economic characteristics of the areas.

Crisis management itself makes the study of mobility patterns in different countries a completely local issue. We believe that the dataset made available to the community will allow a deeper understanding of the particular

a) Devices observed from January 1, 2020, to March 31,2021

b) Average number of data records observed daily

Fig. 2 Devices (**a**) and data point records (**b**) processed per month to build the mobility networks. Data for 2020 and 2021 is shown in green and orange, respectively.

dynamics of a large city in Latin America and the specific ways in which a crisis breaks patterns and creates a new, and differentiated, everyday life.

Methods

For this research study, we procured data from Veraset, a data-as-a-service company that offers a range of datasets, including cell phone location data[45]. The data provided by this company are subjected to anonymization procedures and are presented in a tabular format. Each data point (ping) allows us to ascertain an anonymous user's geographical coordinates (latitude and longitude) at a specific timestamp.

The temporal scope of the dataset used in our study spans from January 1, 2020, to March 31, 2021, and pertains exclusively to geographic records within the borders of Mexico. The dataset for Greater Mexico City comprises a total of 11 billion records from 33 million devices, with approximately 0.4 to 1.2 million devices and records ranging from 10 to 50 million per day, as illustrated in Fig. 2. These data points' temporal resolution extended to the level of seconds. There is a confidentiality agreement with Veraset specifying that publications are restricted to presenting results derived solely from aggregated information or estimated model parameters. Nonetheless, in the interest of transparency and reproducibility, we have made available an illustrative example of the source dataset on the Open Science Framework (OSF)*.

Using data from mobile phone usage holds significant promise for facilitating an accurate estimation of population mobility. Notably, Mexico City, the capital of Mexico situated within the broader metropolitan region known as Greater Mexico City, stands out as the federal entity with the highest cellular telephony adoption. As of 2021, approximately 88.4% of its population was reported to have access to this technology[46].
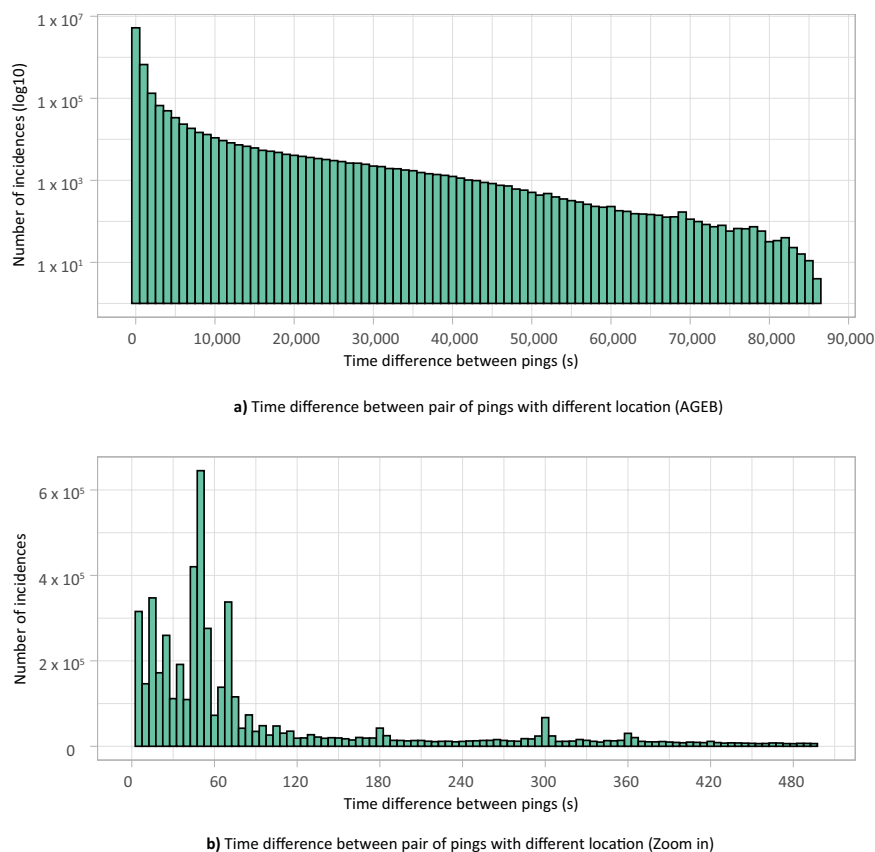
In order to build the mobility networks, it is necessary to locate cell phone devices to previously established geographic units. In the construction of this dataset we use the basic geostatistical areas (AGEBs) defined by the National Institute of Statistics, Geography and Informatics (INEGI), as nodes of the networks. AGEBs are the fundamental unit of the Institute's geostatistical framework and are adjusted to the state and municipal boundaries of the political-administrative division of the country.

Therefore, we aggregate devices into the respective AGEBs where they are observed to create study units. The aggregation process commences by identifying all AGEBs where a device has been recorded, with this identification being performed daily for each device. Subsequently, we aggregate all devices observed at origin nodes and similarly at destination nodes. As a result, our reporting centers on the count of devices observed transitioning from one AGEB to another, represented by edge weights. Furthermore, we define travel time as the period during which a device moves from one AGEB to another between two consecutive time points. Upon analyzing the time taken by each device for this journey, we have determined that, in most cases, it falls within the range of 1 to 120 seconds, as illustrated in Fig. 3. Therefore, we consider all travels lasting at least one second.

The area considered in this study is delimited according to INEGI's definition of the Metropolitan Zone of the Valley of Mexico, which covers about 7,855 km^2 and includes the 16 municipalities of Mexico City, 59 municipalities of the state of Mexico, and one municipality of the state of Hidalgo. Regarding our study units, the area of Greater Mexico City is organized into 5,740 urban AGEBs and 169 rural AGEBs[39].

