www.netmob.org

2026 NetMob



BOOK OF ABSTRACTS

NITERÓI • RJ • BRAZIL

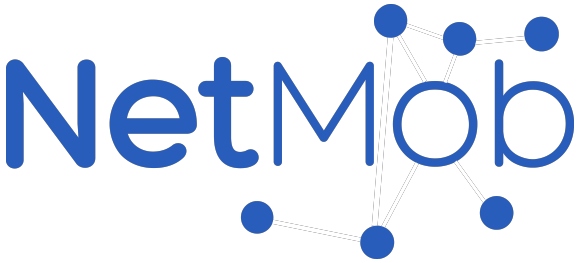
October 14th – 16th

Sponsors

UNICO Qualcomm

Partners

UFMG uff CNPq FAPERJ Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro



10th NetMob

Niterói, RJ, Brazil, October 14-16 2026

Book of Abstracts

Organizers

Jussara M. Almeida, UFMG
Antônio “Guto” Rocha, UFF
Antonio Loureiro, UFMG
Thiago H Silva, UTFPR
Humberto Torres Marques-Neto, PUC Minas
Bruno Pereira dos Santos, UFBA
Vinícius Mota, UFES
Clayson Celes, ITA
Felipe da Cunha, PUC Minas
Steffen Knoblauch, Heidelberg University
Igor Monteiro Moraes, UFF
Célio Vinicius Neves de Albuquerque, UFF

Steering Committee

Vincent Blondel, University of Louvain
Leo Ferres, IDSUDD, ISI Foundation & Telefónica
Marco Fiore, IMDEA Networks Institute & Net AI
Vanessa Frias-Martinez, University of Maryland
Renaud Lambiotte, University of Oxford
Esteban Moro Egidio, Northeastern University
Alex (Sandy) Pentland, MIT
Sveta Milusheva, World Bank
Aline Carneiro Viana, INRIA

Abstracts

Main Conference: Talks

A behavioral census of US cities from interpretable mobility embeddings 1
Hamish Gibbs (Northeastern University, USA), Esteban Moro (Northeastern University, USA)

The power law in human mobility is a mixture artifact: evidence from a pandemic natural experiment 3
Leo Ferres (Universidad del Desarrollo, Chile), Bruno Gonçalves (D4Sci, USA)

The Hyperfractal Dimension: Measuring the Multi-Scale Structure of Urban Mobility 5
Geoffrey Deperle (INRIA Saclay, France), Philippe Jacquet (INRIA, France)

Navigating decision points in mobility data 7
Giuseppe Perona (UC Berkeley, USA), Marta Gonzalez (UC Berkeley, USA)

An Interactive Stop-Detection-Algorithm Tool for Comparison and Visualization 9
Jorge Barreras (University of Pennsylvania, USA), Tina Ni (University of Pennsylvania, USA), Lorenzo Santos (University of Pennsylvania, Philadelphia, USA), Celia Tung (University of Pennsylvania, USA), Darren Wang (University of Pennsylvania, USA)

Fastkit-Mobility: A High-Performance, DataFrame-Agnostic Library for Mobility Analysis 11
Gustavo H. B. dos Santos (Federal University of Technology of Parana, Brazil), Aline Carneiro Viana (Inria Saclay, France), Luca Pappalardo (CNR Pisa, Italy), Thiago H. Silva (Federal University of Technology of Parana, Brazil)

Correcting socioeconomic bias in mobile phone mobility estimates 13
Leo Ferres (Universidad del Desarrollo, Chile), Laetitia Gauvin (IRD, France)

Selection Bias in Phone-Based Migration Statistics: Evidence and Correction Strategies 15
Paul Blanchard (World Bank, France), Sveta Milusheva (World Bank, USA), Francisco Rowe (University of Liverpool, UK), Stefania Rubrichi (Orange, France)

Identifying and characterizing dangerous road sections using vehicle telemetry and OpenStreetMap data 17
Giovana Piorino (Universidade Federal de Minas Gerais, Brazil), Andrea Lanzillotta (Politecnico di Torino, Italy), Luca Vassio (Politecnico di Torino, Italy), Danilo Giordano (Politecnico di Torino, Italy), Ana P. C. da Silva (Universidade Federal de Minas Gerais, Brazil), Marco Mellia (Politecnico di Torino, Italy), Claudio De Tommasi (Unipoltech, Italy), Marco Di Cristina (Unipoltech, Italy)

People-Place Networks Reveal Economic Disparities in United States 19
Baoshen Guo (MIT, USA), Yuzhou Chen (University of Florida, USA), Shenhao Wang (University of Florida, USA)

Amenity Proximity Redistributes Mobility Across Home, Workplace, and City Center: Evidence from Austria 21
Piero Birello (Central European University, Austria), Andrew Renninger (Central European University, Austria), Marton Karsai (Central European University, Austria)

Planning for isolation? The role of urban form and function in shaping mobility in Brasília 23
Andrew Renninger (Central European University, Austria)

Beyond Proximity: Rethinking Urban Accessibility Through Human Mobility Patterns in United States 25
Saumitra Kulkarni (Northeastern University, USA), Esteban Moro (Northeastern University, USA)

Moving Unequally: Phone-Based Mobility Metrics Reveal Socio-Spatial Inequality in Mexico City 27
Randy B. Gallegos-Rodríguez (Centro de Investigación en Computación, Mexico), Tonatiuh Suarez-Meaney (Centro de Investigación en Computación, Mexico), Miguel Jesús T. Ruiz (Centro de Investigación en Computación, Mexico), Maribel R. Hernandez (CINVESTAV-IRA, Mexico)

The parenthood effect in urban mobility 29
Mariana Macedo (Constructor University, Germany), Ronaldo Menezes (University of Exeter, UK), Alessio Cardillo (University of Barcelona, Spain)

Measuring the Daytime POI-visit Impacts of Hybrid Work across U.S. Cities 31
Callie Clark (New York University, USA), Xiaotong Xu (The Hong Kong Polytechnic University, China), Wei Ma (The Hong Kong Polytechnic University, China), Takahiro Yabe (New York University, USA)

A Multi-resolution Mobility-aware Framework for Dynamic Air Pollution Exposure Assessment 33
Bassel Othman (IFPEN, France), Guillaume Sabiron (IFPEN, France)

Social media-based measures of social capital attenuate economic disparities in extreme weather evacuations and recovery 35
Aparimit Kasliwal (UC Berkeley, USA), Carlos Guirado (UC Berkeley, USA), Marta Gonzalez (UC Berkeley, USA)

Prediction of Infection Risk in an Epidemic Contact Network via Temporal Line Graph Representations 37
Mathieu Gilles (Télécom SudParis, France), Zaher Moussawel (Télécom SudParis, France), Miguel Nunez-del-Prado (World Bank, USA), Hugo Alatrista-Salas (De Vinci, France), Vincent Gauthier (Télécom SudParis, France)

Empirical Trajectory Sparsity Biases Mobility-Informed Epidemic Modeling Outcomes and Parameter Estimates 39
Federico Delussu (Technical University of Denmark, Denmark), Jorge Barreras (University of Pennsylvania, USA), Yuan Liao (Lund University, Sweden), Duncan J. Watts (University of Pennsylvania, USA), Laura Alessandretti (Technical University of Denmark, Denmark)

Mobility Networks Reveal Association between Dynamic Food Environment and Chronic Cardiometabolic Diseases across U.S. Cities 41
Yuzhou Chen (University of Florida, USA), Baoshen Guo (MIT, USA), Chang Zhao (University of Florida, USA), Shanting Chen (University of Florida, USA), Shenhao Wang (University of Florida, USA)

Transmission-aware mobility interventions for epidemic control 43
Bibandhan Poudyal (University of Rochester, USA), David S. Panos (University of Rochester, USA), Gourab Ghoshal (University of Rochester, USA)

Urban mobility network centrality predicts social resilience 45
Lin Chen (Northeastern University, USA), Fengli Xu (Tsinghua University, China), Esteban Moro (Northeastern University, USA), Pan Hui (Deutsche Telekom Laboratories, Germany), Yong Li (Tsinghua University, China), James Evans (University of Chicago, USA)

Infrastructure Matters: Quantifying Measurement Bias in Mobile Phone Mobility Data 47
Leo Ferres (Universidad del Desarrollo, Chile), Erick Elejalde (L3S Research Center, Germany), Kyriaki Kalimeri (ISI, Italy), Loreto Bravo (Universidad del Desarrollo, Chile)

Parental Activity Space Across Childcare Stages 49
Shusei Hayashi (Tohoku University, Japan), Takahiro Yabe (New York University, USA), Naoya Fujiwara (Tohoku University, Japan)

Invariant, elastic and plastic: how mobility reorganizes around a residential move 51
Andrew Renninger (Central European University, Austria), Piero Birello (Central European University, Austria), Marton Karsai (Central European University, Austria)

Cross-Metropolitan Transferability of Behavior-based Dependency Networks 53
Boyang Li (New York University, USA), Saumitra Kulkarni (Northeastern University, USA), Morgan Frank (Case Western Reserve University, USA), Esteban Moro (Northeastern University, USA), Takahiro Yabe (New York University, USA)

Do tsunami evacuation drills reproduce real evacuation response? Evidence from paired events in coastal Chile 55
Kyriaki Kalimeri (ISI, Italy), Leo Ferres (Universidad del Desarrollo, Chile)

Comparing Gender-Labeled Urban Co-Visitation Networks Across U.S. Cities Using Google Places 57
Jemal Jilo (Federal University of Technology of Parana, Brazil), Myriam R. B. S. Delgado (Federal University of Technology of Parana, Brazil), Thiago H. Silva (Federal University of Technology of Parana, Brazil)

Leveraging mobility behavior to re-identify users 59
Lucas Felix (Universidade Federal de Minas Gerais, Brazil), Vinícius da F. Vieira (Universidade Federal de São João del rei, Brazil), Anne J. Kouam (INRIA Saclay, France), Alexandre G. Evsukoff (COPPE/UFRJ, Brazil), Aline Carneiro Viana (INRIA Saclay, France), Nadjib Achir (INRIA Saclay, France), Jussara Almeida (Universidade Federal de Minas Gerais, Brazil)

Large Language Models Create an Uneven Informational Layer over Cities 61
Lin Chen (Northeastern University, USA), Guangyuan Weng (Northeastern University, USA), Esteban Moro (Northeastern University, USA)

Do LLM Agents Reproduce Human Mobility Laws? 63
Gustavo H. B. dos Santos (Federal University of Technology of Parana, Brazil), Aline Carneiro Viana (Inria Saclay, France), Thiago H. Silva (Federal University of Technology of Parana, Brazil)

Predictability of Urban Park Visitation Behavior Using Context-Aware Fine-Tuned LLM Agents 65
DongHak Lee (New York University, USA), Xiyu Pan (New York University, USA), Yixuan Hong (New York University, USA), Takahiro Yabe (New York University, USA)

Some Users Are Easier to Re-identify: Evidence from Mobility Profiles and Embedding Geometry 67
Tiziano Martin Bernal (Universidad de San Andrés, Argentina), Anne J. Kouam (INRIA Saclay, France), Mariano Beiró (Universidad de San Andrés, Argentina)

BUBBLE-BLUE: private group messaging over an open Bluetooth mesh, measured on six boards 69
Paul-Maxence Baraton (INRIA Saclay, France), Wilfred Dore (QUALCOMM, France), Nadjib Achir (INRIA Saclay, France), Philippe Jacquet (INRIA, France)

Data Challenge: Talks

Privacy and anonymization of stochastic mobility trajectories from bus payment logs 71
Vincenzo Ricciardelli (Politecnico di Torino), Nikhil Jha (Politecnico di Torino), Luca Vassio (Politecnico di Torino)

Unequal by Design: Bus Network Structure, Social Vulnerability, and Favela Access in Niterói 73
Issa Moussa Diop (Université de Pau et des Pays de l'Adour), Ali Yassin (Université Bourgogne Europe), Randy Brandon Gallegos-Rodríguez (Instituto Politécnico Nacional), Tonatiuh Suárez-Meaney (SEDATU, Mexico), Maribel Hernández-Rosales (CINVESTAV Irapuato)

Planned and experienced accessibility reveal contrasting diagnoses of transit poverty and inequality 75
Egor Kotov (TU Dortmund University; Universitat Pompeu Fabra; Max Planck Institute for Demographic Research), Yuan Liao (Lund University), Leonardo Akinaga Moreira (University of São Paulo), Rafael H. M. Pereira (Institute for Applied Economic Research – Ipea)

Edge on Wheels: Forecasting Bus Fleet Compute with Calibrated Uncertainty 77
Wellington Lobato (Inria Saclay), Nadjib Achir (Inria Saclay), Aline Carneiro Viana (Inria Saclay)

Transit Opportunity Landscapes Under Climate Stress: Spatial Equity, Vulnerability Evaluation, and the Usage–Accessibility Gap 79
Chenchen Sun (The University of Tokyo; Renmin University of China), Yuxi Zhang (The University of Tokyo), Yichuan Zhang (The University of Tokyo), Shuta Okuaki (Pitzer College), Yuya Shibuya (The University of Tokyo)

Impact of provided services and socio-demographic factors on bus accessibility: a Niterói case study 81
Luca Fantin (IMT School for Advanced Studies Lucca; Motion Analytica), Chiara Pugliese (IIT-CNR), Fabio Pinelli (IMT School for Advanced Studies Lucca), Giulio Loddi (IMT School for Advanced Studies Lucca), Chiara Renso (ISTI-CNR), Alessandro Betti (IMT School for Advanced Studies Lucca), Bruno Zamengo (Motion Analytica), Andrea Frasson (Motion Analytica)

Reaching the Stop, Reaching the City: Socio-Spatial Inequality in Public-Transport Access in Niterói (SDG 11.2.1) 83
Kinin Koné (Agence Nationale de la Statistique – ANStat, Côte d'Ivoire), Abdoulaye Doumbia (Agence Nationale de la Statistique – ANStat, Côte d'Ivoire)

Urban Mobility and Demand Dynamics: A Telemetry and Ticketing Baseline for Niterói's Bus Network 85
Evaldo Alves de Brito Júnior (Federal University of Campina Grande), Luiz Sergio Pompeu Alves Filho (Federal University of Campina Grande), Claudio E. C. Campelo (Federal University of Campina Grande)

The Predictability of Bus Bunching and Who Pays for It: Evidence from Niterói, Brazil 87
Hugo Barbosa (University of Exeter), Federico Botta (University of Exeter), Zihao Chen (University of Exeter), Luiz Correia (University of Exeter), Yifan He (University of Exeter), Soroush Karimi (University of Exeter), Ronaldo Menezes (University of Exeter), Diogo Pacheco (University of Exeter), Yuchen Zhou (University of Exeter)

Finite-size-corrected synchronization in a city-scale busvnetwork: crossover, not demonstrated criticality 89
Po-Hsun Lat (Physics, National Taiwan University, Taipei, Taiwan)

Main Conference: Posters

GraniteMob: An AI Framework for Mobility Intelligence in the City of Niterói, Rio de Janeiro, Brazil 91
Krishna M. Bathula (Pace University, USA), Jaiden Roman (Pace University, USA), Amir Abdullozoda (Pace University, USA), Christelle Scharff (Pace University, USA)

Beyond Geographic Proximity: A Methodological Framework Informed by Mobility Data for Assessing Access to Specialized Hospital Services 93
Darcio L. Mario (Ebserh, Brazil), Rodrigo P. Campos (Ebserh, Brazil), Aluizio N. da Silva Junior (Universidade Federal Fluminense, Brazil), Marco de O. da Silva (Universidade Federal Fluminense, Brazil)

Mobile Data for Public Good: Sustainable Pathways within a Changing Mobile Ecosystem . . . 95
Anna Colquhoun (GSMA, UK)

Using UAVs To Improve The Logistics Of Medical Supply Transport In Urban Areas 97
Pedro H. Maia (Pontifical Catholic University of Minas Gerais - PUC Minas, Brazil), Gabriel F. Pereira (Pontifical Catholic University of Minas Gerais - PUC Minas, Brazil), Felipe Belem (Pontifical Catholic University of Minas Gerais - PUC Minas, Brazil), Humberto T. Marques-Neto (Pontifical Catholic University of Minas Gerais - PUC Minas, Brazil)

Scalable multilayer urban social networks using Call Detail Records and census data 99
Paulo G. N. Cançado (Universidade Federal de São João del-Rei, Brazil), Carolina R. Xavier (Universidade Federal de São João del-Rei, Brazil), Alexandre G. Evsukoff (Universidade Federal do Rio de Janeiro, Brazil), Vinícius da F. Vieira (Universidade Federal de São João del-Rei, Brazil)

Daily municipal bus demand forecasting in Niterói: combining weekly recurrence, calendar effects, and operational heterogeneity 101
Matheus Silva (Universidade Federal do Rio e Janeiro, Brazil), Marina Baltar (Universidade Federal do Rio e Janeiro, Brazil), Luís E. Santos (CEFET-RJ, Brazil), (CEFET/RJ, Brazil), Glaydston M. Ribeiro (Universidade Federal do Rio e Janeiro, Brazil), Anderson Marden (Universidade Federal do Rio e Janeiro, Brazil)

Data Challenge: Posters

Detecting bus delays and service anomalies in Niterói with autoencoder latent feature representations 103
Matheus Germano (Federal University of Campina Grande), Claudio E. C. Campelo (Federal University of Campina Grande)

Demand-Weighted Reliability and Diagnostic Observability in Public Transport: A Multi-Source Alignment Approach 105
Manuel Mendoza-Hurtado (University of Cordoba), José Manuel Cuevas-Muñoz (University of Cordoba), Domingo Ortiz-Boyer (University of Cordoba)

Counterfactual Dispatch Regularization Reduces Passenger Waiting in Urban Bus Services . 107
M. Mavin De Silva (Nagaoka University of Technology; The University of Tokyo), Santiago Garcia Gabilondo (The University of Tokyo), Ketian Chen (The University of Tokyo), Shusei Hayashi (Tohoku University), Jue Ma (The University of Tokyo), Yoshihide Sekimoto (The University of Tokyo)

A GPS-Based Framework to Distinguish Road-Network Congestion from Bus Operation in Niterói, Brazil 109
Alexander Bousman (Washington and Lee University; DIS Study Abroad in Scandinavia)

The Impact of Precipitation on Public Transportation in Niterói, Brazil 111
Lucas Toshioka Tenório (University of São Paulo), Fernanda Yumi Assonuma (University of São Paulo), Pedro Henrique Campos Ribeiro (University of São Paulo), Alan Uema Tavares (University of São Paulo), Matheus Teixeira Passos Giarola (University of São Paulo)

Assessing the Effects of Precipitation on Public Transportation Demand Through Eigenvector Analysis 113
Geymerson S. Ramos (Inria, INSA Lyon), Sachit Mishra (Inria, INSA Lyon), Razvan Stanica (Inria, INSA Lyon)

Spatial Analysis of Operational Discrepancies in Niterói's Municipal Bus System 115
Matheus de Souza e Silva (COPPE, Federal University of Rio de Janeiro), Marina Leite de Barros Baltar (COPPE, Federal University of Rio de Janeiro), Luis Eduardo Santos Fortes (COPPE, Federal University of Rio de Janeiro; CEFET/RJ), Anderson Marden de Sousa Silva (COPPE, Federal University of Rio de Janeiro), Glaydston Mattos Ribeiro (COPPE, Federal University of Rio de Janeiro), Cintia Machado de Oliveira (CEFET/RJ), Daniel Kim Imai (COPPE, Federal University of Rio de Janeiro), Lucas Ribeiro Sampaio (COPPE, Federal University of Rio de Janeiro), Bruno Peçanha Simões (COPPE, Federal University of Rio de Janeiro)

Diagnosing Transit Performance from Passive Data: Reliability, Vulnerability and Boarding-Pressure in Niterói, Brazil 117
Swetha Cheemakurthi (Shiv Nadar University, Chennai, India)

A behavioral census of US cities from interpretable mobility embeddings

HAMISH GIBBS[1,*] AND ESTEBAN MORO[1]
[1] Network Science Institute, Northeastern University, Boston, MA, United States
[*] h.gibbs@northeastern.edu

1. ABSTRACT

Large-scale records of human mobility make it possible to measure features of urban life that are absent from standard demographic data: what people do, where they travel, and how they combine activities in their daily life. However, raw mobility traces are highly complex and pose risks to individual privacy, making them challenging to use as a primary source in downstream analysis. In this paper, we use GPS-based mobility data across 20 U.S. cities from 2019 to 2023, constructing activity-time profiles from GPS visits to points of interest and decomposing these profiles into low-dimensional behavioral representations capturing co-occurrence structure in visitation patterns. We design these latent mobility representations to be interpretable, low dimensional, and stable across populations, geographies, and over time, while remaining reliable across spatial scales of aggregation. The identified behavioral components reveal semantic patterns of activity, and can complement standard demographic features such as income, education, age, employment, and race/ethnicity. In downstream evaluations, we find that behavioral components improve predictions of a range of urban outcomes, and generalize across neighborhoods, cities, and over time. Further, we show that mobility embeddings provide a basis for quantifying meaningful shifts in mobility activity under disruptions such as the COVID-19 pandemic, showing how out-of-home activity re-organized under lockdown restrictions, and recovered at different rates, and with modified structure, during the relaxation of COVID-19 restrictions. We also show that changes in out-of-home activity were associated with changes in experiences of urban life, including experienced segregation and patterns of access to urban amenities.

Keywords: human mobility; behavioral census; interpretable machine learning; urban data science

2. INTRODUCTION

Census variables which describe the characteristics of populations resident within a set of administrative geographies are a common resource for urban scientists. These demographic measures are valuable, but they fail to capture the complex nature of urban behavior, which encompasses how people move, access amenities, and organize their lives in cities. Mobility data offer an avenue to measure these behavioral patterns at scale, but carry unique challenges of data complexity, bias, and risks to individual privacy. Research on human mobility has shown that individual movement is regular, structured, and useful for applications from epidemic modeling and economic forecasting to urban planning [1, 2]. Within cities, mobile phone data have been used to estimate experienced segregation and social mixing [3], to identify interpretable latent activity behaviors [4], and to learn embeddings that reveal barriers in urban space [5]. These studies show that mobility traces contain meaningful structure relevant to a broad range of urban processes. However, existing applications of mobility data typically use hand-crafted mobility indicators tailored to specific research domains. It remains unclear whether it is possible to learn robust, validated behavioral regularities which can be used as a reliable behavioral proxy for measurement of urban activity across cities, over time, and under disruptions to mobility behavior.

In this paper, we present a set of low-dimensional behavioral components learned across US cities, providing unsupervised embeddings of common patterns of behavior which recur across diverse populations and can help measure changes in behavior across geographies and over time. Our approach, which transforms large unstructured GPS traces into a small set of interpretable mobility features, provides a meaningful vocabulary to describe shifts in urban activity, reducing the complexity of raw GPS visitation into meaningful labels like "Fast Food visits," "Campus Life," or "Fitness and Outdoors." These interpretable representations can complement standard demographic features with indicators of typical patterns of mobility activity. To ensure validity of identified components, we assess the tradeoff between accuracy and interpretability, and ensure stability across geographies and over time. We also demonstrate the utility of these features in a set of downstream evaluations estimating diverse urban outcomes, and show how they can be used to understand shifts in urban behavior in response to disruption, including the COVID-19 pandemic.

3. METHODS

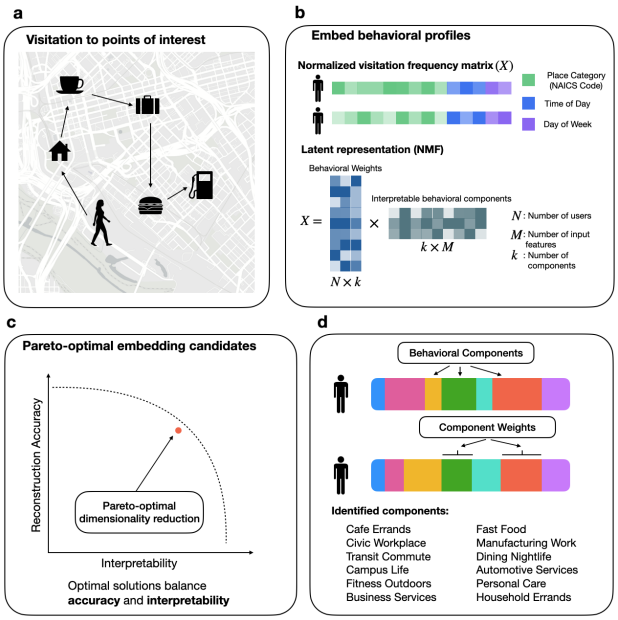


Fig. 1. Raw visits are encoded as activity-time profiles and decomposed into behavioral components using non-negative matrix factorization. Selected representations balance accuracy with interpretability, producing a set of latent behavioral features which describe mobility activity.

We analyze 31B GPS-based mobility records collected by the mobility data aggregator Cuebiq from 20 U.S. cities between 2019 and 2023. For each resident device, we construct an activity-time vector from visits to points of interest, combining venue categories with temporal signatures such as time of day and day of week. These features describe characteristic patterns of urban activity driven by behaviors such as commuting, eating, shopping, running errands, and socializing. We select a representative panel of mobile devices across cities and 6-month time periods, using inverse-penetration-rate sampling to adjust for sampling bias with respect to age, income, race, and population density. Using this demographically representative panel of mobile devices, we apply non-negative matrix factorization (NMF), decomposing activity profiles into additive components which describe co-occurring patterns within the input activity features. We select an appropriate dimensionality reduction using a multi-objective Pareto optimization that balances reconstruction accuracy, component interpretability, and low-rank structure. Our method permits each resident to be represented as an additive mixture of interpretable behavioral components, providing a low-dimensional behavioral embedding for each device which can be aggregated to describe the behavioral profile of neighborhoods. Separately, we validate the mobility embeddings by testing component interpretability,

reconstruction error, within-device split-sample robustness, and transfer of aggregated embeddings across cities and time periods.

To assess the predictive value of these unsupervised mobility embeddings, we aggregate individual behavioral profiles to areas (Census Tracts, ZIP Codes) based on the residence location of mobile devices. We use five-fold cross-validation within the main sample to compare predictive performance for demographics alone (D) and demographics plus behavioral components (D + B) across urban outcomes related to neighborhood structure, civic participation, housing, environment, and economic opportunity. We separately evaluate transfer of the combined D + B model to geographically held-out areas (OOS). As a case study, we explore the impact of the COVID-19 pandemic lockdown in March 2020, and quantify changes in mobility components across neighborhoods within cities, showing how the weight of different components changed as activity routines re-organized in response to pandemic restrictions. Though this case study focuses on the COVID-19 pandemic, the method is general and can serve to reveal shifts in behavior related to policy interventions, external disruptions, or more gradual trends of neighborhood change.

4. RESULTS

Our mobility embedding pipeline identifies optimal decompositions between 10 and 14 components along the accuracy-interpretability Pareto frontier. Identified behavioral components recover recognizable routines, including weekday commuting, campus life, weekend errands, fast food, outdoor recreation, and service-oriented activity. These components appear consistently enough across metropolitan areas to support comparison, while still varying across neighborhoods in ways not reducible to standard demographics.

In downstream evaluations, mobility embeddings complement standard demographics, adding predictive signal across all outcome domains, with particularly large gains across several domains, including civic engagement and housing. The combined model also generalizes across areas: performance in the geographically held-out evaluation closely matches its within-sample cross-validation performance across outcome domains.

Outcome domain	D	D + B	OOS (Full Model)
Civic Engagement (3)	0.41	0.55	0.54 ±0.09
Housing (8)	0.32	0.43	0.48 ±0.06
Economic Opportunity (4)	0.40	0.48	0.48 ±0.21
Built Environment (12)	0.26	0.33	0.33 ±0.10
Economy (3)	0.13	0.24	0.23 ±0.08
Environment (12)	0.16	0.23	0.21 ±0.11

Table 1. Predictive performance by outcome domain. D and D + B report five-fold cross-validation within the main sample for models using demographics alone and demographics plus behavioral components, respectively. OOS reports the D + B model evaluated on geographically held-out areas, with uncertainty across outcomes within each domain. Parentheses give outcome counts.

Mobility embeddings also reveal transitions in urban activity during COVID-19. By measuring changes in individual-level behavioral weights, or the share of individuals' behavioral weights in different mobility components, it is possible to measure changes to the semantic meaning of out-of-home activity, not just the reduction in overall travel volumes during the pandemic.

During the March 2020 lockdown, mobility embeddings show how out-of-home mobility routines shifted away from discretionary activities such as campus life, dining nightlife, and cafe errands, and towards more essential shopping and low-contact amenities such as household errands and fast food. These shifts confirm that mobility embeddings can describe meaningful responses to external shocks, and could be generalized to other processes which alter urban behavior such as urban policy interventions or acute disruptions to urban life like natural disasters. Moreover, using individual behavioral profiles, our results can offer insight into the specific mobility activities driving shifts in experienced segregation and the loss of in-person socializing, serving as a complement to existing understandings of the impact of homeworking and decreases in travel activity during the COVID-19 pandemic.

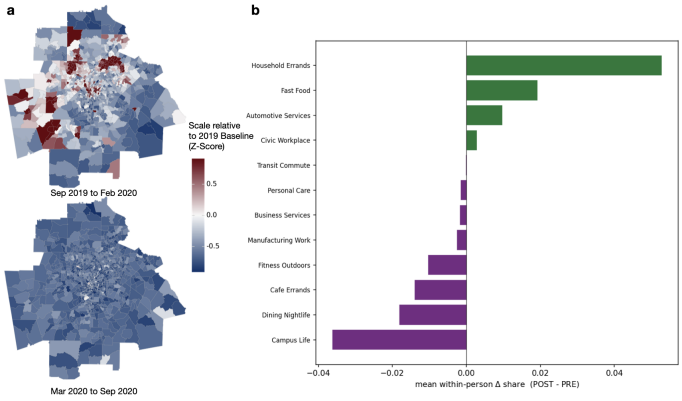


Fig. 2. a) COVID-19 case study showing neighborhood-level change relative to a 2019 baseline in Atlanta; b) component-level shifts between pre- and post-pandemic periods across 20 US cities.

5. DISCUSSION

We present mobility embeddings which provide a low-dimensional representation which translates raw mobility traces into interpretable behavioral components analogous to standard demographic variables, allowing them to be incorporated in urban policy applications much like existing census data. This is useful because many interventions act through behavior: transit investments, zoning changes, campus closures, park improvements, public-health restrictions, or changes in the geography of service areas may be more closely related to where people go and what they do than to their demographic characteristics.

Our approach offers an opportunity to gain insight into the ways that residents access the amenities within a city, revealing otherwise unobserved behavioral dependencies. By translating visits into interpretable behavioral components, we can measure how residents use the city, how those patterns differ across neighborhoods, and how they change during disruption. Our results show that, rather than replace traditional demographic features, mobility embeddings can complement existing urban data sources, providing complementary signal which can be incorporated in downstream planning and evaluations of urban policy.

REFERENCES

1. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779–782 (2008).
2. H. Barbosa, M. Barthelemy, G. Ghoshal, *et al.*, "Human mobility: Models and applications," Phys. Reports **734**, 1–74 (2018).
3. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. Commun. **12**, 4633 (2021).
4. Y. Yang, A. Pentland, and E. Moro, "Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics," EPJ Data Sci. **12**, 1–15 (2023).
5. G. Weng, M. Kim, Y.-Y. Ahn, and E. Moro, "Beyond distance: Mobility neural embeddings reveal visible and invisible barriers in urban space," arXiv preprint arXiv:2506.24061 (2025).

The power law in human mobility is a mixture artifact: evidence from a pandemic natural experiment

LEO FERRES[1,2,*] AND BRUNO GONÇALVES[2,3]

[1] Institute of Data Science, Universidad del Desarrollo, Santiago, Chile
[2] ISI Foundation, Turin, Italy
[3] Data For Science, Inc., New York, NY, USA
[*] lferres@udd.cl

For nearly two decades, human mobility has been read either as scale free or as lognormal. One account sees displacement distributions as heavy tails resembling truncated power laws, traceable to individual Lévy flights. A rival account holds that movement within each spatial container is lognormal and that the aggregate power law is an artifact of mixing containers of different sizes. Both fit the same aggregate data, so the debate has been hard to settle. We use the COVID-19 lockdowns in Chile as a natural experiment that removes long-distance travel and leaves local travel intact, analyzing 2.1 billion displacements from 4.4 million mobile-phone users. The aggregate exponent steepens under lockdown, but reweighting who travels reproduces that shift on its own, so the aggregate cannot separate the accounts. Individual-level tests can. The lognormal classification is stable and the power-law classification is not, and at matched sample size single users' tails reject the power law while equal-sized pooled mixtures pass. The power law of human travel is a feature of aggregation, not of individual movement.

1. INTRODUCTION

How far people travel shapes how disease spreads, how cities grow, and how transport is planned [1]. Across several orders of magnitude, displacement distributions look like a power law, $P(\Delta r) \propto \Delta r^{-\alpha}$ with α between 1.6 and 1.8 [2, 3]. Two explanations compete for its origin. The first (H1) holds that each person moves following a truncated Lévy flight set by a personal radius of gyration r_g [2]; According to this view the heavy tail is a real, scale-free property of individual movement. The second hypothesis (H2) holds that movement within a single spatial scale (a neighborhood, a city, a region) is lognormal, and that the observed power-law behavior is merely a result of how the local population is composed [4]. A mixture of lognormals has a lognormal body and a tail that is hard to distinguish from a power law over a finite sample [5, 6], so under H2 the population-level power law carries no information about any individual being scale free. Both accounts fit aggregate data equally well, which is why the debate has lasted.

2. OUR NATURAL EXPERIMENT

Distinguishing between the two hypotheses requires an intervention that removes one ingredient while keeping everything else the same. The COVID-19 lockdowns supplied one such intervention, suppressing long-distance trips between cities and regions while leaving local movement largely undisturbed, and the two hypotheses make opposite predictions under it. Under H1, the exponent α is a property of the individual process and should remain approximately constant; under H2 the heavy tail is an artifact of long-distance travel, so restricting long-range mobility should increase α. We analyze 2.1 billion displacement records from 4.4 million anonymized mobile-phone users in Chile across three periods set by government rules (pre-lockdown, partial, and full lockdown), measuring each displacement as the great-circle distance between the towers of two consecutive events of a user. The aggregate exponent steepens monotonically, from 1.657 before lockdown to 1.737 under full lockdown (Fig. 1), while the median radius of gyration r_g falls from 7.31 to 2.29 km. Both changes are consistent with H2.

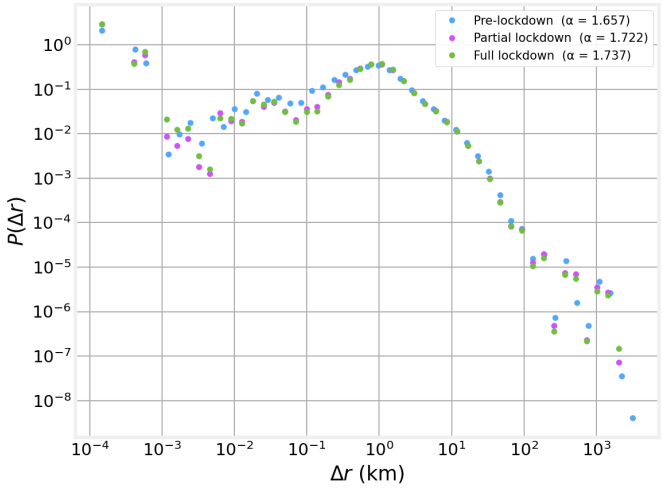


Fig. 1. Aggregate displacement distributions $P(\Delta r)$ by period. The power-law exponent α increases, from 1.66 to 1.74, as the long-distance mobility constraints are increased.

One explanation for the increase is that individuals constrained to travel shorter distances must follow a different distribution even if nothing else about their behavior changed. We test this by resampling unconstrained pre-lockdown users so that their r_g distribution matches the full-lockdown one, then refitting. Every individual displacement is left untouched; only the mix of users changes. The fitted exponent rises from 1.67 to 1.81, past the empirical full-lockdown value of 1.73. Composition alone reproduces and even overshoots the observed steepening, so the aggregate exponent cannot separate the accounts on its own; the test must move to the individual level.

3. INDIVIDUAL LEVEL ANALYSIS

We build a panel of 1.52 million users with at least 50 displacements in each of the three periods and fit a truncated power law and a lognormal to each user-period pair, choosing the better model with a Vuong [7] likelihood-ratio test. The fraction of users displaying lognormal behavior rises from 49.5% before lockdown to 60.9% under full lockdown, while the fraction best fitted by a truncated power law falls from 18.1% to 13.9%. The transitions are clearer still: of the users classified as lognormal before the lockdowns, 78.7% stay lognormal under lockdown,

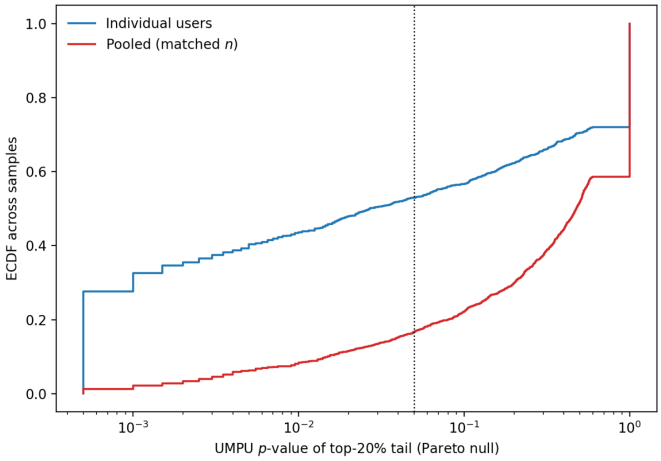


Fig. 2. Matched-size tail test. Maximum-entropy tail-test p-values for the top-20% tail across individual users (blue) and pooled samples of matched size (red). Over 50% of individual tails fall below the 5% line, rejecting the power law, while fewer than 20% of equal-sized pooled mixtures do. The power-law character is an artifact of the mixture, not of individual behavior.

while of those classified as power law only 36.9% remain and 30.9% switch to lognormal. The users who switch are the ones whose travel range was most reduced by the restrictions.

A further possibility is purely statistical: individuals have fewer data points than the population. We address it by matching sample sizes. For 1,537 users with at least 300 pre-lockdown displacements we compare each user's top-20% tail with a pooled population sample of the same size, testing both with a maximum-entropy method [6, 8]. Individual tails reject the power law 52.9% of the time; pooled samples of identical size reject it only 16.7% of the time, and the pooled power-law tail runs about three times longer (Fig. 2). Even at matched sample size, the power-law character comes from the mixture, not from the individual.

4. MODEL-FREE TAIL EVIDENCE

A power law is scale free, so a Pareto tail must remain Pareto however far out the threshold is placed. Above a threshold u, the log-excesses $\log(\Delta r/u)$ of a Pareto tail are exponential, so their coefficient of variation is 1 at every threshold, with no fitting and no choice of $x_{\min}$. Sliding the threshold across the top 500 to 500,000 of two million displacements, the coefficient of variation never settles at 1. It is about 1.25 between 10 and 30 km, where city-level and region-level containers mix, and falls to 0.4 to 0.8 between 50 and 800 km, the value expected inside a single lognormal container, crossing 1 only in passing near 35 to 40 km, close to where standard fitting procedures place $x_{\min}$. The power law is rejected at 97 to 98% of thresholds in all three periods, and the three profiles nearly coincide, so this reflects stable spatial structure rather than pandemic behavior. The extreme tail is not even continuous: the hundred longest displacements take only seven distinct values under full lockdown, down from 67 before. These are inter-city corridors that become unavailable as lockdowns remove inter-city travel, which H2 predicts and H1 cannot explain. Combining the within-level lognormal densities with their observed weights rebuilds the aggregate ($R^2 = 0.98$), and below 5 km the lognormal wins at every threshold. The Lévy prediction of a universal rescaling collapse $F(\Delta r/r_g)$ fails and worsens under lockdown (collapse CV 0.62 $\rightarrow$ 0.76), and the steepening concentrates in wide-ranging users, rising from $+0.051$ in the lowest r_g quartile to $+0.184$ in the highest ($p = 0.0003$).

