www.nature.com/scientificdata

SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

Access to mass rapid transit in OECD urban areas

Vincent Verbavatz[1,2] & Marc Barthelemy[1,3]✉

As mitigating car traffic in cities has become paramount to abate climate change effects, fostering public transport in cities appears ever-more appealing. A key ingredient in that purpose is easy access to mass rapid transit (MRT) systems. So far, we have however few empirical estimates of the coverage of MRT in urban areas, computed as the share of people living in MRT catchment areas, say for instance within walking distance. In this work, we clarify a universal definition of such a metrics - People Near Transit (PNT) - and present measures of this quantity for 85 urban areas in OECD countries – the largest dataset of such a quantity so far. By suggesting a standardized protocol, we make our dataset sound and expandable to other countries and cities in the world, which grounds our work into solid basis for multiple reuses in transport, environmental or economic studies.

Background & Summary

Motorized transport currently accounts for more than 15% of world greenhouse gas emissions[1]. As most humans live in urban areas and two-thirds of world population will live in cities by 2050[2], mitigating car traffic in cities has become crucial for limiting climate change effects[3–6]. Daily commuting is the main driver for passenger car use - about 75% of American commuters drive everyday (U.S. Department of Transportation, Bureau of Transportation Statistics, National Transportation Statistics. Table 1–41 at http://www.bts.gov (2016)) - while alternative transport modes such as public transportation networks are unevenly developed among countries and cities (List of Metro Systems, *Wikimedia Foundation* https://en.wikipedia.org/wiki/List_of_metro_systems, 2020).

Over the last decades, various attempts to assess the environmental impact of car use in cities have emerged from multiple fields, ranging from econometric studies to physics or urban studies[7–11]. A seminal result of transport theory, by Newman and Kenworthy[10], correlated transport-related emissions with a determinant spatial criterion: urban density. Alternatively, Duranton and Turner[11] claimed that public transport services were unsuccessful in reducing traffic, as transit riders lured off the roads are replaced by new drivers on the released roads. Such results, however, crucially lack both theoretical and empirical foundations[12–15] and new research[16] shows that the two main critical factors that control car traffic in cities are urban sprawl and access to mass rapid transit (MRT).

More generally, understanding mobility in urban areas is fundamental, not only for transport planning, but also for understanding many processes in cities, such as congestion problems, or epidemic spread[17,18] for example. But what is a good measure of access to transit? Studies have mainly focused on the number of lines or stops[19–21], length of the network or graph analysis[22–24]. Few works[16,25,26], however, have considered investigating catchment areas of MRT stations, i.e. looking at the share of population living close to MRT stations, for instance within walking distance. Such conditions have however proved to be essential in explaining commuting behaviours and mobility patterns[16].

The most detailed definition of such catchment metrics is the People Near Transit (PNT), and originates from a 2016 publication from the Institute for Transportation and Development Policy (IDTP)[25]. It produces a rigorous dataset of the share of population living close to transit (less than 1 km) for 25 cities in the world (12 in OECD countries). However, definitions of urban areas and rapid transit systems in that dataset are multiple and need to be refined while the number of cities must be expanded.

Hence, in order to expand our global knowledge of urban mobility, we need a common, unified and universal definition of access to public transit as well as sound measures of such a quantity. In this paper, we clarify its definition and propose what is to our knowledge the largest global dataset of PNT.

[1]Université Paris-Saclay, CNRS, CEA, Institut de physique théorique, 91191, Gif-sur-Yvette, France. [2]École des Ponts ParisTech, Champs-sur-Marne, France. [3]Centre d'Etude et de Mathématique Sociales, CNRS/EHESS, 54 Boulevard Raspail, 75006, Paris, France. ✉e-mail: marc.barthelemy@ipht.fr