Using AGEBs allows the efficient cross-referencing of information on the areas, for example, with the Institute's periodic and diverse instruments such as the Population and Housing Census, the National Survey on the Dynamics of Household Relationships, or the National Survey on the Availability and Use of Information Technologies in Households[46–48]. Figure 4 shows some indicators obtained from these surveys for each urban AGEB considered in the region of our analysis. Another analysis that is feasible is to cross-reference data with the marginalization indices reported by INEGI[49].

Once the geographic units of study have been established, we define the mobility network for a given day D_k as a directed and weighted graph $G_k (V, E_k)$, where:

a) Time difference between pair of pings with different location (AGEB)

b) Time difference between pair of pings with different location (Zoom in)

Fig. 3 Distribution of the time difference found between pair of records where a change in location was detected. Most of the crossover between AGEBs occur between pair of records with a time difference below 120 seconds.

- V represents the set of AGEBs within Greater Mexico City.
- An edge $(i, j) \in E_k$ signify mobility, originating from a source node i and terminating at a target node j during the day D_k.
- Mobility in our network is defined based on observing at least one device moving from i to j. The weight assigned to each edge indicates the fraction of observed devices that have moved from node i to node j relative to the total number of observed devices during D_k, i.e.,

$$W_{k_{ij}} = \frac{|D_{k_{ij}}|}{|D_k|} \times 10^6,$$

where $|D_{k_{ij}}|$ is the number of devices that were observed to move from i to j, that is, were observed in i and their next immediate ping after at least one second was in j; and $|D_k|$ is the cardinality of the set of unique device ids observed throughout the day D_k.

As shown, the weights employed within each network G_k undergo a normalization process relative to the volume of devices observed during the day D_k. The normalization procedure constrains the weights $W_{k_{ij}}$ to a range between 0 and 1, which enhances their interpretability and facilitates comparisons between nodes and across networks within the collection, i.e., across movement flows corresponding to different days. The multiplication by a factor of 10^6 is employed to manage numerical values efficiently.

Algorithm 1 was developed to construct a network using daily mobile device location data as its input. As depicted in the algorithm, it accomplishes several key tasks. Firstly, it associates each ping with its respective AGEB. Subsequently, it computes the count of observed trips between different areas. Finally, the algorithm normalizes the data following the previously described method.

The complexity to compute these networks is in $O(nm + l^2)$, where n is the number of pings or records processed, m is the number of AGEBs and l is the number of devices. The quantity of records and devices accessible for network construction varies from day to day. The distribution of records and devices available for network construction is depicted in Fig. 2, with mean values displayed for each month. A minor reduction in the average number of records and devices per day was noted during the early months of 2021. Nevertheless, the normalization of edge weights in the networks enables meaningful comparisons among different days.

Algorithm 1 Mobility network construction.

```
 1: inputs
 2:   date = yyyy-mm-dd
 3:   records = data(device_ID, time_stamp, lattitude, longitude)
 4:
 5: outputs
 6:   mobilitynetwork = table(origin, dest, weight)
 7:
 8:   records_by_date = filter record by date
 9:   devices= select distinct device from records_by_date
10:   number_of_devices= count(devices)
11:
12:   for each: record in records_by_date
13:       record.location = locate ageb based on latitude, longitude
14:   end for each
15:
16:   % init origin-destination matrix
17:   od = empty(origin, dest)
18:
19:   for each: device in devices
20:       sort records_by_date of device by timestamp
21:       If   records_by_date[t].location != records_by_date[t+1].location   then
22:           travel_time = records_by_date[t+1].timestamp - records_by_date[t].timestamp
23:           If  travel_time > 0 seconds  then
24:               append (origin = records_by_date[t].location, dest=records_by_date[t+1].location) to od
25:           endif
26:       endif
27:   end for each
28:
29:   mobilitynetwork = select distinct (origin,dest) from od
30:   for each:
31:       mobilitynetwork.weight = [count(origin,dest in od) *(1e+6) ] / (number_of_devices)
32:   end for each
33:
34:   return mobilitynetwork
```
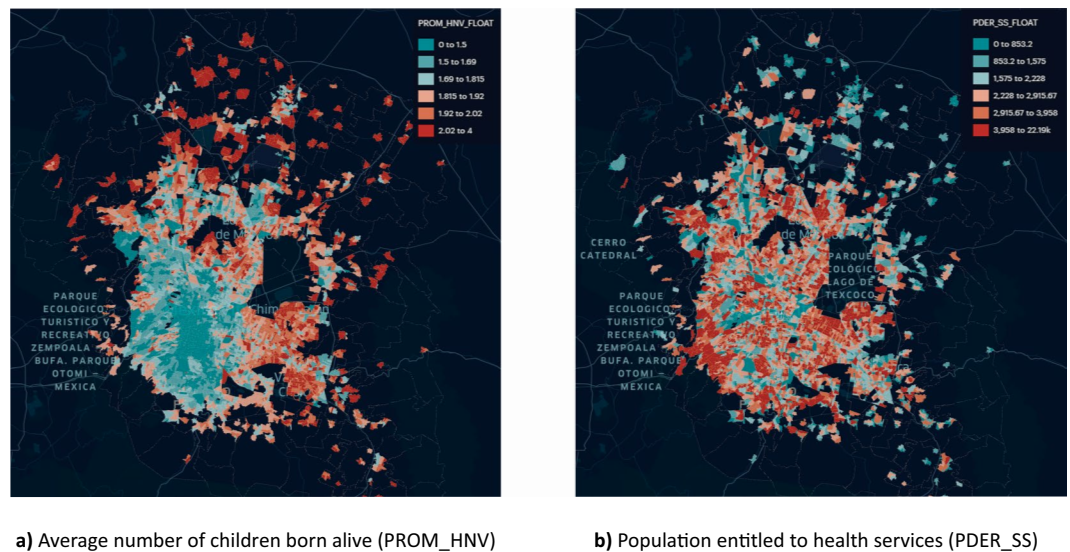
Data Records

The observation period considered in the data covers 456 days, starting January 1st, 2020. Thus, our data consists of 456 mobility networks, each represented by an individual comma-separated values (csv) text file. Each file has an index corresponding to the chronological order of the data; day 1 corresponds to January 1st, 2020, whereas day 456 corresponds to March 31st, 2021. The dataset can be found in the OSF repository: https://doi.org/10.17605/OSF.IO/GWQ6U[50].