5. ROBUSTNESS AND MECHANISM

Finally, our analysis suggests that the shift toward lognormal is not an artifact of data volume or spatial units. Down-sampling each user's pre-lockdown record to their lockdown count moves the lognormal fraction by about four points, an order of magnitude below the shift caused by the lockdown, and splitting the pre-lockdown period into two equally sized halves produces no change at all, with individual exponents correlated at $R^2 = 0.83$. Coarsening the spatial resolution, by rounding distances or remapping towers to a grid, leaves the lognormal preference in place and often strengthens it. A simple mechanism accounts for this: trip length within a container is the product of many positive factors (destination type, routing, availability), and products of positive factors tend to be lognormal, as in Gibrat's law.

6. CONCLUSION

We performed several independent lines of evidence that all corroborate H2: the population-level power law is a result of how the population is composed, not of how individuals move. A Pareto tail covering 14% to 16% of the range at a few thousand points, the size of the datasets originally used in this area, all but vanishes by ten thousand points, while the model comparison prefers lognormal distributions at every sample size, so data set size may explain why earlier work found power-law behavior [5]. On the Chilean data our findings reproduce the spatial analyses of González and colleagues while diverging on the temporal ones, as expected from denser event sampling [2]. They make the container model of Alessandretti and colleagues causal, since removing access to the upper levels changes the aggregate as the container account predicts, and they match the selective loss of long trips found in the German lockdown [9]. Some caveats are in order: the individual exponent shift is small (Cohen's $d = 0.349$), about 40% of users move the other way, and the data come from one country in an unusual period, so the exact magnitudes need not transfer. Chile's staggered Paso a Paso rollout turns this binary contrast into a graded dose-response test that we will develop in future work.

REFERENCES

1. D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," Nature **439**, 462–465 (2006).
2. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779–782 (2008).
3. C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," Nat. Phys. **6**, 818–823 (2010).
4. L. Alessandretti, U. Aslak, and S. Lehmann, "The scales of human mobility," Nature **587**, 402–407 (2020).
5. R. Perline, "Strong, weak and false inverse power laws," Stat. Sci. **20** (2005).
6. M. Bee, M. Riccaboni, and S. Schiavo, "Pareto versus lognormal: A maximum entropy test," Phys. Rev. E **84** (2011).
7. Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," Econometrica **57**, 307 (1989).
8. J. del Castillo and P. Puig, "The best test of exponentiality against singly truncated normal alternatives," J. Am. Stat. Assoc. **94**, 529–532 (1999).
9. F. Schlosser, B. F. Maier, O. Jack, *et al.*, "Covid-19 lockdown induces disease-mitigating structural changes in mobility networks," Proc. National Acad. Sci. **117**, 32883–32890 (2020).

The Hyperfractal Dimension: Measuring the Multi-Scale Structure of Urban Mobility

GEOFFREY DEPERLE[1] AND PHILIPPE JACQUET[1]

[1] Inria Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France
* geoffrey.deperle@inria.com

1. INTRODUCTION : HIERARCHY IN CITIES

Urban mobility does not occur on homogeneous spaces. Road systems are organized through layered hierarchies composed of highways, arterial roads, secondary streets and residential streets. At every scale, a small number of major roads concentrates a large fraction of the mobility flow, while a much larger number of smaller streets supports local mobility. Michael Batty [1] was among the first researchers to describe cities as exhibiting self-similar structures across scales. Large arterial structures tend to reappear recursively within finer street organizations, revealing strong statistical similarities between scales.

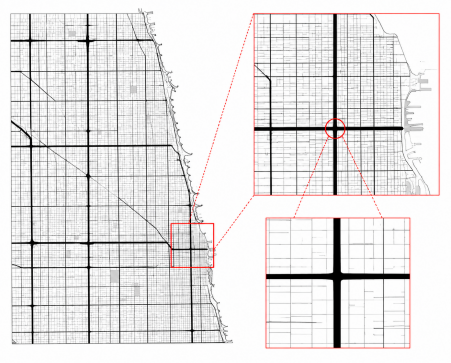


Fig. 1. Map of the city of Chicago showing multi-scale geometry. Large-scale arterial structures reappear at smaller scales inside the downtown district.

This hierarchy is not only geometric. Empirical studies show that traffic intensity, street centrality and road usage often follow power-law distributions [2]. If streets are ranked by decreasing traffic intensity, the traffic T_k carried by the k-th street approximately satisfies

$$T_k \propto T_0 k^{-\alpha},$$

where $\alpha > 0$ is a scaling exponent and T_0 is the traffic of the busiest street.

Such scaling laws suggest that urban mobility possesses an intrinsic multi-scale organization. This raises a natural question: can this hierarchical structure be quantified through a single mathematical quantity? In this work, we introduce the hyperfractal dimension framework to quantify the hidden multi-scale organization of urban mobility.

2. THE HYPERFRACTAL DIMENSION

To quantify the multi-scale organization of urban mobility, we develop a simple recursive construction called the *Manhattan model*.

The construction starts from a large central cross representing a major street intersection. At the next scale, smaller crosses are inserted inside each quadrant, representing secondary mobility structures. The procedure is then repeated recursively at finer scales, generating a hierarchical organization of streets.

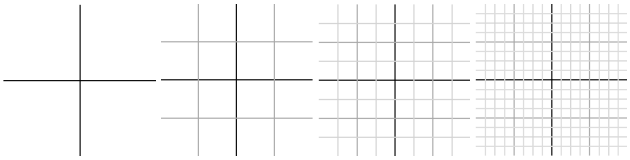


Fig. 2. Recursive construction of the Manhattan mobility model.

Mobility is distributed unevenly across scales: a large fraction of the traffic remains concentrated on major arterial roads, while only a smaller fraction propagates toward finer mobility structures. The hyperfractal dimension denoted by d_F quantifies how mobility is redistributed across scales, from major arterial roads to finer urban streets [3].

In the Manhattan model, this redistribution is controlled by a parameter $q \in (0, 1)$ describing the fraction of mobility recursively transferred toward smaller scales. The associated hyperfractal dimension is $d_F = \dfrac{\log(4/q)}{\log 2} > 2$.

Values of d_F close to 2 correspond to nearly uniform mobility distributions, whereas larger values indicate increasingly hierarchical concentrations on major streets.

The effect of the hyperfractal dimension can be visualized by distributing mobility events (vehicles, trips or parking requests) according to the hierarchical structure of the Manhattan model.

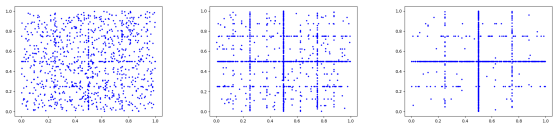


Fig. 3. 1000 simulated mobility events according to the Manhattan model for $d_F = 2.1$, $d_F = 3$, and $d_F = 4.3$.

As d_F increases, mobility progressively concentrates on a smaller number of dominant arterial roads.

3. MEASURING THE HYPERFRACTAL DIMENSION

A. Estimation from aerial images

Previous works estimated the hyperfractal dimension using detailed urban datasets containing both street geometries and traffic intensities [3]. While effective, such approaches require large-scale mobility measurements and substantial preprocessing.

Recently, we developed image-based estimation algorithms showing that aerial night images can provide a surprisingly effective and robust proxy for hyperfractal organization. The key observation is that the hyperfractal dimension does not depend on the exact value of traffic flows, but rather on how mobility is distributed across spatial scales. Major arterial roads dominate urban luminance and recursively generate visible structures at multiple resolutions.

As a consequence, the scaling organization of night luminance contains information about the hierarchical structure of urban mobility and we can estimate the hyperfractal dimension by measuring how typical luminance evolves across spatial scales. The resulting scaling law directly provides an estimate of the hyperfractal dimension.

Applied to Minneapolis, our method experimentally recovers $d_F \approx 2.18$, quantitatively consistent with mobility-based estimations.

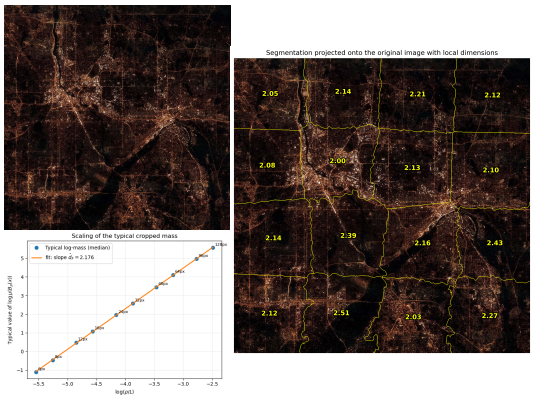


Fig. 4. Satellite nighttime image of Minneapolis (left top), associated scaling-law estimation yielding $d_F = 2.176$ with $R^2 = 0.97$ (left bottom), and neighborhood-scale segmentation based on local hyperfractal dimensions (right).

B. Local hyperfractal regimes inside cities

Real cities are not perfectly homogeneous. Different neighborhoods may exhibit distinct hierarchical organizations depending on density, road structure or mobility usage [4].

The image-based framework therefore naturally extends to local hyperfractal estimation. Instead of assigning a single dimension to the entire city, we estimate local hyperfractal dimensions over different urban districts.

Interestingly, despite local variations, estimated dimensions remain concentrated between 2 and 2.5. Our analysis suggests that Minneapolis remains reasonably approximated by a single effective hyperfractal dimension, while local refinements provide a more precise multi-scale description of urban organization.

This dual perspective is particularly important for mobility modeling. Depending on the desired level of accuracy, a city may either be modeled as a single effective hyperfractal structure or as a composition of locally hyperfractal districts.

4. HOW d_F SHAPES URBAN MOBILITY

Originally introduced to model information propagation in vehicular networks [3], hyperfractal geometry has more recently been applied to parking-search dynamics.

In free-floating car-sharing systems, finding an available parking slot is a central operational issue. While classical models often assume homogeneous parking opportunities, real cities exhibit strong hierarchical mobility structures that can be quantified through the hyperfractal dimension d_F.

In recent work [5], we showed that in a hyperfractal Manhattan model, the average distance $D(\lambda)$ before finding a parking slot satisfies the scaling law

$$\mathbb{E}[D(\lambda)] \sim C\lambda^{-1/d_F},$$

where λ denotes the city-wide parking turnover intensity.

This reveals that parking-search dynamics are mainly governed by the large-scale geometry of the street network rather than by purely local effects. We further showed that the exponent remains robust under random perturbations of street attractiveness such as commercial activity or local congestion. The hyperfractal dimension also influences the navigation process itself. The typical number of turns before parking scales as

$$\text{number of turns} \sim \frac{\log_2 \lambda}{d_F},$$

meaning that cities with larger hyperfractal dimensions tend to produce fewer turns before parking, but longer travelled distances.

Using our estimation framework across several cities, we observe distinct hyperfractal regimes.

City	Paris	Chicago	Minneapolis	Seattle	Adelaide
d_F	2.0	2.09	2.18	2.3	2.7

Table 1. Typical computed hyperfractal dimensions.

For instance, Seattle ($d_F \approx 2.3$) and Adelaide ($d_F \approx 2.7$), despite sharing regular grid-like urban structures, reveal different levels of hierarchical organization. Paris ($d_F \approx 2$), with its dense and highly interconnected urban fabric, lies closer to the nearly uniform regime.



Seattle ($d_F \approx 2.3$) Adelaide ($d_F \approx 2.7$)

In this framework, lower-dimensional cities tend to yield faster parking-distance scaling but require more turns before parking. Conversely, highly hierarchical cities concentrate parking opportunities along dominant roads, leading to fewer turns but slower distance scaling. These observations suggest that the hyperfractal dimension may link urban geometry, mobility flows, and parking-search efficiency.

5. CONCLUSION

The hyperfractal dimension provides a quantitative framework for describing the multi-scale organization of urban mobility. Beyond theory, image-based estimation suggests that hyperfractal structures may be observable in real cities. These results suggest that hyperfractal geometry may provide a bridge between urban form, mobility flows, and parking-search dynamics.

REFERENCES

1. M. Batty and P. Longley, *Fractal Cities: A Geometry of Form and Function* (Academic Press, 1994).
2. B. Jiang, "Street hierarchies: a minority of streets account for a majority of traffic flow," Int. J. Geogr. Inf. Sci. (2009).
3. P. Jacquet and D. Popescu, "Self-similar geometry for ad-hoc wireless networks: Hyperfractals," in *Geometric Science of Information*, (Springer, 2017), Lecture Notes in Computer Science, pp. 522–529.
4. G. Deperle and P. Jacquet, "Generating Random Hyperfractal Cities," in *Geometric Science of Information7th International Conference, GSI 2025, Proceedings, Part I*, (2025).
5. G. Deperle, C. Fricker, P. Jacquet, *et al.*, "Asymptotics of Parking Search in Hyperfractal Networks," in *LIPIcs – Leibniz International Proceedings in Informatics*, (2026).

Navigating decision points in mobility data

GIUSEPPE PERONA[1,*] AND MARTA C. GONZÁLEZ[1,2,3]

[1] Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94704
[2] Department of City and Regional Planning, University of California, Berkeley, CA, USA
[3] Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[*] perona@berkeley.edu

1. INTRODUCTION

Passively collected mobility data can be an immensely useful tool for urban planners, transportation engineers, and, more broadly, any entity with an interest in understanding human mobility behavior. Anyone interested in using this data is faced with the immediate choice of what type of data to purchase, and which provider to purchase it from. Many products exist that present mobility data with various levels of pre-processing. Veraset, for example, sells raw location-based services (LBS) records, but sells add-ons that link this information to home and work locations, trips, and points of interest (POIs). Safegraph focuses on the POI visits product. Replica goes further, providing a table of trips that is representative of a day of travel in the United States, estimated for the entire population. Once a researcher has chosen one or more of these products, more processing of the data they have purchased is necessary in order to make any inference about the world (see Fig. 1).

Both the dataset choice and the pipeline used downstream of that data induce variability into the conclusions drawn from the data. Many of these decisions have no clear correct choice. Rather, the application and characteristics of the data themselves often dictate how these decisions ought to be made. Thus, it is important for researchers to be aware of these degrees of freedom and the consequences they may have on their analyses. Concretely, we trace one policy-relevant quantity end-to-end. We use commute-associated vehicle kilometers traveled (VKT), which is the most important metric the state of California uses to determine environmental impact. VKT sits at the far end of the chain in Fig. 1, so every upstream choice affects the final number.

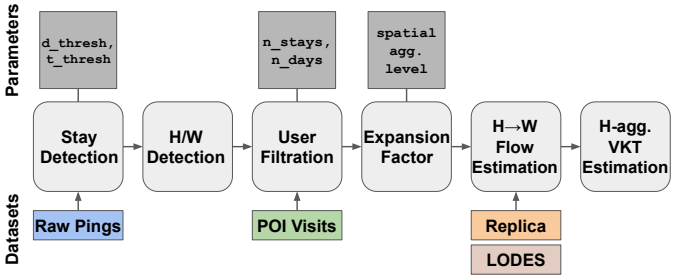


Fig. 1. A mobility data pipeline for comparing commute-associated vehicles kilometers travelled (VKT). Each dataset enters the pipeline at the stage corresponding to its level of pre-processing (bottom row). Parameters subject to researcher choice are shown above the stages where they apply (top row).

The dataset-induced variability may be further split into data collection and processing methodology. For example, consider LBS data. This source consists of trajectories of pings generated by on-device apps that track user locations that are bought by third-party aggregators who merge data sold by different providers. Each aggregator's methodology affects the data differently. For example, each aggregator might purchase data from a different basket of apps, and use different logic for merging data across apps to form a single user's trajectory. Importantly, the mechanics behind these differences are invisible to customers. Furthermore, these differences are difficult to compare across datasets due to the cost of purchasing redundant data from multiple providers. A more detailed discussion of the LBS data collection pipeline and what differences may occur across providers can be found in references [1–3].

Prior work describes sources of variability in mobility data. This literature is fundamental in identifying variability among sources, but [1, 3, 4] stop short of providing a detailed analysis of the data itself. Other work goes further, showing how data from different providers can yield different displacement distributions, and importantly contextualizes these differences in each provider's data collection methodology [2]. This analysis focuses on displacement distributions derived close to the raw data, preceding the deeper processing stages that many research applications require. Related pipelines spanning home detection, trip inference, and spatial aggregation have been built for call detail records (CDRs) and travel survey data [5]. Individual pipeline components have also received dedicated attention, including sensitivity analyses of home and work detection [6, 7]. Related work builds complete pipelines [8–11]; we build on these frameworks and contribute a systematic comparison across datasets and hyperparameter choices.

Our contribution builds on the literature along the following three axes: 1) We quantify dataset and hyperparameter-induced variability across four different sources of mobility data and variability propagation through the processing pipeline. 2) We compare the variability of commute-associated VKT. 3) We offer insights at some key decision points that researchers may use to guide their own choices. We hope to contribute to a standard of how this type of data should be handled.

2. RESULTS

We applied the data processing pipeline in Figure 1 to four datasets under the categories listed in the bottom row of that figure: raw pings products from Cuebiq and Veraset, which enter at the stay detection stage, and Replica and LODES, which enter at the home-to-work flow estimation stage. We include LODES as an administrative data source with collection methodology orthogonal to the others mentioned, which rely heavily on LBS data. Our Cuebiq data covers Texas in November 2019, while our Veraset data covers California in November 2024. We obtained Replica data covering California in Fall 2024, and Texas in Fall 2019, and LODES data from California in 2023 and Texas in 2019.

Before running stay detection on Cuebiq and Veraset, Figure 2 shows us some immediate differences between the two raw pings products. Both datasets show a prominent horizontal band corresponding to their respective location measurement noise floors[1]. Choosing distance thresholds below these noise bands

[1] In the Veraset data, latitude and longitude coordinates are often truncated to four decimal points, pinning much of the data to a $\sim$ 11 meter grid. The Cuebiq data is partially collected using GPS coordinates, and partially using WiFi-based location, which is less accurate, contributing to the near-100 meter band [12].

can cause the stay detection algorithm to miss real stays because of persistent noise-induced breaches of this threshold or classify location jitter between points that truly belong to the same stay as separate stays.

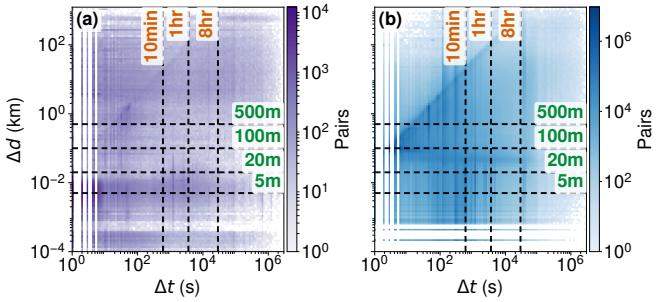


Fig. 2. Joint distribution of change in time (Δt) and change in position(Δd) between successive pings in (a) the Veraset California and (b) the Cuebiq Texas datasets. The Veraset data shows a horizontal band of higher density between about 5 and 20 meters, and the Cuebiq data has a band between 20 and 100 meters. In orange are three time thresholds tested in this analysis. In green are four distance thresholds we tested.

Next, after applying the pipeline from Fig. 1 to the Veraset and Cuebiq data without applying user filtration, we obtained home-work travel flow estimates from the datasets. To further inform the stay detection parameter choice, we found home-work flow correlations for Veraset and Cuebiq versus LODES at the tract level using the stay detection parameters listed in Figure 2. Of the twelve threshold combinations tested, Veraset's tract-level agreement with LODES was flat across $\{100\text{ m}, 500\text{ m}\} \times \{1\text{ hr}, 8\text{ hr}\}$ (Pearson $r = 0.30$ on log-transformed flows), while Cuebiq peaked at 100 meters with a time threshold of either 1 hr or 8 hr ($r = 0.47$).

We also tested various levels of user filtration, and found that home-work flows match LODES benchmarks better at looser levels of filtration. Specifically, we tried filtering users by the number of stays in their data, using minimum threshold values of $0, 5, 20$, and 50 stays, and by the number of days with stays, using threshold values of $0, 3, 10$, and 20 days. We found that Veraset's home-work flow correlation with LODES begins to drop off beyond keeping users with at least 5 stays and 3 days with stays, while Cuebiq's results dropped off beyond at least 20 stays and 10 days with stays. We attribute this effect to sparsity induced by more restrictive filtration, and our general recommendation is to filter out low-quality users only as far as the data quality allows. We also found that coarser spatial aggregation scales (Uber H3 resolution-6 hexagons) give better home-work flow correlations across both Veraset and Cuebiq, though the dropoff induced by filtering remains. Notably, Cuebiq appears to match the LODES benchmark more closely than Veraset.

Returning to VKT, Figure 3 shows that in California[2] different datasets come with vastly different conclusions. It is vital to consider the tradeoffs inherent in each dataset choice in the context of one's application. On one hand, LODES assumes that individuals work at the address on their tax forms, underestimating the number of people who work from home. On the other hand, Replica has no commutes over 300 km. Veraset's estimated VKT falls between the two, with a notably broader and flatter distribution, reflecting greater uncertainty in its home-work VKT estimates compared to the more sharply peaked LODES and Replica distributions. When seeking to estimate private vehicle emissions, these tradeoffs carry different consequences: LODES risks overcounting commute trips, Replica risks undercounting long-distance ones, and Veraset's wider spread suggests less consistency in its underlying estimates. The magnitude of these dataset-induced differences means that policy decisions relying on a single mobility data source carry meaningful uncertainty, underscoring the need to make these tradeoffs explicit rather than treating any one dataset's estimate as ground truth.

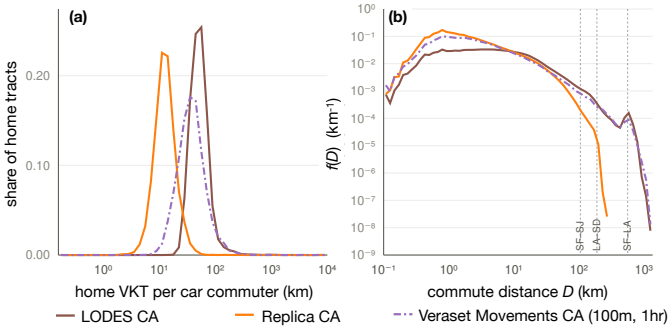


Fig. 3. Comparison of VKT per capita distributions across mobility datasets. (a) The Replica dataset estimates much lower per-tract VKT per capita than the other two datasets. We estimate per-capita VKT by estimating the distance between each home-work census tract pair in our data and multiplying it by the total flow we observe on that home-work tract pair. (b) The distribution of commute distances exposes the 300km cutoff in Replica.

[2]We found similar results in Texas using the Cuebiq data

REFERENCES

1. F. Barreras and D. J. Watts, "The exciting potential and daunting challenge of using GPS human-mobility data for epidemic modeling," Nat. Comput. Sci. **4**, 398 (2024).
2. R. Gallotti, D. Maniscalco, M. Barthelemy, and M. De Domenico, "Distorted insights from human mobility data," Commun. Phys. **7**, 421 (2024).
3. C. Chen, R. Wang, P. Bansal, *et al.*, "From biases to opportunities: Leveraging Location-Based-Service (LBS) data for next-generation transportation planning," Transp. Res. Part C: Emerg. Technol. **182**, 105416 (2026).
4. T. Yabe, M. Luca, K. Tsubouchi, *et al.*, "Enhancing human mobility research with open and standardized datasets," Nat. Comput. Sci. **4**, 469 (2024).
5. L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," Transp. Res. Part C: Emerg. Technol. **58**, 240 (2015).
6. M. Vanhoof, C. Lee, and Z. Smoreda, "Performance and sensitivities of home detection from mobile phone data," .
7. M. Vanhoof, F. Reis, T. Ploetz, and Z. Smoreda, "Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics," J. Off. Stat. **34**, 935 (2018).
8. J. Wolf, S. Schönfelder, U. Samaga, *et al.*, "Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information," Transp. Res. Rec. J. Transp. Res. Board **1870**, 46 (2004).
9. S. Cao, C. Pais, V. Lingenfelter, *et al.*, "A Data Science Framework to Measure Vehicle Miles Traveled by Mode and Purpose," .
10. M. Yang, W. Luo, M. Ashoori, *et al.*, "Big-Data Driven Framework to Estimate Vehicle Volume Based on Mobile Device Location Data," Transp. Res. Rec. J. Transp. Res. Board **2678**, 352 (2024).
11. L. Zhang, S. Ghader, A. Darzi, *et al.*, "Data Analytics and Modeling Methods for Tracking and Predicting Origin-Destination Travel Trends based on Mobile Device Data," Technical Report FHWA-HRT-20-012, United States. Federal Highway Administration (2020).
12. "Privacy Policy | Cuebiq Group, LLC," (2017).

An Interactive Stop-Detection-Algorithm Tool for Comparison and Visualization

FRANCISCO BARRERAS*[1], TINA NI[1], LORENZO SANTOS[1], CELIA TUNG[1], AND DARREN WANG[1]

[1] Department of Computer and Information Science, University of Pennsylvania, 3333 Chestnut Street, Philadelphia, PA 19104
* fbarrer@sas.upenn.edu

As human mobility science applications mature, there are growing calls to validate the algorithms used in data processing [1, 2]. Commercially-available GPS mobility datasets are often temporally sparse and spatially noisy, which can distort algorithmic outputs and produce anomalies [3–6]. One critical processing step is stop detection—identifying segments of a GPS trajectory in which an individual remained stationary—which both reduces dimensionality, and enriches the raw data with mobility semantics. Detected stops, also called *stay-points*, underlie many mobility metrics and downstream tasks such as home attribution, visit identification, and mobility flows, raising concerns that algorithm and parameterization choices bias results [2]. Yet, rigorous comparisons of stop-detection algorithms or guidance on their parameterization remain limited. Ground-truth data for evaluating accuracy and robustness are rarely available, while open-source implementations are available, but scattered across a small number of libraries, including `infostop`, `nomad`, `scikit-mobility`, and `HuMobi`.

We address these barriers by providing open-source implementations of eight stop-detection algorithms in an interactive pedagogical tool that supports direct comparison and evaluation against synthetic ground truth [7]. Our "Stop Detection Dashboard" makes a broader range of algorithms accessible to researchers, explains their distinctive features, and allows users to examine how their outputs change across parameter settings and trajectory characteristics. Visual analytics of movement has similarly combined computational methods with interactive spatial and temporal representations to support exploration of trajectory data [8, 9]. Its front end combines animated maps, timelines, insights about algorithm operations, and interactive parameter controls, allowing users to follow GPS pings through time and observe how different algorithms and parameterizations produce detected stops, errors, and characteristic failure modes.

DATA AND METHODS

Our learning tool combines synthetic trajectory data, open-source Python implementations of multiple algorithms, and an interactive, animated front end for exploring algorithm outputs (Fig. 1d). We generated synthetic GPS trajectories using NOMAD's open-source *garden city* sandbox [7], which simulates mobility within a synthetic city and allows control over trajectory irregularity, temporal sparsity, and horizontal noise. The prototype includes four mobility "diaries" defining the ground-truth stops; for each selected diary and sparsity configuration, we generated 500 sparse, noisy trajectories. We also contribute documented and unit-tested open-source implementations of eight stop-detection algorithms through the `nomad` Python library [10]. These algorithms are organized in two categories. The first are *Sequential* algorithms, implementing an intuitive and straightforward notion of a stop: consecutive pings falling inside a containment area for a minimum duration. We propose 4 implementations of this class of algorithms, that differ on how such containment area is defined as the algorithm is expanding a "candidate stop": (i) as a circle centered (or "anchored") on the first ping of the candidate stop [11]; (ii) as a changing circle centered on the centroid of the pings in the candidate stop [12]; (iii) implicitly, by requiring a maximum (geometrical) diameter on the set of pings, as in Project Lachesis [13]; (iv) as a directly provided polygon, like an H3 hexagon, or administrative boundaries, as in [14]. The second category corresponds to Density-based algorithms, which are *spatio-temporal* adaptations of the well-established DBSCAN [15] *spatial* clustering algorithm, which can effectively handle noise—relaxing the requirement that pings within a stop be consecutive. The four implementations chosen differ primarily on how temporal separation of the resulting clusters is ensured: (i) SeqScan, which implements DBSCAN on growing sequences of pings, adding a cutoff time when a candidate stop is "closed"; (ii) ST-DBSCAN which limits the pings that can be neighbors to those within a temporal threshold, (iii) DBStop, which removes overlaps directly with a maximum-likelihood criterion on conflicted pings; and (iv) HDBStop, a variation of DBStop that implements hierarchical clustering. The two latter implementations are original to the NOMAD project [10], which also includes open-source implementations of the other 6.

Algorithm outputs are precomputed for each trajectory and parameter configuration, and compressed in a cache so that the dashboard can load comparisons quickly without online execution of the algorithms. Performance is evaluated minute-by-minute by spatially joining detected stops to the synthetic buildings: a minute is correct when the detected building matches the ground-truth building specified by the diary, and incorrect when it does not or when no stop is detected. The fraction of correctly classified minutes provides a common accuracy measure across algorithms and trajectories. The dashboard links these precomputed results to an interactive map and timeline, allowing users to move through each trajectory and inspect when GPS pings occur, how detected clusters form, and where stop assignments diverge from ground truth.

RESULTS

The NOMAD Stop Detection Dashboard helps users understand how different stop-detection algorithms transform raw GPS traces into stops or visits and where each algorithm may fail. The dashboard has three main analytical views: *Learn* summarizes each algorithm's assumptions, source, parameters, and an explanatory diagram, with links to notebook demonstrations. *Demo*, visualizes a selected algorithm and parameterization on a synthetic trajectory, showing GPS pings, ground-truth stops, detected clusters, and algorithm-specific features such as anchors and DBSCAN core points, also allowing side by side comparisons of two algorithms. *Aggregated Metrics* shows performance across a large sample of trajectories generated from a common ground truth, as a function of algorithm parameters, including mean performance and showing the best parameter (and performance) of each trajectory. Users can select individual trajectories from this view and inspect them in Demo to understand the stop-level failures underlying aggregate patterns.

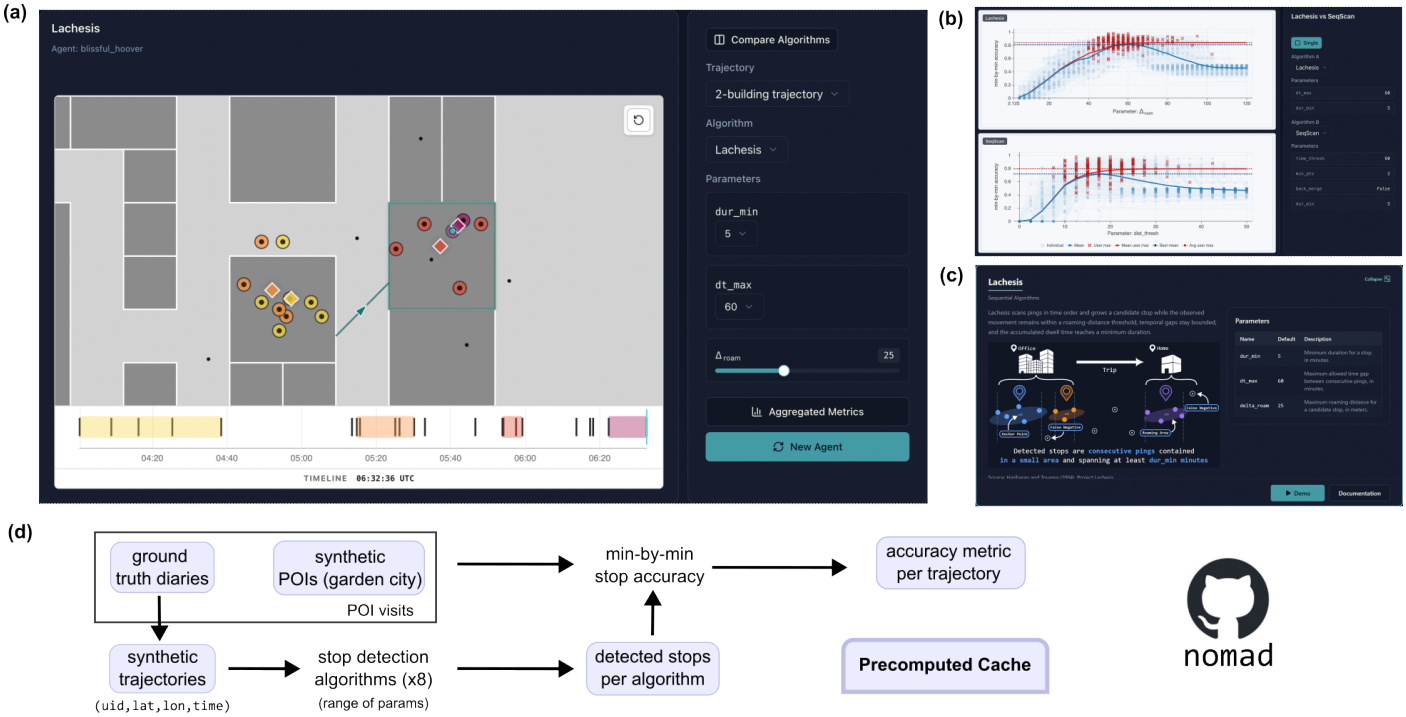


Fig. 1. Overview of the NOMAD Stop Detection Dashboard. (a) The Demo view synchronizes a trajectory map and timeline, showing raw GPS pings, ground-truth stops, detected stops, and algorithm-specific features while users vary parameters or compare algorithms on the same trajectory. **(b)** The Aggregated Metrics view shows minute-level stop-detection accuracy across parameter settings and trajectories; individual trajectories can be highlighted and opened in the Demo view for inspection. **(c)** The Learn view summarizes each implemented algorithm with a method diagram, parameter descriptions, and source. **(d)** Synthetic trajectories with known ground-truth diaries are processed offline using eight stop-detection implementations; detected stops and accuracy metrics are precomputed for interactive exploration in the dashboard.

As a use case, Fig. 1c shows side-by-side aggregated metrics of accuracy for 500 trajectories across the effective range of scale parameters for Lachesis and SeqScan. This view illustrates two common failure modes: when the scale parameter is too small, a stop may be fragmented into several detected stops; when it is too large, stops corresponding to different buildings may be merged. The trajectory-level optima (red markers) also show how preferred parameter settings vary across trajectories.

DISCUSSION AND AVAILABILITY

By providing visual and quantitative feedback on how algorithmic choices affect detected stops and accuracy, the dashboard allows researchers to inspect stop-detection failure modes and parameter sensitivity directly. Limited ground-truth data, scattered open-source implementations, and limited guidance on parameter choices make such comparisons difficult. The interactive maps and timelines, performance metrics, and algorithm explanations, support direct exploration of stop-detection methodologies before downstream analyses.

Additional algorithms can be incorporated through NOMAD's public Python library, which accepts community contributions. Support for arbitrary user datasets remains future work; integrations with Leaflet or ArcGIS could reuse the existing visualization architecture for real trajectories. The source code is available at https://github.com/Watts-Lab/nomad-stop-detection-dashboard, and the tool is publicly available at https://nomad.seas.upenn.edu/dashboards/stop-detection.

Funding This work was possible with the support of the National Science Foundation, NSF award 2341932.

REFERENCES

1. S. de Sojo, L. Lucchini, O. D. Langle-Chimal, *et al.*, Comput. Environ. Urban Syst. **126**, 102397 (2026).
2. F. Barreras and D. J. Watts, Nat. Comput. Sci. pp. 1–14 (2024).
3. S. Bähr, G.-C. Haas, F. Keusch, *et al.*, Soc. Sci. Comput. Rev. **40**, 212 (2022).
4. A. Graser, J. Locat. Based Serv. **15**, 89 (2021).
5. D. McCool, P. Lugtig, and B. Schouten, Transportation **51**, 297 (2024).
6. S. A. Sanchez, H. Gibbs, T. Yabe, *et al.*, arXiv preprint arXiv:2601.22330 (2026).
7. T. H. Li and F. Barreras, arXiv preprint arXiv:2412.00913 (2024).
8. N. Andrienko and G. Andrienko, Inf. visualization **12**, 3 (2013).
9. S. Dodge and E. Noi, Cartogr. Geogr. Inf. Sci. **48**, 353 (2021).
10. F. Barreras, T. Li, F. Delussu, and A. Mondragon, "Nomad: Network for open mobility analysis and data," GitHub repository (2025). Version 0.2.0.
11. Y. Ye, Y. Zheng, Y. Chen, *et al.*, "Mining individual life pattern based on location history," in *2009 tenth international conference on mobile data management: Systems, services and middleware,* (IEEE, 2009), pp. 1–10.
12. K. Zhao, S. Tarkoma, S. Liu, and H. Vo, "Urban human mobility data mining: An overview," in *2016 IEEE International Conference on Big Data (Big Data),* (IEEE, 2016), pp. 1911–1920.
13. R. Hariharan and K. Toyama, "Project lachesis: parsing and modeling location histories," in *International Conference on Geographic Information Science,* (Springer, 2004), pp. 106–124.
14. E. Pepe, P. Bajardi, L. Gauvin, *et al.*, Sci. data **7**, 230 (2020).
15. M. Ester, H.-P. Kriegel, J. Sander, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *kdd,* , vol. 96 (1996), pp. 226–231.

Fastkit-Mobility: A High-Performance, DataFrame-Agnostic Library for Mobility Analysis

GUSTAVO H. SANTOS[1,2,*], ALINE CARNEIRO VIANA[2], LUCA PAPPALARDO[4], AND THIAGO H. SILVA[1,3]

[1] UTFPR, Curitiba, Brazil
[2] INRIA, Paris, France
[3] University of Toronto, Toronto, Canada
[4] ISTI-CNR, Pisa, Italy
[*] gustavohenriquesantos@alunos.utfpr.edu.br

1. INTRODUCTION

Over the past decades, massive spatiotemporal mobility datasets have become widely available, containing observations from GPS positioning systems, cellular network records, and geotagged social media posts. These datasets have fueled extensive research on human mobility, advancing mobility understanding and applications ranging from computational epidemiology to urban planning and transportation engineering.

To support reproducible and standardized mobility analytics, dedicated open-source Python libraries, such as *Scikit-Mobility* [1] or *Trackintel* [2], have been developed to provide reusable tools for trajectory preprocessing and analysis. While *Trackintel* adopts a hierarchical movement data model originating from transport planning and emphasizes semantic trajectory processing (e.g., activity detection, trip generation, and travel mode prediction), *Scikit-Mobility* [1] has become the *de facto* library for quantitative human mobility analytics. It provides a comprehensive framework for mobility metrics computation, mobility-flow analysis, synthetic mobility generation, and privacy analysis.