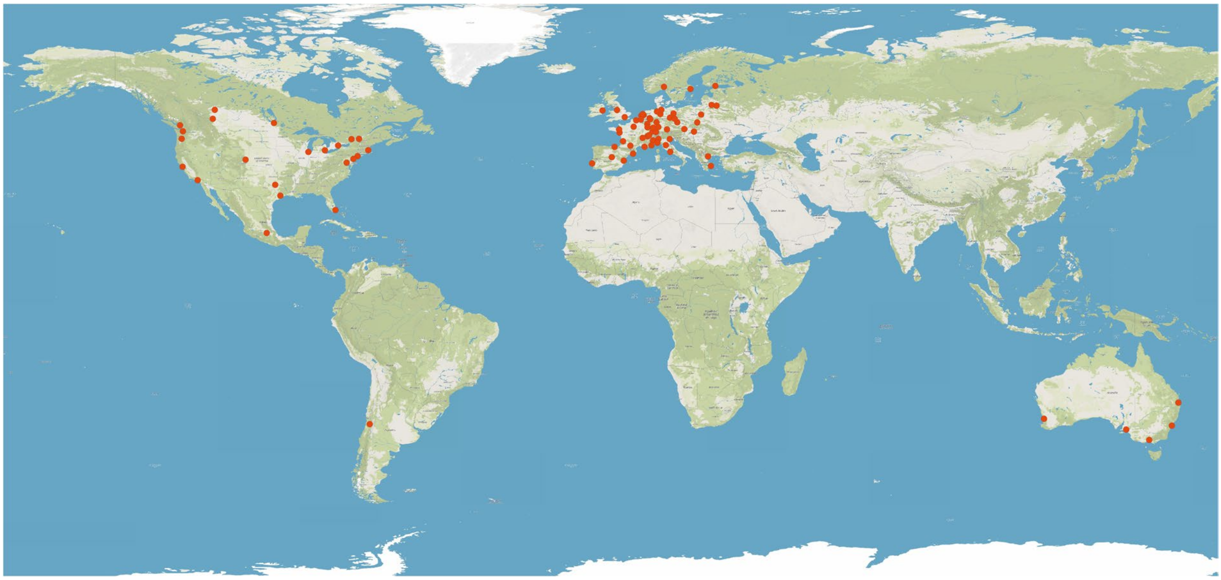

Fig. 1 The 85 OECD cities for which we found data are mostly found in Europe and in North America (MapTiler Basic and MapTiler Topo. https://www.maptiler.com, 2020).

City	Country	Population	500 m PNT (%)	1000 m PNT (%)	1500 m PNT (%)
Basel	Switzerland	528811	57.78	80.15	86.96
Bilbao	Spain	986042	56.84	76.79	83.52
Geneva	Switzerland	592893	50.44	74.68	85.07
London	United Kingdom	11754700	43.09	72.56	85.8
Zurich	Switzerland	1329898	42.7	68.18	82.09

Table 1. Population Near Transit values: Share of population living within catchment area from a MRT station at thresholds 500 m, 1000 m and 1500 m. Top 5 cities with easiest (1000 m) access to MRT.

Our analysis uses functional urban areas (FUA) in OECD countries, a consistent definition of cities across several countries[27]. We restrict our measures to mass rapid transit, usually referring to high-capacity heavy rail public transport, to which we added light rails and trams. In our sense, mass rapid transit thus encompasses:

- Tram, streetcar or light rail services.
- Subway, Metro or any underground service.
- Suburban rail services.

Buses are not comprised in that definition. In contrast with[25], we do not exclude any form of commuting trains based on station spacing or schedule criteria. As we detail it in the Method section, we identify services and corresponding stops with the General Transit Feed Specification (GTFS), a common format for public transportation schedules and associated geographic information (GTFS Static Overview. https://developers.google.com/transit/gtfs, 2020).

Crossing open-access information from public transport agencies in OECD urban areas with population-grid estimates of world population[28], we publish here a list of 85 OECD cities (see Fig. 1) for which we were able to compute the People Near Transit (PNT) levels defined as the share of urban population living at geometric distances of 500 m, 1,000 m and 1,500 m from any MRT station in the agglomeration:

$$\mathrm{PNT}(d) = \frac{\text{population s. t. euclidean minimum distance } < d}{\text{total population}} \tag{1}$$

where $d = 500, 1000, 1500$.