Each node in the network represents an AGEB, defined in accordance with the National Institute of Statistics, Geography and Informatics (INEGI). Each network file contains an edge list of the corresponding network, using three columns. The first two columns indicate the origin and destination nodes, respectively. The third column is a integer representing the edge weight, i.e., the number of trips observed during the day between the connected AGEBs. As described above, the weight of the networks has been normalized.

When the number of observed trips between AGEBs is zero for a given day, the network may contain isolated nodes. These nodes are excluded from the network's edge list, indicating that not all the csv files comprising the dataset have an identical number of rows.

Finally, in the dataset we have included information on each of the geographic zones considered in the networks. The `agebs_ZMVM.csv` file indicates, for each zone:

- `IDGEO`: A unique key for each AGEB. It is formed by concatenating the key associated with the entity, municipality, locality (in the case of urban AGEBs), and a specific AGEB key. IDGEO corresponds to 'CVEGEO', an identifier commonly encountered in Mexican databases.
- `ID_ENT`: The key of the federal entity according to INEGI.
- `ID_MUN`: The key of the municipality to which the AGEB corresponds.
- `ID_LOC`: In the case of urban AGEBs, the key of the locality to which they belong. This data does not exist for rural AGEBs.
- `ID_AGEB`: The key of the AGEB according to INEGI.
- `Entity`: Name of the federal entity to which the AGEB belongs.

a) Average number of children born alive (PROM_HNV)   **b)** Population entitled to health services (PDER_SS)

Fig. 4 Distribution of devices in the source dataset correlates to population reported in each AGEB. The devices detected on a day were assigned to a home AGEB, contabilized and compared to the population reported for each AGEB on the 2020 National Population and Housing Census[47]. (**a**) The Spearman correlation of this comparison on a pre-pandemic day (2020-01-26) is shown for detail, as well as (**b**) the correlation of these values across the first two months of 2020. (**c**) Distribution across the population in different socioeconomic strata unveils a higher concentration of AGEBs and devices within the medium and high socioeconomic strata. (**d**) The correlation to the population reported by census, but grouped by reported socioeconomical level of AGEBs, reveals an overrepresentation of AGEBs characterized by a high socioeconomic level.

- `Municipality`**:** Name of the municipality to which each AGEB belongs.
- `Type`**:** If the AGEB is rural or urban.
- `Long`**:** Longitude of the centroid of the considered AGEB. This allows us to locate each node of our networks spatially.
- `Lat`**:** Latitude of the centroid of the considered AGEB. This allows us to locate each node of our networks spatially.
- `Pob_MUN`**:** Total population of the municipality to which the AGEB belongs, according to the National Population and Housing Census carried out in 2020[47]. Since this is a data corresponding to the municipality, and not to particular AGEBs, it is repeated in all the AGEBs that make up a municipality.
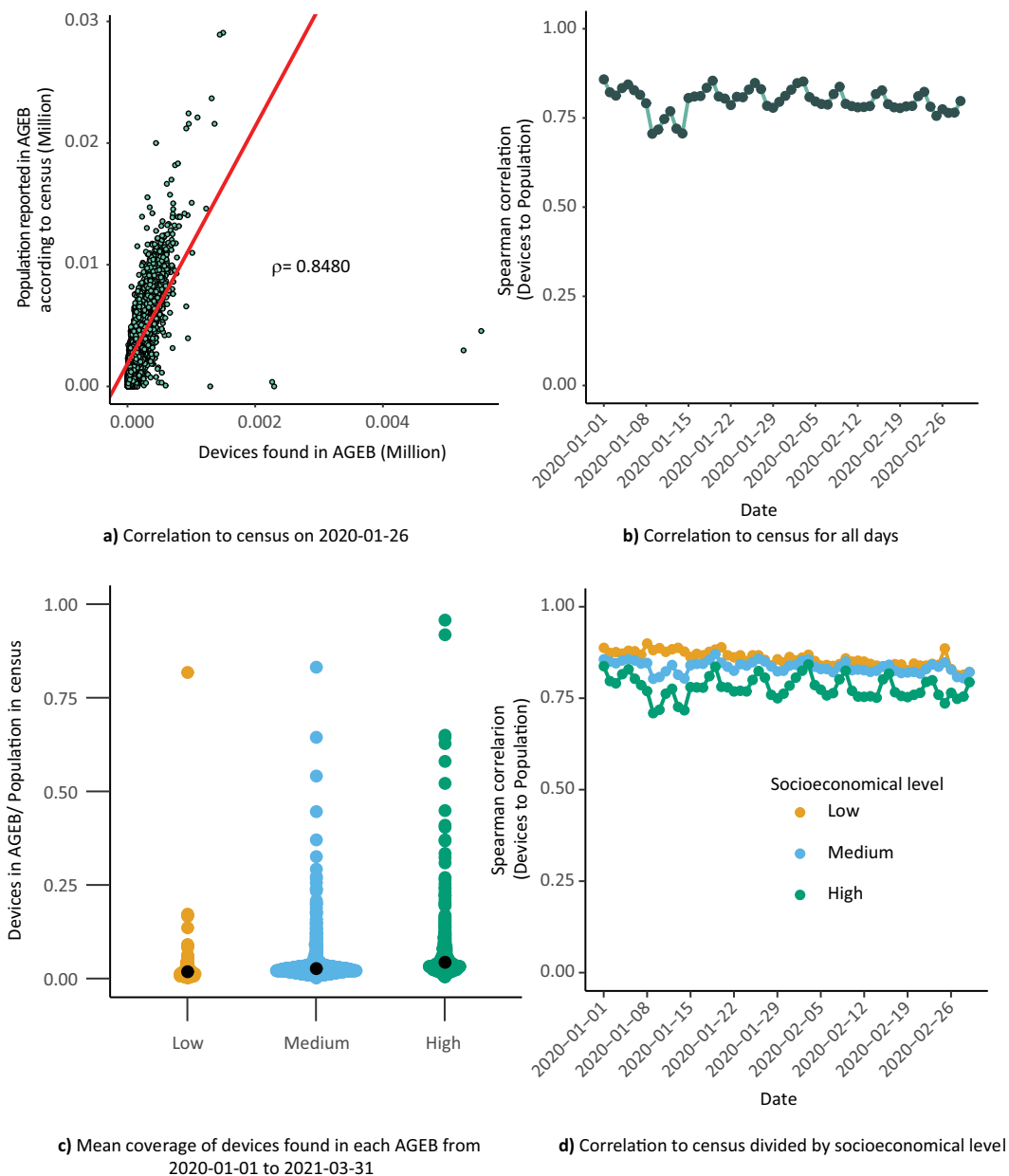
Technical Validation