Given its broad adoption and comprehensive functionality, *Scikit-Mobility* has become the reference library for human mobility analysis, in particular, within the *NetMob* community. Improving its scalability therefore benefits a wide range of existing mobility-analysis workflows. Achieving this objective, however, requires addressing the scalability limitations of its current implementation. Built on Python and Pandas, *Scikit-Mobility* relies largely on single-threaded algorithms and dataframe operations that become a bottleneck when processing datasets containing millions of records. This paper presents *Fastkit-Mobility*[1], a dataframe-agnostic open-source library for scalable mobility analytics. Its main contributions are:

- We provide a drop-in replacement for *Scikit-Mobility* that preserves its API while improving performance and scalability.
- We introduce a Rust-based execution engine with parallelism and zero-copy data access for efficient mobility analytics.
- We enable backend portability across modern dataframe systems via *Narwhals*, allowing execution on diverse pipelines.
- We demonstrate substantial speedups and memory reduction on large-scale mobility datasets, while maintaining functional equivalence with *Scikit-Mobility*.

2. METHODOLOGY

Fastkit-Mobility is inspired by *Scikit-Mobility* [1] and preserves its existing mobility analysis workflows while significantly improving performance. To this end, *Fastkit-Mobility* re-implements the full *Scikit-Mobility* API with compatible function signatures, enabling a seamless transition for existing users and codebases.

Beyond performance improvements, *Fastkit-Mobility* broadens accessibility through a dataframe-agnostic architecture. It achieves this by integrating Narwhals [3], a compatibility layer that enables seamless support across multiple dataframe libraries. This design allows the package to operate on various platforms while fully leveraging the native optimizations of the underlying dataframes.

While maintaining familiar Python interfaces, the library leverages a Rust backend to enable efficient, parallel execution of mobility algorithms. Because most of these algorithms analyze or aggregate data at the individual level, the tasks are inherently parallelizable, yielding substantial performance improvements.

To minimize its memory footprint, the Rust backend accesses dataframes using zero-copy operations. Rather than creating intermediate copies or converting data formats—as occurs in some *Scikit-Mobility* functions—it operates directly on the native memory layout of the underlying dataframe (e.g., NumPy or Arrow). Eliminating the computational overhead associated with data duplication and memory management not only conserves memory but further accelerates execution times.

Most suites within the library (excluding generation) offer two computation paths: *indexed* and *presorted*. The presorted path assumes data is ordered by user unique identifiers (UID) and timestamps, ensuring contiguous memory allocation for individual user records. This enables two key hardware optimizations: enhanced cache locality, which efficiently batches data from RAM to the CPU cache, and Single Instruction, Multiple Data (SIMD) vectorization. By applying operations—such as haversine distance calculations—concurrently across multiple data points, SIMD yields hardware-dependent speedups of 2x to 8x. However, because the presorted path requires a computationally expensive preliminary sorting step, the library also provides an indexed path. This alternative avoids global sorting by extracting and locally sorting row indices per user, computing the data via memory indirection. While presorting is strongly recommended for massive datasets (exceeding one billion rows), the indexed path may offer faster end-to-end execution depending on the broader workflow.

The privacy suite leverages Narwhals to execute native dataframe operations, such as joins, instead of custom single-threaded Python loops, to efficiently quantify privacy risks across various attack scenarios while significantly outperforming the single-threaded custom loops of *Scikit-Mobility*.

For mobility generation models, which simulate synthetic behavior rather than analyzing empirical user data, the same performance optimizations are applied. Additionally, to support large-scale simulations, we address the quadratic memory scaling of Origin-Destination (OD) matrices by implementing a Least-Recently Used (LRU) thread-safe caching mechanism with the moka crate. By storing only a user-configurable maximum of N rows, this strategy tightly constrains the memory footprint to available RAM, regardless of the tessellation grid size.

Beyond re-implementing *Scikit-Mobility*, *Fastkit-Mobility* extends its analytical capabilities in three directions. First, it intro-

[1] https://github.com/gefgu/fastmob

Table 1. Performance Comparison: *Fastkit-Mobility* vs. *Scikit-Mobility* across Evaluation Suites. Results report the Mean ± Standard Deviation across 5 independent runs to capture variance. Speed-ups and memory reductions are formatted as (Pandas / Polars) where applicable. Evaluations were performed on a Brightkite-4M dataset (or a 690-row toy dataset for Privacy) and the real-world GeoLife dataset. Benchmarks span three architectures: Intel Core i7-10700 (8C/16T), Intel Xeon Silver 4316 (20C/40T), and Ryzen 9 9950X3D (16C/32T).

Suite	Dataset	Mode	Speed-up (Mean ± SD) [Pandas / Polars]			Mem. Reduc. (± SD) [Pandas / Polars]
			Core i7	Xeon	Ryzen	
Preprocessing (n=4)	Brightkite-4M	Presorted	540±378× / 723±266×	438±422× / 659±211×	298±216× / 521±92×	85±27% / 35±21%
		Indexed	518±389× / 624±247×	435±421× / 551±115×	199±148× / 346±117×	85±27% / 31±21%
	GeoLife	Presorted	161±12× / 261±54×	165±50× / 306±10×	112±12× / 219±56×	73±09% / 79±07%
		Indexed	150±19× / 248±22×	165±58× / 303±20×	77±11× / 118±48×	73±09% / 79±07%
Measures Ind. (n=16)	Brightkite-4M	Presorted	714±604× / 798±461×	504±307× / 745±456×	220±122× / 299±169×	94±25% / 84±20%
		Indexed	247±150× / 287±208×	143±77× / 173±95×	136±83× / 160±97×	81±28% / 42±31%
	GeoLife	Presorted	141±157× / 190±191×	188±225× / 273±300×	145±191× / 167±199×	51±49% / 48±48%
		Indexed	142±168× / 198±211×	192±237× / 282±312×	85±142× / 92±147×	33±48% / 25±39%
Measures Coll. (n=6)	Brightkite-4M	Presorted	228±247× / 508±693×	150±133× / 636±719×	95±81× / 279±277×	11±59% / 91±13%
		Indexed	268±305× / 654±850×	146±132× / 650±749×	86±71× / 276±261×	11±59% / 91±13%
	GeoLife	Presorted	128±192× / 876±1423×	78±113× / 742±1075×	54±78× / 295±434×	32±37% / 86±17%
		Indexed	128±192× / 876±1423×	78±113× / 742±1075×	54±76× / 311±424×	32±37% / 86±17%
Privacy (n=8)	Toy (690)	Indexed	27833±23012× / 28560±21523×	23413±20441× / 31086±25729×	47404±35970× / 54541±37358×	4±04% / 4±04%
Models (n=5)	10k locations/agents		86±48×	135±72×	99±52×	56±50%

duces motif classification [4] for extracting structural patterns from daily mobility trajectories. To support efficient comparison, motifs containing up to six nodes are mapped to canonical 64-bit identifiers. Second, the library introduces fitting algorithms for fundamental mobility laws, enabling straightforward evaluation of jump lengths and radius of gyration [5], log-normal visit distributions [4], and distance-frequency laws [6]. Finally, it extends the evaluation framework with statistical measures for comparing mobility datasets, including the Wasserstein distance for numerical and ordinal metrics and the Jensen-Shannon divergence for categorical distributions, facilitating rigorous validation of synthetic mobility models against empirical data.

3. LIBRARY EVALUATION

Table 1 details the substantial performance gains achieved by *Fastkit-Mobility* across the evaluated suites. Certain legacy modules, such as *measures-evaluation* and I/O functions, are excluded from this comparison because they primarily wrap external Python libraries rather than implementing custom mobility logic.

The benchmarking methodology was strictly controlled across all suites. For the preprocessing and measures modules, we executed five iterations on a 4-million-record subset of the Brightkite LBSN dataset and Geolife GPS dataset, verifying that all outputs were numerically equivalent within a 10^{-5} tolerance. Brightkite took 3.8 ± 0.3s and GeoLife took 9.2 ± 0.3s for the presorting step. The privacy suite was tested on a 690-row dataset (the original *Scikit-Mobility* toy dataset repeated 30 times). This small size is due to the combinatorial complexity of the problem, with *Scikit-Mobility* taking 390.57 ± 321.18 s per run, making it difficult to run larger scenarios. In comparison, *Fastkit-Mobility* takes around 0.0120 ± 0.0020 s. Generative models were evaluated using a 10,000-location tessellation, scaling to 10,000 agents for agent-based models. Execution speeds were measured using Python's time.perf_counter, explicitly isolating algorithmic runtime by excluding data I/O overhead. Memory profiling was conducted using Memray[7] to track allocations across both the Python and Rust runtimes, with the table reporting the percentage reduction in peak memory consumption. While we report only Pandas and Polars results for space limitations, we also tested with PyArrow, Modin, DuckDB, and Dask dataframes, and the speed-ups are comparable.

Shifting computational bottlenecks to a compiled Rust backend yields performance gains between 2 and 3 orders of magnitude, peaking at a 54,541x improvement for privacy risk operation. This acceleration stems directly from replacing interpreted Python loops with optimized native execution. However, scaling behavior depends tightly on the interplay between hardware architecture and spatiotemporal data characteristics. CPUs with high core counts excel on LBSN datasets featuring numerous users with sparse point distributions, whereas processors with faster single-core speeds favor GPS datasets with fewer users but denser trajectory sequences. Two architectural bounds influence these execution times: the Indexed mode inherently includes index-construction overhead, and user-level parallelization limits scaling on datasets with a constrained user space, such as GeoLife (182 users). Finally, while Scikit-Mobility baseline times inherently fluctuate across hardware, utilizing Polars via Narwhals consistently maximizes speed-ups and minimizes memory consumption by leveraging its Arrow memory model and parallelized dataframe operations.

4. CONCLUSION

By significantly accelerating the computation of mobility metrics, synthetic data generation, and privacy risk evaluations, *Fastkit-Mobility* empowers researchers to analyze massive datasets with vastly reduced execution times. Furthermore, its optimized resource utilization lowers the barrier to entry, enabling advanced mobility analysis even on constrained hardware. Future development will expand the available metrics and law-fitting algorithms, further establishing *Fastkit-Mobility* as a scalable platform for mobility research.

ACKNOWLEDGMENTS

CNPq (processes 314603/2023-9, 441444/2023-7, 444724/2024-9).

REFERENCES

1. L. Pappalardo *et al.*, "scikit-mobility: A python library for the analysis, generation, and risk assessment of mobility data," J. Stat. Softw. .
2. H. Martin *et al.*, "Trackintel: An open-source python library for human mobility analysis," Comput. Environ. Urban Syst. (2023).
3. M. Gorelli and contributors, "Narwhals: Extremely lightweight compatibility layer between dataframe libraries," (2024). Python package.
4. C. M. Schneider, V. Belik, T. Couronné, *et al.*, "Unravelling daily human mobility motifs," J. The Royal Soc. Interface **10**, 20130246 (2013).
5. M. C. González *et al.*, "Understanding individual human mobility patterns," Nature (2008).
6. Schläpfer *et al.*, "The universal visitation law of human mobility," Nature (2021).
7. Bloomberg L.P., "Memray," (2022). GitHub repository.

Correcting socioeconomic bias in mobile phone mobility estimates

LEO FERRES [1,3] AND LAETITIA GAUVIN[2,3,*]

[1] Institute of Data Science, Faculty of Engineering, Universidad del Desarrollo, Santiago, Chile
[2] IRD, UMR 215 Prodig, 5, course of Humanities, F-93 322, Aubervilliers Cedex, France
[3] ISI Foundation, Turin, Italy
[*] laetitia.gauvin@ird.fr

1. INTRODUCTION

Mobile phone data have transformed the study of human mobility providing unprecedented spatial and temporal resolution [1, 2]. These data have supported applications ranging from urban planning [3] to epidemic modelling [4–6] and poverty estimation [7].

However, mobile phone data are not necessarily representative of the general population. Early studies primarily attributed this bias to socioeconomic disparities in mobile phone ownership in low-penetration settings [8, 9], showing that these biases can propagate into downstream analyses such as epidemic models [10, 11]. Even though representativeness bias has been extensively documented across mobile phone datasets [12–14], comparatively little attention has been devoted to developing statistical methods that explicitly correct for these biases. Most existing approaches typically rely on simple rescaling or reweighting to improve representativeness [15, 16].

More sophisticated correction methods have emerged in related contexts. Difference-in-differences frameworks have been used to estimate sampling bias and infer relative population changes without requiring ground-truth data [17], and methods for correcting temporal bias have been proposed for GPS-based mobility data [18]. However, these approaches focus primarily on temporal dynamics.

In other fields, Multilevel Regression and Post-stratification (MRP) has become a standard approach for correcting sampling bias in non-representative data [19]. Post-stratification alone has also been applied to indicators derived from mobile phone surveys [20–22] but these studies do not incorporate the multilevel regression component that allows information to be shared across demographic and geographic groups. Yet, post-stratification alone cannot reliably estimate outcomes for sparsely sampled or unobserved areas, nor can it disentangle the respective contributions of socioeconomic and geographic factors to representativeness bias.

n this study, we apply multilevel regression and post-stratification (MRP) to correct for representativeness bias in call detail record (CDR)-based mobility estimates derived from mobile phone data in Chile. In a first specification, we exploit the socioeconomic information available in the CDR data; we then demonstrate a geography-only variant that can be used when individual socioeconomic data are not available. Applying these approaches to the radius of gyration, we show that MRP yields substantially different estimates compared with naïve averages computed directly from CDR data.

2. DATA

We use anonymized Call Detail Records (CDRs) from Telefónica Chile (Movistar), one of the country's three largest mobile network operators, covering May-July 2016. For each call, the data include the serving cell tower, timestamp, call metadata (incoming/outgoing and duration), user gender, and a socioeconomic group (GSE) assigned by the operator. We restrict the analysis to Chilean nationals with a valid GSE classification and at least one call per day on average, following standard filtering procedures to obtain reliable mobility estimates [23].

Home locations are inferred as the most frequently used antenna between 20:00 and 06:00, following established practice [24]. Individual mobility is summarized by the radius of gyration (ROG) [1]. We compute monthly ROG separately for weekdays and weekends and model its logarithm to account for the right-skewed distribution of mobility distances.

Socioeconomic information is drawn from two sources. Individual-level GSE labels are provided together with the CDR data. Population-level GSE distributions, used for bias correction, are derived from the ISMT index (Indice Socio Material Territorial),[25], an indicator developed by the Observatorio de Ciudades to assess socioeconomic conditions and quality of life in both urban and rural areas of Chile, based on census data. The geographic unit of analysis is the distrito censal (329 census districts), each nested within one of the 47 comunas of the Santiago Metropolitan Region.

3. METHODS AND RESULTS

To correct for representativeness bias in CDR-derived mobility estimates, we apply multilevel regression and post-stratification (MRP). MRP first models mobility as a function of demographic and geographic characteristics and then reweights model predictions to match the demographic composition of the target population. The analysis is performed on cell-level summaries defined by the cross-classification of distrito censal, socioeconomic group (GSE), gender, month, and weekday/weekend. For each cell, we compute the mean log radius of gyration (ROG), which serves as the response variable. Let y_{ijkmt} denote the mean log ROG for distrito i, GSE group j, gender k, month m, and weekday indicator t. We fit the Bayesian multilevel model

$$\begin{aligned} y_{ijkmt} = \mu + \beta_j^{\text{GSE}} + \beta_k^{\text{gender}} + \beta_{jk}^{\text{GSE}\times\text{gender}} \\ + \beta_t^{\text{weekday}} + \beta_m^{\text{month}} + \alpha_{c(i)}^{\text{comuna}} + \varepsilon_{ijkmt} \end{aligned} \quad (1)$$

where fixed effects account for socioeconomic status, gender, their interaction, and temporal variation, while comuna-level random intercepts capture unobserved geographic heterogeneity. The residual variance is scaled by the number of users in each cell so that more precisely estimated cell means receive greater weight. Models are fitted in Bambi [26] using the PyMC backend [27] with NumPyro (four chains, 2,000 post-warmup draws). The post-stratification frame consists of all combinations of distrito, GSE, and gender (3,282 cells, representing 6.2 million individuals). Cell populations are estimated as $N_j = N_{\text{distrito}} P(\text{GSE} \mid \text{distrito}) P(\text{gender} \mid \text{comuna})$, where district populations come from the 2017 census and GSE distributions are derived from the ISMT index. Because GSE is defined at the area level, GSE and gender are conditionally independent

by construction. For each posterior draw, corrected estimates of the radius of gyration for any target population are obtained by averaging the predicted values using the census cell counts

$$\frac{\sum_{j\in D} N_j \exp(\tilde{y}_j^{(s)})}{\sum_{j\in D} N_j}, \tag{2}$$

with uncertainty propagated across posterior draws. Although the framework above uses individual-level SES labels available in the CDR data, many CDR datasets do not contain this information. To evaluate the applicability of MRP in such settings, we also fit a reduced geography-only MRP model that excludes individual SES while retaining the geographic hierarchy and post-stratification. In case the socioeconomic composition is spatially structured, geographic information acts as a proxy for socioeconomic composition, allowing MRP to recover part of the representativeness bias even without individual SES labels. To test the robustness of the estimates, we used cross-validation. Figure 1 shows average radius of gyration estimates by socioeconomic group before correction, after full MRP correction using individual SES information, and after the geography-only MRP correction.

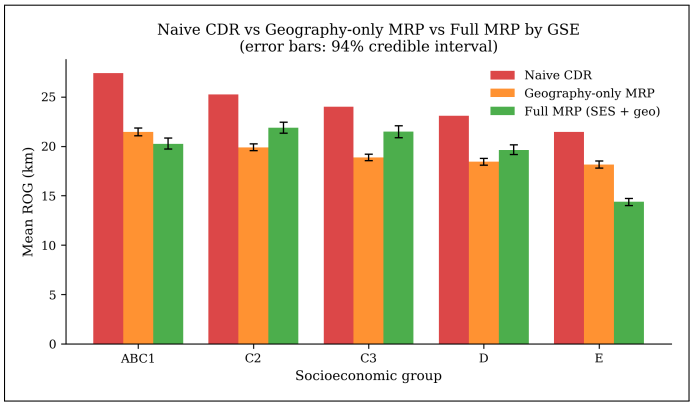


Fig. 1. Radius of gyration by GSE group for three estimators: naive CDR average (red), geography-only MRP.

MRP reduces the naïve CDR estimate of average radius of gyration by approximately 17%. While the full model yields the largest correction, the geography-only variant captures a substantial share of the bias, indicating that accurate correction is possible even when individual socioeconomic information is unavailable, as long as socioeconomic segregation is reflected in geographic patterns. These results demonstrate that MRP offers a practical and principled framework for correcting representativeness bias in CDR-derived mobility estimates, making the approach applicable to the many mobile phone datasets that lack user-level socioeconomic information.

Data Availability Statement The raw data analyzed in this study cannot be made publicly available for privacy reasons.

REFERENCES

1. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779 (2008).
2. H. Barbosa, M. Barthelemy, G. Ghoshal, *et al.*, "Human mobility: models and applications," Phys. Reports **734**, 1 (2018).
3. M. Lenormand, T. Louail, O. G. Cantú-Ros, *et al.*, "Influence of sociodemographics on human mobility," Sci. Reports **5**, 10075 (2015).
4. A. Wesolowski, N. Eagle, A. J. Tatem, *et al.*, "Quantifying the impact of human mobility on malaria," Science **338**, 267 (2012).
5. M. Tizzoni, P. Bajardi, A. Decuyper, *et al.*, "On the use of human mobility proxies for modeling epidemics," PLoS Comput. Biol. **10**, e1003716 (2014).
6. N. Gozzi, M. Tizzoni, M. Chinazzi, *et al.*, "Estimating the effect of social inequalities on the mitigation of covid-19 across communities in santiago de chile," Nat. Commun. **12** (2021).
7. J. Blumenstock, G. Cadamuro, and R. On, "Predicting poverty and wealth from mobile phone metadata," Science **350**, 1073 (2015).
8. J. Blumenstock and N. Eagle, "Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda," in *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, (2010).
9. A. Wesolowski, N. Eagle, A. M. Noor, *et al.*, "The impact of biases in mobile phone ownership on estimates of human mobility," J. Royal Soc. Interface **10**, 20120986 (2013).
10. T. Chin, B. G. Cordova, E. Laber, *et al.*, "Bias in mobility datasets drives divergence in modeled outbreak dynamics," Commun. Med. **5**, 45 (2025).
11. R. Gallotti, D. Maniscalco, M. Barthelemy, and M. De Domenico, "Distorted insights from human mobility data," Commun. Phys. **7** (2024).
12. Z. Zhao, S.-L. Shaw, Y. Xu, *et al.*, "Understanding the bias of call detail records in human mobility research," Int. J. Geogr. Inf. Sci. **30**, 1738 (2016). _eprint: https://doi.org/10.1080/13658816.2015.1137298.
13. C. Liu, P. Holme, S. Lehmann, *et al.*, "Nonrepresentativeness of human mobility data and its impact on modeling dynamics of the covid-19 pandemic: Systematic evaluation," JMIR Form. Res. **8**, e55013 (2024).
14. C. Cabrera and F. Rowe, "A systematic machine learning approach to measure and assess biases in mobile phone population data," (2025). ArXiv:2509.02603 [stat].
15. A. Wesolowski, N. Eagle, A. M. Noor, *et al.*, "The impact of biases in mobile phone ownership on estimates of human mobility," J. The Royal Soc. Interface **10**, 20120986 (2013).
16. J. Meppelink, J. Van Langen, A. Siebes, and M. Spruit, "Beware Thy Bias: Scaling Mobile Phone Data to Measure Traffic Intensities," Sustainability **12**, 3631 (2020).
17. G. Pestre, E. Letouzé, and E. Zagheni, "The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates," The World Bank Econ. Rev. **34**, S89 (2020).
18. S. A. Sanchez, H. Gibbs, T. Yabe, *et al.*, "Correcting temporal bias in mobility data using time-use surveys," (2026).
19. W. Wang, D. Rothschild, S. Goel, and A. Gelman, "Forecasting elections with non-representative polls," Int. J. Forecast. **31**, 980 (2015).
20. D. A. Sánchez-Páez, B. Masquelier, A. Menashe-Oren, *et al.*, "Measuring under-5 mortality and fertility through mobile phone surveys: an assessment of selection bias in 34 low-income and middle-income countries," BMJ Open **13**, e071791 (2023).
21. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large US cities," Nat. Commun. **12** (2021).
22. T. Yabe, B. García Bulle Bueno, X. Dong, *et al.*, "Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters," Nat. Commun. **14** (2023).
23. L. Pappalardo, F. Simini, S. Rinzivillo, *et al.*, "Returners and explorers dichotomy in human mobility," Nat. Commun. **6**, 8166 (2015).
24. M. Vanhoof, F. Reis, T. Ploetz, and Z. Smoreda, "Assessing the quality of home detection from mobile phone data for official statistics," J. Off. Stat. **34**, 935 (2018).
25. M. R. Araya, *ismtchile: Calculating Socio Material Territorial Index* (2023). R package version 2.1.5. Proyecto Índice Socio Material Territorial. Proyecto Fondecyt DIMES, Centro de Investigación Urbano-Territorial, Observatorio de Ciudades, Universidad Católica de Chile.
26. T. Capretto, C. Piho, R. Kumar, *et al.*, "Bambi: a simple interface for fitting Bayesian linear models in Python," J. Stat. Softw. **103**, 1 (2022).
27. J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," PeerJ Comput. Sci. **2**, e55 (2016).

Selection Bias in Phone-Based Migration Statistics: Evidence and Correction Strategies

PAUL BLANCHARD,[1,*] SVETA MILUSHEVA,[1] FRANCISCO ROWE[2] ,STEFANIA RUBRICHI[3]

[1]Development Impact Group, World Bank, Washington, D.C., USA; [2]University of Liverpool, Liverpool, UK; [3]Orange Labs, Châtillon, France
* pblanchard@worldbank.org

1. INTRODUCTION

MNO data create opportunities to produce granular migration statistics, particularly in developing countries where traditional data sources are scarcer [1, 2]. However, those are often met with caution: the non-random nature of phone ownership raises concerns about selection bias. These biases challenge the statistical validity of phone-based migration indicators. While previous studies have proposed ad hoc correction strategies [3, 4], the application of formal statistical frameworks for inference from non-probability samples remains limited.

The objective of this study is twofold. First, it documents the magnitude of selection bias in migration statistics inferred from phone users across 29 countries in Sub-Saharan Africa using nationally representative Demographic and Health Surveys (DHS). Second, it provides a formal quasi-randomization framework as a statistical foundation for understanding and evaluating commonly used post-stratification corrections. While geographic post-stratification is intuitive and widely used to scale phone-based statistics to the target population, the framework makes explicit the selection mechanism and the conditional independence assumption under which the resulting estimator is unbiased. This provides a principled basis for assessing when geographic adjustment can remove selection bias and for interpreting the discrepancies that remain when the available stratification variables do not fully capture the selection mechanism. This method is applied to adjust migration statistics derived from samples of phone users, testing alternative stratification designs. The approach is first tested using phone user subsamples from DHS, where the availability of full population data enables a direct assessment of bias reduction. It is then applied to migration estimates derived from actual MNO data, allowing us to examine how different stratification strategies affect the resulting estimates.

2. STATISTICAL FRAMEWORK

Migration estimates based on phone users rely on a non-probability sample and are therefore potentially biased. In the quasi-randomization framework, the observed population of users is treated as a non-probability sample generated by an implicit selection mechanism. Inclusion probabilities are modeled as a function $\pi(X_i)$ of observable characteristics X_i for any individual i in the target population. Under the conditional independence assumption, i.e. when selection is independent of migration outcomes given X_i, the following estimator is unbiased:

$$\hat{Y}^{P,qr} = \sum_{i \in S^P} w_i \, y_i \qquad (1)$$

Where y_i is a migration dummy, $w_i = {}^{1}/_{\pi(X_i)}$, and S^P a sample of phone users. In practice, with MNO data, X_i is limited to the home location that is typically calculated as the most frequent location during night hours, so the correction method reduces to post-stratification. Dividing the population into strata indexed by k, with known population sizes N_k and sample counts n_k, the selection function is defined as $\pi_i = n_k/N_k$ for any individual i belonging to stratum k. This yields a well-known Horvitz-Thompson reweighted estimator equivalent to applying stratum-level scaling factors $\delta_k = N_k/n_k$, which is typically used to approximate population-level statistics. Importantly, the validity of this estimator rests on the conditional independence assumption. In practice, this assumption is unlikely to hold exactly when home location is the only observable characteristic: phone ownership and migration may also vary with other unobserved individual characteristics. Geographic post-stratification can reduce selection bias when geographic strata capture systematic differences between phone users and the target population, but does not guarantee eliminating it. The proposed framework makes this assumption explicit, while the DHS empirical exercise quantifies the remaining bias.

3. DATA AND METHODS

The analysis combines two complementary empirical components. The first relies on DHS conducted in 29 Sub-Saharan African countries. These nationally representative surveys provide both migration outcomes and individual-level phone ownership, allowing us to treat phone owners as simulated phone-user samples while retaining information on the full target population. We first compare migration estimates among phone users with those from the full population to quantify selection bias and then apply alternative geographic post-stratification schemes to the phone-user samples and compare the adjusted estimates with the known population values. Formally, the selection bias is defined as:

$$bias = \mathbb{E}[\hat{Y}^P - \hat{Y}] \qquad (2)$$

Where $\hat{Y}^P$ is a migration estimator derived from a sample of phone users and $\hat{Y}$ is the same migration estimator derived from the full sample. Repeated resampling allows us to estimate both the magnitude and the distribution of the bias, thereby enabling a rigorous assessment of its statistical significance.

Then, the quasi-randomization approach is applied on each individual country using a post-stratification on three levels: region, zone (rural vs. urban), and region×zone. For each stratification design, the adjusted estimates are compared with the true values (i.e. migration estimates derived from the full sample) to assess bias reduction.

The second component applies the quasi-randomization framework to CDR-based migration statistics in Senegal. Given the granularity of the estimates, the analysis tests a wider array of stratification schemes: region, department,

district, district×zone, Voronoi. Because true population migration rates are not directly observed in MNO data, this analysis cannot quantify bias reduction directly. However, it allows to examine the behavior of adjusted estimates with finer stratification levels and highlights a bias-variance tradeoff.

4. RESULTS

Survey-based descriptive results show widespread but uneven mobile phone diffusion across the region: in 75% of sampled countries, at least half of the adult population owns a phone. Nonetheless, substantial disparities persist both across and within countries, driven primarily by rural-urban differences and gender gaps. Internal migration rates also vary widely, ranging from 2.2% in Angola to 17.6% in Uganda. Comparisons between phone users and non-users reveal a systematic, positive but moderate selection bias, with phone users consistently exhibiting higher migration rates.

The DHS analysis across the 29 countries reveals that the bias is mostly positive and statistically significant: phone users exhibit a systematically higher migration propensity compared to the full population. At the national-level, the size of the bias remains modest, from 0 p.p. to 2.7 p.p. (Figure 1), even though this represents up to 38% in relative terms (Madagascar). On average, post-stratification adjustments consistently reduce bias in migration estimates, with the largest improvements observed under finer geographic stratification, at the region×zone-level (Figure 2).

Quasi-randomization experiments using MNO data in Senegal yield two main insights. First, stratification design strongly influences migration estimates, with meaningful differences observed across stratification levels. Second, finer stratification increases statistical uncertainty, highlighting the bias-variance tradeoff: while finer stratification levels mechanically enrich the selection model and decrease bias, they also increase local variance through the emergence of strata with high population-to-user ratios.

5. CONCLUSION

This study combines large-scale evidence on selection bias across 29 Sub-Saharan African countries with a formal conceptual framework for inference from non-probability samples. By making explicit the selection mechanism and assumptions underlying geographic post-stratification, the framework provides a theoretically grounded basis for assessing when such corrections can reduce bias and for interpreting remaining discrepancies. By documenting the structure and magnitude of migration biases across 29 Sub-Saharan African countries, it provides one of the most comprehensive assessments to date of the reliability of phone-based migration measurement. On the other hand, the Senegal MNO application illustrates the practical trade-off between finer geographic adjustment and statistical uncertainty. These findings provide guidance for integrating MNO data into migration measurement and developing bias-aware population statistics from digital trace data.

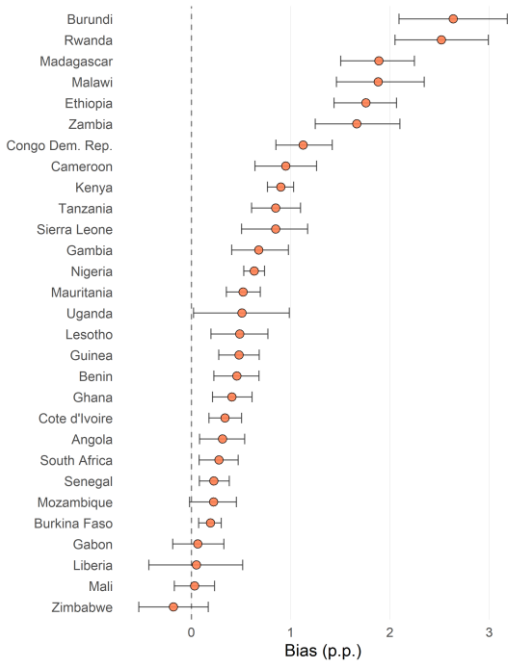


Fig. 1. Migration bias estimations based on DHS data.

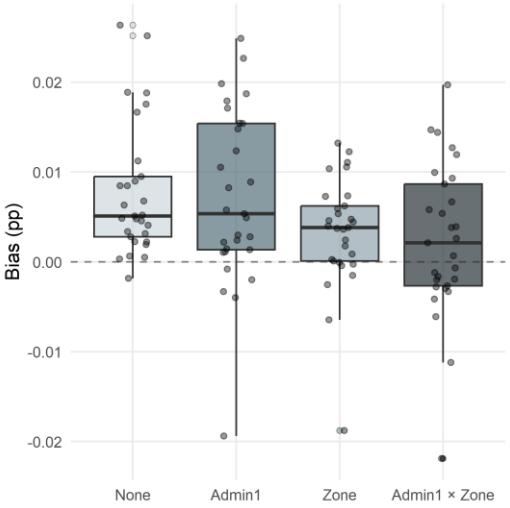


Fig. 2. Migration bias for adjusted estimates across 29 countries, with various levels of stratification .

Note: Each dot represents a country; boxplots summarize the distribution across countries.

References

1. Blumenstock, J. E. (2012): "Inferring Patterns of Internal Migration from Mobile Phone Call Records : Evidence from Rwanda," Information Technology for Development, 18, 107–125.
2. Blanchard, P. and S. Rubrichi (2025): "A highly granular temporary migration dataset derived from mobile phone data in Senegal," Scientific Data, 12, 1051.
3. Mazzoli, M., B. Diechtiareff, A. Tugores, W. Wives, N. Adler, P. Colet, and J. J. Ramasco (2020): "Migrant mobility flows characterized with digital data," PLOS ONE, 15, e0230264.
4. Meppelink, J., J. Van Langen, A. Siebes, and M. Spruit (2020): "Beware Thy Bias: Scaling Mobile Phone Data to Measure Traffic Intensities," Sustainability, 12, 3631.

Identifying and characterizing dangerous road sections using vehicle telemetry and OpenStreetMap data

Giovana Piorino[1], Andrea Lanzillotta[2], Luca Vassio[2,*], Danilo Giordano[2], Ana Paula Couto da Silva[1], Marco Mellia[2], Claudio De Tommasi[3], and Marco Di Cristina[3]

[1] Computer Science Department, Universidade Federal de Minas Gerais, Brazil
[2] Politecnico di Torino, Italy
[3] UnipolTech, Italy
[*] luca.vassio@polito.it

1. INTRODUCTION

Identifying the areas of a city that pose potential risks to drivers is important for improving road safety and supporting urban planning [1]. By detecting locations, i.e., **dangerous road sections**, where drivers repeatedly experience abnormal vehicle dynamics, i.e, **dangerous events**, it becomes possible to identify areas affected by infrastructure deficiencies, such as potholes and pavement irregularities, as well as road design and traffic control elements, including speed bumps, intersections, and signage. These insights can support data-driven decisions to improve road safety, optimize infrastructure investments, and assess the impact of interventions.

We propose the framework illustrated in Figure 1 to identify potentially dangerous road sections using car telemetry provided by UnipolTech[1] and OpenStreetMap (OSM) data available in [2]. Briefly, vehicle telemetry time-series measurements related to dangerous events are extracted and preprocessed. These events are then aggregated to identify dangerous road sections. OSM data support this identification and characterization by providing road network information and contextual attributes, such as road type and speed limits. The resulting dangerous road sections are ranked according to their relative frequency, visualized on the road network, and further analyzed through temporal comparisons and clustering techniques.

2. DANGEROUS EVENTS FROM VEHICLE TELEMETRY

We validate our methodology with telematics data from private insurance users. Telematics boxes are permanently installed in each car and collect data to provide insurance services. Each telematics box includes a GNSS sensor and an accelerometer. When the accelerometer reading exceeds a predefined threshold (1.5g in our case) on any axis, a dangerous event is triggered. In this case, the telematics box logs positions and accelerations over an interval spanning 30 seconds before and after the triggering instant, which we call the climax, allowing the reconstruction of the trajectory and the behavior close to the event. Real crashes that led to an insurance claim are removed from the dataset to reduce de-anonymization risks. Moreover, the company has access to anonymized sample positions from vehicles during normal conditions. As a case study, we focus on the events triggered by cars registered in the city of Lucca, Italy. The dataset covers the period from October 2025 to March 2026 for 8,455 vehicles.

3. EXTRACTING DANGEROUS TRAJECTORIES

For each dangerous event, we consider a temporal window of ± 4 seconds around the climax. We clean the data by excluding trajectories with fewer than 9 data points and low-quality signals. The trajectories are further enhanced by improving the GNSS positions through Kalman filters applied to accelerometer data. By continuously fusing the absolute, but potentially noisy, spatial coordinates from the GNSS with the high-frequency inertial dynamics captured by the accelerometers, the filter recursively minimizes the overall positional uncertainty [3].

Finally, we constrain the trajectories to real roads obtained from OSM. The road network of the study area is retrieved as a directed graph from OSM. We then apply Mappymatch[2], a Hidden Markov Model-based algorithm, to match raw GNSS trajectories to this road graph. This preserves the sequential coherence of the trajectories and enables the extraction of road-level attributes related to the graph edges and their OSM metadata. We obtain 5,829 distinct dangerous trajectories for analysis, each providing a unique climax point. Notice that the same climax position can be shared by multiple trajectories.

4. DANGEROUS ROAD SECTIONS RANK, EVOLUTION, AND VISUALIZATION

After extracting dangerous trajectories, we rank dangerous road sections based on their climax positions. A simple count of dangerous events per section would be biased towards roads with the highest traffic flows. Therefore, we use telematics data from sampled vehicle positions during normal operations to estimate traffic flows on each road. We assign each vehicle position to the closest 30-meter road section using the OSM road network. We estimate the relative frequency of dangerous locations in each section by dividing the number of climax points by the estimated flow in the same section. We consider only sections with enough passages to properly estimate flows. We further normalize this value by the maximum observed level in the whole study area, obtaining a danger level between 0 and 1. Table 1 shows the top-5 road sections by relative danger level for one week in 2025.

By comparing danger levels in 2025 and 2026, we identify the locations whose danger level changes the most (increased or decreased). This is important for studying cases in which municipalities modify traffic circulation. To compare the two periods fairly, we normalize them by the same maximum observed value so the two years share one scale. We then rank sections by the absolute difference in danger levels. Table 2 compares the top-5 zones with the largest changes in danger level between 2025 and 2026. A positive Δ means the zone became less dangerous in 2026. Applying a Poisson-based margin to isolate significant temporal variations, 75.9% of locations remained stable, 3.4% increased in danger, and 20.7% decreased significantly. To mitigate outlier effects, the analysis in the tables is restricted to locations with at least 10 occurrences for both dangerous events and sampled passages.

To visualize dangerous locations according to their danger

[1] https://www.unipoltech.com/site/en/who-we-are/

[2] https://mappymatch.readthedocs.io/en/latest/

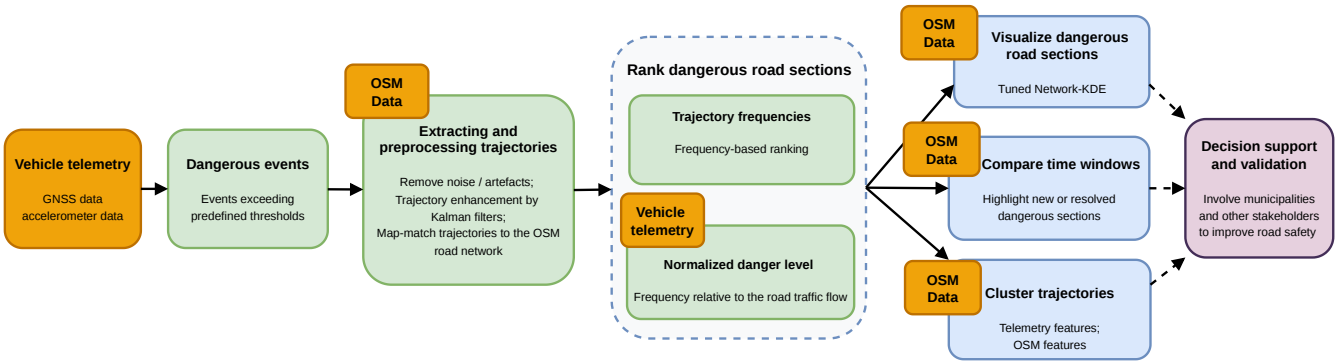


Fig. 1. Overview of the proposed pipeline for dangerous road section analysis from vehicle trajectories and OSM data.

Table 1. Top-5 dangerous sections (2025).

Position (lat, lon)	Road	Normalized danger level
(43.8644, 10.5247)	Via dell'Acqua calda	1.0000
(43.8198, 10.5229)	Via dei Boschi	0.8548
(43.8638, 10.4874)	Via della Pieve di S. Stefano	0.6452
(43.8594, 10.5415)	Via dei Capannoresi	0.6415
(43.8574, 10.5395)	Via delle Pierone	0.6273

Table 2. Top-5 changes in danger level between 2025 and 2026.