We display on Tables 1 and 2 the 5 cities with easiest access to MRT (largest PNT) and the 5 cities with scarcest access to MRT (smallest PNT).

We also provide for each city the population grid-maps with corresponding MRT access level, i.e. grid-maps of MRT catchment areas at different distances with population in each grid. As an example, Fig. 2 shows the 1000 m catchment area of MRT stations in Paris.

City	Country	Population	500 m PNT (%)	1000 m PNT (%)	1500 m PNT (%)
Winnipeg	Canada	846133	0	0	0
Detroit	United States	4263202	0.1	0.18	0.31
Houston	United States	6706227	0.98	2.28	3.54
Miami	United Sates	5964846	1.26	3.51	5.65
Dallas	United States	7294931	1.18	4.05	7.64

Table 2. Population Near Transit values: Share of population living within catchment area from a MRT station at thresholds 500 m, 1000 m and 1500 m. 5 cities with poorest (1000 m) access to MRT.

Methods

Residential populations for FUA. Our analysis relies on the 2015 residential population estimates mapped into the global Human Settlement Population (GHS-POP) project[28]. This spatial raster dataset depicts the distribution of population expressed as the number of individuals per cell on a grid of cells 250 m long. Residential population estimates for the target year 2015 are provided by CIESIN GPWv4.10[29] and were disaggregated from census or administrative units to grid cells.

We downloaded population tiles that cover land on the globe in the Mollweide projection (EPSG:54009) and in raster format (.tif files). These raster data are made of pixels of width 250 m with associated value the number of people living in the cell. We processed the downloaded tiles with **Python** 3.7.6 (Python Language Reference, version 3.7.6 available at http://www.python.org) and package **gdal** 3.0.2 (GDAL/OGR, Geospatial Data Abstraction software Library, *Open Source Geospatial Foundation* https://gdal.org) to convert the raster files into vectorized shapefiles. The resulting shapefiles are comprised of polygons with field value the population in each polygon. Since the polygonization process merges adjacent pixels with common value into single polygons, populations for each polygon must be recomputed from polygon area and density through the simple following rule:

$$\mathrm{Pop}_{polygon} = \mathrm{Pop}_{pixel} \times \frac{\mathrm{Area}_{polygon}}{\mathrm{Area}_{pixel}} \tag{2}$$

where $\mathrm{Area}_{pixel} = 250 \times 250 = 62500\,\mathrm{m}^2$. This leaves us with a list of 224 shapefiles of population that cover land area on earth.

By intersecting the resulting shapefiles with OECD shapefiles delineating Functional Urban Areas (FUA) in OECD countries[27] (reprojected into Mollweide projection), we can build a population-grided dataset of cities in OECD countries.

These resulting files are the population substrates used for measuring population living close to MRT stations.

Extracts of MRT stations from GTFS files. A common and *de facto* standard format for public transportation schedules and associated geographic information is the General Transit Feed Specification (GTFS Static Overview. https://developers.google.com/transit/gtfs, 2020).

A GTFS feed is a collection of at least six CSV files (with extension.txt) contained within a.zip file. It encompasses general information about transit agencies and routes in the network, schedule information such as trips and stop times and geographic information for stops (geographic coordinates).

The three main objects we require are:

- Routes: distinct routes in the network of a certain type. A route is a (one-direction) regular line, for instance a metro or bus line. The route types we use are (https://developers.google.com/transit/gtfs):

  - Tram, Streetcar, Light rail. Any light rail or street level system within a metropolitan area.
  - Subway, Metro. Any underground rail system within a metropolitan area.
  - Rail. Used for intercity or long-distance travel.
  - Cable tram. Used for street-level rail cars where the cable runs beneath the vehicle, e.g., cable car in San Francisco.