The code for the creation of the networks was developed within the team following standard peer review practices. Once the dataset was constructed, three different analyses were implemented to explore the validity and possible limitations of the data.

Data representativity.   To validate the representation of our source dataset to the actual population in Greater Mexico City, a comparison to the data reported by AGEB in the 2020 National Population and Housing Census[47] was conducted. The devices available for a given date were assigned to a home AGEB according to their first record of the day.

When comparing the number of devices assigned to each AGEB to the population reported on the census for each AGEB, an acceptable correlation was found as shown in Fig. 5.

Correlation was further inspected to asses if there was a bias by other sociodemographic factors. To inspect the representation of various socioeconomic strata, the AGEBs were categorized into three groups based on their designated socioeconomic status calculated from INEGI[51]. Greater Mexico City encompasses 5545 AGEBs, with 509, 3396, and 1640 falling into low, medium, and high socioeconomic strata, respectively. The average population residing in AGEBs classified as low socioeconomic status is approximately 2,300 individuals, while AGEBs categorized as medium and high socioeconomic strata have populations averaging about 4,100 and 3,400 residents, respectively.

The fraction of unique devices assigned to a home AGEB is determined relative to the population of the corresponding AGEB. Upon examination, it becomes evident that AGEBs with high and medium socioeconomic status concentrate a larger number of devices, as depicted in Fig. 5. However, when evaluating the correlation between the number of devices observed in home AGEBs in relation to the census data, an improved correlation is observed in AGEBs falling within the low and medium socioeconomic levels. This enhanced correlation contrasts with a lower correlation observed in AGEBs with high socioeconomic status, possibly attributed to

a) Correlation to census on 2020-01-26

b) Correlation to census for all days

c) Mean coverage of devices found in each AGEB from 2020-01-01 to 2021-03-31

d) Correlation to census divided by socioeconomical level

Fig. 5 Urban basic geographic units (AGEBs) considered in our analysis of Greater Mexico City correspond to census block groups. The intensity of color in each region is linked to an indicator from the Population and Housing Census conducted by the National Institute of Statistics, Geography, and Informatics (INEGI) in Mexico[47]. The example illustrates the convenience of building the networks following regions defined by INEGI, as this enriches the information on each zone and takes advantage of public access instruments designed by the Institute.

an overrepresentation of devices in such AGEBs, potentially due to individuals in this socioeconomic status possessing more than one device.

Comparison to reference data. Google generated Community Mobility Reports during the COVID-19 crisis to illustrate changes in population movement trends across different regions of the world. These openly accessible data consist of reports for various countries and provinces, highlighting the relative changes in recorded visits to different categories of places: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas.

For each category, the reference point was the observed dynamics over five weeks, from January 3rd to February 6th. The median value for each day of the week was calculated during this time frame. Subsequently, from February 15th to December 31st, the relative change observed in comparison to the baseline was established.
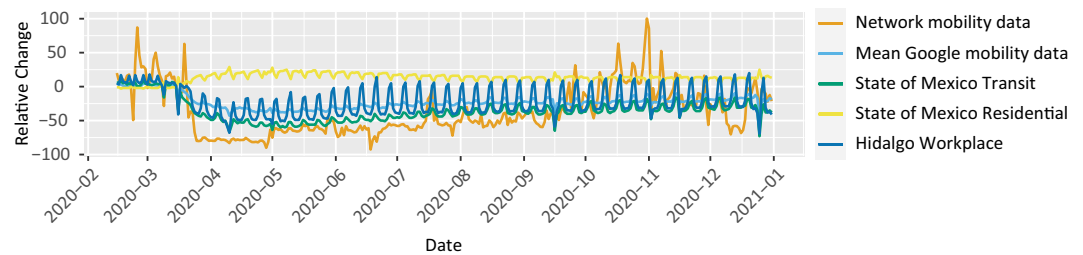
Fig. 6 Comparison between the overall mobility observed in mobility networks (the summation of all weights in the network corresponding to each day) and Google's mobility reports. Google adopts a distinct segmentation approach, quantifying visits to places within specific categories. This figure contrasts our data, featuring three examples of Google's reports related to Residential points in the State of Mexico, Transit stations in the State of Mexico, and Workplaces in the State of Hidalgo. Additionally, an average of Google's reports across categories and the states comprising the Greater Mexico City region (Hidalgo, State of Mexico, and Mexico City) is presented.