Position (lat, lon)	Road	Δ normalized danger level
(43.8322, 10.5054)	Via Bonagiunta Urbiciani	-0.356
(43.8442, 10.4476)	Via della Polveriera	+0.256
(43.8378, 10.5444)	Via del Marginone	-0.209
(43.8571, 10.5604)	Via dell'ave Maria	-0.187
(43.8574, 10.5395)	Via delle Pierone	+0.178

level, we show them on an OSM road network through network-KDE [4] using optimized sparse matrices. The continuous road network is discretized into linear segments called lixels, while the bandwidth defines the spatial radius of influence for each event. After parameter tuning, a robust combination of 10 m lixels and a 30 m bandwidth with a quartic discontinuous kernel was selected. Figure 2 shows the 3-D distribution of danger levels over the map of Lucca obtained with network-KDE.

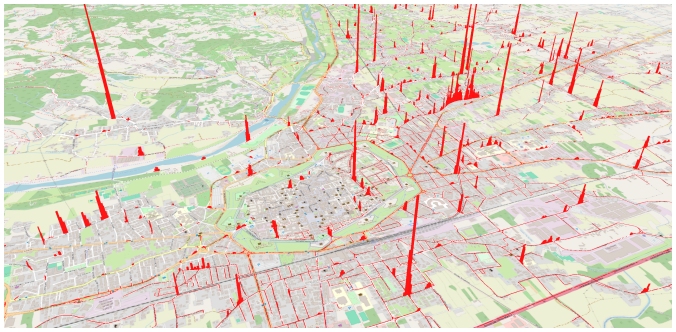


Fig. 2. 3-D OSM network-KDE visualization of danger levels obtained from the climax positions of dangerous trajectories.

5. CLUSTERING AND CHARACTERIZATION

We further analyze the characteristics of dangerous trajectories and road sections. Our goal is to discover recurring patterns in how dangerous events unfold, considering kinematic and route-structure aspects, so that route-level risk patterns can be identified and communicated to municipalities. We therefore cluster the trajectories based on telemetry-related features, such that trajectories within a group share a common risk profile. Following established driving-behavior literature [5], we summarize each trajectory using statistical descriptors (minimum, maximum, mean, and standard deviation) for speed, acceleration, and jerk.

We perform PCA on these features and then k-means clustering. Here we report the results by using k equal to 2, which led to the best silhouette values. We use OSM data to further characterize the clusters (e.g., type of road, speed limit, tortuosity). Figure 3 shows the different characteristics of the speed profiles during the dangerous events for the two identified clusters. The two clusters show a coherent contrast between a slower, more dynamically and geometrically complex profile (Cluster 0) and a faster, smoother, and more direct profile (Cluster 1).

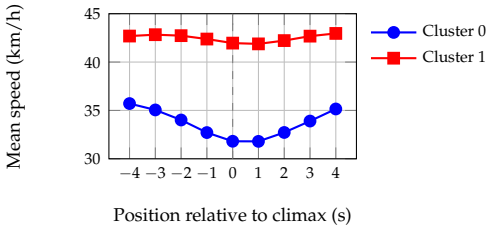


Fig. 3. Mean speed profile per cluster across the event window.

Overall, the proposed framework turns large-scale vehicle telemetry into interpretable road-section risk indicators that can support targeted inspection, maintenance, and traffic-management decisions.

Acknowledgment. We thank UnipolTech (Unipol Group) for providing access to anonymized and aggregated vehicle telematics data.

Disclosures. The authors declare no conflicts of interest.

Data availability. Map data are copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org. Telematics data are not publicly available and are the property of UnipolTech.

REFERENCES

1. World Health Organization, *Global Status Report on Road Safety 2023*.
2. OpenStreetMap contributors, https://www.openstreetmap.org (2017).
3. M. N. Cahyadi, T. Asfihani, H. F. Suhandri, and R. Erfianti, "Unscented kalman filter for a low-cost gnss/imu-based mobile mapping application under demanding conditions," Géod. Geodyn. **15**, 166 (2024).
4. Z. Xie and J. Yan, "Kernel density estimation of traffic accidents in a network space," Comput. environment urban systems **32**, 396 (2008).
5. A. K. Kushwaha, H. Kumar, H. Arora, and B. A. Kumar, "A deep embedded clustering method for location-specific driving safety profiling using trajectory data," Sci. Reports **15**, 41635 (2025).

People-Place Networks Reveal Economic Disparities in United States

BAOSHEN GUO[1], YUZHOU CHEN[2], AND SHENHAO WANG[2,*]

[1] Singapore-MIT Alliance for Research and Technology
[2] Department of Urban and Regional Planning, University of Florida
[*] shenhaowang@ufl.edu

1. INTRODUCTION

A decade of research has shown that human mobility encodes economic outcomes: network diversity predicts development [1], phone metadata predicts poverty [2], and mobility patterns reveal income segregation beyond residential geography [3]. This evidence increasingly points to higher-order people-place interactions—a person's economic exposure depends not only on the places they visit, but on who else visits those places and where those people go in turn. Yet prior work typically reduces mobility to a single-mode projection, a people–people or a place-place network [4], or to aggregated mobility indicators [5], and studies these views in isolation, leaving the joint people–place interaction structure, and how economic information is distributed across it, largely unexplored.

Here we introduce **PP-Net**, a large people–place network connecting individual persons to the places they visit across major US cities [5], together with a people-place graph neural network (GNN) algorithm to characterize the interactions among people and places. Through stacked message passing, PP-Net captures direct people-place visits at the first hop, recovers the people–people co-location and place–place dependency networks at the second layer, and encodes higher-order mobility-community structure at deeper layers. We find that the resulting embeddings strongly predict multiple socioeconomic indicators—educational attainment, income, and poverty. The same embeddings identify urban demographic geography and directly predict each individual's experienced income segregation, an individual-level resolution beyond prior place-based measures, demonstrating that PP-Net can provide an interpretable and scalable instrument for understanding economic disparities in cities.

2. METHOD

A. The people-place mobility network

We represent a city's People-Place mobility network as a weighted bipartite graph $\mathcal{G} = (\mathcal{P} \cup \mathcal{L}, \mathcal{E})$, in which one vertex set $\mathcal{P}$ indexes anonymized *people* and the second set $\mathcal{L}$ indexes *places* (points of interest). An edge $(u, l) \in \mathcal{E}$ is drawn whenever person u is observed to visit place l in the Veraset GPS record [6], and carries an integer visit count c_{ul}. Every person carries demographics-free mobility statistic features, comprising the visit count, the number of distinct places visited, the entropy of the visited category , and the radius of the visited locations. Each place carries a descriptor x_l that combines its location with the semantics of the POI category, name, and related address information. We build PP-Net using 4 months of data (January to April 2025) for Boston (5.1M persons, 123K places, 4,148 CBGs, 105M visits), Chicago (11.9M, 207K, 6,994, 176M)

B. Bipartite People-Place Message Passing and Representation

Bipartite People-Place Message Passing: We update the people and place embeddings through bipartite message passing that alternates people-to-place and place-to-people aggregation:

$$h_u^{(\ell+1)} = \sum_{l:(u,l)\in\mathcal{E}} \frac{c_{ul}}{\sqrt{d_u}\sqrt{d_l}} h_l^{(\ell)}, \quad (1)$$

$$h_l^{(\ell+1)} = \sum_{u:(u,l)\in\mathcal{E}} \frac{c_{ul}}{\sqrt{d_u}\sqrt{d_l}} h_u^{(\ell)}, \quad (2)$$

, where d_u and d_l are the weighted degrees of u and l, and $h_u^{(0)}, h_l^{(0)} \in \mathbb{R}^d$ are linear projections features. Successive depths therefore capture progressively higher-order mobility structure on each side of the graph (Table 1).

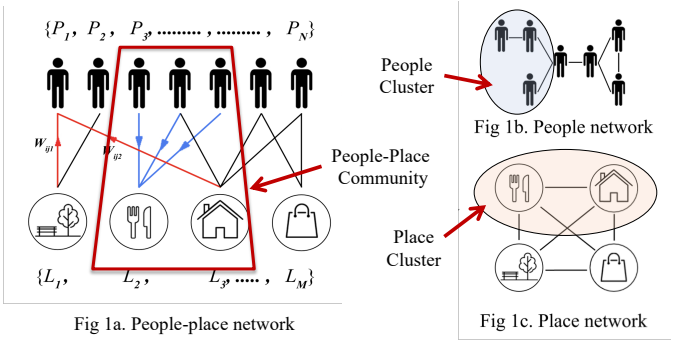


Fig 1a. People-place network

Fig. 1. Diagram of People-Place Network; Fig 1a. Visualization of the People-Place bipartite network. The red and blue arrows represent the place-to-people and people-to-place message passing. Fig 1b, c. Projection of the People-Place network to people and place networks. The concept of community is also operationalized as people-place, people, and place clusters.

Table 1. Mobility structure captured at each message-passing depth of the people–place graph. P: people; L: place.

Depth	People side h_u	Place side h_l
1	P←L directly visited places	L←P visitor composition
2	P←L←P shared-place encounters	L←P←L co-visited places
3	P←L←P←L mobility communities	L←P←L←P functional place clusters

For the final people and place latents, to capture the information of all layers and avoid over-smoothing, we take the layer-combination readout of LightGCN [7], a convex weighting of all intermediate representation and have $h_u = \sum_{\ell=0}^{L} \alpha_\ell h_u^{(\ell)}$ and $h_l = \sum_{\ell=0}^{L} \alpha_\ell h_l^{(\ell)}$.

Self-Supervised People-place Network representation We train the model to obtain the node representation in a fully self-supervised manner. A link-reconstruction objective fits observed visits, jointly capturing the fine-grained people–place interaction

behavior and the structure of the bipartite network. We also add two structural contrastive objectives:

(i) Place-side contrast: Two places are behaviorally similar if the same people frequent them. We therefore form positive pairs (i, j) of places co-visited by a shared person (excluding hub places) and, with an InfoNCE loss, pull their embeddings together against n sampled negatives.

(ii) People-side contrast: Symmetrically, two people may *encounter* one another through a shared place. Following the mobility-encounter view of urban mixing, we form positive pairs (u, v) of people who visit a common place, and contrast them against sampled negatives.

CBG representation from different view of People-place network: With the network trained, we freeze the node embeddings and build census-block-group (CBG) representations by mobility-based mean pooling from two views: (1) a *residential* (people-side) profile z_b^{ppl}, obtained by mean pooling the embeddings of all residents of b; (2) an *ambient* (place-side) profile z_b^{plc}, obtained by mean pooling all places located in b; and their *fused* concatenation $z_b^{\text{fuse}} = [z_b^{\text{ppl}} \| z_b^{\text{plc}}] \in \mathbb{R}^{2d}$. These embeddings are used for downstream socioeconomic prediction tasks.

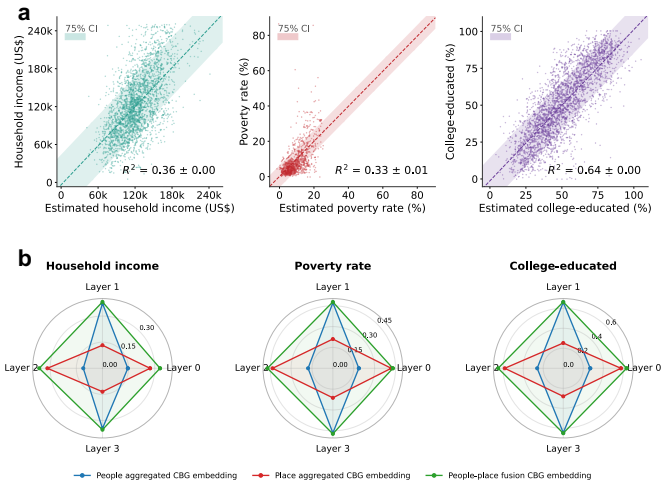


Fig. 2. Fig 2a shows estimated versus observed values for three socioeconomic indicators in Boston, from a linear ridge regression on the frozen CBG embedding. Fig 2b resolves the socioeconomic prediction (R^2) by message-passing depth: each radar panel covers one indicator, with the four axes as layers 0-3 and the three curves as the people-aggregated, place-aggregated, and fused CBG embeddings.

3. RESULTS

A. Predicting Economic Status via People-Place Latents

Fig. 2a shows that a linear probe on the frozen CBG embeddings can predict multiple socioeconomic indicators, recovering the share of college-educated residents best ($R^2 = 0.64 \pm 0.00$), followed by median household income (0.36 ± 0.00) and poverty rate (0.33 ± 0.01).

The per-layer decomposition (Fig. 2b) reveals where this signal resides: the people- and place-side embeddings peak at complementary depths. People-side representations are most predictive after an *odd* number of hops—once a person's embedding has absorbed the peers they encounter and the communities they belong to—while place-side representations peak after *even* hops, once a venue has taken on the profile of the places it is co-visited with. Because the two sides are strongest at different depths, their fusion z^{fuse} matches or exceeds both at every layer, tracing the outer envelope of the radar and confirming that people and place offer non-redundant views of the underlying socioeconomic geography.

B. Reveal Urban Disparities with PP-Net Representations

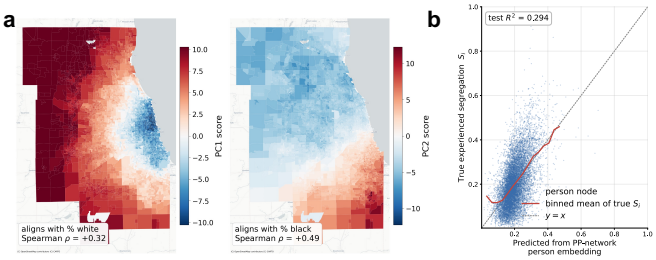


Fig. 3. Fig 3a shows that the unsupervised principal components of PP-Net embeddings can recover racial geography in the embedding space. The first two principal components of the aggregated people-side CBG embedding for Chicago align with the white (Spearman ρ = +0.32) and black population (Spearman ρ = +0.49) shares. Fig 3b illustrates that the learned embeddings of each person have the predictive power of individual experienced segregation [3]. The embedding recovers individual-level segregation with $R^2 = 0.29$.

Fig. 3a shows that an unsupervised PCA of the aggregated people-side CBG embedding yields leading components that align with racial composition-PC1 with the share of white residents, and PC2 with the share of black residents-so racial segregation is already separated in the embedding space . The same embedding also captures *individual* experience segregation [3]: a linear probe on each person's embedding predicts their experienced-segregation index S_i on held-out people ($R^2 = 0.29$; Fig. 3b). Together, these results show that the PP-Net representations effectively capture the distribution of individual-level disparities across the city.

REFERENCES

1. N. Eagle, M. Macy, and R. Claxton, "Network diversity and economic development," Science **328**, 1029 (2010).
2. J. Blumenstock, G. Cadamuro, and R. On, "Predicting poverty and wealth from mobile phone metadata," Science **350**, 1073 (2015).
3. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. communications **12**, 4633 (2021).
4. T. Yabe, B. García Bulle Bueno, M. R. Frank, *et al.*, "Behaviour-based dependency networks between places shape urban economic resilience," Nat. human behaviour **9**, 496 (2025).
5. F. Xu, Q. Wang, E. Moro, *et al.*, "Using human mobility data to quantify experienced urban inequalities," Nat. Hum. Behav. **9**, 654 (2025).
6. Veraset, "Veraset movement data," Distributed by Dewey Data Inc. (2026).
7. X. He, K. Deng, X. Wang, *et al.*, "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, (2020), pp. 639–648.

Acknowledgment. This research is supported by the National Research Foundation (NRF), Singapore, under its Mens, Manus, and Machina programme of the SMART Centre.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [6].

Amenity Proximity Redistributes Mobility Across Home, Workplace, and City Center: Evidence from Austria

PIERO BIRELLO[1], ANDREW RENNINGER[1], AND MÁRTON KARSAI[1,2,*]

[1] Department of Network and Data Science, Central European University, Vienna, Austria
[2] HUN-REN Rényi Institute of Mathematics, Budapest, Hungary
[*] karsaim@ceu.edu

1. INTRODUCTION

The 15-minute city concept has generated extensive research on how amenity proximity shapes urban mobility [1–3]. Much of the existing literature, however, treats the home location as the sole activity anchor. In this study, we analyze GPS traces from over 250,000 individuals across the six largest Austrian cities and show that home, workplace, and city center jointly structure individual mobility — each accounting for at least 10% of all observed Point of Interest (POI) visits in every city. At the individual level, we find that many people concentrate most of their activity around either the home or the work anchor, while the city center serves as an additional pole.

The choice of a dominant anchor between home and work is strongly associated with the relative amenity supply around each. Crucially, this effect is driven largely by the aggregate supply differential rather than by proximity to specific amenity categories: in most cities, individuals exploit the better-supplied anchor broadly across amenity types rather than switching anchors selectively by destination category.

These findings suggest that the 15-minute city framework — residentially focused as it is — overlooks the workplace as an equally consequential activity anchor and the city center as an additional pole. Moreover, since amenity abundance redirects activity toward the better-served anchor, the socioeconomic consequences of local amenity supply may operate through spatial sorting across anchor locations rather than through residential neighborhood effects alone.

2. DATA AND METHODS

We analyze GPS trajectories from over 6,000,000 individuals in Austria, preprocessed through a pipeline of stop detection and home and work location inference. We retain individuals for whom both home and workplace are successfully inferred as distinct locations within the administrative boundary of one of the six largest cities: Vienna, Graz, Linz, Salzburg, Innsbruck, and Klagenfurt. We selected egos who are observed for minimum one week, resulting in a sample of approximately 250,000 individuals. Stops are matched to POIs, which we source from OpenStreetMap. The city center is located as the mode of a city's kernel density estimate of POIs. For each user, we attribute stops to the home (H), work (W), or city center (C) anchor by running a multi-source Dijkstra search on a proximity graph over stops, seeded at the three anchors; each stop is assigned to the closest reachable anchor, with stops far from all attributed to noise. This allows us to count the number of POI visits that user i makes around anchor $A \in \{H, W, C\}$, n_i^A.

For each user i, anchor A, and POI category (tag) t, we compute the Haversine distance from A to the nearest POI of category t, yielding distances d_{it}^A. The weighted proximity of anchor A to the POI landscape, $d_{\mathrm{POI},i}^A = \sum_t f_t \, d_{it}^A$, summarizes the amenity supply around A as the expected distance to the nearest relevant POI, with relevance calibrated by observed population-level visitation frequencies f_t [3]. We quantify the POI-supply differential between two anchors A and B, x_i^{AB}, as the normalized difference in $d_{\mathrm{POI},i}$. For instance, $x_i^{HW} > 0$ indicates that the home anchor is better supplied with POIs relative to work and vice versa.

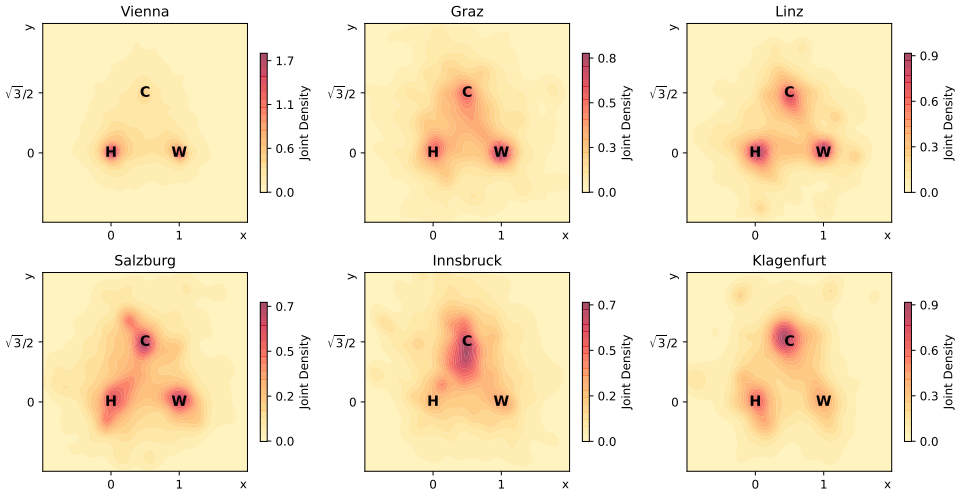


Fig. 1. Gaussian Kernel Density Estimate of visits to POIs – pooled over all users in Austrian cities – in the affine-transformed space defined fixing home (H), workplace (W) and city center (C) to a shared coordinate triple. Home, workplace and city center represent the peaks of distinct modes of the density, having varying importance across cities.

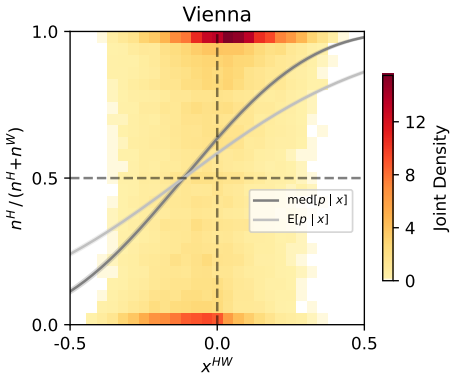


Fig. 2. Effect of relative home-work POI proximity on anchor usage in Vienna. We show the joint distributions of x^{HW} against the fraction of POI visits around home vs. the workplace. Overlaid curves show the estimated median and mean tendencies for a given x^{HW}, at the population level. We find a strong association: as the home anchor becomes relatively better supplied, the fraction of POI-visits around home increases steeply.

We estimate the effect of relative POI supply on anchor usage by fitting a Dirichlet-Multinomial regression model of users' preferences over the three anchors, controlling for a set of user- and anchor-specific covariates. Our main coefficient of interest, β^{HW}, measures how strongly the supply differential x_i^{HW} shifts anchor usage toward home ($\beta^{HW} > 0$) or work ($\beta^{HW} < 0$), with larger magnitudes indicating a stronger effect. To test whether this shift reflects overall POI supply or proximity to specific amenity categories, we fit an analogous model at the level of user–category pairs: here β^{HW} again captures the effect of the overall supply differential x_i^{HW}, while a second coefficient δ^{HW} captures the additional effect of the category-specific differential x_{it}^{HW}, computed analogously to x_i^{HW} from category-level distances d_{it}^{A}.

3. RESULTS

Home, work and city center as mobility anchors. Figure 1 visualizes POI visits across the six cities relative to the three anchors. We adopt a user-specific affine reparametrization of the plane, in which each individual's home, workplace, and city center are mapped onto a fixed shared triplet. The resulting distribution is clearly three-modal, with sharp peaks at the home, workplace, and city-center coordinates; density decays with distance from each anchor, with corridors of elevated density connecting anchor pairs. This picture establishes home, workplace, and city center as dominant, though city-specific, spatial anchors for amenity visitation. In particular, the city-center peak is more pronounced in smaller cities and nearly absent in Vienna.

POI supply shapes preferential usage of the anchors. We observe that many users concentrate most of their activity around either the home or work anchor. We turn to our Dirichlet-Multinomial model to explain what drives the asymmetry. Results indicate that relative anchor-POI proximity is the dominant predictor of home vs. work anchor usage, with estimated coefficients $\hat{\beta}$ ranging from 0.80 to 1.99 across cities: as x_i^{HW} increases — that is, as the home anchor becomes relatively better supplied with amenities — the fraction of POI visits near home rises steeply. Figure 2 shows this effect for Vienna. We also estimate the indifference point, representing the x^{HW} value at which home and work are equally visited. The indifference point is negative in all cities, indicating that even when the workplace is modestly better supplied than home (x_i^{HW} slightly below zero), individuals still tend to exploit the home anchor more on average. Nevertheless, once the supply differential is large enough, individuals anchor around the better supplied place.

Next, focusing at the POI-category level, we ask whether anchor usage reflects aggregate supply — the average proximity across all POI categories — or selective exploitation of anchors based on proximity to specific categories (e.g., visiting coffee shops near work but restaurants near home). As shown in Figure 3, the coefficient β on the user-level average proximity exceeds the category-specific coefficient δ in most cities (four of six), with both coefficients significant throughout. This indicates that anchor exploitation is driven at least as much — and typically more — by overall amenity supply as by category-specific proximity: individuals who live closer to more POIs than available at their workplace tend to anchor around home more broadly across all categories, rather than selectively for only the categories for which home happens to be better supplied.

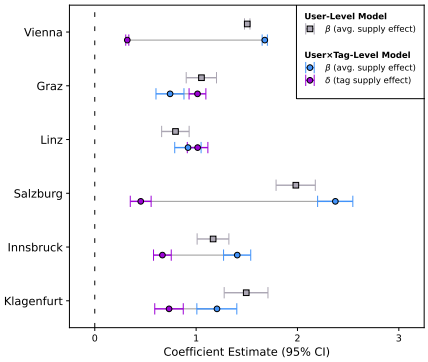


Fig. 3. Estimated coefficients (95% CI) of relative anchor-POI proximity x^{HW} on home-work anchor usage, across cities. Gray squares show the average supply effect β from the user-level model. Blue circles and purple circles show, respectively, the average supply effect β and the tag-specific supply effect δ from the user×tag-level model. The average supply effect exceeds the tag-specific effect in most cities, indicating that anchor exploitation is driven largely by aggregate POI supply.

Conclusions. Overall, our findings indicate that the 15-minute city framework's focus on residential proximity captures only part of how amenity supply shapes urban mobility. Human mobility is structured around multiple anchors, and the POI supply around them is a strong determinant of their usage. Such selection is largely an aggregate-supply-driven phenomenon, rather than simply a category-specific differentiation. This suggests that the socioeconomic consequences typically attributed to local amenity supply may instead operate through where activity concentrates — home, work and the city center —, with implications for interventions that target residential neighborhoods alone.

REFERENCES

1. E.-J. Liu and X.-Y. Yan, "A universal opportunity model for human mobility," Sci. Reports **10**, 4657 (2020).
2. S. Heroy, I. Loaiza, A. Pentland, and N. O'Clery, "Are neighbourhood amenities associated with more walking and less driving? Yes, but predominantly for the wealthy," Environ. Plan. B: Urban Anal. City Sci. **50**, 958 (2023).
3. T. Abbiasov, C. Heine, S. Sabouri, *et al.*, "The 15-minute city quantified using human mobility data," Nat. Hum. Behav. **8**, 445 (2024).

Planning for isolation? The role of urban form and function in shaping mobility in Brasília

ANDREW RENNINGER[1,*]

[1] Department of Network and Data Science, Central European University
[*] renningera@ceu.edu

1. INTRODUCTION

Brasília is a rare urban laboratory: a capital city built almost at once, with a modernist core designed around separation of functions and movements, and a metropolitan periphery that grew through a different logic of necessity. This juxtaposition produces an extreme version of a familiar urban problem. Segregation is not only where people live, but what they experience as they move through the city: the offices, services, retail, leisure, and chance encounters that constitute everyday life. Recent work on "experienced segregation" shows that mobility can either soften residential divides (by mixing activity spaces) or harden them (by enabling selective exposure and avoidance). When a city is planned around enclaves, corridors, and discontinuities, does urban form itself make segregation easier to reproduce through movement? We study segregation at the scale across the blocks, districts, and morphological "cells" that structure access and encounter. We link spatial signatures of the built environment to GPS activity spaces and measure how groups sort. The result is a dynamic view of Brasília's social geometry: who goes where, and whether the city's form turns motion into mixing.

2. DATA AND METHODS

We characterise Brasília's urban form using spatial signatures [1]. We construct an enclosed tessellation from 208,758 building footprints, generating morphological cells organised into a hierarchy of enclosures (524 districts, 3,494 neighbourhoods, 23,919 blocks), before computing morphometrics quantifying form, coverage, density, vegetation, amenities, then apply clustering to produce morphotypes — recurring combinations of urban form that we label as spatial signatures. Mobility data come from GPS traces for 331,934 anonymised users across Brasília, consolidated into 27.4 million stays. Users are assigned to income quintiles using residential location and census data. We define place segregation S_α for each tessellation cell as the evenness of the income distribution of its visitors, following existing work [2], and individual experienced segregation S_i for each user as the mean segregation of all places they visit. We compute exposure profiles as the proportion of each income class in a user's copresence set.

3. RESULTS

Brasília's planned urban form produces stark spatial inequalities in both mobility and experienced segregation. Residents of the poorest income quintile travel farther than the richest (Figure 1A), consistent with "compelled mobility" from peripheral settlements to centrally located employment. Despite this extended mobility, the poor do not encounter more diverse populations: individual experienced segregation (S_i) is highest among Q5 residents (Figure 1B). The richest quintile thus travels less but self-sorts more effectively into homogeneous activity spaces. This pattern is reinforced by residential composition: 64.2% of Superquadra residents and 94.8% of Planned Suburb residents belong to the top quintile, while 71.1% of Dense Poor Irregular residents are much poorer. Exposure profiles confirm the insularity of these enclaves — Q5 residents' encounters are 52.1% with other Q5 individuals, while Q1 residents encounter other Q1 individuals 44.0% of the time.

Place segregation (S_α) varies systematically across morphotypes. The Civic Core — Brasília's planned commercial centre — is the most integrated morphotype, followed by Mid-Rise, Mixed-Use areas (0.41; Figure 1D). The modernist Superquadra falls in the middle, attracting a visitor base 5.1× more likely to be Q5 than Q1. The most segregated morphotypes are Dense Poor Irregular and Planned Suburb, the latter despite — or because of — its high amenity provision and low density. Neighbourhood maps (Figure 1E) illustrate this gradient. Segregation also intensifies at finer spatial scales: mean S_α rises from 0.33 at the district level (L1 enclosures) to 0.39 at the neighbourhood level (L2) and 0.55 at the block level (L3), confirming that experienced segregation is fundamentally a microscopic phenomenon. Critically, 77.2% of users experience lower segregation through mobility than their home neighbourhood alone would predict ($S_i <$ home S_α), yet this ameliorating effect is insufficient to overcome the structural sorting embedded in Brasília's morphology (Figure 1F).

We attempt to distinguish between mobility and segregation that emerges from the structure of the city using a modified "intervening opportunities" model, which distributes trips counterfactually according to the amenities around an individual's home and work. This allows us to compute expected segregation: segregation that would occur if people typically visited locations nearby home and work. While much of the patterns we observe can be attributed to where people live and work, observed S_i is still 10% higher than our modelled expectation, and S_α is 13% higher. Significant predictors of segregation below expectation are amenity mix and network centrality: neighbourhoods with fewer restaurants and bars see more segregation, as do neighbourhoods that are isolated on the street network.

4. DISCUSSION

Brasília's paradox is that mobility does not guarantee mixing. Poor residents travel farther, yet the richest experience the most homogeneous activity spaces, consistent with selective exposure and enclave-based activity spaces. Segregation also sharpens with spatial resolution, indicating that separation is produced at fine spatial scales where permeability and through-movement vary sharply. Taken together, these results are consistent with a role for urban form in amplifying experienced segregation beyond what residential location alone would predict.

REFERENCES

1. Arribas-Bel, D., & Fleischmann, M. Spatial Signatures—Understanding (urban) spaces through form and function. *Habitat International* **128**, 102641 (2022).
2. Moro, E., et al. Mobility patterns are associated with experienced income segregation in large US cities. *Nature Communications* **12**, 4633 (2021).

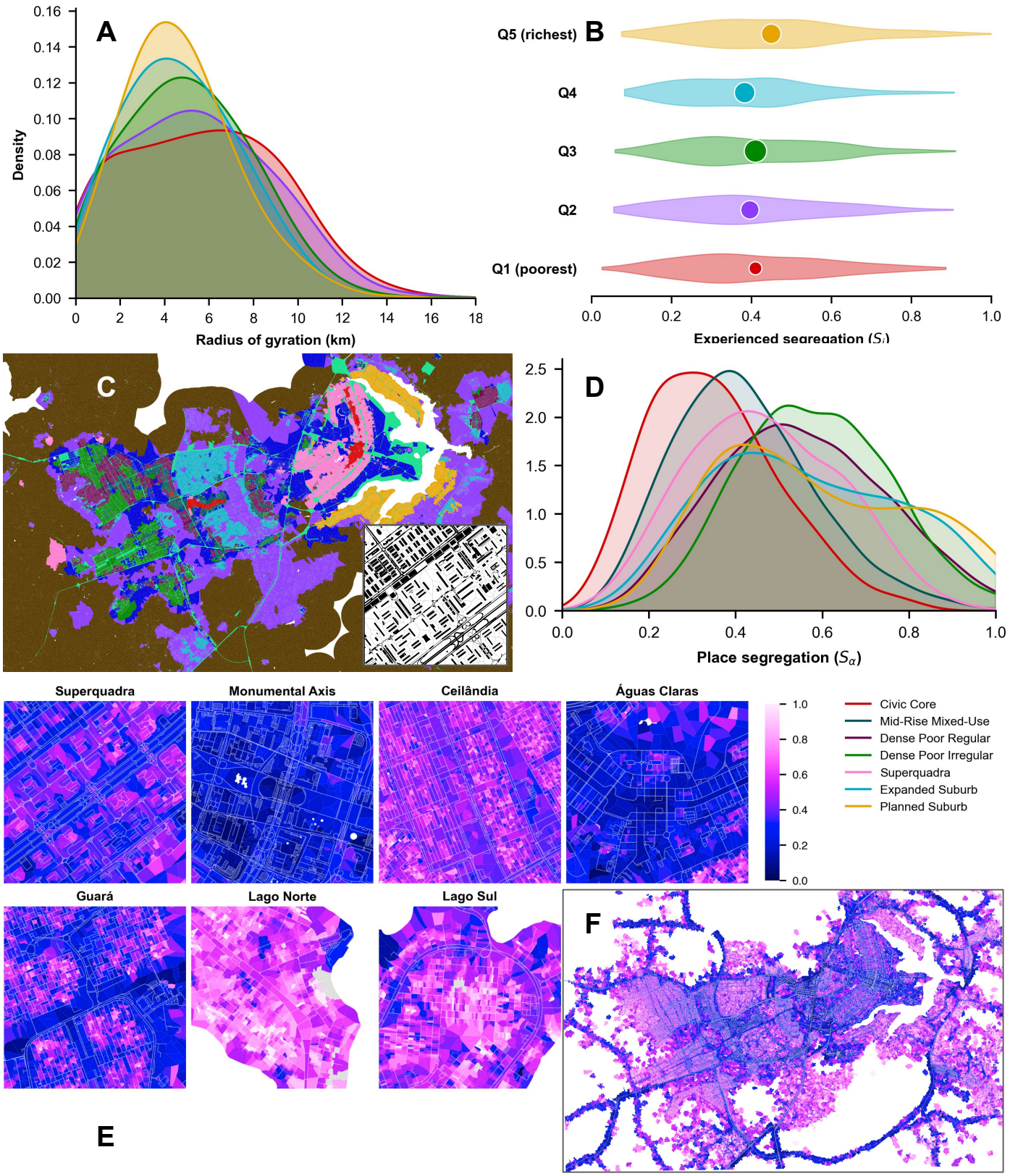


Fig. 1. Spatial signatures of experienced segregation in Brasília. A Radius of gyration by income quintile; the poorest travel farther than the richest (median 4.2 versus 4.0 km). **B** Individual experienced segregation S_i by income class; wealthy residents experience higher segregation than others (mean $S_i = 0.45$ versus 0.38–0.41, $p < 0.001$), reflecting sorting into homogeneous activity spaces. **C** Morphotypic map of Brasília classified into 12 spatial signatures, with an inset showing the underlying form in a superquadra. **D** Place segregation (S_α) by morphotype; the Civic Core is most integrated (median 0.33), while Dense Poor Irregular (0.58) and Planned Suburb (0.55) are most segregated. The modernist Superquadra (0.46) falls between, attracting visitors 5.1 times more likely to be Q5 than Q1. **E** Place segregation S_α across selected neighbourhoods, from the integrated high-rise, mixed-use environment of Águas Claras to the segregated informality of Ceilândia. **F** S_α across the full metropolitan area, revealing a segregated periphery surrounding an integrated civic core.

Beyond Proximity: Rethinking Urban Accessibility Through Human Mobility Patterns in United States

SAUMITRA KULKARNI[1 *] AND ESTEBAN MORO[1,2 **]

[1] Network Science Institute, Northeastern University, Boston, MA 02115, USA
[2] Department of Physics, Northeastern University, Boston, MA 02115, USA
[*] kulkarni.saumi@northeastern.edu
[**] e.moroegido@northeastern.edu

Urban accessibility is typically measured by proximity to home, but this overlooks how people navigate cities and use their time. Using mobility data from 1.86 million individuals across 11 U.S. metro areas, we show that access to daily necessities is structured around a small set of key locations—containers—such as home, work, and other routine destinations. We introduce the offer-meets-demand ratio to measure alignment between available POIs to visit and actual visited POIs. Most individuals rely on multiple containers, with the second container often linked to job locations. Our findings highlight a dual nature of accessibility and call for urban policies grounded in human behavior.

1. INTRODUCTION

Traditional measures of accessibility to daily necessities in cities are based on proximity, either in terms of physical distance or time, from home. Planning and policy-making paradigms like the "15-minute-city" [1] or "food-desserts" [2] are based on this notion, assuming that close access to home ensures effective use. However, such models often overlook how people actually organize their routines and daily necessities across urban spaces. For example, only 14% of daily consumption trips in the US happen within 15-minute walking distance from home [3], and residents in large urban areas in the US travel a median of 6.7km from home to consume fast food [2]. In this study, we use high-resolution mobility data to investigate whether and how people access daily necessities beyond their immediate residential environment, and which locations actually serve as anchors for their daily routines.

2. DATA AND METHODS

In this study, we use individual-level mobility data from 1.86 million anonymized smartphones across 11 U.S. urban areas over a six-month period (October 2016 - March 2017). We identified visits to daily necessities by extracting the stays of people from mobility data around a large collection of points of interest (POI) obtained from Foursquare. We only focus on eight types of POIs as daily necessities that belong to the following categories: Banks, Coffee Shops, Grocery Stores, Gyms, Laundries, Pharmacies, Post Offices, and Salons.

We define the offers available to an individual as the set of the k-nearest POIs of a given category surrounding specific locations—referred to as containers (here we use $k = 10$) [4, 5]. Containers are locations that play a meaningful role in an individual's daily routine. We identify them by ranking all visited locations based on the total time an individual spends there, with the home location always designated as the primary (1^{st}) container. Subsequent containers may include workplaces, shopping centers, or other visited places in the city (see Fig. 1.A). To quantify how well these containers and their associated offers account for actual behavior, we define the offer-meets-demand ratio (ϕ^n) as the proportion of visits made by an individual to POIs within the combined offer sets of their first n containers, relative to their total visits to that POI category. A value of $\phi^n = 1$ indicates that all of an individual's visits are to nearby offers, while lower values suggest that many visits occur outside these offers. The case ϕ^1 corresponds to traditional accessibility measured only around home. We use the 2016-2020 5-year American Community Survey (ACS) to obtain population estimates, and the LEHD Origin-Destination Employment Statistics (LODES) data on the number of jobs at a census block group (CBG) level across 11 metro areas under consideration.

3. RESULTS AND DISCUSSION

We begin by examining whether individuals actually visit the POIs available near their homes. While the median distance to the nearest POI of a specific category from home is just 700 meters (IQR: 0.39–1.36 km), the median distance people travel to visit a POI of the same category is significantly greater—about 7 km (IQR: 3.07–15.19 km; see Fig.1.B). This indicates that, on average, individuals routinely bypass nearby options for their daily activities. To explore this further, we compute the offer-meets-demand ratio ϕ^1 around the 1^{st} container (home) for each individual. Under the assumption that individuals spend most of their time at home, we would expect high ϕ^1 values if proximity from home was the primary driver of behavior. Instead, we find three distinct groups: those who exclusively visit POIs near home ($\phi^1 = 1$, around 10% of them), those who never do ($\phi^1 = 0$, almost 45% of them), and a broad spectrum of individuals in between ($0 < \phi^1 < 1$, the rest; see Fig.1.C).