  Our definition of MRT excludes bus and ferry types:

  - Bus. Used for short- and long-distance bus routes.
  - Ferry. Used for short- and long-distance boat service.

- Trips: trips are associated to a route and define a particular and scheduled trip between specific stations. For instance, the first train of the day is a trip.
- Stops: stops are geographic locations of the stops, stations and their amenities within the transit system. Stops are organized into a parent station and their amenities (e.g. platforms or exits).

Joining in this order the four tables **routes.txt**, **trips.txt**, **stop_times.txt** and **stops.txt** allows us to bind stops with their associated route types. We can thus discriminate between bus stops and metro stops and thereby select objects according to our definition of MRT.

In a nutshell each GTFS file can be processed to produce localized and route-typed stops.

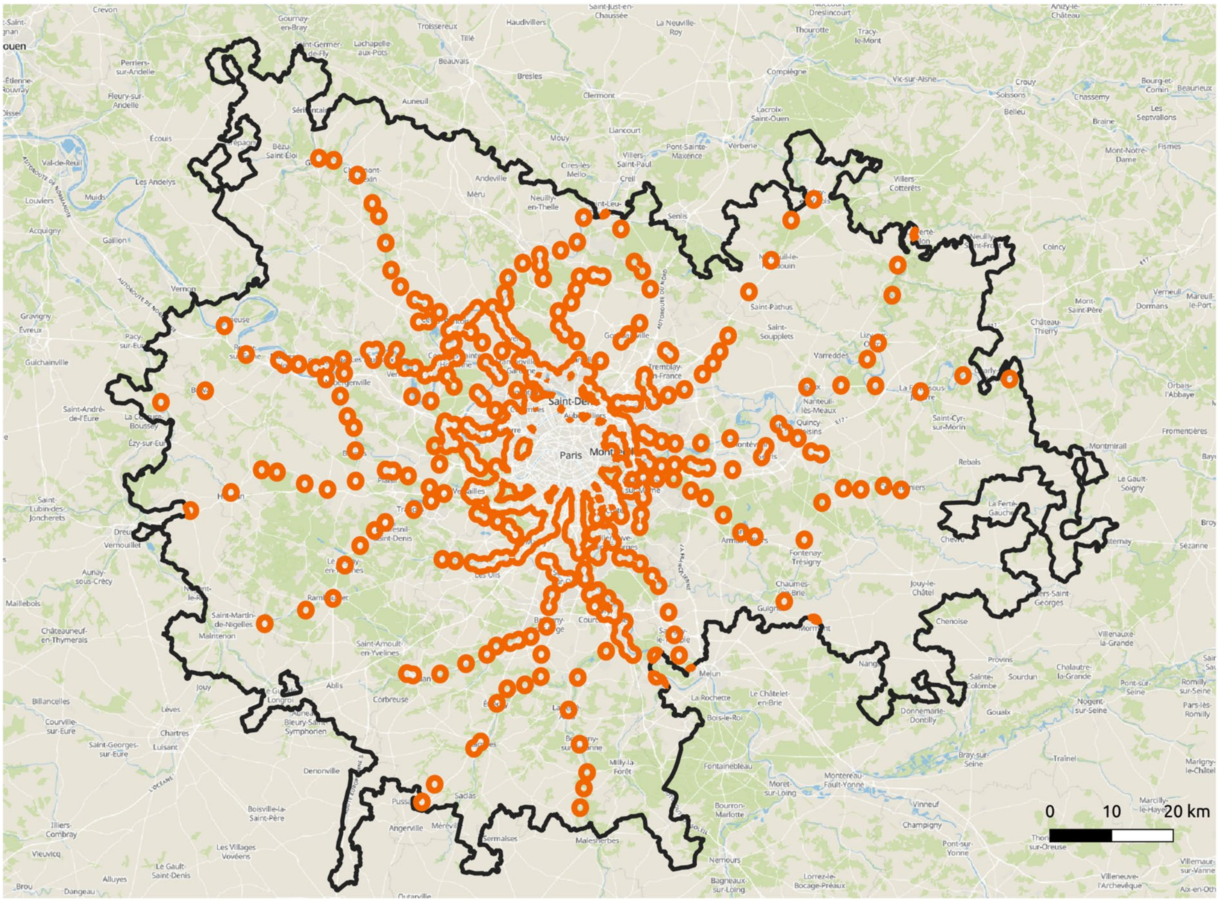

Fig. 2 1000 m catchment areas of MRT stations (in orange) in Paris functional urban area (boundaries are in black) (MapTiler Basic and MapTiler Topo. https://www.maptiler.com, 2020).

Measure of People Near Transit (PNT). In order to measure PNT within urban areas, we must bind transit systems with their respective FUAs. We need to retrieve - and merge - all available GTFS files pertaining to a specific urban area and make sure that no rapid transit agency is excluded in the process.

Most GTFS files for cities in the world are collected by the OpenMobilityData platform (https://transitfeeds.com, 2020). For each city in our dataset, we cross-checked the OpenMobilityData with Wikipedia local network information (List of Metro Systems, https://en.wikipedia.org/wiki/List_of_metro_systems) to ensure that we considered all agencies of rapid transit within the urban area.

For some European countries (Germany, France), GTFS files were not availaible on OpenMobilityData and had to be retrieved from other sources (*GTFS für Deutschland* https://gtfs.de/ and *Open platform for French public data* https://www.data.gouv.fr). We also note that GTFS format is not common in South Korea, Japan and in the United Kingdom where we only found GTFS data for Manchester area on OpenMobilityData (https://transitfeeds.com) while we directly used station coordinates for London (*Transport for London*. TFL Station Locations available at https://data.london.gov.uk/dataset/tfl-station-locations).