To compare our data, we replicated Google's calculation, considering the same time frame as a reference and calculating the daily median within our dataset (as shown in Fig. 6). However, when evaluating our data against Google's reports, three crucial points must be considered. First, Google's reports for Mexico are organized by state, not by municipality or AGEBs. This limitation hinders an appropriate comparison for the Greater Mexico City area, which, as detailed earlier, encompasses municipalities from three different states (Mexico City, State of Mexico, State of Hidalgo). Secondly, Google employs particular categories to quantify the difference between the baseline and the analyzed period of 2020; specifically, they calculate the relative change in recorded visits to locations corresponding to retail and recreation, grocery and pharmacy, parks, transit stations, workplaces and residential. In contrast, our study does not identify locations serving as pivot points to measure changes in population dynamics. Lastly, Google's report documentation recommends caution when contrasting data from different regions, as local differences could lead to inappropriate comparisons.

The change in territorial division and measurement focus prevents a direct comparison of our data with the information in these reports. Figure 6 illustrates three examples contrasting our data with that for states partially forming part of Greater Mexico City. Due to different categories of points of interest, these data records vary significantly among themselves. The three examples correspond to Residential points in the State of Mexico, Transit stations in the State of Mexico, and Workplaces in the State of Hidalgo.

We computed the average of Google's reports for all categories within the three federal entities that comprise the metropolitan area of Greater Mexico City. This average is also shown in Fig. 6, where we observe a Spearman correlation coefficient of 0.77 with our data. Veraset's data collection methodology (Fig. 2) could potentially translate into data fluctuations within our networks; in the comparison with Google's data, a significant change can be spotted from October 2020 onward.

Considering the distinct nature of both datasets in terms of design, our data appears reasonably compatible with Google's reference data. Furthermore, our dataset surpasses the capabilities of Google's reports in capturing mobility patterns in terms of granularity, as detailed intra-mobility analyses can be conducted within Greater Mexico City or any user-defined region. In this way, the dataset reported in the present work enables the examination of inter-mobility patterns between regions and facilitates the discernment of changes over short periods, which makes it an invaluable resource for mobility studies.

Event detection. To estimate the quality of the data, a time series analysis of different metrics in the network collection was performed. Metrics considered for this analysis were total node and edge counts, density, average path length, strong connected components and node eccentricity. Figure 7 shows a graphical representation of these metrics for the network collection. These and additional network statistics are available as a table on this publication's repository[50]. As expected, these metrics fluctuate across time, and further analysis of these changes may help understand the mobility patterns underlying. For example, we decided to build a series that considers the total sum of weights for each day in the collection networks. Then, we used the PersistAd model of the ADTK library in Python, which detects anomalies using a previous time window as a reference; a point is labeled as anomalous if the change in value it represents with respect to the median of the previous period is abnormally large. For our experiment we consider a seven-day window and the option to detect anomalies related to both positive and negative changes.

Figure 8 shows the time series obtained and the dates that, according to the selected model represent an anomaly. As can be seen, the algorithm identified January 14, 2020, the period from March 26 to 30, 2020 and December 24th, 2020 as anomalous values.

The date identified in January, which corresponds to an increase in mobility, could be related to the return to normal activities (work, school) after the winter vacation period. On the other hand, the algorithm identifies a noticeable decline in mobility during March 2020. The dates identified in this period coincide with the outbreak of COVID-19 cases in the country and with the start of the National Healthy Distance Program (Jornada Nacional de Sana Distancia) implemented by the federal government in Mexico, which officially began on March 23rd[52]. Finally, the algorithm identifies a new increase in mobility in the city at the end of the year; these changes
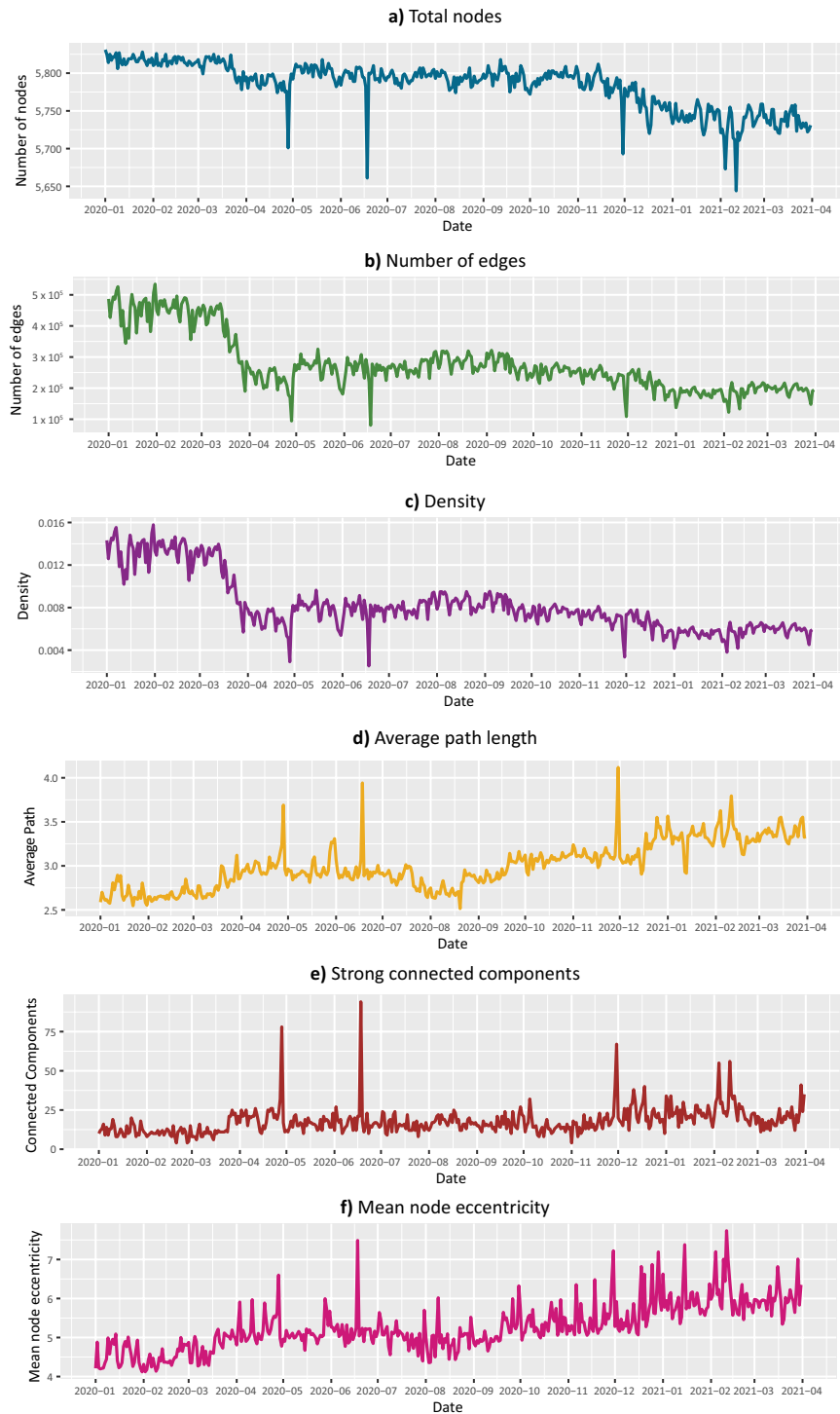
Fig. 7 Mobility network's properties change across time. Basic network statistics were calculated for the networks corresponding to each day between January 2020 and March 2021.