If not around home, then where do people do most of their daily activities? We define an individual's 2^{nd} container as the

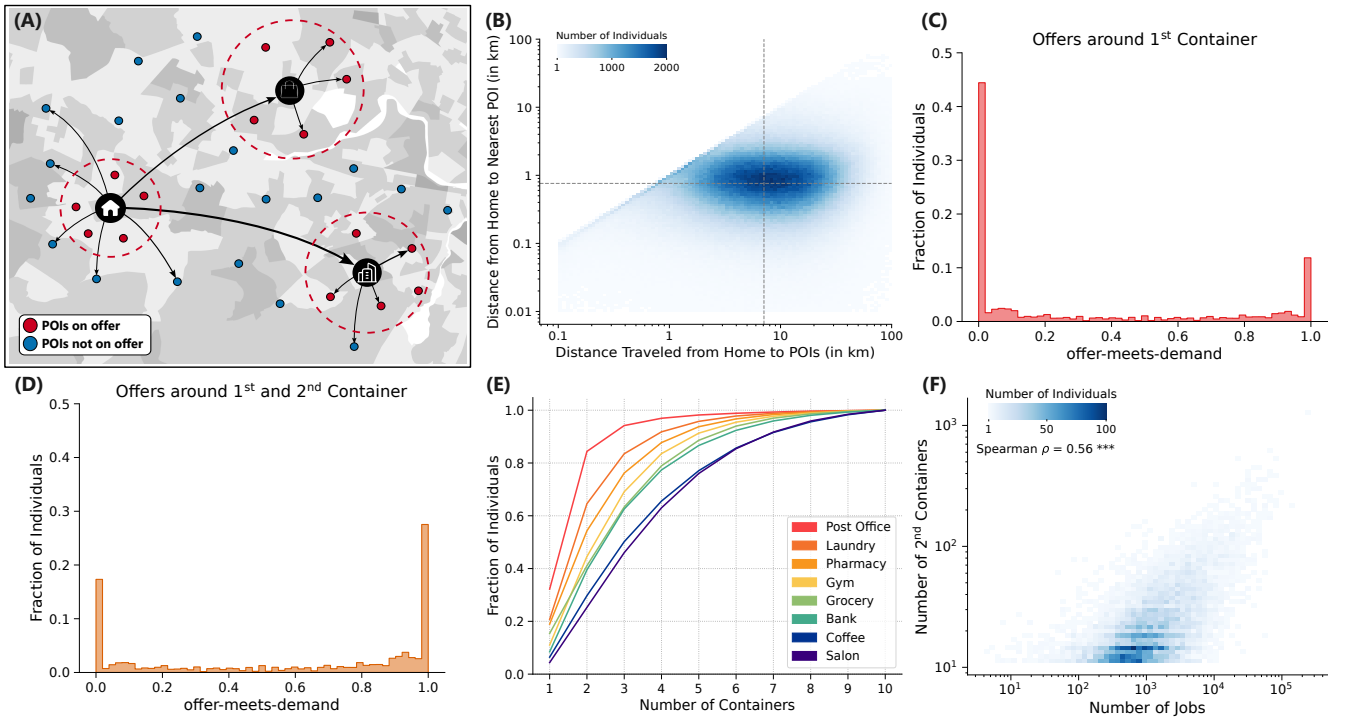


Fig. 1. **(A)** Illustration of the concept of offers around different containers ($k = 5$ and $n = 3$ here) for an individual. Blue points represent all POIs of a specific category in the city, while red points indicate POIs included in the offer set around each container. The offer-meets-demand ratio ϕ^n measures the proportion of visits to these offers out of total visits. Thickness of arrows roughly indicate frequency of visits. **(B)** Heatmap showing the distance from individuals' homes to the nearest POIs of categories under study versus the actual distance traveled to visit POIs of same categories. Grey dotted lines indicate median values. **(C)** Distribution of offer-meets-demand ϕ^1 for individuals with only offers around their 1st container (i.e. homes). **(D)** Distribution of offer-meets-demand ϕ^2 for same individuals with offers around their 1st containers and 2nd containers (i.e. place where they spend most of their time outside home). **(E)** Fraction of individuals with $\phi^n = 1$ as a function of number of containers included, shown separately for different POI categories. **(F)** Relationship between the number of 2nd containers and the number of jobs in each census block group.

location where they spend the most time outside of home. When we include the 10 nearest POIs around this 2nd container to the set of offers, the offer-meets-demand ratio ϕ^2 increases, and for many individuals, it approaches 1 (see Fig. 1.D). This suggests that individuals who do not visit nearby POIs around home often access them near other (next) key locations in their routines. We extend this analysis by progressively adding offers around each individual's ranked containers and computing ϕ^n for increasing values of n. As shown in Fig. 1.E, the fraction of individuals with $\phi^n = 1$ grows with the number of containers, but the rate of increase depends on the POI category. For instance, only 3 containers are needed to capture $\geq 75\%$ of individuals visiting Post Offices, Laundries, or Pharmacies—activities with relatively consistent patterns. In contrast, categories like Salons and Coffee Shops require 5 containers to reach similar coverage, reflecting more varied behavior. Grocery Stores, Gyms, and Banks fall in between, typically requiring 4 containers.

What are the containers that people organize their time around? Our results show that access to daily necessities is shaped not just by residential location but by a small set of recurring places in people's routines. For example, people seem to shop for groceries around four different containers (on average). The nature of these containers varies across individual lifestyles, but common patterns emerge. Notably, areas that contain a large number of individuals' 2nd containers tend to have higher concentrations of jobs (see Fig. 1.F), indicating that workplaces are key anchors in organizing access to daily necessities beyond the home. Ongoing work investigates the demographic and urban drivers of this behavior and whether individuals reorganize their offer-meets-demand in response to both exogenous shocks, such as crisis response policies, and endogenous shocks, such as changes in home or work location.

Our results reveal a dual nature of accessibility in cities—shaped by both where people live and where they routinely spend their time. Measuring access solely from home overlooks this important behavioral aspect. Importantly, identifying where these containers cluster across the city can inform decisions on where to deploy services, amenities, or policy interventions to maximize their impact. By incorporating the actual patterns of people's routines into accessibility metrics, we move toward more realistic, equitable, and effective urban planning.

REFERENCES

1. C. Moreno, Z. Allam, D. Chabaud, *et al.*, Smart cities **4**, 93 (2021).
2. B. García Bulle Bueno, A. L. Horn, B. M. Bell, *et al.*, Nat. Commun. **15**, 2291 (2024).
3. T. Abbiasov, C. Heine, S. Sabouri, *et al.*, Nat. Hum. Behav. **8**, 445 (2024).
4. L. Alessandretti, U. Aslak, and S. Lehmann, Nature **587**, 402 (2020).
5. L. Pappalardo, F. Simini, S. Rinzivillo, *et al.*, Nat. Commun. **6**, 8166 (2015).

Moving Unequally: Phone-Based Mobility Metrics Reveal Socio-Spatial Inequality in Mexico City

RANDY BRANDON GALLEGOS RODRÍGUEZ[1], TONATIUH SUÁREZ-MEANEY[2], MIGUEL JESÚS TORRES RUIZ[1], AND MARIBEL HERNÁNDEZ-ROSALES[3*]

[1] *Instituto Politécnico Nacional, Centro de Investigación en Computación, UPALM-Zacatenco, Mexico City, Mexico*
[2] *Mexico's Ministry of Agrarian, Territorial, and Urban Development (SEDATU), Mexico City, Mexico*
[3] *Center for Research and Advanced Studies, CINVESTAV Irapuato Unit, Guanajuato, Mexico*
[*] *maribel.hr@cinvestav.mx*

1. INTRODUCTION

Mobile phone data have become a valuable source for characterizing the relationship between urban mobility and socioeconomic inequality. Previous studies show that location-level mobility indicators can capture meaningful socioeconomic structure: in the Chilean urban system, pulse (describing the temporal rhythm of residents' outward mobility) was shown to align with a socioeconomic gradient across residential locations [1]; in the Rio de Janeiro Metropolitan Area, entropy of visitors and radius of attraction (characterizing the diversity and spatial reach of incoming flows) were found to be associated with composite development indices [2]. By additionally inferring home and primary daytime locations, we bring general daytime mobility and anchor-directed movement (likely correlated with commuting) together in a single case study, allowing us to test whether pulse and anchor-directed movement carry distinct socio-spatial signatures. This extension to Mexico City (CDMX) is particularly relevant because socioeconomic and mobility relationships are context-dependent and vary across cities and urban systems [3].

2. METHODS AND RESULTS

Data. We used a high-resolution geolocated mobile phone dataset from Veraset, a data provider, covering Mexico City (CDMX) from January 1 to March 20, 2020, immediately before Mexico's national COVID social-distancing measures began on March 23, 2020 [4]. Mobility indicators were aggregated to urban Basic Geostatistical Areas (AGEBs), an official INEGI census geography used by CONAPO to report urban marginalization [5, 6]. The device panel was filtered to devices with sufficiently repeated presence in CDMX to support reliable home inference (at least 5 distinct days observed); applying the multi-night stay procedure described below to this filtered panel assigns a confirmed home location to 265*k* devices. Subsequent resident-based mobility indicators were computed from this confirmed-resident sample, prioritizing the reliability of spatial assignment over raw device coverage.

Home and primary daytime location (PDL) inference. Each GPS ping in the dataset was assigned to its containing AGEB. For home inference, we define a nightly *stance* as the AGEB where a user spent the most distinct hours between 00:00–06:00 among CDMX, accepting only strict winners (ties yield no stance). Home is the modal stance AGEB across all weeknights, accepted when a user accumulates $\geq$5 weeknight stances and the modal AGEB accounts for $\geq$60% of a user's weeknight stances. Among users who received a home assignment, the mean share of nights spent in the inferred home AGEB is 0.93, indicating high internal consistency of the procedure. The PDL is inferred using the same logic over 11:00–16:00 on weekdays, with the consistency threshold relaxed to $\geq$40% to accommodate greater daytime variability; this AGEB likely corresponds to a workplace, school, or other regularly visited anchor point. Device coverage (confirmed-home devices relative to census population) is uneven across IM: mean coverage rises from 3.1% (most marginalized quintile) to 6.8% (least marginalized), and from 2.5% to 5.4% when also requiring a confirmed PDL; we address this with a coverage covariate in

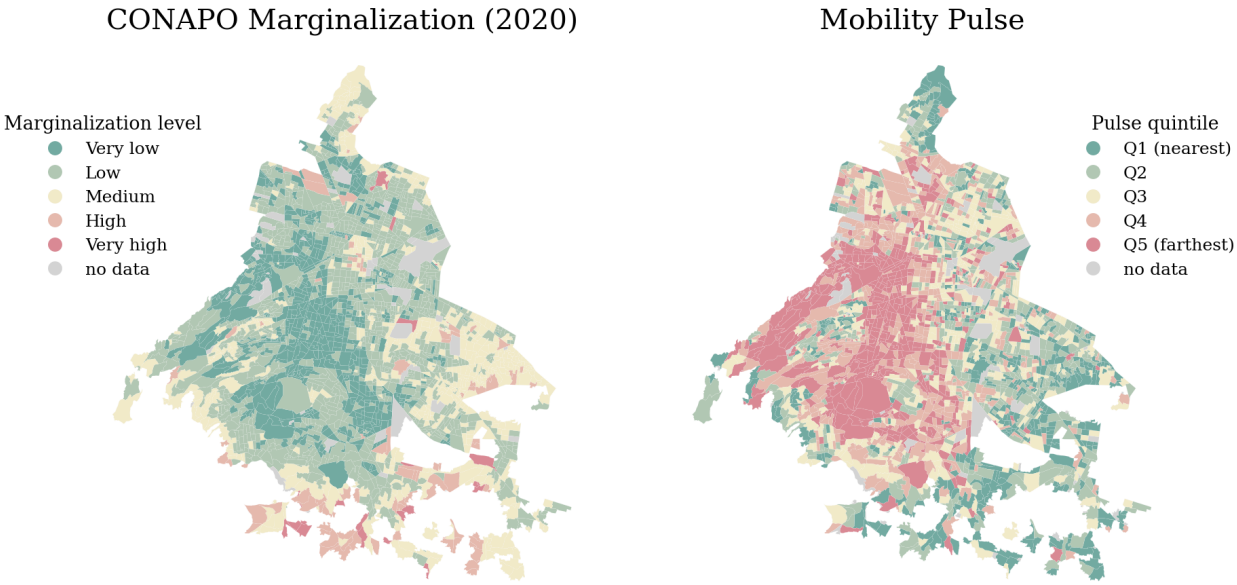


Fig. 1. Choropleth maps of CONAPO marginalization grade and daytime pulse quintile at AGEB level, Mexico City (2020).

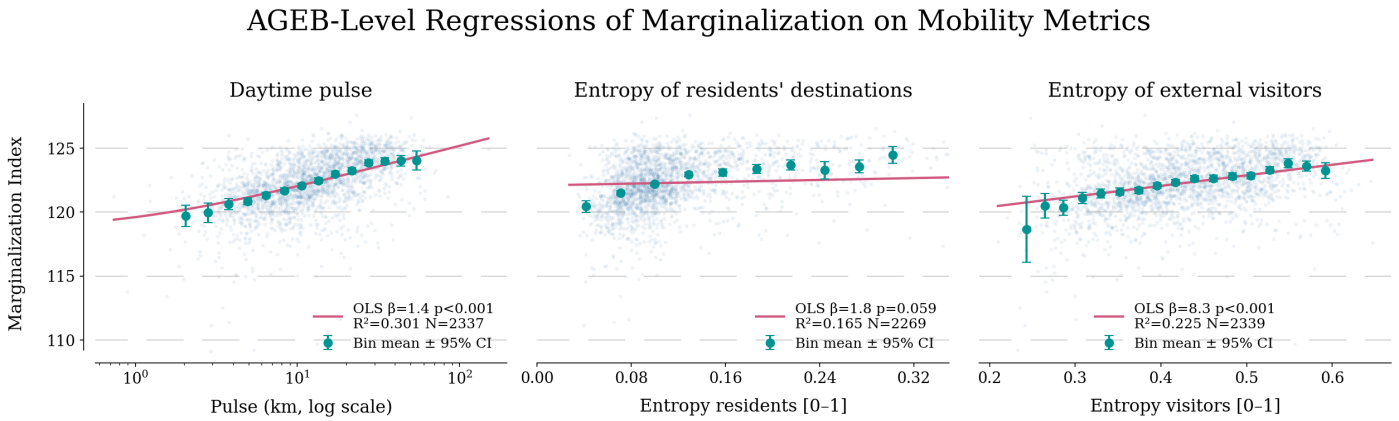


Fig. 2. OLS regression of CONAPO marginalization (higher = less marginalized) on daytime pulse, entropy of residents' destinations, and entropy of external visitors across CDMX AGEBs.

the regression framework below.

Mobility metrics. All metrics are aggregated to AGEB level across all days (excluding public holidays), with the exception of distance to the PDL and share of local anchoring. *Pulse* measures the mean distance residents travel from their home AGEB centroid at each hour, averaged within users across days and then across residents of each AGEB. *Entropy of residents' destinations* captures how diversely residents distribute across destination AGEBs at each hour, using the modal destination per user per hour normalised by $\ln N$, where N is the number of distinct CDMX AGEBs recorded as a destination in the dataset. *Entropy of visitors* and *radius of attraction* characterise each AGEB from the perspective of incoming flows, restricted to external visitors (home AGEB $\neq$ destination AGEB): the former measures the diversity of visitors' home origins; the latter their mean home-to-destination distance. Both entropy metrics are normalised to $[0, 1]$. Distance to the PDL is the haversine distance between a user's home and the PDL centroids, defined only when the two differ; share of local anchoring is the share of residents whose home AGEB and PDL coincide.

Regression framework and results. We estimate the association between each mobility metric and IM using ordinary least squares (OLS) regression at the AGEB level with heteroscedasticity-robust standard errors (HC1). Since device coverage was found to be uneven across AGEB's IM, all regressions include a log-transformed coverage covariate (confirmed devices per AGEB relative to census population, or relative to raw footfall for visitor-side metrics) to control for representative bias. We additionally verified these associations under alternative weighting schemes (device-weighted least squares) and temporal scopes (weekday-only and weekend-only), finding consistent directional associations throughout. To assess sign stability, we apply a 1,000-resample bootstrap and report whether the 95% confidence interval of the slope coefficient excludes zero. Pulse is the strongest predictor of IM ($R^2 = 0.301$, bootstrap-stable), with higher daytime mobility associated with lower socioeconomic disadvantage (Fig. 2). The remaining mobility metrics also show significant, non-null associations with IM. While pulse captures general daytime movement, distance to the PDL and the share of residents whose home and anchor AGEB coincide are both negatively associated with IM, suggesting that anchor-directed trips capture a distinct dimension of mobility from general movement. However, our pipeline's primary-daytime-location inference disproportionately excludes highly marginalized AGEBs (estimated inclusion probability roughly 50% vs. 80% across the marginalization range), possibly reflecting lower mobile data of residents in these areas leading to sparser daytime location pings; while representativeness concerns affect all six metrics to some degree, this particular retention gap is most pronounced for the PDL-based metrics.

3. CONCLUSIONS AND FUTURE WORK

Our results show that GPS-derived mobility metrics capture meaningful dimensions of mobility in México City extending their application beyond prior studies. Pulse emerges as the strongest predictor of the CONAPO marginalization index ($R^2 = 0.301$), potentially reflecting greater access to private transport and more discretionary daytime travel in less marginalized areas.

Beyond the statistical association, this pattern carries a social reading: residents of more marginalized AGEBs not only travel less over the course of the day but are also more likely to remain anchored within their own home AGEB and, when they do leave it, to travel farther to reach their primary daytime location, three distinct signals of the same underlying constraint on access to the rest of the city. Such restricted mobility can itself reinforce socio-spatial inequality by limiting exposure to jobs, services, and opportunities concentrated elsewhere, a dimension not directly captured by a purely economic marginalization index.

Future work will test whether these patterns generalize across Mexican cities and assess alternative explanations, including car access, transit accessibility, urban centrality, and employment density.

REFERENCES

1. M. Lenormand and H. Samaniego, "Uncovering the Socioeconomic Structure of Spatial and Social Interactions in Cities," URBAN SCIENCE **7** (2023).
2. M. Lenormand, H. Samaniego, J. Chaves, *et al.*, "Entropy as a Measure of Attractiveness and Socioeconomic Complexity in Rio de Janeiro Metropolitan Area," ENTROPY **22** (2020).
3. H. Barbosa, S. Hazarie, B. Dickinson, *et al.*, "Uncovering the socioeconomic facets of human mobility. arxiv 2020," arXiv preprint arXiv:2012.00838 (2012).
4. Secretaría de Salud, "Jornada Nacional de Sana Distancia," https://www.gob.mx/cms/uploads/attachment/file/541687/Jornada_Nacional_de_Sana_Distancia.pdf (2020). Government of Mexico. Accessed: 2026-06-16.
5. Instituto Nacional de Estadistica y Geografia, *Principales resultados por AGEB y manzana urbana: Censo de Poblacion y Vivienda 2020*, INEGI, tercera edicion ed. (2021). Accessed: 2026-06-16.
6. Consejo Nacional de Población (CONAPO), "Índices de marginación 2020," https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372 (2021). Accessed: 2026-06-16.

The parenthood effect in urban mobility

MARIANA MACEDO[1,2,3,*], RONALDO MENEZES[4,5], AND ALESSIO CARDILLO[6]

[1] Constructor University, Bremen, Germany; [2] Northeastern University London, London, UK; [3] Khoury College of Computer Sciences, Northeastern University, Boston, USA; [4] University of Exeter, Exeter, UK; [5] Federal University of Ceará, Fortaleza, Brazil; [6] University of Barcelona and UBICS, Barcelona, Spain; [*] mmacedo@constructor.university

1. INTRODUCTION

Urban mobility between residential and workplace zones across metropolitan areas has been studied for a long time using the toolbox of network science [1]. The vast majority of mobility models available consider travellers as indistinguishable, using an 'average citizen' approximation. However, our society is intrinsically *diverse*, with travellers distinguishable, for instance, according to attributes like gender, age, and socioeconomic status [2].

We investigate how two life-changing events—parenthood and marriage—which alter an individual's priorities, routines, and ultimately their interactions with the environment, reshape urban mobility patterns.

By leveraging the American Community Survey data combined with OpenStreetMap amenity information, we analyse how parenthood and marriage reshape urban experiences across several metropolitan areas in the USA. We employ Shannon entropy to compute the *mobility diversity* and quantify the accessibility to amenities based on commuting patterns. Then, we characterise metropolitan areas according to mobility diversity and *cost* (*i.e.*, the commuting travel time). Finally, we compute the difference in mobility diversity and cost between parents and non-parents, and married and non-married travellers.

2. DATA AND METHODS

Our data are a combination of mobility data obtained from the US Census Bureau and data on amenities obtained via OpenStreetMap. We have considered only trips related to work made by people older than 15. Our data were collected using the United States Census Bureau under the 2019 edition of the so-called American Community Survey (ACS). The variables collected by the ACS are: the state, city, metropolitan region, individual (traveller) characteristics (*e.g.*, gender, age, marital status), mobility characteristics (*e.g.*, work location, commuting travel time), and expansion factors. Expansion factors are used to ensure that the sample is representative of the whole population.

Individuals are classified as married if they report being currently married (MSP $= 1$ for married, spouse present, or MSP $= 2$ for married, spouse absent). The 'married' category likely includes not only legally married individuals but also a portion of unmarried cohabiting partners. Individuals are classified as non-married if they report being widowed, divorced, separated, or never married. For the parent category, individuals are classified based on the number of children (NOC > 0).

For each urban area, the space is divided into zones. We consider 17 US metropolitan areas, spanning 6 to 23 zones and 0.8 to 9.6 million weighted trips; the full list and their characteristics are reported in Ref. [3]. However, the spatial tessellations of home and work zones provided by the US census do not perfectly align. In particular, multiple home zones often map onto a single work zone, requiring us to collapse some home zones. This preliminary step avoids inconsistencies in the comparison of differences in residential and mobility concentration, although it considerably reduces the number of available zones. We decided to use the metropolitan areas classified as such by the US census as our reference, as they are well-recognised functional areas in which people are more likely to work.

We can use Shannon entropy to compute the *mobility diversity* (*i.e., diversity of accessibility to amenities*) by travellers of type Y, M^Y, over an area A_i^Y, as:

$$M^Y = -\frac{1}{\log_2 N_Z} \sum_{i=1}^{N_Z} p_i^Y \ \log_2 \frac{A_i^Y}{p_i^Y}\,, \qquad (1)$$

where N_Z is the number of zones, the probability p_i^Y corresponds to the ratio between the product of the number of amenities in zone i, n_i, and the number of travellers of type Y whose destination zone is i, T_i^Y (*i.e.*, $n_i^Y = n_i\ T_i^Y$), and its sum over all zones $N^Y = \sum_{i=1}^{N_Z} n_i^Y$.

3. RESULTS

We employ *mobility diversity* [M, see Eq. (1)] measure to quantify accessibility to amenities based on commuting patterns (Fig. 1A) and observe that all urban areas display few differences across distinct groups of travellers (*e.g.*, married). Then, we characterise metropolitan areas according to mobility diversity, M, and *cost* (commuting travel time), C, (Fig. 1B). The plot suggests that (except for Houston) metropolitan areas with higher values of M (*e.g.*, Baltimore) are those with higher values of C.

Finally, we compute the difference in mobility diversity, ΔM, and cost, ΔC, between parents and non-parents (Fig. 1C), and married and non-married travellers (Fig. 1D). We observe variation across metropolitan areas, with some showing higher costs for parents (*e.g.*, Houston) and others showing the opposite trend (*e.g.*, Chicago). Concerning married people, metropolitan areas like New York have smaller costs for non-married people, whereas metropolitan areas like Baltimore display the opposite trend.

All the differences observed in panels C and D of Fig. 1 raise the following question: 'Are these differences meaningful, or are they just the product of noise or chance?' To this aim, we designed five null hypotheses/models:

NM1: We keep the trips (*i.e.*, their origin, destination, and time), but we shuffle the household arrangement feature (*i.e.*, married/non-married or parent/non-parent).

NM2: We keep the trips' origin, destination, and travellers' type, but the travel time is extracted from a probability distribution obtained by fitting the empirical values.

NM3: We keep the trips' origin, time, and travellers' type, but the travel distance is extracted from a probability distribution obtained by fitting the empirical travel distances.

NM4: We keep the trips' origin, destination, and travellers' type, but we shuffle the travel time feature.

NM5: We keep the trips' origin, time, and travellers' type, but we pick a random travel destination.

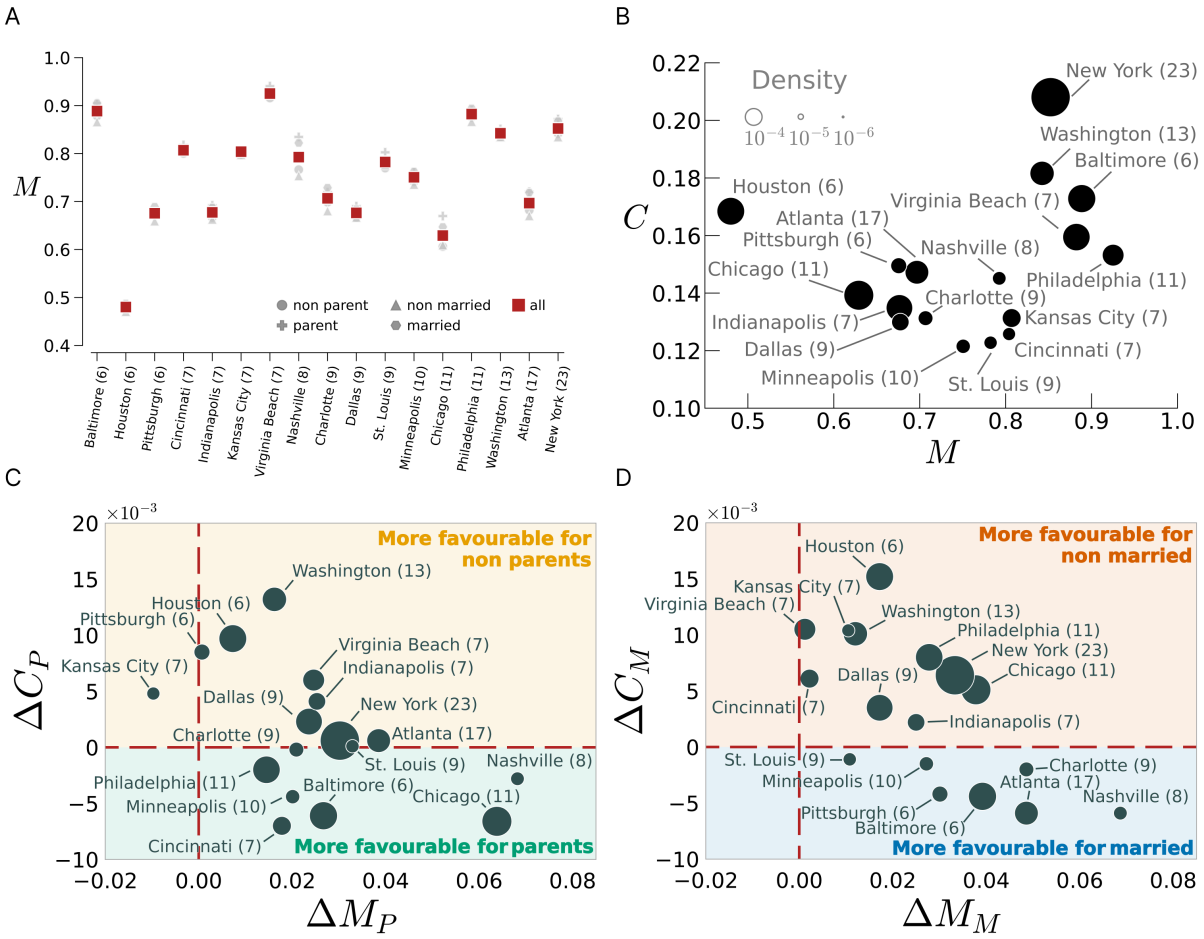


Fig. 1. Characterisation of mobility diversity and cost for parent and married travellers. (A) the mobility diversity, M, for different household arrangements; (B) the average value of travel cost, C, and mobility diversity, M; (C) the differences in cost and mobility diversity between parents and non-parents $(\Delta C_P, \Delta M_P)$; and (D) between married and non-married travellers $(\Delta C_M, \Delta M_M)$.

For each null case, we average the results over an ensemble of 5,000 realisations by bootstrapping 80% of the data. NM1 examines whether the cost differences observed, ΔC, are due exclusively to the type of travellers (*e.g.*, parent) and their distribution across the metropolitan area. NM2 and NM3 test whether the distributions of travel time and distance explain most of the differences. NM4 aims to understand whether the differences are due to the 'speed' at which trips occur. Finally, NM5 checks whether the differences are due to the 'spatial tessellation' by which zones are defined. More details are available in [3].

The statistical significance of these differences varies across null models, with some cities showing robust differences across all five null models ($p < 0.05$), and others showing sensitivity to specific assumptions about travel patterns. Overall, in 10 out of 17 metropolitan areas considered, the empirical differences are positive ($\Delta C_P > 0$), suggesting a general tendency toward higher mobility costs for parents, though the robustness of this pattern varies by metropolitan area.

4. CONCLUSION

Our analysis demonstrates that cities vary significantly in accommodating different household arrangements: some better accommodate single individuals (Houston, Virginia Beach) or married people (Atlanta, Baltimore), whereas others favour parents (Cincinnati, Chicago). Parents and married individuals face different mobility costs and amenity access patterns compared to their counterparts. In most cities, non-parents and unmarried individuals spend less time travelling, whereas parents and married individuals live closer to essential amenities (Fig. 1C and D).

These insights together contribute to research advocating for nuanced, equitable urban planning, advancing discourse by promoting tailored design strategies addressing diverse demographic needs rather than one-size-fits-all approaches.

Funding. A.C. acknowledges financial support from the Ramón y Cajal program (Agencia Estatal de Investigación) through the grant RYC2023-044587-I. A.C. acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación, through project No. PID2024-158120NB-C22.

Data Availability Statement. Data underlying the results presented in this study are available in Ref. [4].

REFERENCES

1. H. Barbosa, M. Barthelemy, G. Ghoshal, *et al.*, "Human mobility: Models and applications," Phys. Reports **734**, 1 (2018).
2. M. Macedo, L. Lotero, A. Cardillo, *et al.*, "Differences in the spatial landscape of urban mobility: gender and socioeconomic perspectives," Plos one **17**, e0260874 (2022).
3. M. Macedo, R. Menezes, and A. Cardillo, "The parenthood effect in urban mobility," npj Complex. **3**, 23 (2026).
4. https://github.com/marianagmmacedo/parenthood_effect_mobility.

Measuring the Daytime POI-visit Impacts of Hybrid Work across U.S. Cities

XIAOTONG XU[†1,2], CALLIE CLARK[†2,3], WEI MA[1], AND TAKAHIRO YABE[2,4,*]

[1] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hong Kong SAR, 999077
[2] Center for Urban Science and Progress, Tandon School of Engineering, New York University, 370 Jay Street, New York, 11201
[3] Marron Institute of Urban Management, New York University, 370 Jay Street, New York, 11201
[4] Department of Technology Management and Innovation, Tandon School of Engineering, New York University, 2 Metrotech, New York, 11201
[†] These authors contributed equally to this work.
[*] takahiroyabe@nyu.edu

1. INTRODUCTION

Cities historically formed around the concentration of work, especially office work, their spatial structure shaped by the tradeoff between commuting costs and land prices [1, 2]. As highways lowered those costs, employment and population decentralized and the metropolitan form dispersed [3]. This dispersal spread daily activity across the metropolitan area, where point-of-interest (POI) visitation data now traces where people spend time beyond their residence [4].

The Covid-19 pandemic triggered a widespread shift to remote and hybrid work, reducing how often employees commute to a workplace. This change has persisted: in the United States, the share of full workdays performed from home rose from 7% in 2019 to 60% in 2020 and stabilized near 28% by mid-2023 [5], while across 27 countries full-time employees averaged 1.5 WFH days per week (2021 to 2022) [6]. Hybrid work, dividing the workweek between home (WFH) and the office, varies widely across firms and occupations in required days and hours, making its spatial impacts heterogeneous. The effects of this shift are being studied at the level of urban form, including rents, migration, and the "donut effect" [7], yet its effect on individual POI-visit behavior during work hours remains unstudied, along with downstream questions such as experienced segregation [8].

2. METHOD

Using anonymized, high-resolution mobile phone location data provided by Cuebiq from 2019 and 2022 across eight major U.S. MSA regions[9], we identify individual-level hybrid work patterns. By further integrating the SafeGraph POI dataset[10], we examine their visitation behaviors to different categories of POIs.

A. Home and Work Location Identification

The first step to identifying hybrid workers is to label each user's home and work location. We use 3 months of data from April to June in 2019 and 2022, evaluating stops between 10pm and 6am to determine home location and 8am to 4pm for work locations. Due to privacy considerations, all stops within a user's work or home census block group (CBG)(as determined by the data provider) are uplevelled to a random point within the CBG, so we cannot use building type or exact POI to inform identification.

For home locations, we evaluate all stops within the time period where the user spends at least four hours. We require the user to spend at least 15 nights across 8 unique weeks in that CBG to qualify as a home candidate. Then we calculate the total dwell time and total nights for each location, selecting the location with the highest combination.

Once we establish the home location, we require a user's work to be in a different CBG, considering all weekday stops lasting at least two hours outside the home CBG. Given the inconsistent nature of hybrid work, we adapt established work detection algorithms [11] by lowering thresholds to capture work locations visited as infrequently as once a week. For a location to be a work candidate, at least 10% of observed weekday activity (excluding home dwell time) must be spent there, and a user must stop there at least 4 times over 4 unique weeks. We then select the location with the highest combination of dwell time and visits. After applying these algorithms, we have 1,264,707 cuebiq users across all 8 MSAs with both home and work locations. We validate our identified work locations using LEHD [12] data and find correlation values range between 0.64 to 0.76.

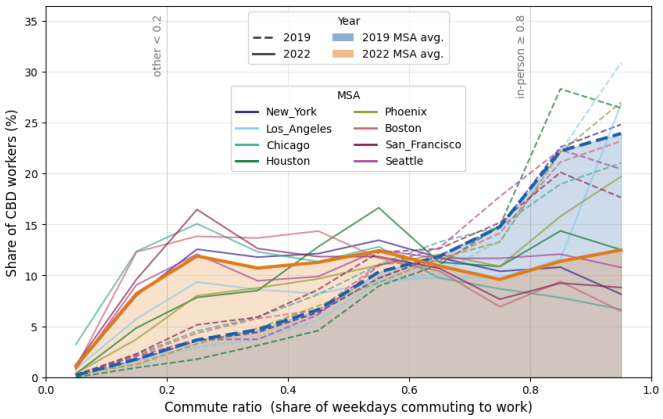


Fig. 1. Commute ratio distribution for CBD workers in 2019 and 2022.

B. Hybrid Worker Detection

Next we identify hybrid workers using their weekday work visitation patterns. To address the wide range of work schedules, we limit our subset to those who work in the Central Business District (CBD) and live in a CBG with over 75% "white collar" workers (defined using ACS census data on occupation type). Although this reduces the sample size to 32,128 users, it also reduces the risk of mislabeling in-person workers with hours different than 9-5pm (ie service workers) as hybrid workers.

Next, we look at all weekdays where a user is observed and identify those with a work commute. To ensure we are not conflating WFH with time off, we remove any day with more than two hours spent in a "non-work" POI during 9-5pm. When defining 'non-work' POIs we exclude those that could facilitate work like cafes and any POI with the following keywords in their name: 'Public Library', 'Branch Library', 'cowork', 'cafe', 'coffee'. Finally, we calculate a commute ratio as the proportion of weekdays a user commutes to work. Comparing the commute ratio in 2019 and 2022 in Figure 1, we see a consistent trend of lower ratios in 2022, indicating that we are successfully capturing the shift in commuting patterns.

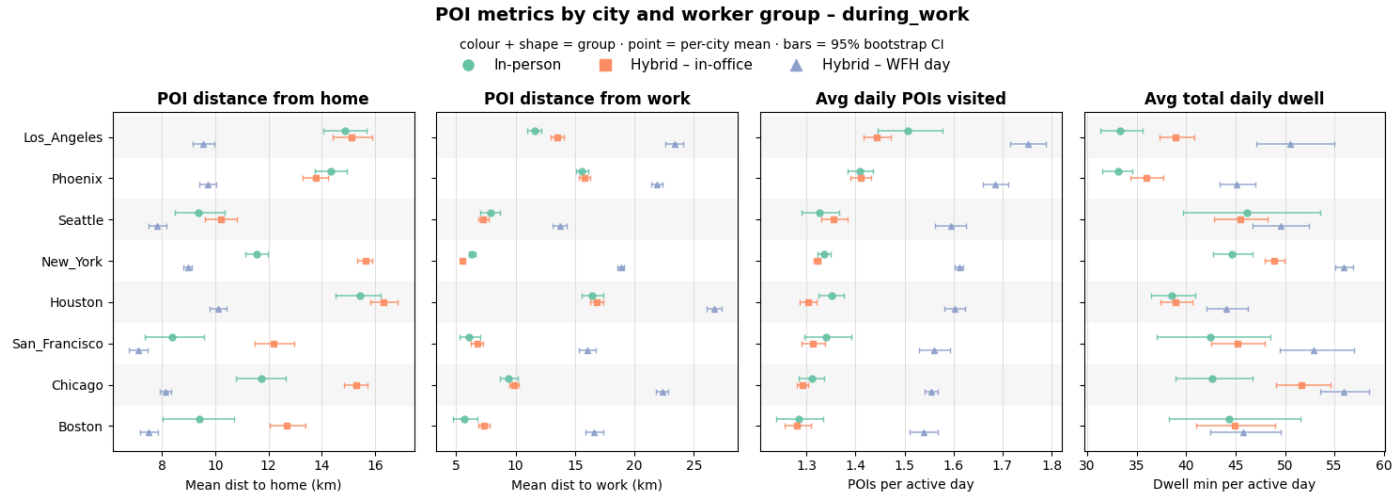


Fig. 2. POI Visitation Metrics by day-type with 95% CI. Average metrics are calculated per-user (and per day for visits/dwell time).

3. POI VISITATION TRENDS

We define an in-person worker as a user with a commute ratio greater than 0.8, and a hybrid worker as one with a ratio between 0.2 and 0.8 (purposefully excluding fully remote workers). These values yield an average hybrid rate for white collar workers of 74% (79% in NYC), compared to 77% in a NYC 2022 survey of office workers [12]. Next, we consider three day types: in-person worker in office, hybrid worker in office, and hybrid worker WFH. For each, we evaluate all visits to POIs in ten categories: beauty, errands, fitness, grocery, pharmacy, retail, restaurants, bars, cafes, and parks. We find that hybrid workers behave significantly differently on WFH days versus in-office days (Figure 2).