We were thus left with a list of 85 urban areas in the world for which we had complete, reliable and extensive data. From route-typed stop coordinates within that dataset, we can extract MRT stops (excluding buses and ferries) and buffer - still using gdal - catchment areas for several distance thresholds: 500 m, 1000 m and 1500 m. Intersecting the resulting buffers with the population-grided shapefiles gives us the total population living within catchment areas, that can be expressed as a share of the total urban area population resulting in the value of the PNT metric. Our results are shown in Online-only Table 1.

Data Records

The Data Record of PNT in OECD urban areas is available online on *Figshare*[30].

PNT levels at distance thresholds: 500 m, 1 000 m and 1 500 m for the 85 Functional Urban Areas are shown on the Online-only Table 1. The list of transit agencies for each city is online along with PNT statistics (**mrt_access.csv**)[30].

We also provide, for each city, grid-maps of population at different distances from MRT(**pops_close_to_MRT.zip**)[30].

The Tables read as follows: Basel urban area has 528811 inhabitants, of which 57.78% live within 500 m of a MRT station, 80.15% within 1000 m and 86.96% within 1500 m.

City	Country	Population in[25]	Types in[25]	PNT Share (%) in[25]	Our PNT Share (%)	Our PNT Share (%) with[25] criteria	Comments about[25]
Barcelona	Spain	3200000	Metro + LRT	76	65	74	Urban core and not FUA 1–suburban trains are *de facto* included in[25]
Boston	US	4650000	Metro + LRT	15	31	17	Excludes suburban trains
Chicago	US	9500000	Metro	14	13	13	/
London	UK	10000000	Metro + LRT + Suburban Rail	61	73	/	Some suburban trains are excluded in[25]
Los Angeles	US	13000000	Metro + LRT + BRT	11	9	/	We exclude Bus Rapid Transit
Madrid	Spain	5500000	Metro + LRT	76	60	72	Urban core and not FUA 1–suburban trains are *de facto* included in[25]
New York	US	19800000	Metro + LRT	35	45	34	Excludes suburban trains
Paris	France	12000000	Metro + Tram + Suburban Rail	50	63	/	Some tramlines are excluded in[25]
Rotterdam	Netherlands	1200000	Metro + LRT	55	41	50	Urban core and not FUA 1–suburban trains are *de facto* included in[25]
Vancouver	Canada	2300000	Metro	19	23	23	/
Washington	US	5800000	Metro	12	8	12	Urban core and not FUA

Table 3. Comparison of MRT Share from the IDTP report[25] with our estimations for 11 OECD cities. Discrepancies at first glance can be explained by different delineations of cities or transit systems. Applied on the same entities, results are similar.