in movement were most likely related to the Christmas holidays. The mobility between the different AGEBs on the identified dates is shown in Fig. 9.

Beyond the specific dates pinpointed by the algorithm, a significant reduction in overall mobility within the region is noticeable starting from March 2020. This example serves as an intuitive validation of the data quality underpinning the networks and the network construction process. It appears that the data effectively capture significant shifts in population mobility that occurred in Mexico following the health contingency.

Furthermore, this example may also prompt questions that could be explored through in-depth analysis of the dataset. Utilizing more advanced tools for complex network analysis or cross-referencing these data with other sources of AGEB-related information could unveil more intricate patterns. Such an approach would allow
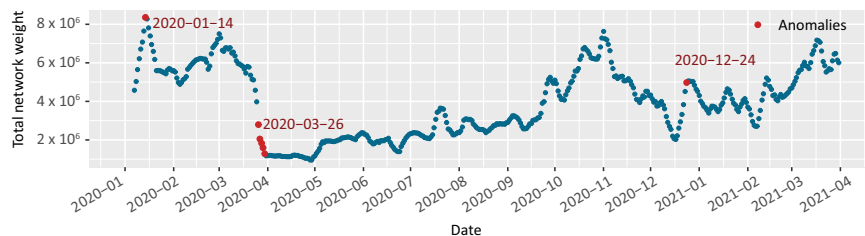
Fig. 8 Total sum of edge weights of the networks in each mobility network. This value reflects the general displacement observed in Greater Mexico City on the indicated date. Dates detected as anomalies have been marked in red.

a) January 14, 2020

b) March 27, 2020

c) March 30, 2020

d) December 24, 2020
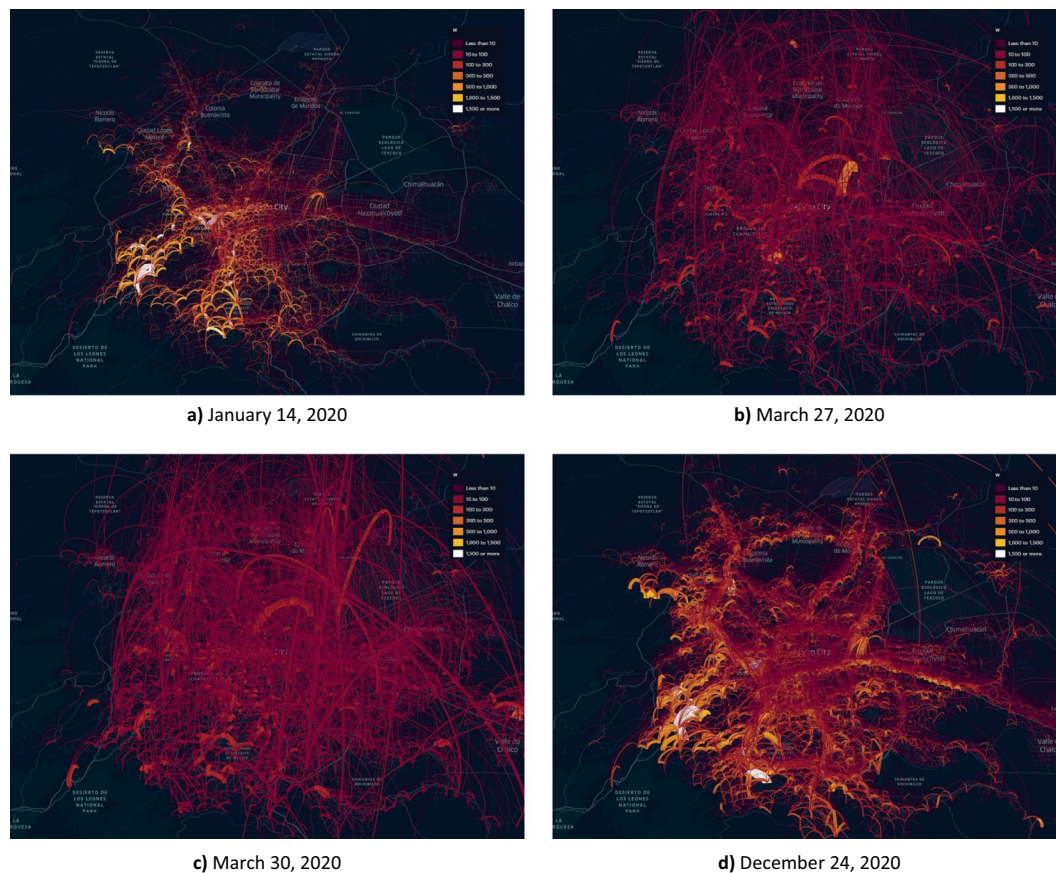
Fig. 9 Mobility between zones (AGEBs) of Greater Mexico City on days identified as anomalies during the time period between January 1, 2020 and March 31, 2021. January 14 shows an increase in mobility possibly related to the return to normal activities after the winter vacations. In contrast, the period of March 27–30 shows a clear decrease in population movement, following the initiation of government-recommended security measures due to the pandemic. Movement increases again in December, specially on the occasion of Christmas.