POI distance from home during work hours (9am-5pm) is higher for in-office workers, and distance to work is lower. These expected trends further validate that we are capturing hybrid workers accurately. Hybrid workers show similar POI visitation to in-person workers when in office, but significantly higher visitation rates and total dwell time when working from home, a trend that holds for the majority of the 8 cities.

Next we compare each hybrid worker's daily average visitation rate in-person versus WFH for each POI category. When working from home, hybrid workers are more likely to visit utilitarian POIs, indicating they may perform other activities during the work day. We also see an increase in cafe visits, demonstrating that workers spend recreational time outside their home during work hours. However, trends vary for parks and restaurants, potentially due to the spatial form of the city and POI concentration. Evaluating after-work (6pm-midnight) behavior, we find no significant or consistent trends, indicating that WFH may have little impact on after-work activities. Overall, these findings demonstrate that the shift to hybrid work impacts the spatial and temporal economic activity in POIs, particularly in areas with high rates of hybrid work.

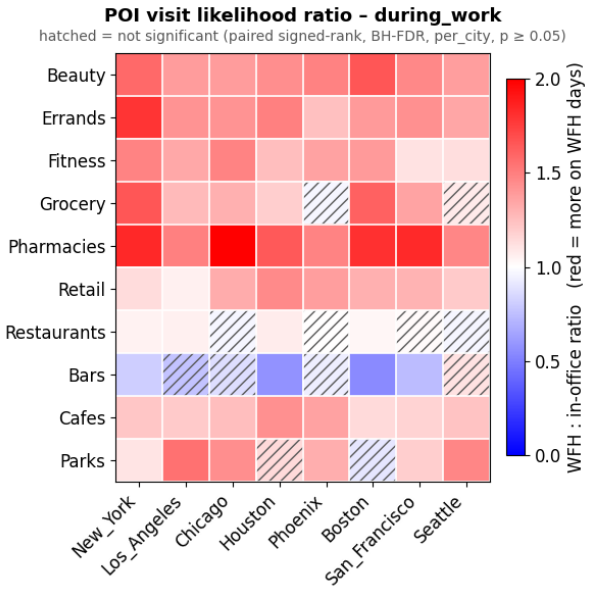


Fig. 3. Hybrid workers' POI visit ratio by day-type. Differences are computed per worker using paired Wilcoxon signed-rank with Benjamini-Hochberg correction within city.

REFERENCES

1. A. Marshall, *Principles of Economics* (Macmillan, London, 1890).
2. W. Alonso, *Location and Land Use: Toward a General Theory of Land Rent* (Harvard University Press, Cambridge, MA, 1964).
3. N. Baum-Snow, "Did highways cause suburbanization?" The Q. J. Econ. **122**, 775 (2007).
4. D. R. Davis, J. I. Dingel, J. Monras, and E. Morales, "How segregated is urban consumption?" J. Political Econ. **127**, 1684 (2019).
5. J. M. Barrero, N. Bloom, and S. J. Davis, "The evolution of work from home," J. Econ. Perspect. **37**, 23 (2023).
6. C. G. Aksoy, J. M. Barrero, N. Bloom, *et al.*, "Working from home around the world," Brookings Pap. on Econ. Activity **2022**, 281 (2022).
7. A. Ramani and N. Bloom, "The donut effect of COVID-19 on cities," Working Paper 28876, National Bureau of Economic Research (2021).
8. S. Athey, B. Ferguson, M. Gentzkow, and T. Schmidt, "Estimating experienced racial segregation in US cities using large-scale GPS data," Proc. National Acad. Sci. **118**, e2026160118 (2021).
9. Cuebiq, "Cuebiq mobility data for research," https://www.cuebiq.com/about/data-for-good/. Accessed: 2025-06-27.
10. SafeGraph, "Places dataset," https://www.safegraph.com/data-examples/places-data (2022). Accessed: 2025-06-27.
11. S. De Sojo, L. Lucchini, O. D. Langle-Chimal, *et al.*, "Howde: A validated algorithm for home and work location detection," Comput. Environ. Urban Syst. **126**, 102397 (2026).
12. U.S. Census Bureau, "State_wac_s000_jt00_{2019, 2022}.csv," LEHD Origin-Destination Employment Statistics (2019, 2022). Workplace Area Characteristics, multiple states. Accessed on 6/30/2025.

A Multi-resolution Mobility-aware Framework for Dynamic Air Pollution Exposure Assessment

BASSEL OTHMAN[1,*] AND GUILLAUME SABIRON[1]

[1] IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France
[*] bassel.othman@ifpen.fr

1. INTRODUCTION

Human exposure to air pollution is strongly driven by daily mobility patterns. Conventional exposure assessments rely on residential locations or static population distributions, overlooking the continuous redistribution of individuals between home, work, and transport infrastructures. Recent mobility modeling approaches increasingly rely on traffic simulation and emission models to represent these dynamics.

However, existing modeling chains remain dependent on specific traffic and emission model combinations, limiting transferability and introducing structural uncertainties. Recent work has shown that differences in mobility and emission modeling can lead to significant variations in spatial emission patterns under identical demand conditions [1].

Systemic approaches have been proposed to better integrate mobility, environmental fields, and population dynamics within unified exposure frameworks, highlighting the importance of consistent spatio-temporal representations of mobility [2]. Moreover, recent integrated environmental assessments have shown that modeling chain choices can significantly influence exposure indicators in urban contexts [3].

Mobility-induced exposure variability motivates frameworks able to exploit heterogeneous mobility data sources. This work investigates how heterogeneous mobility representations can be coupled with spatio-temporal air-quality fields to quantify population-level and individual exposure variability.

2. MOBILITY-AWARE EXPOSURE FRAMEWORK

The proposed framework couples mobility representations with spatio-temporal pollutant concentration fields to estimate exposure at different levels of mobility granularity, as illustrated in Fig. 1. The air quality modeling chain (traffic, emissions, dispersion) is assumed to be pre-computed and follows established methodologies described in [4].

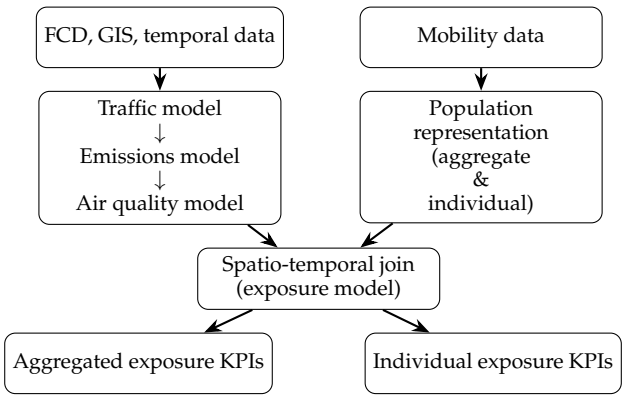


Fig. 1. Framework coupling air-quality outputs with heterogeneous mobility representations (population distributions and trajectories) to estimate exposure. Air-quality modeling (traffic, emissions, dispersion) follows [4].

Let $C(\mathbf{x}, t)$ denote the spatio-temporal pollutant concentration field. Depending on mobility data availability, exposure is estimated either from aggregated population distributions or individual trajectories. Both regimes rely on the same spatio-temporal coupling between mobility and concentration fields but produce different exposure observables: the former characterizes population-level exposure, whereas the latter captures individual exposure variability.

Aggregated mobility regime. When mobility is represented at an aggregated level, the population is described by a dynamic spatial distribution $P(z, t)$, representing the number of individuals located in zone z at time t. Such dynamic distributions can be obtained from origin–destination flows derived from mobile phone data, population counts, or other aggregated mobility products. The resulting indicator corresponds to a population exposure burden:

$$E_{\text{agg}} = \sum_t \sum_z P(z,t)\, \bar{C}(z,t)\, \Delta t \qquad \textbf{(1)}$$

where $\bar{C}(z, t)$ denotes the spatially averaged concentration in zone z at time t. This regime produces population-level exposure indicators based on spatial aggregation of mobility and concentration fields.

Individual mobility regime. When mobility is available at the individual level, each agent i follows a trajectory $\mathbf{x}_i(t)$, and exposure is computed along these trajectories:

$$E_i = \int_{t_0}^{t_1} C(\mathbf{x}_i(t), t)\, dt \qquad \textbf{(2)}$$

where E_i is expressed in concentration-time units (e.g., $\mu g\, m^{-3}\, h$ when using hourly NO_2 concentrations) and $\mathbf{x}_i(t)$ is the trajectory of individual i. This formulation preserves inter-individual variability and enables the analysis of exposure distributions within the population, including heterogeneity and percentile-based indicators.

3. AGGREGATED EXPOSURE RESULTS

In the aggregated mobility regime, the proposed framework produces time-resolved population exposure indicators by coupling dynamic population distributions with spatio-temporal concentration fields. This enables the characterization of how population movements redistribute exposure across the urban space over the course of the day.

Results are illustrated using a synthetic population evolving over the Paris metropolitan area and hourly pollutant concentration fields generated by the air-quality modeling chain.

Figure 2 illustrates the spatial density of individuals exposed above an AQI threshold (AQI $>$ 20) at 17:00 using synthetic mobility data in Paris. The highest densities occur near major traffic corridors, where traffic emissions and population presence combine to create exposure hotspots. These hotspots result from the temporal alignment between elevated pollutant concentrations and mobile populations, illustrating why static residential distributions can misrepresent exposure.

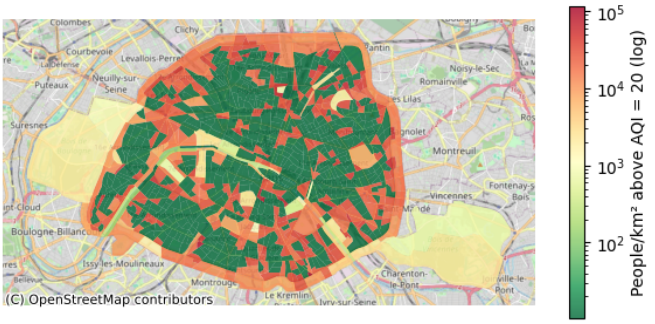


Fig. 2. Spatial distribution of population exposed to AQI > 20 at 17:00 in synthetic Paris data (log scale).

4. INDIVIDUAL EXPOSURE RESULTS AND POPULATION HETEROGENEITY

In the individual mobility regime, exposure is computed along trajectories, enabling full population exposure distributions and fine-grained variability analysis. Figure 3 shows the cumulative AQI-based indicator distribution over a synthetic population (100 histogram bins), revealing exposure variability induced by heterogeneous mobility patterns. For visualization purposes, AQI values derived from the concentration fields are used as an illustrative normalized indicator. The resulting cumulative AQI-based metric should therefore be interpreted as a relative exposure indicator rather than a physical dose. The distribution exhibits substantial inter-individual variability, with a standard deviation of $\sigma = 84.3$ AQI·h, indicating that heterogeneous mobility patterns can lead to markedly different cumulative exposure levels within the same urban environment.

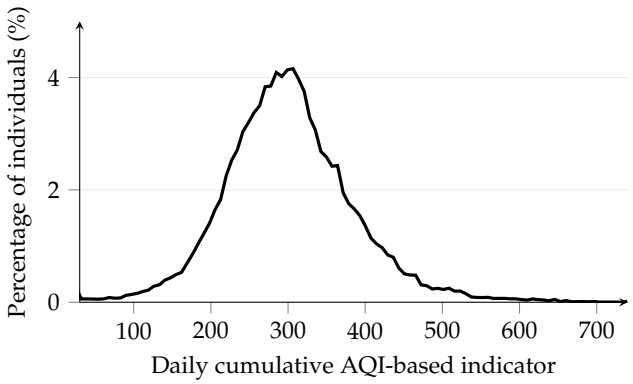


Fig. 3. Distribution of cumulative AQI-based indicator across a synthetic population, illustrating the variability captured by trajectory-based exposure modeling.

To further illustrate the framework capabilities, an example individual is analyzed in detail. This agent exhibits a cumulative daily AQI-based indicator value of 431 AQI·h and lies approximately at the 94th percentile of the indicator distribution shown in Fig. 3, illustrating a high-exposure individual. Figure 4 presents the temporal evolution of this individual's hourly exposure together with an illustrative threshold (AQI = 20).

The profile reveals a prolonged period of exposure above the threshold, with a marked increase during daytime followed by a gradual decrease in the evening. Beyond cumulative exposure, trajectory-based modeling also enables complementary indicators such as the duration spent above a given pollution threshold, which can distinguish individuals experiencing sustained moderate exposure from those experiencing shorter periods of higher exposure.

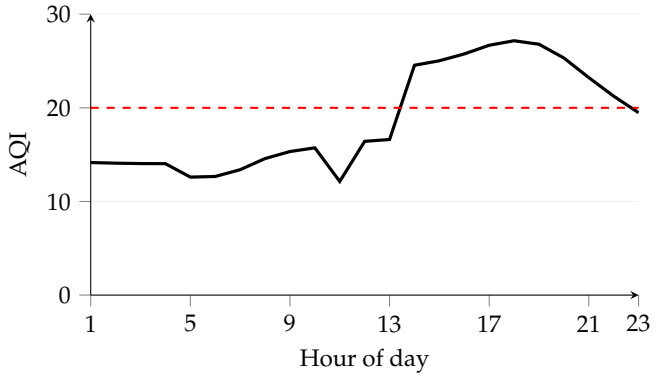


Fig. 4. Hourly AQI-based indicator profile of a synthetic agent with a cumulative value of 431 AQI·h. The dashed line indicates the illustrative threshold (AQI = 20).

Overall, these results demonstrate the added value of trajectory-based exposure modeling in capturing intra-population variability and identifying highly exposed individuals that cannot be resolved under aggregated mobility representations due to the loss of trajectory-level observability.

5. TOWARDS MODE-AWARE EXPOSURE MODELING

A natural extension of the framework is mobility-segment exposure modeling, where fine-scale concentration gradients along road networks can be resolved from trajectories.

In this context, transport-mode effects could be represented through a simple correction factor:

$$C_{\text{eff}} = k_{\text{mode}} \cdot C, \tag{3}$$

where k_{mode} accounts for mode-dependent micro-environment effects.

Future work will use this formulation to quantify mode- and segment-specific exposure contributions and assess mobility strategies aimed at reducing exposure during active travel.

6. CONCLUSION

We presented a mobility-aware framework for dynamic exposure assessment coupling mobility representations with high-resolution air-quality simulations. The framework supports aggregated and trajectory-based mobility data while preserving a common spatio-temporal coupling mechanism and producing exposure indicators adapted to each representation. It provides a unified basis for exposure assessment and future extensions toward mode-aware modeling.

REFERENCES

1. M. Tirico, V. Le Bescond, D. Sengelin, *et al.*, "Multi-scale comparison of modeling chains for traffic-related air pollutant emission assessment," Sustain. Cities Soc. **129**, 106354 (2025).
2. A. Can, P. Gastineau, B. Chaix, *et al.*, "Systemic approach for the assessment of the impact of urban mobility on exposures to environmental pollutions," in *TaP International Conference*, (2025).
3. M. Tirico, Y. Gao, D. Sengelin, *et al.*, "Modeling traffic-related air and noise pollution: Multi-criteria assessment case study around schools," Transp. Res. Part D: Transp. Environ. **149**, 105029 (2025).
4. G. Sabiron, S. Jay, L. Agelas, *et al.*, "From road traffic to virtual air quality sensors: An integrated mesoscopic modelling framework," Transp. Pollut. (TaP) Conf. (2025).

Social media-based measures of social capital attenuate economic disparities in extreme weather evacuations and recovery

APARIMIT KASLIWAL[1], CARLOS GUIRADO[1], AND MARTA C. GONZÁLEZ[1,2,3,*]

[1] Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA
[2] Department of City and Regional Planning, University of California, Berkeley, CA, USA
[3] Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[*] martag@berkeley.edu

Natural disasters impose severe costs on communities. In the United States alone, 2024 saw weather and climate disasters totaling US$182.7 billion in losses, including several hurricanes [1]. Behavioral responses to these disasters, such as evacuation patterns, vary across racial and socioeconomic groups, with departure and return times differing significantly by group [2]. Furthermore, outcomes for damaged buildings diverge sharply by income level: lower-income neighborhoods are less likely to rebuild, while higher-income areas not only rebuild but tend to improve relative to their pre-disaster state [3].

Measuring behavioral responses to disasters at scale has been made possible by the expansion of high-resolution digital mobility traces and social media data. Prior work has revealed spatiotemporal decay in mobility patterns in the vicinity of affected regions and in the days surrounding impact [4] and destination-dependent disruptions in visits to points-of-interest. Mobility patterns obtained through Location-Based Services (LBS) data have also revealed that education level drives disparities in behavioral response beyond income alone, as higher-educated individuals are more likely to stay indoors under increasing exposure to wildfire smoke [5]. Disaster periods also increase experienced segregation [6], as COVID-19 behavioral changes decreased the income diversity of urban encounters and amplified existing income-based segregation [7]. These findings motivate our use of social media data as a lens through which to quantify community connectivity and how it shapes behavioral responses to disasters across large populations and multiple events.

Resilient response during post-disaster recovery is enabled by the structure of the underlying social network, quantified through social capital which is defined as the resources embedded in social relationships [8]. Social capital is commonly categorized into bonding, bridging, and linking capital [9]. Bonding capital reflects ties within close-knit groups (e.g., family and neighbors), bridging capital reflects ties across different social groups (e.g., acquaintances and coworkers), and linking capital reflects ties to formal institutions and authorities (e.g., government). Large-scale digital social media data have made it possible to measure these dimensions of social capital at fine spatial resolution [10, 11]. Bridging capital becomes particularly consequential during disasters since cross-class ties enable access to information, resources, and opportunities that lie outside one's immediate social circle [8]. Furthermore, social ties and mobility patterns are deeply intertwined, as social relationships shape where people move while mobility creates opportunities to form and maintain social ties [12], a coupling that becomes particularly salient during disasters when evacuees tend to displace toward socially connected destinations [13].

Despite these recent advances, the role of social capital in shaping actual recovery outcomes after disasters, across large populations and multiple events, remains largely unexplored. Existing studies often rely on limited survey-based samples, census-derived static measures of social capital, or are restricted to single disaster events [14]. It remains an open question how network-based measures of social capital, derived from social media data, systematically stratify recovery outcomes that are now observable at scale through novel data sources. Here, we combine large-scale Location-Based Services (LBS) mobility data across five major hurricane events with social capital measures derived from online social network data to assess the role of social capital in post-disaster recovery. We use social capital data developed from Meta's social network, with aggregated metrics at the zip code level from the Social Capital Atlas [10].

This work makes three contributions to the existing literature. First, we introduce novel social media data [10] to understand disaster response and, motivated by the difference in data granularity, compare its aggregated social capital metrics with measures derived from individual-level Call-Detail Records (CDR). To construct the social network from CDR data, we build an undirected network in which each node is a subscriber with a successfully inferred home location, and an edge is included if the pair exchanged at least a given number of total interactions (voice calls plus SMS messages) over the full study period, spanning December 2018 through to June 2019. Second, our findings draw on five separate natural disaster events in the United States, and while the observed responses are broadly consistent across disasters, their strength and significance vary somewhat by event and displacement threshold, suggesting the results are not an artifact of any single disaster.

We also quantify two measures of bonding social capital: clustering, which captures the prevalence of triadic closure within neighborhoods, and support ratio, which measures the proportion of within-neighborhood friendships that share a third mutual friend within the same neighborhood. We center our analysis on economic connectedness (EC), a measure of bridging social capital that quantifies the prevalence of cross-class friendship ties between lower- and higher-socioeconomic-status individuals and has been shown to predict long-run economic upward mobility at the neighborhood level [10]. Factors such as education, social isolation, housing conditions and policies, and other place-specific characteristics may also influence evacuation behavior; however, our analysis focuses on EC as the primary measure of social capital. Because EC is strongly correlated with income, we residualize EC by regressing it on income and use the resulting residuals as a measure of cross-class connectedness that is orthogonal to income. This residual captures the variation in EC that cannot be explained by income. Specifically, we estimate the regression as in equation 1 below:

$$\mathrm{EC}_i = \alpha_0 + \alpha_1 \,\mathrm{Income}_i + \varepsilon_i, \tag{1}$$

where EC_i denotes the Economic Connectedness of zip code i as obtained from the Social Capital Atlas [10], ε_i denotes the

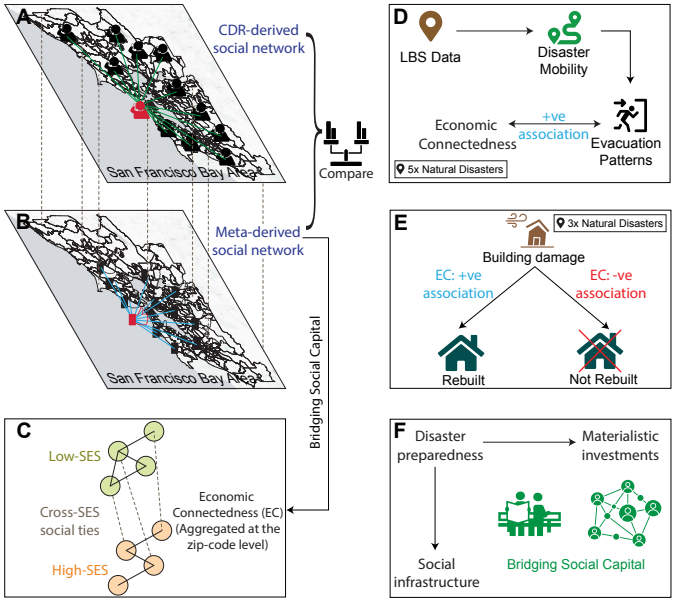


Fig. 1. From social networks to disaster resilience. A: Construction of social networks from Call Detail Records (CDR) based on calls and messages exchanged between anonymized users. **B:** Social network structure derived from Meta social media data, with network metrics compared across the two data sources in the San Francisco Bay Area. **C:** Economic Connectedness (EC), a measure of bridging social capital, quantifies the extent of cross-class social ties within a social network. **D:** Evacuation patterns derived from LBS data across five natural disasters are positively associated with EC at the zip code level. **E:** The rebuilding of damaged buildings following disasters is positively associated with EC. **F:** Beyond material investments in infrastructure, disaster preparedness may also entail the development of social infrastructure, such as parks and community spaces, that foster cross-class ties.

income-residualized measure of EC for zip code i, the standardized version of which we denote by $\widetilde{\mathrm{EC}}_i^{\mathrm{resid}}$. Third, this residualization isolates variation in EC that is not explained by baseline income differences, after which we find that higher EC is associated with elevated evacuation rates, with a positive association whose statistical significance varies across the five disasters and displacement thresholds we study.

To quantify the association between bridging social capital and evacuation behavior, we estimate the following specification at the zip code level, separately for each storm s and displacement threshold $\tau \in \{10, 50, 100\}$ km, capturing how this association varies across different spatial scales of evacuation:

$$\bar{y}_{i,\tau} = \beta_0 + \beta_\tau \widetilde{\mathrm{EC}}_i^{\mathrm{resid}} + \delta\, \mathrm{Income}_i + u_i, \qquad (2)$$

where $\bar{y}_{i,\tau}$ is the zip code evacuation rate, $\widetilde{\mathrm{EC}}_i^{\mathrm{resid}}$ is the standardized income-residualized EC, and Income_i is median household income. Specifically, a user is classified as an evacuee if detected away from home for at least eight hours during the nighttime within the three-day post-impact window. We find that EC is consistently positively associated with evacuation rates, with significant coefficients at the 5% level in 8 of the 15 reported storm-threshold combinations. The variation in the significance level of these coefficients might potentially be because of factors specific to certain disasters or due to limitations of the data itself.

Interestingly, $\widetilde{\mathrm{EC}}^{\mathrm{resid}}$ is positively associated with rebuilding and negatively associated with abandonment of damaged buildings post-diaster Figure 1E, explaining a share of variance in recovery outcomes comparable to that of median household income. For long-term recovery, we estimate a weighted least squares (WLS) model for each storm, with observations weighted by the number of damaged buildings per zip code:

$$\tilde{y}_i = \beta_0 + \beta_1 \widetilde{\mathrm{Income}}_i + \beta_2 \widetilde{\mathrm{EC}}_i^{\mathrm{resid}} + u_i, \qquad (3)$$

where $\tilde{y}_i$ is the standardized recovery outcome, $\widetilde{\mathrm{Income}}_i$ is standardized median household income, and $\widetilde{\mathrm{EC}}_i^{\mathrm{resid}}$ is the income-residualized EC measure. External factors not captured in this specification, such as changes in property ownership, may also contribute to observed differences in long-term recovery.

From a policy perspective, while risk and hazard maps capture physical exposure, they miss an important social dimension of urban resilience. The geospatial distribution of bridging social capital could therefore complement existing risk stratification frameworks and inform targeted investments in social infrastructure in communities where cross-class ties are systematically sparse [15]. We note three limitations of our work: the CDR network is geographically constrained to the San Francisco Bay Area, and social capital metrics do not capture potential changes in social ties during or after the extreme weather event [8]. Furthermore, the analysis is conducted at zip code resolution and our findings are associational: we do not uncover the underlying mechanisms through which cross-class ties are leveraged to enable evacuation or rebuilding post-disasters.

REFERENCES

1. NOAA National Centers for Environmental Information, "U.s. billion-dollar weather and climate disasters," (2025). Dataset and annual summary statistics.
2. H. Deng, D. P. Aldrich, M. M. Danziger, *et al.*, "High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns," Humanit. Soc. Sci. Commun. **8** (2021).
3. T. Huang, C. Zanocco, Z. Wang, *et al.*, "Built environment disparities are amplified during extreme weather recovery," Nature **648** (2025).
4. W. Li, Q. Wang, Y. Liu, *et al.*, "A spatiotemporal decay model of human mobility when facing large-scale crises," Proc. National Acad. Sci. **119** (2022).
5. C. Berestycki and M. K. Chen, "Behavioral responses to wildfire smoke: Insights from smartphone location data," Proc. National Acad. Sci. **123** (2026).
6. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. Commun. **12** (2021).
7. T. Yabe, B. G. B. Bueno, X. Dong, *et al.*, "Behavioral changes during the covid-19 pandemic decreased income diversity of urban encounters," Nat. Commun. **14** (2023).
8. D. P. Aldrich, *Building Resilience: Social Capital in Post-Disaster Recovery* (The University of Chicago Press, Chicago, 2012).
9. D. Kyne and D. P. Aldrich, "Capturing bonding, bridging, and linking social capital through publicly available data," Risk, Hazards amp; Crisis Public Policy **11**, 61–86 (2019).
10. R. Chetty, M. O. Jackson, T. Kuchler, *et al.*, "Social capital i: measurement and associations with economic mobility," Nature **608** (2022).
11. M. Bailey, R. Cao, T. Kuchler, *et al.*, "Social connectedness: Measurement, determinants, and effects," J. Econ. Perspect. **32** (2018).
12. J. L. Toole, C. Herrera-Yaqüe, C. M. Schneider, and M. C. González, "Coupling human mobility and social ties," J. The Royal Soc. Interface **12**, 20141128 (2015).
13. V. Raipat and T. Yabe, "Social homophily is associated with evacuation destination choice and long-term displacement decisions after the marshall fire in the united states," Humanit. Soc. Sci. Commun. (2026).
14. V. Raipat, D. Aldrich, and T. Yabe, "Disaster-induced behavioral change restructures social networks toward bonding ties," (2026).
15. T. Fraser, O. Awadalla, H. Sarup, and D. P. Aldrich, "A tale of many cities: Mapping social infrastructure and social capital across the united states," Comput. Environ. Urban Syst. **114**, 102195 (2024).

Prediction of Infection Risk in an Epidemic Contact Network via Temporal Line Graph Representations

MATHIEU GILLES[1,*], ZAHER MOUSSAWEL[1,*], MIGUEL NUNEZ–DEL-PRADO[2], HUGO ALATRISTA-SALAS[3], AND VINCENT GAUTHIER[1]

[1] SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
[2] The World Bank, Washington D.C., 20433, U.S.A.
[3] De Vinci Research Center, De Vinci Higher Education, Paris, France
[*] Both authors contributed equally.

1. INTRODUCTION

This study describes a machine learning pipeline for predicting individual COVID-19 infection risks using digital contact tracing (DCT) data from Peru. The core innovation is the Time-Decayed Line Graph (TDLG) representation, which transforms temporal contact patterns into network embeddings. Combined with Graph Neural Networks, the approach achieves a 20-fold improvement over random baseline predictions, despite extreme class imbalance (268:1 ratio) and sparse data adoption (∼5% app penetration).

The primary objective of this study is to identify individuals at high risk of COVID-19 infection before clinical diagnosis occurs, enabling public health officials to intervene with targeted isolation, testing, or vaccination campaigns [1]. The research addresses three critical technical challenges that fundamentally constrain infection risk prediction in real-world contact tracing scenarios. First, the study processes large-scale sparse networks with 1.15 million nodes and 12.4 million contact edges from the Peruvian mobile application "Perú en tus manos," demanding computationally efficient methods capable of processing massive graphs without excessive memory requirements. Thus, the 12 M^2-scale adjacency matrix exists implicitly as a factored product and is only accessed through matrix-vector products during the Lanczos iterations. Second, epidemic transmission patterns are inherently non-stationary, with model performance varying dramatically across different phases of disease spread.

This temporal non-stationarity requires validation strategies that carefully account for temporal dependencies and precludes standard cross-validation approaches that assume independent samples. The dataset spans from April 5 to June 5, 2020, during the early phase of Peru's COVID-19 epidemic, and includes approximately 4,308 confirmed positive cases among the 1.15 million app users.

2. METHODOLOGY

In the present work, we develop a four-stage machine learning pipeline.

First, we adapt the *Time-Decayed Line Graph (TDLG)* representation [2] to epidemic contact networks. Contacts shorter than 15 minutes are discarded before TDLG construction and feature extraction. Each remaining contact event is represented as a node in the line graph. Two events i and j sharing an individual are connected according to

$$A_{ij} = w_i w_j \exp\left(-\frac{\Delta t_{ij}^2}{2\sigma_t^2}\right), \qquad w_i = \text{sigmoid}\left(\frac{d_i - 900}{300}\right),$$

where $\Delta t_{ij} = |t_i - t_j|$ is measured in seconds, d_i is the duration of contact i, and $\sigma_t = 7200\,\text{s}$ (2 h). Event pairs separated by more than $3\sigma_t = 6\,\text{h}$ are discarded.

A truncated eigendecomposition of the sparse weighted adjacency matrix ($k = 16$) yields contact-level spectral embeddings. User embeddings are then obtained by averaging incident contact embeddings using the duration-dependent weights w_i. The line graph is never constructed densely: candidate pairs are generated only among contacts sharing a user within the 6-hour temporal support and stored as a sparse CSR matrix, with at most 500 chronologically ordered contacts considered per user.

At scale, the TDLG is materialized sparsely: a representative construction contained 50.8 million line-graph nodes and 713.7 million non-zero entries (density 5.53×10^{-7}). Experiments ran on two Intel Xeon Gold 6226R CPUs with 755 GB RAM and an NVIDIA Quadro RTX 4000 (8 GB).

The model combines two feature categories: 16 spectral features derived from the TDLG eigendecomposition, capturing the structural properties of temporal contact patterns, and 11 manual graph features extracted from the dataset, including degree, total and maximum contact duration, number of distinct contact days, contact-time regularity, contact-duration coefficient of variation, first-contact time, approximate betweenness centrality, average exposure to confirmed cases, and the diagnosed-neighbor rate. In the third stage, the GNN applies two parallel graph-convolution passes on the original and transposed edge index, then concatenates the outputs to obtain a node embedding that is symmetric with respect to contact direction on the undirected network. Different Graph Neural Network variants (GCN [3], GraphSAGE [4], and GATv2 [5]) are evaluated; the choice of GNN backbone had a comparatively limited effect in the reported experiments. Finally, *temporal validation* was performed using a 7-fold sliding-window cross-validation strategy [6], with a training window of 15 consecutive days, a test window of 7 days, and a stride of 7 days. An individual is labeled positive when their recorded diagnosis date follows their first observed contact within the analysis window. The available outcome is the recorded diagnosis date rather than the unobserved infection time, so infection-to-diagnosis delays are not modeled explicitly.

3. RESULTS

This section reports Average Precision (AP), area under the ROC curve (AUC), and F_1 performance. We begin by examining how temporal encoding of contact networks influences model performance, then evaluate how epidemic phase and contact characteristics affect prediction accuracy. Finally, we report the quantitative metrics of our best-performing model configuration.

Temporal Encoding. Classical baselines on the manual features alone barely exceed the 0.37% prevalence: logistic regression reaches AP = 0.43% and LightGBM 0.33%; a GCN deprived of TDLG embeddings performs equally poorly (AP = 0.58%), confirming that the convolution operator alone cannot extract a usable signal from static features under this level of imbalance. Adding *TDLG* embeddings to the GCN with the six basic graph

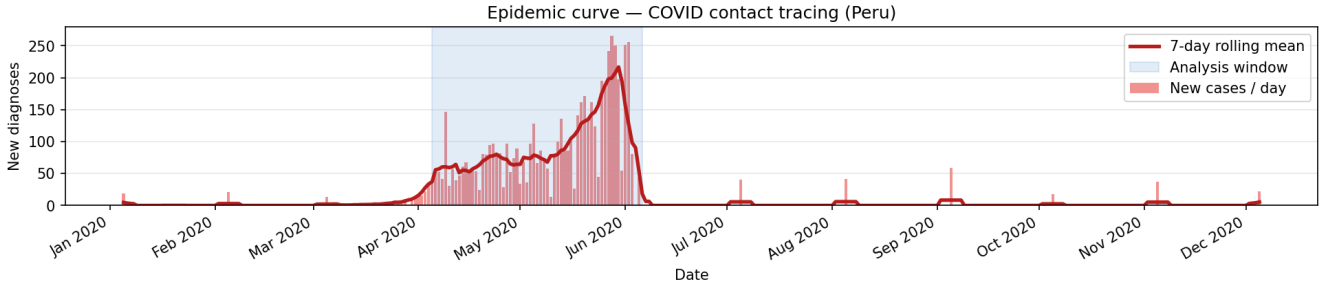


Fig. 1. Daily confirmed cases in the Peruvian DCT dataset. Shading marks the April 5–June 5 analysis window; *average precision* AP peaks at 14.8% on May 4–11 and falls to 0.2% in the final fold, where only 15 positive cases remain.

features (no 15-min contact filter) brings AP to 5.71% (15.4× lift); applying the 15-minute filter raises AP to 6.83% (a 19.6% relative gain), and combining *TDLG* with the full set of 11 manual features yields the best configuration at AP = 7.52% (20.3× lift over prevalence). Switching the GCN backbone to GraphSAGE (AP = 6.98%) or GATv2 (AP = 6.21%) changes AP by less than 0.8 pp, within the dispersion of the cross-validation procedure. These small differences suggest that once the *TDLG* representation provides the necessary temporal signal, the specific mechanism by which the neural network processes this signal becomes less critical than initially expected.

Concerning the **Epidemic Phase Dependency**, the model varies sharply across the seven temporal folds, revealing a critical dependency on epidemic phase and transmission dynamics (Figure 1). AP peaks at 14.8% on the week of May 4–11, near the local maximum of community transmission, and degrades to 4.7% during the declining phase in mid-May (test window May 18–25). Most strikingly, in the final test window (June 1–5) only 15 confirmed positive cases survive, and AP drops to 0.2%, approaching baseline random performance. This indicates that the predictive signal requires a sufficient density of active transmission chains, and that any deployment should include phase-aware calibration.

For the **Contact Duration Filtering**, we found that a 15-minute threshold significantly improved model performance, increasing Average Precision by approximately 20%. This finding has both technical and epidemiological significance. From a signal processing perspective, removing short contacts reduces noise and allows the model to focus on more meaningful exposure patterns. From an epidemiological perspective, this result aligns with World Health Organization guidance indicating that meaningful exposure duration is approximately 15 minutes or longer. Finally, the model provides a tunable risk score whose

Configuration	AP (%)	AUC (%)	F1 (%)
Logistic Regression	0.43	58.5	1.5
LightGBM	0.33	50.3	0.7
GCN (no TDLG)	0.58	58.5	1.3
GCN + TDLG, $d_{\min} = 5$ min	5.71	73.9	14.9
GCN + TDLG, $d_{\min} = 15$ min	6.83	73.6	16.3
GraphSAGE + TDLG	6.98	73.2	14.0
GATv2 + TDLG	6.21	74.8	15.8
GCN + TDLG + 11 features (Ours)	**7.52**	**74.1**	**16.5**

Table 1. Ablation over seven sliding-window folds. Random-baseline AP = 0.37%. TDLG-based configurations substantially outperform classical baselines, while GNN-backbone differences remain small.

decision threshold can be adapted to operational constraints. At the max-F_1 threshold ($\tau = 0.97$) the model emits 393 alerts to detect 127 true cases (precision P = 32.3%, recall R = 14.3%), a 104× lift over the test-set prevalence (0.31%). An operationally usable zone spans recall 20–30%: 1,893 alerts at R = 20% (P = 9.4%, ×30 enrichment) or 14,378 alerts at R = 30% (P = 1.9%, ×6). Beyond R = 30% the precision-recall curve collapses, reaching R = 40% requires 28,560 alerts for 356 cases. The decision threshold is therefore not a hyperparameter to be tuned during training but an operational lever, set by the testing capacity that public-health agencies are willing to commit.

4. CONCLUSION

This research shows that COVID-19 infection risk can be predicted from temporal contact-network structure despite extreme class imbalance and incomplete observations. The full TDLG-based pipeline achieves a 20.3× AP lift over prevalence, while differences between GNN backbones remain comparatively small. Performance nevertheless varies strongly with epidemic phase. At the max-F_1 operating point, precision reaches 32.3% at 14.3% recall, a 104× enrichment over test-set prevalence, supporting capacity-aware prioritization of public-health interventions.

Generalizability remains limited by the single-country, early-COVID-19 setting and ∼5% app penetration, which may induce demographic, socioeconomic, and geographic selection biases. Validation across other populations, diseases, and DCT systems, together with subgroup analyses and real-time feature reconstruction, is therefore needed before deployment. In addition, the TDLG embeddings needs to be tested to determine whether temporally-leaky features are responsible for the marginal gain.

REFERENCES

1. C. Li, V. Gauthier, M. Nunez-del Prado, *et al.*, "Assessing the usefulness of digital contact tracing using real-world contact data," Sci. Reports (2026).
2. S. Chanpuriya, R. A. Rossi, S. Kim, *et al.*, "Direct embedding of temporal network edges via time-decayed line graphs," arXiv preprint arXiv:2210.00032 (2022).
3. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*, (2017).
4. W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, (2017).
5. S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" in *International Conference on Learning Representations (ICLR)*, (2022).
6. C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," Comput. Stat. & Data Anal. **120**, 70 (2018).

Empirical Trajectory Sparsity Biases Mobility-Informed Epidemic Modeling Outcomes and Parameter Estimates

FEDERICO DELUSSU[1], FRANCISCO BARRERAS[2], YUAN LIAO[1], DUNCAN J. WATTS[2], AND LAURA ALESSANDRETTI[1]

[1] *Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark*
[2] *Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA*
[*] *Corresponding author: lauale@dtu.dk*

1. INTRODUCTION

GPS mobility data are increasingly used in epidemic modeling. By allowing the construction of co-location networks and population flows, these data provide richer insights into the relationship between mobility and epidemic spreading [1–3]. These trajectories typically exhibit high temporal sparsity because data collection is opportunistic and tied to phone use [4, 5]. Despite growing awareness of this limitation, mobility-informed epidemic models have largely overlooked the analysis and treatment of biases derived from such sparsity, raising concerns about the robustness of downstream inferences. In this paper we introduce a principled framework to quantify the impact of trajectory sparsity on key epidemic modeling outcomes across different levels of data missingness. We improve upon previous "data amputation" frameworks used in related work [6, 7] by introducing empirical sparsity patterns on highly complete trajectories, allowing the comparison of biased statistics against their ground-truth values. Our approach leverages a highly complete dataset that exhibits both near-complete and sparse GPS trajectories. In addition to evaluating the effect of sparsity, we propose and evaluate correction methods based on inverse probability weighting of contact durations before epidemic model calibration, which are shown to reduce bias and parameter misspecification.