Technical Validation

The most thorough and exhaustive measure of PNT in urban areas in existing literature is a 2016 report from the Institute for Transportation and Development Policy[25]. To validate our results and our methodology, we compared them with those results.

Out of the 12 OECD cities considered in[25], 11 are in our dataset: 5 in the United States, 2 in Spain, 1 in Canada, 1 in France, 1 in the United Kingdom and 1 in the Netherlands (see Table 3). Unfortunately, we found no data in the remaining city: Seoul.

Out of these 11 cities, we had at first glance similar results for only two cities: Chicago (13% for both) and Vancouver (19% vs 23%). The discrepancies observed for the other cases stem from different definitions of cities and from the different transit systems that were taken into account. While we work with Functional Urban Areas (FUA) only, the authors of[25] mix two different definitions of cities: FUA and urban cores. By applying our method to urban cores and not functional urban areas, we found the same or similar results for Barcelona, Madrid, Rotterdam and Washington (see Table 3).

Also, the authors of[25] considered a definition of the LRT (Light Rail Transit) and Suburban Rail that depends on station spacing and schedule criteria. We didn't choose this definition and for Boston and New York, we had therefore to exclude suburban trains - while keeping the definition of FUA - in order to retrieve results similar to those of Table 3. In contrast, the study[25] took into account the Bus Rapid Transit for Los Angeles, that we decided to exclude. Finally, in Paris the authors of[25] considered that the so-called RER trains were comprised in Suburban Rail, but not Transilien trains, while we included both systems in our analysis.

The conclusion here is that for similar definitions for cities and transit systems, we obtain similar results, validating our method and calculations. In order to facilitate the comparison across future studies, we would recommend using the definition of cities given by Functional Urban Areas since it is very commonly used and already unified for OECD countries. Concerning transit systems, we think that it is more relevant and also verifiable to consider transit systems based on their types (Rail versus Road) rather that on spacing and schedule criteria that are specious and less universal. Hence, in comparing our results with results from the IDTP report[25] and after checking on Table 3 that our methodology is correct, we decided to keep our unmodified estimations for the considered cities, despite the discrepancies with[25].

For other cities in the dataset we have unfortunately found no existing data to compare with. Thus, we hope for future research to test and expand our estimations and results.

Usage Notes

Easy code and hints are given on *Gitlab* (https://gitlab.iscpif.fr/vverbavatz/mrt-access-project).

We strongly recommand using GDAL (GDAL/OGR, Geospatial Data Abstraction software Library, *Open Source Geospatial Foundation* https://gdal.org, 2020) to handle geographic data with Python.

Code availability

Detailed code generating the database can be accessed from the source code hosted via *Gitlab* (https://gitlab.iscpif.fr/vverbavatz/mrt-access-project).

Received: 15 April 2020; Accepted: 30 July 2020;
Published online: 08 September 2020

References

1. Herzog, T. World greenhouse gas emissions in 2005. *World Resources Institute* (2009).
2. *United Nations, Department of Economic and Social Affairs, Population Division*. World Urbanization Prospects: The 2014 Revision. Highlights ST/ESA/SER.A/352 (2014).