us to contrast the disruption that daily life experienced in different sectors of the city, providing valuable insights into the varying impacts of the pandemic.

It is important to underscore that our data, much like other datasets of a similar nature, are subject to limitations imposed by the data collection and sampling strategy employed by the data provider. Nevertheless, based on our exploration, the dataset can prove valuable for analyzing population mobility within the Greater Mexico City metropolitan area.

Usage Notes

The dataset is available in CSV format, making it easily accessible without the need for specialized software. While importing the data, it is important to note that the networks are directed and weighted.

To enhance the richness of the information presented within the networks, we recommend utilizing databases provided by the Mexican National Institute of Statistics, Geography and Informatics (Instituto Nacional

de Estadística, Geografía e Informática de México - INEGI). The definition of localities within our dataset aligns with the division proposed by INEGI, ensuring compatibility with results from various tools and resources available on their website[39,53].

Code availability

DuckDB and Python were employed for constructing the mobility networks. The entire network dataset, along with the code utilized for network construction and supplementary tables can be found at: https://doi.org/10.17605/OSF.IO/GWQ6U[50].

References

1. Liu, X., Gong, L., Gong, Y. & Liu, Y. Revealing travel patterns and city structure with taxi trip data. *Journal of transport Geography* **43**, 78–90 (2015).
2. Cao, J. *et al.* Resolving urban mobility networks from individual travel graphs using massive-scale mobile phone tracking data. *Cities* **110**, 103077 (2021).
3. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *nature* **453**, 779–782 (2008).
4. Hamedmoghadam, H., Ramezani, M. & Saberi, M. Revealing latent characteristics of mobility networks with coarse-graining. *Scientific reports* **9**, 7545 (2019).
5. Nanni, M., Tortosa, L., Vicent, J. F. & Yeghikyan, G. Ranking places in attributed temporal urban mobility networks. *Plos one* **15**, e0239319 (2020).
6. Gardiner, O. & Dong, X. Mobility networks for predicting gentrification. In *Complex Networks & Their Applications IX:* Volume 2, *Proceedings of the Ninth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS* 2020, 181–192 (Springer, 2021).
7. Çolak, S., Lima, A. & González, M. C. Understanding congested travel in urban areas. *Nature communications* **7**, 10793 (2016).
8. Amézquita-López, J., Valdés-Atencio, J. & Angulo-Garcia, D. Understanding traffic congestion via network analysis, agent modeling, and the trajectory of urban expansion: a coastal city case. *Infrastructures* **6**, 85 (2021).
9. Gonzalez-Vidal, A., Terroso-Sáenz, F. & Skarmeta, A. Parking availability prediction with coarse-grained human mobility data. *CMC-Computers Materials & Continua* **71**, 4355–4375 (2022).
10. Parmar, J., Das, P. & Dave, S. M. Study on demand and characteristics of parking system in urban areas: A review. *Journal of Traffic and Transportation Engineering (English Edition)* **7**, 111–124 (2020).
11. Nykvist, B. & Whitmarsh, L. A multi-level analysis of sustainable mobility transitions: Niche development in the uk and sweden. *Technological forecasting and social change* **75**, 1373–1387 (2008).
12. Zhang, T., Duan, X. & Li, Y. Unveiling transit mobility structure towards sustainable cities: An integrated graph embedding approach. *Sustainable Cities and Society* **72**, 103027 (2021).
13. Tran, M. & Brand, C. Smart urban mobility for mitigating carbon emissions, reducing health impacts and avoiding environmental damage costs. *Environmental Research Letters* **16**, 114023 (2021).
14. Banister, D. Cities, mobility and climate change. *Journal of transport geography* **19**, 1538–1546 (2011).
15. Takabatake, T., Shibayama, T., Esteban, M., Ishii, H. & Hamano, G. Simulated tsunami evacuation behavior of local residents and visitors in kamakura, japan. *International journal of disaster risk reduction* **23**, 1–14 (2017).
16. Zhou, Y., Wang, J. & Sheu, J.-B. On connectivity of post-earthquake road networks. *Transportation Research Part E: Logistics and Transportation Review* **123**, 1–16 (2019).
17. Pintér, G., Nadai, L., Bognár, G. & Felde, I. Evaluation of mobile phone signals in urban environment during a large social event. In *2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, 000247–000250 (IEEE, 2018).
18. Pintér, G., Nádai, L. & Felde, I. Analysis of mobility patterns during a large social event. In *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 000339–000344 (IEEE, 2018).
19. Gu, Y., Fu, X., Liu, Z., Xu, X. & Chen, A. Performance of transportation network under perturbations: Reliability, vulnerability, and resilience. *Transportation Research Part E: Logistics and Transportation Review* **133**, 101809 (2020).
20. Liu, H. & Li, Y. Smart cities for emergency management. *Nature* **578**, 515–516 (2020).
21. Nishino, A., Kodaka, A., Nakajima, M. & Kohtake, N. A model for calculating the spatial coverage of audible disaster warnings using gtfs realtime data. *Sustainability* **13**, 13471 (2021).
22. Yao, Y. *et al.* Analysis of network structure of urban bike-sharing system: A case study based on real-time data of a public bicycle system. *Sustainability* **11**, 5425 (2019).
23. Vogel, M. *et al.* From bicycle sharing system movements to users: a typology of vélo'v cyclists in lyon based on large-scale behavioural dataset. *Journal of Transport Geography* **41**, 280–291 (2014).
24. Gorji, M.-A., Akbarzadeh, M. & Shetab-Boushehri, S.-N. Evaluation and improvement of the urban transportation networks resilience in short-term non-recurring traffic congestion: a novel graph connectivity-based criteria. *Transportation Engineering* **10**, 100152 (2022).
25. Chen, S. *et al.* Traffic dynamics on complex networks: a survey. *Mathematical Problems in Engineering* **2012** (2012).
26. Háznagy, A., Fi, I., London, A. & Nemeth, T. Complex network analysis of public transportation networks: A comprehensive study. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 371–378 (IEEE, 2015).
27. Wang, L.-N., Wang, K. & Shen, J.-L. Weighted complex networks in urban public transportation: Modeling and testing. *Physica A: Statistical Mechanics and its Applications* **545**, 123498 (2020).
28. Rudskoy, A., Ilin, I. & Prokhorov, A. Digital twins in the intelligent transport systems. *Transportation Research Procedia* **54**, 927–935 (2021).
29. Kong, X. *et al.* Big trajectory data: A survey of applications and services. *IEEE Access* **6**, 58295–58306 (2018).
30. Prommaharaj, P., Phithakkitnukoon, S., Demissie, M. G., Kattan, L. & Ratti, C. Visualizing public transit system operation with gtfs data: A case study of calgary, canada. *Heliyon* **6** (2020).
31. Fontanelli, O. *et al.* Human mobility patterns in mexico city and their links with socioeconomic variables during the covid-19 pandemic. Preprint at https://arxiv.org/abs/2208.06456 (2022).
32. Fontanelli, O. *et al.* Intermunicipal travel networks of mexico during the covid-19 pandemic. *Scientific Reports* **13**, 8566 (2023).
33. Castelán-Sánchez, H. G. *et al.* Comparing the evolutionary dynamics of predominant sars-cov-2 virus lineages co-circulating in mexico. *Elife* **12**, e82069 (2023).
34. Rao, N. Migration, Education and Socio-Economic Mobility. (Routledge, 2010).
35. Hawelka, B. *et al.* Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* **41**, 260–271 (2014).
36. Yin, J., Soliman, A., Yin, D. & Wang, S. Depicting urban boundaries from a mobility network of spatial interactions: A case study of great britain with geo-located twitter data. *International Journal of Geographical Information Science* **31**, 1293–1313 (2017).