2. DATA AND METHODS

Our experiments rely on adding empirical temporal sparsity to near-complete anonymized GPS trajectories from approximately 1,000 students at the Technical University of Denmark (DTU) campus [8] exhibiting both fixed-interval measurements for some users, which provide long highly complete sub-trajectories, while other users exhibit the sparse opportunistic sampling typical of commercial datasets. We define a trajectory as "nearly-complete" when fewer than 5% of hours over the study period lack data. Using this definition, we selected 363 near-complete trajectories for a period of 28 days, from 8 Feb to 7 Mar 2014, while sparse gap sequences were sampled from trajectories in the corresponding sparsity ranges 10–20%, 20–30%, 30–40%, 40–50%, and 50–60% of hours without data. In this setup, near-complete trajectories provide a baseline for key epidemic modeling outcomes (EMOs), while sparse trajectories provide realistic missingness patterns that we impose on the baseline to measure bias under prescribed percentages of sparsity. These EMOs are the probability of an outbreak (defined as >5% of the population infected), $p_{\texttt{outbreak}}$, the total infected population ($I_{\texttt{tot}}$), the peak incidence $I_{\texttt{peak}}$, its timing $t_{\texttt{peak}}$, and the day of last infection (or end of study period) $t_{\texttt{tot}}$.

Due to constraints on the duration of the study period, we use a simple stochastic *Susceptible-Infected-Removed* (SIR) model with ad-hoc ground-truth parameters chosen to produce an epidemic regime of moderate intensity ($\overline{R}_0 = 1.92$). This compartmental model simulates spread over daily co-location networks, weighted by the minutes each pair of users spend in the same Geohash-8 tile, or adjacent. For calibration experiments, we use the ground-truth caseload curve as the target and minimize RMSE via Bayesian kernel fitting.

Our "sparsification" procedure introduces missingness by sampling randomly, with replacement, trajectories with a target level of sparsity. Then, for each hour without data in the sampled sparse trajectory, we remove the records of the near-complete trajectory for that same hour. This approach results in an empirical data missingness mechanism which can capture realistic gap distributions and correlations with time in ways that missingness completely at random (MCAR) cannot. For each sparsity level, we sample 50 realizations of the missingness patterns and, within each, run 100 epidemic simulations, yielding an ensemble of 5,000 epidemic curves per configuration.

Finally, because calibration to incidence data is standard in epidemic modeling, we ask whether it is sufficient to address trajectory-sparsity bias or if it can only recover target outcomes at the cost of misspecifying transmission parameters. We therefore compare two modeling settings: *oracle* models, which fix the epidemic parameters to isolate the effect of the contact network, and *calibrated* models, which fit the parameters to the target caseload curve. To improve on calibration alone, we also evaluate the application of a well-established method from survey nonresponse, Inverse Probability Weighting (IPW), used to debias the epidemic network before calibration. The IPW correction to the daily contact duration between users i and j is given by:

$$\widehat{C}_{ij}(t) = \sum_h C_{ij}(h) \cdot \frac{1}{1-p_i^h} \cdot \frac{1}{1-p_j^h}, \qquad \textbf{(1)}$$

where $C_{ij}(h)$ are the observed minutes in contact during hour h, and p_i^h is the probability that a record is missing for user i during hour h. These probabilities are estimated via maximum likelihood in 48 hour-by-week-period buckets: 24 hours crossed with weekday/weekend status.

3. RESULTS

We quantify how trajectory sparsity biases epidemic outcomes describing outbreak occurrence and epidemic intensity. Under the oracle model, where epidemic parameters are fixed at their ground-truth values, sparsity systematically attenuates simulated spread (Fig. 1a–b). In the 50–60% sparsity regime, $I_{\texttt{tot}}$ falls by over 70%, $I_{\texttt{peak}}$ by nearly 80%, and $p_{\texttt{outbreak}}$ by about 33% relative to the near-complete baseline. These effects follow directly from degradation of the contact network: at 50–60% sparsity, average daily contacts fall from 65,766 to 23,309 and average contact duration falls by 23%.

We then ask whether this bias is mitigated by two steps commonly available to modelers: IPW rescaling of observed contacts and calibration to an incidence curve. IPW, a classical survey-nonresponse method, gives more weight to observations with lower probability of being observed—under an assumed missingness distribution. Calibration does not directly correct trajectory sparsity, but it is part of the standard mobility-informed

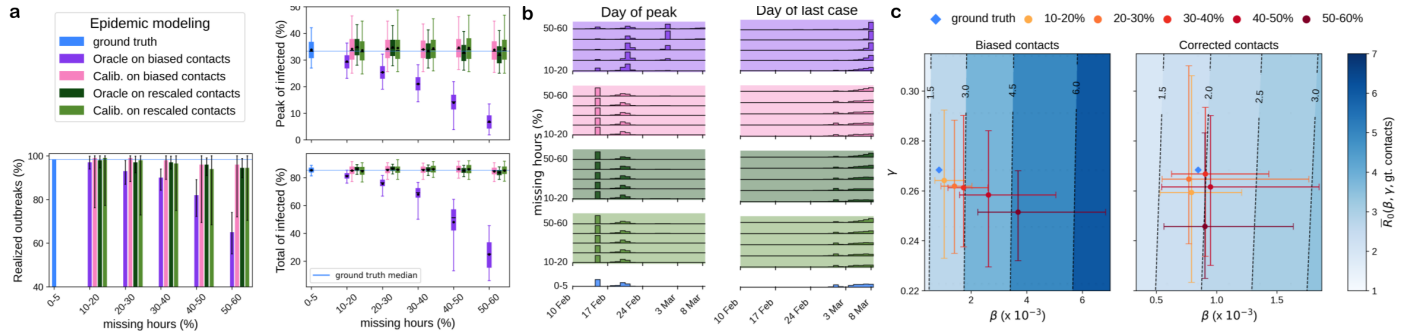


Fig. 1. Debiasing approaches mitigate errors in epidemic simulations. (a) Realized outbreak probability (bottom left), peak daily infections (top right), and total infections (bottom right). Boxplots display the IQR with whiskers marking the 95% CI. Vertical lines denote the median and dots indicate the mean. **(b)** Occurrence of peak and last day of incidence curve. **(c)** The ground-truth simulation parameters (blue diamond) are compared to the estimated parameters derived from calibration on biased contacts (left) and rescaled contacts (right). Dots indicate the median and error bars denote the 95% CI over missingness draws. The underlying heatmap displays the average $\overline{R}_0$ computed from the ground-truth contacts over the (β, γ) parameter space.

epidemic-modeling pipeline and could, in principle, absorb sparsity-induced error in fitted caseload outcomes. Both approaches correct much of the bias in $I_{\texttt{tot}}$ and $I_{\texttt{peak}}$, although $p_{\texttt{outbreak}}$ remains underestimated at high sparsity, consistent with the fact that weighting observed contacts cannot restore entirely missing edges (Fig. 1a–b). Calibration can therefore recover the target caseload curve by absorbing trajectory-sparsity bias into the fitted transmission parameters, allowing the bias in the mobility inputs to go unnoticed. Figure 1c shows the extent to which this recovery comes from inflating the transmission parameters: in the 50–60% regime, calibration on uncorrected contacts yields $\overline{R}_0 = 4.77$, compared with a ground truth of 1.92. Applying IPW before calibration keeps the inferred reproduction number much closer to ground truth, $\overline{R}_0 = 2.11$ (Fig. 1c).

4. DISCUSSION

A rigorous treatment of trajectory data-missingness bias has been largely overlooked in recent human mobility literature. This gap might be due to two limitations. First, many epidemiological studies integrating human mobility data begin from already-aggregated inputs, in which the sparsity in the population contact patterns conflates with the sparsity induced by missing trajectory data [1, 9–11]. This conflation makes sparsity bias correction infeasible, since inputs requiring correction may be indistinguishable from inputs with no missingness-induced bias. It also limits studies quantifying bias by pruning or adding edges to already aggregated flow networks, and observing changes in epidemic outcomes [12, 13].

Our study addresses this gap by isolating and quantifying the biases in EMOs stemming from trajectory sparsity. We do so by adapting established methods from survey nonresponse: using data amputation to generate sparsified counterfactuals that can be compared against a near-complete baseline, and IPW as a viable method to correct such biases. The ground-truth values for the reproduction number and other parameters are typically not known a priori, which is a limitation of our "oracle" simulations. Nonetheless, these results have serious implications for practitioners who calibrate models on one mobility sample and apply the fitted parameters to another, or who borrow transmission parameters from the clinical literature. Namely, simulations on sparser (than training) data will considerably underestimate the intensity of the epidemic in the simulations.

The oracle and calibrated analyses therefore address different practical settings. Oracle models isolate the consequences of trajectory sparsity when epidemic parameters are held fixed, whereas calibrated models represent the standard case in which parameters are inferred from incidence data. Our results show that calibration can preserve fitted epidemic outcomes by absorbing contact-network bias into the transmission parameters, while applying IPW before calibration reduces this misspecification considerably.

Our results are limited by the size and composition of the near-complete DTU baseline, the short simulation window, and the use of GPS-derived co-location as an approximation to true proximity. Nonetheless, we believe that the core mechanism at play—sparsity degrading the connectivity of the network in ways that underestimate epidemic spread—likely generalizes, and corrections are warranted for more realistic applications. The full manuscript extends these results by showing that randomized missingness assumptions can underestimate the magnitude of sparsity-induced bias, and by demonstrating the correction on a separate anonymized commercial GPS sample provided by Cuebiq. In particular, a Cuebiq sample from Ithaca, NY, was analyzed. Although high sparsity prevented the near-complete benchmark to replicate the results, applying the same correction method reduced the variance in total-infected predictions: the 95% CI narrowed from ((1.2, 11.3)%) to ((1.3, 8.7)%), and the IQR from ((4.4, 7.8)%) to ((4.3, 6.8)%). Future work should extend weights-based correction to aggregated mobility inputs, where privacy safeguards often preclude direct work with individual-level trajectories.

Funding This work was supported by NSF award 2341932.

REFERENCES

1. S. Chang, E. Pierson, P. W. Koh, *et al.*, Nature **589**, 82 (2021).
2. C. O. Buckee, S. Balsari, J. Chan, *et al.*, Science **368**, 145 (2020).
3. F. Barreras and D. J. Watts, Nat. Comput. Sci. pp. 1–14 (2024).
4. S. Bähr, G.-C. Haas, F. Keusch, *et al.*, Soc. Sci. Comput. Rev. **40**, 212 (2022). Publisher: SAGE Publications Inc.
5. D. McCool, P. Lugtig, and B. Schouten, Transportation **51**, 297 (2024).
6. I. Barnett and J.-P. Onnela, Biostatistics **21**, e98 (2020).
7. C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, Science **327**, 1018 (2010).
8. A. Stopczynski, V. Sekara, P. Sapiezynski, *et al.*, PloS one **9**, e95978 (2014).
9. Y. Kang, S. Gao, Y. Liang, *et al.*, Sci. data **7**, 1 (2020).
10. M. K. Savi, A. Yadav, W. Zhang, *et al.*, PLOS Digit. Health **2**, e0000233 (2023).
11. F. Vanni, D. Lambert, L. Palatella, and P. Grigolini, Sci. reports **11**, 23286 (2021).
12. R. Gallotti, D. Maniscalco, M. Barthelemy, and M. De Domenico, Commun. Phys. **7**, 421 (2024).
13. F. Schlosser, V. Sekara, D. Brockmann, and M. Garcia-Herranz, arXiv preprint arXiv:2112.12521 (2021).

Mobility Networks Reveal Associations between Dynamic Food Environment and Chronic Cardiometabolic Diseases across U.S. Cities

YUZHOU CHEN[1], BAOSHEN GUO[4], SHANTING CHEN[3], CHANG ZHAO[2,*], AND SHENHAO WANG[1,*]

[1] Department of Urban and Regional Planning, University of Florida, Gainesville, FL 32611
[2] Agronomy Department, University of Florida, Gainesville, FL 32611
[3] Department of Psychology, University of Florida, Gainesville, FL 32611
[4] Singapore-MIT Alliance for Research and Technology
[*] shenhaowang@ufl.edu; changzhao@ufl.edu

1. INTRODUCTION

Residential neighborhood indicators are widely used to analyze urban health, but they only capture static residential status. Large-scale mobility studies have demonstrated that the dynamic mobility data can capture inequalities in physical activities [1], experienced socioeconomic segregation [2–4], and public health disparities across the United States [5–8]. These studies demonstrate a shift from static analysis based on residential locations toward a dynamic approach using human mobility data to understand economy, society, and health.

However, a key measurement challenge remains: mobility data can represent multiple dimensions of daily exposure, from residential neighborhoods and visited destinations to activity spaces and socially stratified environments. Existing studies often examine these dimensions separately, leaving unclear which mobility-derived measures are most informative for explaining physical and mental health outcomes. Meanwhile, existing studies do not fully use the advantage of fine-grained mobility data to capture people-place dynamics that connect individuals with the multiple places shaping their daily exposures. Therefore, this study uses anonymized phone trajectory data to link home tracts, visited destinations, and experienced urban contexts. We ask whether mobility-derived exposure features provide additional predictive information for tract-level cardiometabolic prevalence beyond residential food-environment measures, and which interpretable exposure channels recur across outcomes.

2. DATA AND METHODS

The empirical design uses census tracts as the unit of analysis. Health outcomes are CDC PLACES 2025 release census-tract crude prevalence estimates for adult obesity, diabetes, high blood pressure, and coronary heart disease, derived primarily from 2023 BRFSS data through small-area estimation. Mobility features are derived from Dewey mobility data observed from January 1 to April 30, 2025 and aggregated from home census tracts to visited places and destination tracts. Given the temporal mismatch and ecological design, the analysis is interpreted as predictive association rather than short-term causal health effects.

We design four mobility-based experience channels. M1 captures activity-space extent, including out-of-tract ratio, unique tracts visited, mean travel distance, and radius of gyration. M2 captures destination behavior, including grocery, fast-food, convenience, park, fitness, and nightlife visit shares. M3 captures experienced environmental context, including experienced healthy-food density, unhealthy-food density, food quality, and physical-activity-related density. M4 captures mobility-mediated socioeconomic exposure, including income exposure, upward-mobility ratio, mobility segregation, and individual mobility segregation. Together, these features distinguish where residents travel, what destinations they visit, what environments they experience, and what socioeconomic contexts they encounter.

For each health outcome y_i in tract i, we estimate a residential baseline model and a family of mobility-augmented models:

$$y_i = \beta_0 + X_i^S \beta_S + X_i^M \beta_M + \gamma_c + \epsilon_i, \quad (1)$$

where X_i^S denotes static residential features, X_i^M denotes one or more M1–M4 mobility-exposure blocks, and γ_c denotes city fixed effects. Continuous predictors are standardized using the training split only, and predictive performance is evaluated by held-out tract-level R^2. We compare the residential baseline, individual mobility blocks, the full M1–M4 specification, raw POI-category mobility models, and random-forest robustness models. Because nearby tracts may share mobility patterns and health prevalence, random tract-level validation may overestimate spatial generalization; spatially blocked and leave-one-city-out validation are left for future work.

3. RESULTS

We selected four models to summarize the overall experimental performance, as shown in Table 1. Model A uses residential food-related POI densities as a static baseline. Model B uses the structured M1–M4 mobility metrics in an OLS model. Model C uses raw POI-category mobility visit shares, with weekday/weekend and day/night features merged into all-time category-level exposures. Model D uses the M1–M4 mobility metrics in a random-forest model. Across the four focal outcomes, mobility-augmented models outperform the static residential baseline. Model B increases test R^2 from 0.36 to 0.72 for obesity, 0.26 to 0.69 for diabetes, 0.25 to 0.57 for high blood pressure, and 0.22 to 0.40 for Coronary heart disease (CHD). The raw category-level mobility model performs similarly for obesity and high blood pressure, while the random-forest model obtains the highest predictive performance.

The coefficient heatmaps identify two patterns in the association between mobility-based exposure and chronic health outcomes (Figs. 1 and 2). In Fig. 1, food-related mobility measures are among the largest predictors across obesity, diabetes, and high blood pressure. Grocery visit share is positively associated with these outcomes, while food visit quality and experienced healthy food density show negative coefficients. Measures of activity-space extent and socioeconomic exposure also contribute substantially. "Out-of-tract travel" and "income exposure" are negatively associated with several outcomes, whereas mean travel distance and individual mobility segregation show positive associations, especially for high blood pressure. These results suggest that mobility captures both behavioral destination choices and broader socioeconomic exposure contexts.

Table 1. Test R^2 for representative models and focal cardiometabolic outcomes. Values are rounded from the pooled three-city model comparison with city fixed effects.

Model	Obesity	Diabetes	High blood pressure	CHD
A: Neighborhood POI (density)	0.36	0.26	0.25	0.22
B: M1–M4 mobility metrics (OLS)	0.72	0.69	0.57	0.40
C: Full mobility based POI (share&dwell)	0.70	0.57	0.60	0.32
D: M1–M4 mobility metrics (RF)	0.82	0.77	0.70	0.54

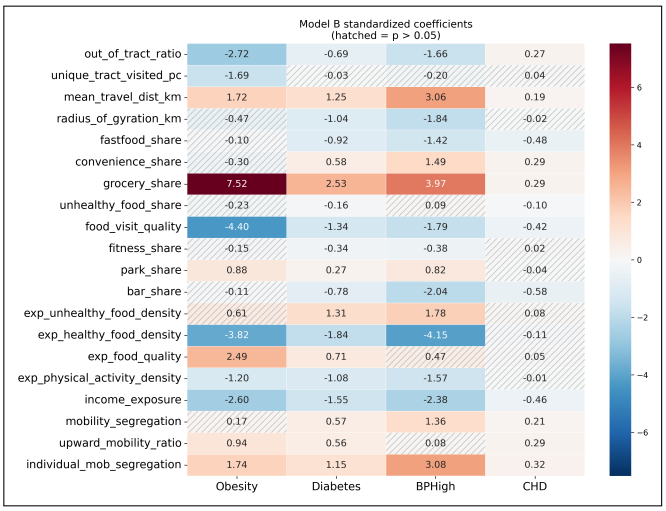


Fig. 1. Standardized OLS coefficients from Model B, the structured M1–M4 mobility framework. Rows are ordered by activity-space extent (M1), destination behavior (M2), experienced environmental context (M3), and mobility-mediated socioeconomic exposure (M4). Hatched cells indicate $p > 0.05$.

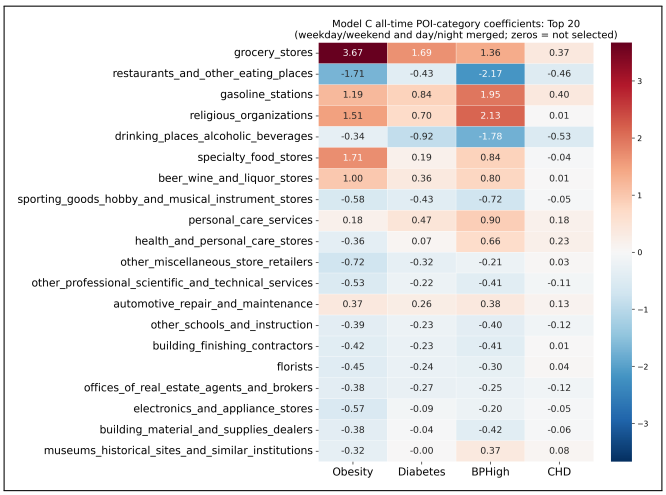


Fig. 2. LASSO robustness check from Model C using all-time raw POI-category visit shares. Weekday/weekend and day/night mobility features are merged by POI category; zero cells indicate features not selected for a given outcome.

The results from LASSO model are shown in Fig. 2. It demonstrates that grocery stores, restaurants, and other eating places, gasoline stations, religious organizations, drinking places, and specialty food stores are among the most informative POI categories for health outcomes. The strongest selected coefficients are concentrated for obesity and high blood pressure, while CHD exhibits weaker and less stable coefficients. The heterogeneity observed across different health outcomes suggests that mobility features should not be simplistically interpreted as a single pathway of health exposure. Rather, it encompass multiple dimensions of urban exposure, including food access, destination-related behaviors, and interactions through socioeconomic strata.

4. DISCUSSION

This study shows that mobility-derived exposure features contain substantial predictive information for health outcomes. The comparison suggests that residential context remains informative, while mobility features add signals about behavioral destinations, experienced environments, and socioeconomic exposure. The results should be interpreted as cross-sectional ecological associations rather than causal effects: mobility patterns may reflect omitted within-city characteristics such as age structure, employment, car ownership, income sorting, or residential segregation. Future work will extend the analysis to additional cities, improve temporal alignment between mobility and health measures, and evaluate spatially blocked and leave-one-city-out validation.

5. DATA AVAILABILITY STATEMENT

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. T. Althoff, R. Sosic, J. L. Hicks, *et al.*, "Large-scale physical activity data reveal worldwide activity inequality," Nature **547**, 336 (2017).
2. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. Commun. **12**, 4633 (2021).
3. S. Athey, B. Ferguson, M. Gentzkow, and T. Schmidt, "Estimating experienced racial segregation in us cities using large-scale gps data," Proc. National Acad. Sci. **118**, e2026160118 (2021).
4. H. Nilforoshan, W. Looi, E. Pierson, *et al.*, "Human mobility networks reveal increased segregation in large cities," Nature **624**, 586 (2023).
5. S. N. Zenk, A. J. Schulz, S. A. Matthews, *et al.*, "Activity space environment and dietary and physical activity behaviors: A pilot study," Health & Place **17**, 1150 (2011).
6. T. Althoff, H. Nilforoshan, J. Hua, and J. Leskovec, "Large-scale diet tracking data reveal disparate associations between food environment and diet," Nat. Commun. **13**, 267 (2022).
7. R. Xu, Q. C. Nguyen, P. S. Hovmand *et al.*, "Integrating human activity into food environments can better predict cardiometabolic diseases in the united states," Nat. Commun. **14**, 7137 (2023).
8. B. Garcia Bulle Bueno, A. L. Horn, B. M. Bell, *et al.*, "Effect of mobile food environments on fast food visits," Nat. Commun. **15**, 2291 (2024).

Transmission-aware mobility interventions for epidemic control

BIBANDHAN POUDYAL[1], DAVID SORIANO PAÑOS[2,3], AND GOURAB GHOSHAL[1,4,*]

[1] Department of Physics & Astronomy, University of Rochester, Rochester, NY, 14627, USA
[2] Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain
[3] GOTHAM lab, Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza, Spain
[4] Department of Computer Science, University of Rochester, Rochester, NY, 14627, USA
[*] gghoshal@pas.rochester.edu

1. INTRODUCTION

This abstract summarizes results recently published in *Communications Physics* [1], where we investigated how mobility-targeted interventions designed for one epidemic may affect another.

Human mobility shapes the spatial propagation of infectious diseases and has therefore become a natural target for non-pharmaceutical interventions [2–4]. During the COVID-19 pandemic, mobility restrictions and flow reshuffling were widely used to reduce contact rates in dense urban areas, often at substantial social and economic cost [5–7]. These interventions, however, are typically evaluated for a single pathogen or a single transmission mechanism. Cities often face multiple epidemic risks at once. In particular, airborne diseases (ABDs) and vector-borne diseases (VBDs) respond to mobility through different mechanisms: ABDs are driven by direct human contact, whereas VBDs depend on exposure to spatially localized vector populations.

Here we ask a simple question: can a mobility intervention designed to reduce vulnerability to one disease increase vulnerability to another? We address this question using Santiago de Cali, Colombia, as an empirical test case. Cali is divided into 22 communes, with commuting flows reconstructed from urban travel surveys, population sizes taken from census data, and mosquito abundance estimated from the container index for *Aedes aegypti*. We combine these data with a metapopulation framework for ABD and VBD vulnerability [8, 9]. Vulnerability is defined as the inverse epidemic threshold, so lower values indicate reduced epidemic risk.

The empirical framework integrates three complementary datasets: origin–destination commuting flows derived from the Cali mobility survey, census-derived residential population distributions across the city's 22 communes, and entomological estimates of mosquito abundance obtained from the container index for *Aedes aegypti*. These data are coupled to previously developed metapopulation models describing airborne and vector-borne transmission, allowing us to compare how identical mobility perturbations affect epidemic vulnerability under different transmission mechanisms.

2. CONTRASTING EFFECTS OF MOBILITY INTERVENTIONS

We first evaluate a mobility-targeted intervention previously proposed for airborne disease control: redistributing flows from high-density population hotspots toward neighboring suburban areas while preserving the total outflow from each hotspot [9]. Hotspots are identified using the LouBar method [10, 11]. For each randomized realization of the intervention, we recompute epidemic vulnerability and compare it to the empirical baseline.

The intervention behaves as expected for ABDs. By moving people away from dense population hotspots, it dilutes contact opportunities and reduces vulnerability by roughly 20%. The same intervention has the opposite effect for VBDs, increasing vulnerability by nearly 50%. As shown in Fig. 1(a,b), a policy that is beneficial for one transmission mechanism can therefore be detrimental for another, even though it acts on the same empirical mobility network.

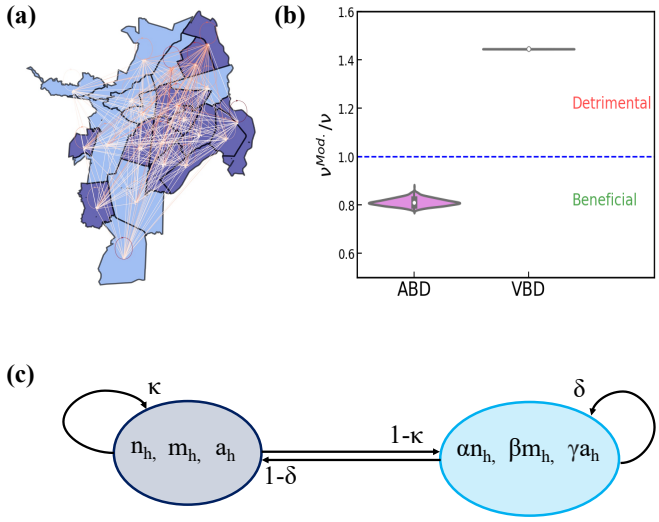


Fig. 1. Empirical paradox and coarse-grained model. In Cali, hotspot-to-suburb mobility reshuffling reduces vulnerability to airborne disease but increases vulnerability to vector-borne disease. The hub–leaf model aggregates hotspots and suburbs into two effective patches, preserving population, area, vector abundance, and mobility asymmetries.

3. MECHANISM: DILUTION VERSUS EVACUATION

To explain this trade-off, we coarse-grain the empirical mobility network into a two-patch hub–leaf metapopulation. The hub aggregates population hotspots and the leaf aggregates suburban regions. Let $\alpha = n_l/n_h$, $\beta = m_l/m_h$, and $\gamma = a_l/a_h$ denote the leaf-to-hub ratios of population, vector abundance, and area, respectively. Mobility is controlled by κ, the fraction of hub residents remaining in the hub, and δ, the fraction of leaf residents remaining in the leaf. The corresponding coarse-grained representation is shown in Fig. 1(c).

The coarse-grained model yields closed-form expressions for ABD and VBD vulnerability as functions of $(\alpha, \beta, \gamma, \kappa, \delta)$. For ABDs, vulnerability is minimized by homogenizing the effective population distribution. This is a dilution mechanism: lowering density in the hub reduces direct contact rates. For VBDs, the mechanism differs because vectors remain spatially localized. Moving people out of vector-rich areas can reduce exposure, but removing too many residents can also increase the exposure of

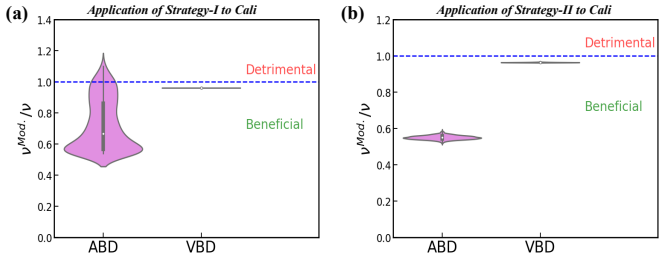


Fig. 2. Model-informed mobility interventions simultaneously reduce airborne and vector-borne vulnerability in Cali. Strategy I, based on balancing hotspot and suburb flows, is strongly beneficial for VBDs and mostly beneficial for ABDs. Strategy II sets $\kappa = 1/(\gamma+1)$ and $\delta = \gamma/(\gamma+1)$ and reduces vulnerability for both disease classes in the empirical Cali mobility network.

the remaining or incoming population. VBD control therefore requires a balance between evacuation and dilution.

This distinction explains why the ABD-optimized intervention does not generalize. It approximates a strong evacuation of the hub, which helps direct human-to-human transmission but can increase exposure to vector-mediated transmission. The analytical model instead predicts a joint optimum in which the mobility parameters are matched to spatial scale,

$$\kappa = \frac{1}{\gamma+1}, \qquad \delta = \frac{\gamma}{\gamma+1}. \tag{1}$$

This regime lies within the beneficial range for VBDs while also minimizing ABD vulnerability.

4. A JOINT INTERVENTION FOR CALI

We next apply the predicted intervention back to the full commuting network of Cali rather than to the coarse-grained model alone. For Cali, the aggregate suburb-to-hotspot area ratio is $\gamma = 1.43$, which fixes the target values of κ and δ in Eq. Eq. (1). We then reshuffle empirical mobility flows to match these aggregate constraints while preserving normalized mobility volumes. Across randomized realizations, the modified network reduces vulnerability for both ABDs and VBDs. The effects of the two model-informed intervention strategies are summarized in Fig. 2. While Strategy I substantially improves VBD outcomes and remains largely beneficial for ABDs, Strategy II, derived from the analytical model, consistently reduces vulnerability for both disease classes.

5. DISCUSSION

Our results show that mobility interventions are not disease agnostic. A strategy that reduces vulnerability to airborne transmission can increase vulnerability to vector-borne transmission because the two disease classes respond differently to the spatial redistribution of people. The relevant distinction is not simply whether mobility is increased or decreased, but how mobility reshapes exposure across population density, area, and vector abundance.

The results in Fig. 2 demonstrate that these analytically derived mobility principles remain effective when applied to the full empirical commuting network rather than only to the coarse-grained model, providing a practical route from theoretical insight to implementable mobility interventions.

The central lesson is that the effectiveness of mobility interventions depends fundamentally on how human movement interacts with the underlying transmission mechanism. Rather than optimizing interventions for individual pathogens, mobility policies should account for the broader transmission ecology of the city. As urban areas increasingly face the co-circulation of multiple infectious diseases, coarse-grained mobility models provide an interpretable framework for designing interventions that remain robust across pathogen classes.

A more detailed theoretical analysis, derivations, and additional numerical experiments are presented in the published version of this work [1].

Funding. GG acknowledges partial support through Grant No. 62417 from the John Templeton Foundation. GG and BP also acknowledge support from the University of Rochester. D.S.-P. acknowledges financial support through grants JDC2022-048339-I and PID2021-128005NB-C21 funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR".

Disclosures. The authors declare no conflicts of interest.

Data availability. Data on the spatial distribution of residents and entomological indices are publicly available from the municipality of Cali. Mobility survey data are available from the authors upon reasonable request.

REFERENCES

1. B. Poudyal, D. Soriano Paños, and G. Ghoshal, "Contrasting and comparing the efficacy of mobility-targeted interventions on airborne and vector-borne diseases," Commun. Phys. **8**, 250 (2025).
2. H. Barbosa, M. Barthelemy, G. Ghoshal, *et al.*, "Human mobility: Models and applications," Phys. Reports **734**, 1 (2018).
3. R. E. Baker, A. S. Mahmud, I. F. Miller, *et al.*, "Infectious disease in an era of global change," Nat. Rev. Microbiol. **20**, 193 (2022).
4. D. Brockmann, "Human mobility and spatial disease dynamics," Rev. nonlinear dynamics complexity **2**, 1 (2009).
5. A. Lison, N. Banholzer, M. Sharma, *et al.*, "Effectiveness assessment of non-pharmaceutical interventions: lessons learned from the covid-19 pandemic," The Lancet Public Health **8**, e311 (2023).
6. N. Perra, "Non-pharmaceutical interventions during the covid-19 pandemic: A review," Phys. Reports **913**, 1 (2021).
7. M. Pangallo, A. Aleta, R. M. del Rio-Chanona, *et al.*, "The unequal effects of the health–economy trade-off during the covid-19 pandemic," Nat. Hum. Behav. **8**, 264 (2024).
8. D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, *et al.*, "Vector-borne epidemics driven by human mobility," Phys. Rev. Res. **2**, 013312 (2020).
9. S. Hazarie, D. Soriano-Paños, A. Arenas, *et al.*, "Interplay between population density and mobility in determining the spread of epidemics in cities," Commun. Phys. **4**, 191 (2021).
10. R. Li, P. Richmond, and B. M. Roehner, "Effect of population density on epidemics," Phys. A: Stat. Mech. its Appl. **510**, 713 (2018).
11. A. Bassolas, H. Barbosa-Filho, B. Dickinson, *et al.*, "Hierarchical organization of urban mobility and its connection with city livability," Nat. communications **10**, 4817 (2019).

Urban mobility network centrality predicts social resilience

LIN CHEN[1,*], FENGLI XU[2,3], ESTEBAN MORO[4], PAN HUI[1,5,*], YONG LI[2,3,*], AND JAMES EVANS[6,7,*]

[1] Dept. of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2] Dept. of Electronic Engineering, Tsinghua University, Beijing, China
[3] BNRist, Tsinghua University, Beijing, China
[4] Network Science Institute, Northeastern University, Boston, MA, USA
[5] The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[6] Santa Fe Institute, Santa Fe, NM, USA
[7] Knowledge Lab and Dept. of Sociology, University of Chicago, Chicago, IL, USA
[*] panhui@ust.hk, liyong07@tsinghua.edu.cn, jevans@uchicago.edu

1. INTRODUCTION

Cities are vibrant hubs whose venues, from restaurants and markets to schools and parks, act as "social anchors" that draw diverse people together to build ties, generate innovation, and improve well-being [1, 2]. Urban shocks, increasingly frequent under climate change and urbanization, threaten these interactions. Existing work documents aggregate responses, namely declining visitation [3] and heightened socio-demographic segregation [4], but these trends obscure striking venue-level heterogeneity: some places maintain or even increase both visitor volume and demographic diversity while others collapse under the same event. Recent work emphasizes that shock impacts propagate across interdependent venues [5], yet we still lack a framework that anticipates socially resilient venues from *pre-shock* mobility structure and explains why certain places sustain both visitation and social mixing. Filling this gap is essential for crisis-response strategies that protect social cohesion, mitigate adversarial impacts, and accelerate post-shock recovery [6].

We study social resilience, a venue's capacity to maintain both visitation and demographic mixing, using SafeGraph mobility data from 15 U.S. metropolitan statistical areas (MSAs) covering more than 103 million residents, 66 venue types, and three shocks: the COVID-19 pandemic, the 2020 Hurricane ETA, and the 2018 California Wildfire.

2. METHODS

Visitation and NAICS metadata come from SafeGraph Patterns and Places; demographics from the 2019 ACS 5-year estimates at the census block group (CBG) level. We map points of interest to 66 harmonized venue sectors, aggregating 4-digit NAICS categories to their 3-digit parent when coverage is insufficient, and split each MSA's CBGs into five equal-population income groups. Experienced income segregation S at a venue is the normalized deviation of its visitor income mix from an equal-share ideal [2]; segregation change (ΔS) and mobility change (ΔM) are relative differences between matched in-shock and pre-shock windows of equal length. Inspired by revealed comparative advantage in economic dependency [7], we build a sector network whose edge weight $P_{i,j} = \max\{P(R_i \mid R_j),\, P(R_j \mid R_i)\}$ is the conditional co-preference of CBGs (a CBG prefers sector i when its relative visit frequency $R_{m,i} = f_{m,i}/f_i > 1$), and compute each sector's eigenvector centrality on this weighted network. Social resilience is modeled with standardized OLS regressions that add centrality on top of pre-shock baselines; core/peripheral comparisons use two-sided Wilcoxon signed-rank tests on within-MSA medians, with effect sizes r.

3. RESULTS

Divergent responses across venues. Against the general trend of rising segregation and falling visitation, 36.28–53.01% of venues show *reduced* segregation (i.e., increased mixing) and 21.04–38.55% show *increased* visitation during shocks. These responses correlate systematically with venues' NAICS sectors: during COVID-19, for example, only 9.9% of *General Merchandise Stores* but 52.7% of *Data Processing Services* venues fall in the region of severe segregation deterioration, a more than fivefold difference. The sectoral orderings are moreover highly consistent across cities (mean Spearman 0.87, *std* 0.02 for segregation; 0.69, *std* 0.06 for mobility), indicating a generalizable structure rather than city-specific noise and pointing to differentiated venue roles in the urban mobility landscape.

A mobility network and its centrality. To quantify this disparity, we build a venue-sector network for each MSA from residents' revealed visitation preferences (Methods). The network is clearly core–peripheral (Fig. 1a): core sectors satisfy essential, common needs (e.g., restaurants and grocery stores), while peripheral sectors serve specialized or discretionary needs (e.g., information and entertainment services). Eigenvector centrality captures a recursive notion of importance, scoring sectors highly when they connect to other well-connected sectors, and thus summarizes each sector's structural prominence in resident mobility pathways.

Centrality predicts resilience. Sector centrality is strongly and negatively associated with both segregation change and mobility change across every city and shock (Fig. 1b–g): core venues sustain population mixing and visitation far better than peripheral ones. Crucially, this is not merely a proxy for intuitive baselines. In stepwise OLS models (Philadelphia, COVID-19), pre-shock segregation, pre-shock mobility, and a location-based bridge index [8] explain little of the variance in segregation change ($R^2 \leq 0.24$); adding centrality nearly doubles it to $R^2 = 0.46$, with a unit increase in centrality corresponding to a 0.0709-unit decrease in segregation change ($p < 0.001$). For mobility change, baselines reach $R^2 < 0.09$ while centrality raises fit to $R^2 = 0.34$. Across MSAs and shocks, centrality lifts explanatory power by more than 80% over the best baseline, establishing pre-shock network position as a robust predictor of resilience.