3. Dodman, D. Blaming cities for climate change? An analysis of urban greenhouse gas emissions inventories. *Environ. Urban.* **21**, 185–201 (2009).
4. Glaeser, E. L. & Kahn, M. E. The greenness of cities: Carbon dioxide emissions and urban development. *J. Urban Econ.* **67**, 404–418 (2010).
5. Oliveira, E. A., Andrade, J. S. Jr. & Makse, H. A. Large cities are less green. *Sci. Rep.* **4**, 13–21 (2014).
6. Newman, P. G. The environmental impact of cities. *Environ. Urban.* **18**, 275–295 (2006).
7. Creutzig F. *et al.* Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Sciences* 112.20, 6283–6288 (2015).
8. Pumain D. Scaling laws and urban systems (Sante Fe Institute, 2004).
9. Barthelemy M. The structure and dynamics of cities. *Cambridge University Press* (2016).
10. Newman P. G. & Kenworthy J. R. Cities and automobile dependence: An international sourcebook (Gower Publishing, 1989).
11. Duranton, G. & Turner, M. A. The fundamental law of road congestion: Evidence from US cities. *Am. Econ. Rev.* **101**, 2616–52 (2011).
12. Buchanan, M. The benefits of public transport. *Nat. Phys.* **15**, 876 (2019).
13. Anderson, M. L. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *Am. Econ. Rev.* **104**, 2763–96 (2014).
14. Litman, T. Evaluating rail transit benefits: A comment. *Transp. Policy* **14**, 94–97 (2007).
15. Baum-Snow N., Kahn M. E. & Voith R. Effects of urban rail transit expansions: Evidence from sixteen cities, 1970–2000. *Brookings-Wharton papers on urban affairs*, 147–206 (2005).
16. Verbavatz, V. & Barthelemy, M. Critical factors for mitigating car traffic in cities. *Plos One* **14**, e0219559 (2019).
17. Dalziel, B. D., Pourbohloul, B. & Ellner, S. P. Human mobility patterns predict divergent epidemic dynamics among cities. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130763 (2013).
18. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences* **106**, 21484–21489 (2009).
19. Fouracre, P., Dunkerley, C. & Gardner, G. Mass rapid transit systems for cities in the developing world. *Transp. Rev.* **23**, 299–310 (2003).
20. Gallotti, R. & Barthelemy, M. Anatomy and efficiency of urban multimodal mobility. *Sci. Rep.* **4**, 1–9 (2014).
21. Gallotti, R. & Barthelemy, M. The multilayer temporal network of public transport in Great Britain. *Sci. Data* **2**, 1–8 (2015).
22. Musso A. & Vuchic V. R. Characteristics of metro networks and methodology for their evaluation. *National Research Council, Transportation Research Board* (1988).
23. Gattuso, D. & Miriello, E. Compared analysis of metro networks supported by graph theory. *Netw. Spat. Econ.* **5**, 395–414 (2005).
24. Derrible, S. & Kennedy, C. The complexity and robustness of metro networks. *Physica A* **389**, 3678–3691 (2010).
25. Marks M., Mason J. & Oliveira, G. People near transit: Improving accessibility and rapid transit coverage in large cities, *Institute for Transportation and Development Policy* (2016).
26. Singer G. & Burda C. Fast Cities: A comparison of rapid transit in major Canadian cities (Pembina Institute, 2014).
27. Dijkstra L., Poelman H. & Veneri P. The EU-OECD definition of a functional urban area. *OECD Regional Development Working Papers, 2019/11, Éditions OCDE, Paris* (2019).
28. Florczyk A. J. *et al. GHSL Data Package 2019*. ISBN 978-92-76-13186-1 (Publications Office of the European Union, 2019).
29. *Center for International Earth Science Information Network (CIESIN)—Columbia University*. Gridded population of the world, version 4 (GPWv4): population density (2016).
30. Verbavatz, V. & Barthelemy, M. People Near Transit (PNT). *figshare* https://doi.org/10.6084/m9.figshare.12013020.v4 (2020).

Acknowledgements
V.V. thanks the École nationale des ponts et chaussées for their financial support. This material is based upon work supported by the Complex Systems Institute of Paris Île-de-France (ISC-PIF).

Author contributions
V.V. and M.B. designed the study, V.V. acquired the data, V.V. analyzed and interpreted the data, V.V. and M.B. and wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.B.

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.