37. Mirzaee, S. & Wang, Q. Urban mobility and resilience: exploring boston's urban mobility network through twitter data. *Applied Network Science* **5**, 1–20 (2020).

38. Martinez, V. R., Escalante, M. A., Beguerisse-Diaz, M., Garduño, E. & Gonzalez, V. M. Understanding human behavior in urban spaces using social network data: a mobility graph approach. *International Journal of Web Services Research (IJWSR)* **13**, 54–66 (2016).

39. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Marco Geoestadístico Nacional, septiembre 2019 (National Geostatistical Framework, September 2019). https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463776079 (2019).

40. Guidotti, E. A worldwide epidemiological database for COVID-19 at fine-grained spatial resolution. *Scientific Data* **9**, 112 (2022).

41. Liu, X. F., Xu, X.-K. & Wu, Y. Mobility, exposure, and epidemiological timelines of COVID-19 infections in China outside Hubei province. *Scientific data* **8**, 54 (2021).

42. Ponce-de Leon, M. *et al.* COVID-19 Flow-Maps an open geographic information system on COVID-19 and human mobility for Spain. *Scientific Data* **8**, 310 (2021).

43. Pepe, E. *et al.* COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data* **7**, 230 (2020).

44. Kang, Y. *et al.* Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scientific data* **7**, 390 (2020).

45. Veraset. https://www.veraset.com/ (2023). Retrieved October 4th, 2023.

46. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (National Survey on the Availability and Use of Information Technologies in Households). https://www.inegi.org.mx/programas/dutih/2021/ (2021).

47. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Censo de Población y Vivienda (Population and Housing Census). https://www.inegi.org.mx/programas/ccpv/2020/ (2020).

48. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (National Survey on Household Relationship Dynamics). https://www.inegi.org.mx/programas/endireh/2021/ (2021).

49. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Índices de marginación (marginalization indices). https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372 (2020).

50. Flores-Garrido, M. *et al.* Mobility Networks for Greater Mexico City. *Open Science Framework* https://doi.org/10.17605/OSF.IO/GWQ6U (2023).

51. Urrutia-Hernández, R.M. Clasificación por nivel socioeconómico de las regiones geográficas de México. (Cinvestav, 2019).

52. Gobierno de México. Jornada Nacional de Sana Distancia (National Healthy Distance Campaign). https://www.gob.mx/salud/hospitalgea/documentos/jornada-nacional-de-sana-distancia (2020).

53. INEGI (Instituto Nacional de Estadística, Geografía e Informática). Glosario (Glossary). https://www.inegi.org.mx/app/glosario/.

Acknowledgements

Author contributions

All authors participated in the creation of the dataset and in the preparation of this manuscript. M.F.G.: data validation and preparation of the first draft of this work. G.da. & M.H.R.: methodological design for the construction of the networks. P.G.: implementation and backend. A.M. & A.H.: backend development. E.C.B.: analysis and figure design. M.H.R.: project management and direction, research design.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.H.-R.

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.