Core vs. peripheral venues. Core (top-10 centrality) and peripheral (bottom-10) sectors show *no* significant difference in spatial distribution (radius of gyration $p = 0.59$; minimum enclosing radius $p = 0.79$; entropy $p = 0.51$), so resilience does not stem from location but from how and when venues are used. Core venues have shorter visit distances (median 7.36 vs. 9.26 km; $r = 0.89$), serve more neighborhoods (median 49 vs. 13), shorter dwell

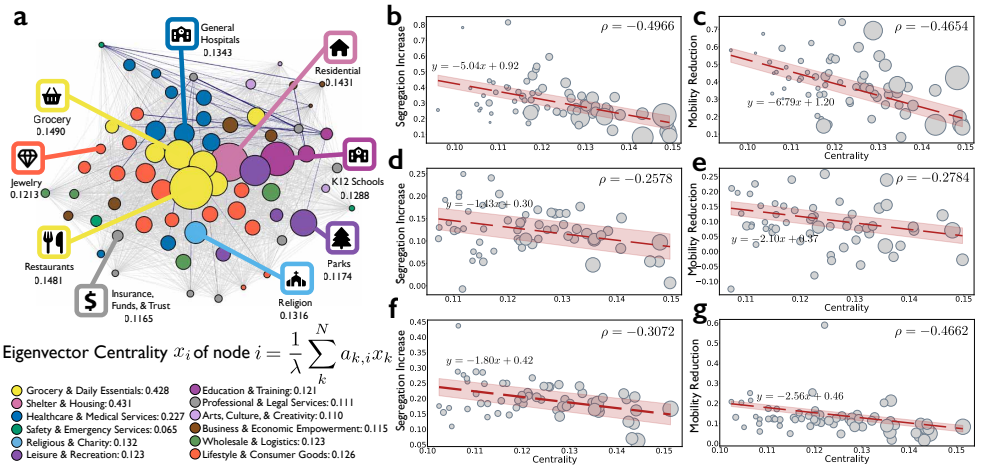


Fig. 1. Mobility-network centrality predicts social resilience. **a**, Pre-shock venue-sector mobility network for the Philadelphia MSA; node size is total visitation, color is broad category, and the value by each sector is its eigenvector centrality. **b,d,f**, Sector centrality vs. relative segregation change in Philadelphia (COVID-19), Riverside (2018 California Wildfire), and Miami (2020 Hurricane ETA). **c,e,g**, Sector centrality vs. relative mobility change for the same cases. Lines are OLS fits; shaded regions are 95% CI.

times (26.0 vs. 65.8 min), and steadier round-the-clock demand (higher hourly-popularity entropy, 4.17 vs. 3.70; r=0.89). These gaps widen during shocks: travel distances rise more for peripheral venues (+5.80%, p=0.002) than core ones (+2.41%, p=0.131), and both serve fewer neighborhoods while core venues still reach the most. Together these patterns imply a dual social mechanism: core venues sustain neighborhood-level ties through frequent, brief, task-oriented encounters that keep visitor turnover high, while peripheral venues support deeper, purpose-driven engagement within specialized social circles.

4. DISCUSSION

Core and peripheral venues provide complementary flavors of urban interaction. Core venues such as grocery stores and restaurants support the frequent "weak ties" [9] and public-life encounters that Jane Jacobs described as the "ballet" of the sidewalk [10], sustaining neighborhood connectivity and collective well-being. Peripheral venues, reached through longer, purpose-driven trips often motivated by pre-existing ties [11], host deeper engagement around focused interests [12]. During shocks the balance shifts toward low-intensity core interaction, a collective adaptation that can strengthen the local mutual-aid networks central to disaster recovery [13]. Bridging urban mobility with economic theory distinguishing staple from discretionary goods [14, 15], we synthesize these findings into a *well-and-pool* analogy: in crisis, citizens concentrate scarce mobility on essential core venues, like water collecting in a deep well; when mobility is abundant, they spread across specialized peripheral venues, like water pooling after rain. This lens yields actionable guidance: planners should safeguard core-venue accessibility through robust, inclusive transportation [16], while directing targeted support to peripheral venues to restore specialized community afterward. Limitations include U.S.-only coverage, CBG-level income proxies, and phone-based mobility that can differentially sample groups [17]. Overall, pre-shock mobility structure offers a robust, broadly applicable lens for anticipating and managing urban social resilience.

REFERENCES

1. L. M. Bettencourt, *Introduction to urban science: evidence and theory of cities as complex systems* (MIT Press, 2021).
2. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. communications **12**, 4633 (2021).
3. S. Chang, E. Pierson, P. W. Koh, *et al.*, "Mobility network models of covid-19 explain inequities and inform reopening," Nature **589**, 82 (2021).
4. T. Yabe, B. G. B. Bueno, X. Dong, *et al.*, "Behavioral changes during the covid-19 pandemic decreased income diversity of urban encounters," Nat. communications **14**, 2310 (2023).
5. T. Yabe, B. García Bulle Bueno, M. R. Frank, *et al.*, "Behaviour-based dependency networks between places shape urban economic resilience," Nat. human behaviour **9**, 496 (2025).
6. L. Chen, F. Xu, Z. Han, *et al.*, "Strategic covid-19 vaccine distribution can simultaneously elevate social utility and equity," Nat. Hum. Behav. **6**, 1503 (2022).
7. C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, "The product space conditions the development of nations," Science **317**, 482 (2007).
8. H. Nilforoshan, W. Looi, E. Pierson, *et al.*, "Human mobility networks reveal increased segregation in large cities," Nature **624**, 586 (2023).
9. M. S. Granovetter, "The strength of weak ties," Am. journal sociology **78**, 1360 (1973).
10. J. Jacobs, *The Death and Life of Great American Cities* (Random House, 1961).
11. E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2011), pp. 1082–1090.
12. Y. Lin, C. B. Frey, and L. Wu, "Remote collaboration fuses fewer breakthrough ideas," Nature **623**, 987 (2023).
13. D. P. Aldrich, *Building resilience: Social capital in post-disaster recovery* (University of Chicago Press, 2012).
14. T. Veblen, *The theory of the leisure class* (Routledge, 2017).
15. J. A. Parker, N. S. Souleles, D. S. Johnson, and R. McClelland, "Consumer spending and the economic stimulus payments of 2008," Am. Econ. Rev. **103**, 2530 (2013).
16. Z. Xu and S. S. Chopra, "Interconnectedness enhances network resilience of multimodal public transportation systems for safe-to-fail urban mobility," Nat. communications **14**, 4291 (2023).
17. N. Grigoropoulou and M. L. Small, "Are large-scale data from private companies reliable? an analysis of machine-generated business location data in a popular dataset," Soc. Sci. Comput. Rev. **43**, 1087 (2025).

Infrastructure Matters: Quantifying Measurement Bias in Mobile Phone Mobility Data

LEO FERRES[1,2], ERICK ELEJALDE[3], KYRIAKI KALIMERI[2], AND LORETO BRAVO[1]

[1] *Faculty of Engineering, Universidad del Desarrollo, Santiago, Chile*
[2] *ISI Foundation, Turin, Italy*
[3] *L3S Research Center, Leibniz University Hannover, Hannover, Germany*

Mobile phone records have transformed the study of human mobility by providing continuous observations of population movement at unprecedented temporal and spatial scales [1]. They have become a standard data source for transportation planning, epidemic forecasting, disaster response, migration analysis, and urban science [2, 3]. Most analyses implicitly assume that, after correcting for operator market share or demographic representativeness [4], the observed trajectories constitute an unbiased sample of the underlying mobility process. This assumption overlooks a fundamental property of the sensing infrastructure itself: mobile phone data in the form of XDR and CDR are not direct observations of human locations but measurements produced by a heterogeneous communication network whose spatial resolution varies dramatically across space. In effect, cellular infrastructure is not merely the source of the data, it is part of the measurement instrument.

The consequences of this observation have received surprisingly little quantitative attention [5]. Cellular networks are engineered to maximize communication performance rather than spatial sampling uniformity. Towers are concentrated where demand is highest, resulting in dense deployments in urban centers and sparse coverage in suburban and rural regions. Because users are localized relative to serving antennas, the probability that a movement is observed depends on the network's local geometry. Short trips that cross several urban sectors may be readily detected, whereas movements of comparable length in rural environments remain completely invisible because they occur within the coverage area of a single tower. Consequently, the observation process itself is spatially heterogeneous, introducing systematic rather than random errors into mobility measurements. These errors propagate directly into origin–destination matrices, displacement distributions, population estimates, and every downstream model that relies upon them [6].

In this work, we focus on a largely unexplored question: How much of the urban–rural differences observed in mobile phone mobility data reflect genuine human behavior, and how much are artifacts of heterogeneous sensing resolution?

To answer this question, we develop an infrastructure-aware framework that explicitly models the measurement process induced by the cellular network before estimating mobility. Rather than treating each tower as an isotropic Voronoi cell, we reconstruct directional antenna sectors using tower azimuth, antenna height, and sectorization information to obtain a realistic approximation of network coverage [7]. From this representation, we derive a local detection model that quantifies the probability that trips of different lengths are observable given the effective spatial resolution surrounding each site. Fine-resolution population counts from the national census are subsequently redistributed onto a regular H3 grid using dasymetric interpolation informed by sector coverage and road network density, thereby providing spatially explicit estimates of population exposure. Finally, heterogeneous observation probabilities are incorporated into origin-destination estimation through inverse probability weighting (IPW), while a Fay-Herriot (FH) small-area estimator stabilizes corrected flows in regions where observations remain sparse. The individual components of the framework are established techniques [4, 6, 8–11]), but their integration around an explicit model of the observation process provides a unified methodology for quantifying infrastructure-induced measurement bias.

We evaluate the framework using one of the largest mobility datasets assembled for Chile, comprising a catalog of 22,836 antenna sectors associated with 2,600 base transceiver stations, detailed 2024 census cartography covering more than 66,000 census blocks [12], and approximately 931 million XDR-derived trips collected over a 6.5-week observation period in the Región Metropolitana of Santiago. The study area encompasses one of the most densely instrumented metropolitan regions in South America while simultaneously containing extensive rural municipalities, making it an ideal environment for investigating how infrastructure heterogeneity affects mobility inference.

Our results demonstrate that the observation process varies by orders of magnitude across the same metropolitan area. Modeling directional antenna sectors improves the network's effective spatial resolution from a median of 894 m under conventional tower-point Voronoi tessellations to 299 m, representing a threefold increase in localization precision (Figure 1). More importantly, the inferred detection threshold is highly heterogeneous: the distance required for a trip to have a 50% probability of being observed ranges from approximately 16 m in the urban core to over 2.5 km at the most isolated rural sites. These findings indicate that identical human movements have fundamentally different probabilities of appearing in the dataset depending only on the surrounding infrastructure.

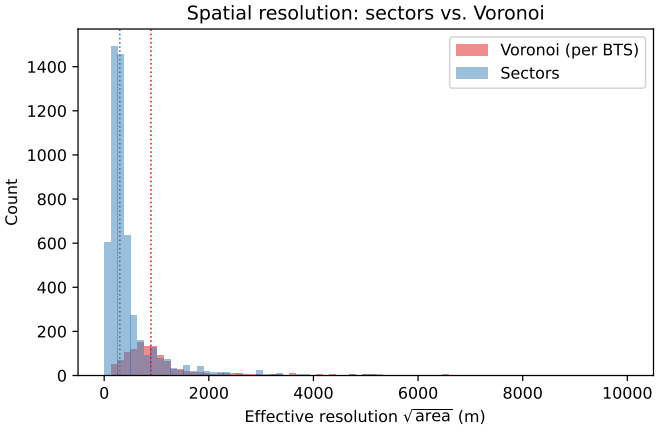


Fig. 1. Effective spatial resolution ($\sqrt{\text{area}}$) for sector polygons (blue) versus tower-point Voronoi cells (red). Dotted lines indicate medians: 299 m for sectors, 894 m for Voronoi, representing a 3.0× gain.

Table 1. Three-way comparison for selected comunas

Comuna	Raw trips	IPW trips	FH trips	γ
Santiago (urban)	73.9M	71.8M	71.4M	0.71
Maipú (suburban)	31.4M	27.9M	27.4M	0.58
Pirque (rural)	6.1M	9.3M	9.5M	0.28

The impact of this heterogeneity is substantial. Approximately 40% of all matched cell transitions occur within the same physical tower site. While many correspond to technology handoffs, nearly 100 million transitions represent genuine movements between directional sectors, recovering short-distance displacements with a median estimated length of 429 m that are completely absent from conventional tower-level analyses.

The three-way comparison in Table 1 illustrates the correction pipeline at work. For Santiago (urban), the IPW and Fay–Herriot estimates are close to the raw counts, because detection is near-complete and the data are reliable. For Pirque (rural), the IPW correction uplifts flows by 51%, and the Fay–Herriot model slightly reinforces this adjustment (+2%), reflecting the low shrinkage weight ($\gamma = 0.28$) that balances the noisy direct estimate against the structural model.

In general, after correcting for heterogeneous observation probabilities, mobility originating from rural municipalities increases by 50–73%, whereas flows from densely instrumented urban areas are moderately down-weighted (Figure 2). These corrections demonstrate that part of the apparent dominance of urban mobility observed in raw mobile phone data is attributable not to human behavior itself but to unequal sensing capabilities across the communication network.

Beyond the specific corrections proposed here, we believe the broader implication is conceptual. Mobile phone datasets should be viewed as observations generated by an imperfect spatial sensor whose characteristics vary across the landscape. Just as remote sensing studies account for sensor resolution and uncertainty before interpreting satellite imagery, mobility studies should explicitly model the spatial observation process induced by telecommunications infrastructure. Ignoring this process risks confounding infrastructure density with human mobility itself, particularly when comparing urban and rural regions or evaluating spatial inequalities. Treating network geometry as an explicit component of the measurement model therefore provides a principled foundation for improving the robustness and comparability of mobility analyses across different countries, operators, and deployment environments. Because the proposed framework requires only a tower catalog, a population layer, and standard mobile phone records, it is readily transferable to other settings and enables the establishment of infrastructure-aware corrections as a standard preprocessing step in mobile phone mobility analysis.

Funding. This research was supported by FONIS grant SA24I0124 to L.F.. L.F. and K.K. also acknowledge financial support from the Lagrange Project of the Institute for Scientific Interchange Foundation (ISI Foundation), funded by the Fondazione Cassa di Risparmio di Torino (Fondazione CRT).

Disclosures. The authors declare no conflicts of interest.

Data availability. The tower catalog is proprietary and cannot be shared. The 2024 Chilean census cartography is publicly available from the Instituto Nacional de Estadísticas (INE). The OpenStreetMap road network is open data. The `mobilens` library and analysis notebooks are available at https://github.com/leoferres/mobilens.

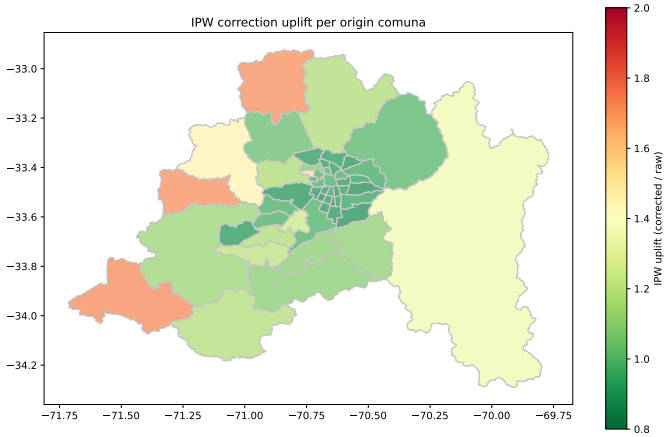


Fig. 2. IPW correction uplift (corrected / raw trip ratio) per origin comuna. Rural comunas in the southwest and south of R13 receive 50–73% more trips after correction; the urban core is slightly downweighted ($\sim$0.86$\times$).

REFERENCES

1. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779 (2008).
2. M. Haraguchi, A. Nishino, A. Kodaka, *et al.*, "Human mobility data and analysis for urban resilience: A systematic review," Environ. Plan. B: Urban Anal. City Sci. **49**, 1507 (2022).
3. N. Kostandova, C. Schluth, R. Arambepola, *et al.*, "A systematic review of using population-level human mobility data to understand SARS-CoV-2 transmission," Nat. Commun. **15**, 1 (2024).
4. O. Järv, H. Tenkanen, and T. Toivonen, "Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation," Int. J. Geogr. Inf. Sci. **31**, 1630 (2017).
5. R. Gallotti, D. Maniscalco, M. Barthélemy, and L. Alessandretti, "Distorted insights from human mobility data," Commun. Phys. **7**, 421 (2024).
6. D. Pfeffermann, "New important developments in small area estimation," Stat. Sci. **28**, 40 (2013).
7. E. Graells-Garrido, O. Peredo, and J. García, "Sensing urban patterns with antenna mappings: The case of santiago, chile," Sensors **16**, 1098 (2016).
8. D. J. Briggs, J. Gulliver, D. Fecht, and D. M. Vienneau, "Dasymetric modelling of small-area population distribution using land cover and light emissions data," Remote. Sens. Environ. **108**, 451 (2007).
9. D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," J. Am. Stat. Assoc. **47**, 663 (1952).
10. J.-C. Deville and C.-E. Särndal, "Calibration estimators in survey sampling," J. Am. Stat. Assoc. **87**, 376 (1992).
11. R. E. Fay and R. A. Herriot, "Estimates of income for small places: An application of James–Stein procedures to census data," J. Am. Stat. Assoc. **74**, 269 (1979).
12. L. Pappalardo, G. Cornacchia, V. Navarro, *et al.*, "A dataset to assess mobility changes in Chile following local quarantines," Sci. Data **10**, 6 (2023).

Parental Activity Space Across Childcare Stages

SHUSEI HAYASHI[1,*], TAKAHIRO YABE[2,3], AND NAOYA FUJIWARA[1,3,4]

[1] Graduate School of Information Sciences, Tohoku University, Sendai, Japan
[2] Department of Technology Management and Innovation, Tandon School of Engineering, New York University, Brooklyn, NY, USA
[3] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan
[4] PRESTO, Japan Science and Technology Agency, Saitama, Japan
[*] hayashi.shusei.s7@dc.tohoku.ac.jp

1. INTRODUCTION

Japan has one of the world's lowest fertility rates. The total fertility rate hit a record low of 1.15 nationwide and 0.96 in Tokyo in 2024 [1]. Childcare support has centered on expanding nursery facilities. Their number grew 63% from 2013 to 2022, and waiting lists reached zero in ∼80% of municipalities by 2020, so the marginal effect of further provision is now limited [2]. The next target must thus lie beyond nursery provision, in the childrearing burden on parents. In high-income countries, however, the effect of this burden on fertility has grown more complex, depending not on its monetary or time cost alone but on *when* and *in what form* it falls [3].

Yet it remains unclear in which everyday behaviors and childcare stages that burden appears. Time-diary studies document how this burden shifts with a child's developmental stage but rely on self-reports and cannot observe *where* it unfolds [4]. Large-scale GPS data can fill this gap but carry no household information, requiring a method to identify parents. Our prior work [5] provides this foundation, identifying parents from drop-off/pick-up behavior at early-childhood facilities and showing that they stay closer to home than non-parents. We extend it to nine prefectures over four years and ask how parents' activity space changes across the childcare stages before, during, and after enrollment. This is a first step toward identifying the stages at which childrearing households most need support.

2. DATA AND METHOD

We use large-scale GPS location data collected with user consent and provided by Blogwatcher Inc. [6], covering nine prefectures in Japan over approximately four years (January 2019–January 2023), with a base population of more than 30 million users. Points-of-interest (POI) are taken from Zenrin's "Telepoint" database, and geospatial data from the Geospatial Information Authority of Japan. Following our prior work [5], we identify parents by drawing a 30 m-radius polygon around each early-childhood facility and treating as a parent any individual with at least 10 location pings inside it on at least 5 distinct days during drop-off/pick-up hours (07–10 and 15–19). Per fiscal year, we retain the 40% of facilities with the fewest unique IDs, since a high count indicates that many non-parents also stay there. We also drop individuals appearing in two or more polygons, and those with any nighttime ping (23–04) inside a polygon, who are likely nearby residents or facility staff. Since enrollment and graduation fall on Japanese fiscal year (FY) boundaries, identification is performed per FY, yielding 83,869 parents (154,833 person-years). Each parent-year falls into one of three childcare stages around enrollment: *in* for a year containing drop-off/pick-up (the child is enrolled), and *pre* or *post* for years before the first or after the last *in* year. We analyze FY2020 (Apr. 2020–Mar. 2021) and FY2021 (Apr. 2021–Mar. 2022), in which all three stages co-occur, comprising 43,555 person-years after matching in FY2020 (*pre* 10,790 / *in* 21,588 / *post* 11,177) and 40,168 in FY2021 (*pre* 6,063 / *in* 22,375 / *post* 11,730).

As controls we take non-parents aged 20–49 with no facility stays, matched one-to-one to parents on home municipality, gender, age band, and within-year stay count; this single matched group serves as the baseline in all figures below. Matching on stay count ensures the differences below are not an artifact of parents having more GPS records. We characterize mobility by two indices. The radius of gyration (r_g) [7] captures the extent of a person's activity space. Location capacity (C) is the number of places within a trailing 30-day window that a person visits at least twice and for at least ten minutes per week. Both indices are computed daily over a 30-day sliding window in 1-day steps and aggregated to one median per person-year. We compare groups with the Mann–Whitney test, reporting Cliff's δ for the parent−non-parent difference. Cliff's δ is the difference between the probabilities that a random parent exceeds a random non-parent and vice versa, ranging from -1 to $+1$.

3. RESULTS

First, we compare the within-year time series of activity range between parents and non-parents at each stage (Fig. 1). The parent curves for all three stages lie consistently below the non-parent baseline in both years. Time variation such as seasonality and the contraction during the first COVID-19 wave is synchronized between groups, indicating that both groups respond to the same external shocks. This confirms the prior finding [5] that parents have a narrower activity space.

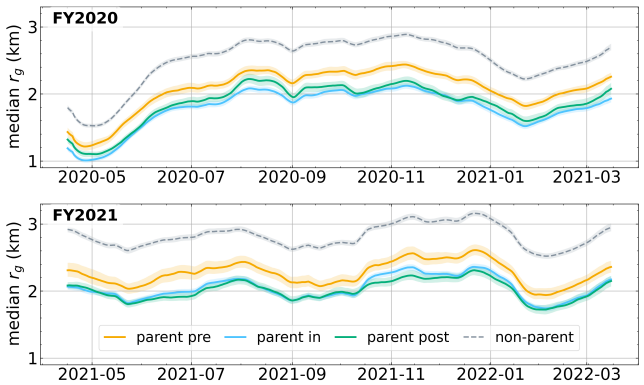


Fig. 1. Within-year time series of the median r_g for parents at each childcare stage and for the matched non-parent baseline. Each curve is a 7-day moving average; shaded bands are 95% confidence intervals from a person-level cluster bootstrap.

Broken down by prefecture (Fig. 2), this δ deepens with the childcare stage in most prefectures, but the effect is not uniform, being significant only in the most populous ones. Whether the deeper narrowing in large cities reflects tighter constraints or simply a denser neighborhood where daily needs are met nearby cannot be told from the group difference alone.

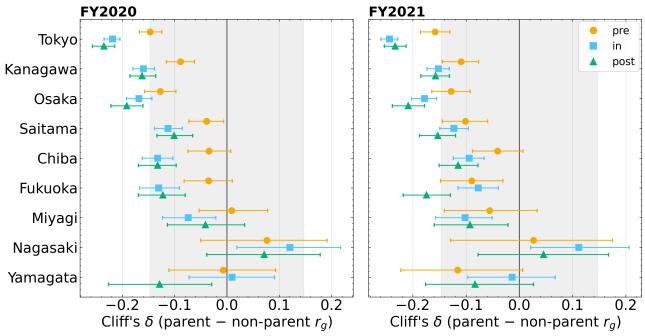


Fig. 2. Forest plot of the parent − non-parent difference in r_g, measured by Cliff's δ, for each prefecture and childcare stage, with prefectures ordered by population. Negative values indicate that parents have a smaller r_g. Error bars are bootstrap 95% confidence intervals, a point being significant when its interval excludes zero, and the shaded band marks the negligible-effect range ($|\delta| < 0.147$).

In contrast, C moves in the opposite direction to r_g. Whereas the parent−non-parent gap in r_g widens as the childcare stage advances, the gap in C is largest in *pre*, where parents visit more places than non-parents, and then returns toward the non-parent level through *in* and *post* (Fig. 3). Decomposed by prefecture (Fig. 4), the positive gap is directionally consistent in the *pre* and *in* stages, but it weakens in *post* and is small (median $\delta = +0.056$ across stages and prefectures), far less robust than the r_g narrowing. A coherent reading is that in *pre*, when children are very young, parents must visit a growing set of places, raising C. As children grow, the number of regularly visited places returns to the non-parent level, yet the activity space remains narrow.

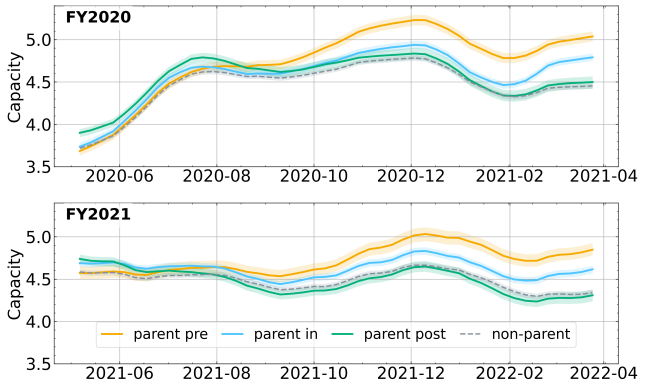


Fig. 3. C over the same window and groups as Fig. 1.

4. DISCUSSION

Across nine prefectures and four years, parents' activity space (r_g) is robustly narrower than non-parents', and the gap persists after equalizing residence, attributes, and record volume. As the childcare stage advances, this narrowing deepens, most strongly in the largest cities. The number of places parents visit moves the other way, with the surplus in *pre* fading to the non-parent level by *post*. Taken together, what sets parents apart is not how many places they visit but how widely those places spread.

These findings carry several caveats. The cross-sectional design captures a stationary difference attached to parenthood, not the net effect of enrollment and graduation. Stages are compared between groups within each year, not by following individuals through transitions, so contrasts may partly reflect composition. Because the non-parent baseline is pooled in coarse age bands, part of the later narrowing may reflect parental aging. Parent identification is not externally validated, and the representativeness of the underlying user base relative to the general population remains unquantified. Both years also overlap the COVID-19 pandemic, so a disproportionate parental shift to remote work could inflate the narrowing.

Even so, the picture is clear. The constraints on parents' mobility are not uniform but sharpen as childcare advances. As nursery-centered support reaches its ceiling, the next target of intervention requires identifying at which stage these constraints bind most tightly, whether a narrow range is a burden or a convenience, and how the childrearing burden divides between fathers and mothers. Pursuing these questions can turn a descriptive narrowing into concrete guidance on where and when households need support.

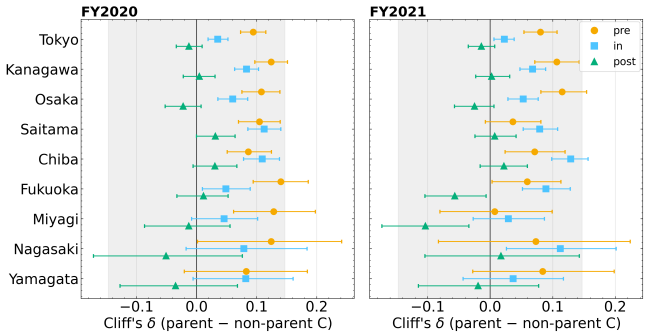


Fig. 4. Prefecture-level Cliff's δ for C, as in Fig. 2.

Funding. Japan Society for the Promotion of Science (JSPS) KAKENHI (JP24K01108, JP24K03007); Japan Science and Technology Agency (JST) PRESTO (JPMJPR21RA).

Acknowledgment. This study was supported by the Joint Research Program Nos. 1273 and 1586 of CSIS, UTokyo (geocoded telephone directory database "Telepoint Pack DB (Yellow pages)" by ZENRIN Co., Ltd.). The authors also thank Blogwatcher, Inc. for providing the GPS location data used in this analysis.

Disclosures. The authors declare no conflicts of interest.

Data availability. The GPS location data underlying the results presented in this paper are not publicly available due to privacy restrictions and are provided under a research-use-only agreement with Blogwatcher, Inc. The POI data from Zenrin Co., Ltd. (Telepoint Pack DB) used in this study are also subject to licensing restrictions. Other geospatial datasets such as road boundaries and administrative boundaries are publicly available from the Geospatial Information Authority of Japan and the Statistics Bureau of Japan.

REFERENCES

1. Ministry of Health, Labour and Welfare, Overview of Vital Statistics Monthly Report, Annual Total, 2024. https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai24/index.html (2025).
2. T. Unayama, RIETI Discussion Paper 23-J-030 (2023).
3. M. Doepke, A. Hannusch, F. Kindermann, and M. Tertilt, in *Handbook of the Economics of the Family*, vol. 1 (North-Holland, 2023), p. 151.
4. A. Kalil, R. Ryan, and M. Corey, Demography **49**, 1361 (2012).
5. S. Hayashi, T. Yabe, and N. Fujiwara, ISPRS Ann. Photogramm. Remote. Sens. Spatial Inf. Sci. **X-4/W7-2025**, 51 (2025).
6. Blogwatcher, https://www.blogwatcher.co.jp/ (Accessed 6 July 2026).
7. M. C. González, C. A. Hidalgo, and A.-L. Barabási, Nature **453**, 779 (2008).

Invariant, elastic and plastic: how mobility reorganises around a residential move

ANDREW RENNINGER[1,*], PIERO BIRELLO[1], AND MARTON KARSAI[1]

[1] *Department of Network and Data Science, Central European University, Quellenstraße 51, 1100 Vienna, Austria*
[*] *renningera@ceu.edu*

1. INTRODUCTION

Individual mobility is famously regular: people return to a few places on a stable rhythm, and the laws describing that regularity hold across populations and technologies [1]. Regularity alone cannot say what the mobility is anchored to. When someone travels the same distances every week for a year, the distance could belong to the person or to the place, and the record on its own will not tell us which. A residential move pulls them apart. It swaps the environment while holding the person fixed, the design economics uses to read the effect of place off those who change it [2]. We turn that design on behaviour, asking which mobility signatures a move leaves unchanged, which it disturbs and then restores, and which it resets to a durably new level: invariant, elastic and plastic. The answer bears on experienced segregation, where daily movement rather than residence decides whether people who live apart also spend their days apart [3].

2. DATA AND METHODS

We use GPS traces from a commercial panel covering Austria, filtered to a horizontal accuracy of 100 m, and detect home and workplace with HoWDe, which infers each from where a device rests by night and by day. A device is a mover when its detected home shifts between two stable locations, and the changepoint dates the move. Extending detection across the full trace yields **87,035 movers**, more than half again the cohort available in a single year, each matched to a non-mover on trace volume and active days so that seasonality and secular drift are differenced out rather than attributed to the move. We split movers by the city they leave or enter, taking Vienna as the first case of a definition we carry to other cities: *into city* (12,165), *out of city* (13,396) and *within city* (23,117). To read exposure we follow the places a person visits rather than the one they sleep in, attaching to each visited place the income mix of its visitors and to each home its own income band; a mover *crosses a boundary* when that band changes.

3. WHAT A MOVE CHANGES

Range is plastic, and its *direction* is set by the destination (Fig. 2A). Movers into a city arrive from a wide rural or suburban orbit and contract onto a compact one, 6.9 to 3.1 km; movers out expand nearly threefold, 2.8 to 7.7 km; within-city movers hold at 2.6 km. The same environment change expands some orbits and compresses others; moving predicts neither, and the destination decides which. Radius of gyration captures this because it is a distance, whereas counts of distinct places fall mechanically when an orbit compresses below the grid, so we do not rely on them. Rhythm stays nearer invariant. Exploration dips for every cohort in the fortnight around the move, when behaviour and sampling are both disturbed, then recovers within weeks and settles slightly above baseline for movers who change environment (Fig. 2B). Income homophily is elastic, disturbed around the move and largely restored, falling most

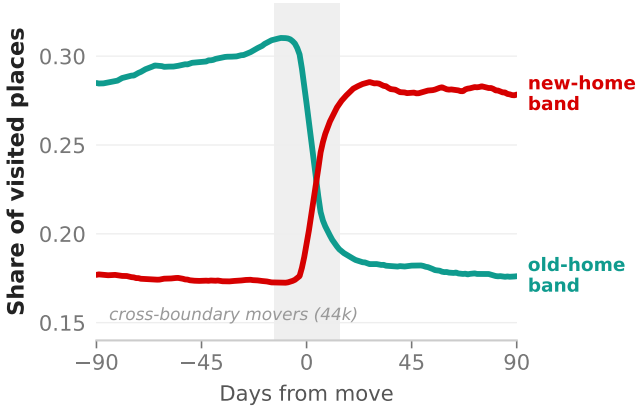


Fig. 1. Crossing an income boundary shifts exposure from the old-home band to the new. For movers whose new home changes income band, the median share of visited places in the old-home band (teal) and the new-home band (red), against days from the move; the shaded fortnight is the move window. Exposure crosses over within weeks and does not return.

and recovering least for those who leave a city for an unfamiliar social mix (Fig. 2C). Matched controls hold flat throughout, so the reorganisation belongs to the move and not to the calendar.

Exposure changes most, and the change is durable. For movers who cross an income boundary, the share of visited places in the old-home band falls from 0.30 to 0.18 as the new-home band rises from 0.17 to 0.28, a crossover completed within weeks of the move and sustained thereafter (Fig. 1). Crossing an income boundary changes the mix of people among whom a mover spends their days, and it endures. Two cautions bound these claims. Movers run elevated on several dimensions before they move, so we cannot yet separate people who move because they are mobile from people made mobile by moving; the within-city cohort, which returns to baseline, is the internal control that lets us treat that selection as a signal rather than a confound. And ninety days is a short horizon on which to call anything durable.

4. A MODEL OF THE MOVE

To read the move as a mechanism, we give each mover an agent whose movement obeys the laws of mobility, a preference to return balanced against a rate of exploration with the draw of each destination scaled by its opportunities [4], fit to the person's own trace from before the move. Calibrated this way, the agents match the observed spread of how people move (Fig. 3). Swapping only the home and work anchors and re-running the agent reproduces the observed signatures: the expansion of movers who leave a city and the stability of those who stay follow from the anchor swap alone, the person held fixed. Arrivals are the exception: they contract faster than an anchor swap can

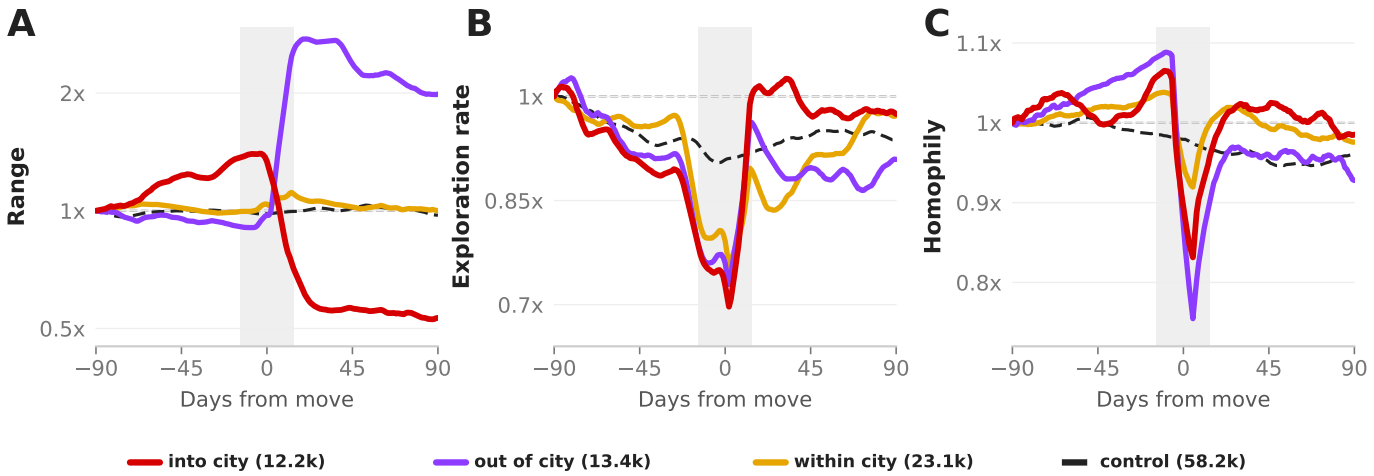


Fig. 2. A move reorganises mobility, and the destination sets the direction. Cohort medians against days from the move, averaged over a centred week; each cohort is indexed to its own level three months before the move (1×, dashed), so panels read as change relative to before; the shaded fortnight is the move window, where sampling also thins; matched controls are the secular baseline. **A**, Range is plastic and directional: movers into a city more than halve their orbit (6.9 to 3.1 km) while movers out nearly triple it (2.8 to 7.7 km), and within-city movers hold near 2.6 km. **B**, Exploration is elastic: it dips for everyone around the move and recovers, ending highest for those who change environment. **C**, Income homophily is disturbed and largely restored, dropping most for movers out of a city into an unfamiliar social mix. Visit entropy and raw place counts are omitted: over a fixed spatial grid they fall mechanically as an orbit compresses, a resolution artefact rather than behaviour.

produce, and the agent reproduces them only once the prior activity space is cut. A move into a city severs the old orbit; it does not simply re-anchor onto a new one.

The model also fixes the pace. It settles movers who leave a city into their wider orbit within days, and brings arrivals onto their compact one only over about a month, the slower relaxation of a routine rebuilt around new anchors. Because the agent is calibrated to each person and driven by mechanism, it runs forward under conditions that never occurred: a rezoning or a new development enters through the anchors, and the exposure it would produce is read off the simulation before it is built.

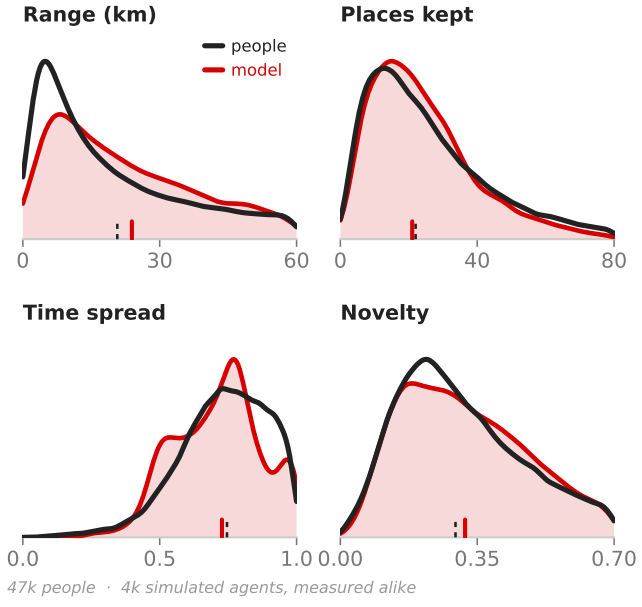


Fig. 3. The agents move like the people. Distributions of four mobility measures for 47,000 residents (dark) and 4,000 simulated agents (red), each agent drawn from the measured spread of traits and seen through the same sparse sampling before it is measured. Range, places kept and the spread of time across places line up; novelty runs a shade high. This is the generative model whose home and work anchors we swap to simulate a move.

5. DISCUSSION

Mobility reorganises around a move: rhythm tracks the person, range and exposure track the place, and the destination decides how. Exposure is the dimension a move rewrites most, and the mover's design turns it from a description into a quantity an intervention can be tested against, so experienced segregation acquires a policy variable that can be measured in advance. Vienna is the first city carried through; the cohort definition and the agent transfer unchanged to any city with the same traces.

REFERENCES

1. M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," Nature **453**, 779 (2008).
2. R. Chetty and N. Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," Q. J. Econ. **133**, 1107 (2018).
3. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large US cities," Nat. Commun. **12**, 4633 (2021).
4. C. Song, T. Koren, P. Wang, and A.-L. Barabási, "Modelling the scaling properties of human mobility," Nat. Phys. **6**, 818 (2010).

Cross-Metropolitan Transferability of Behavior-based Dependency Networks

BOYANG LI[1,2], SAUMITRA KULKARNI[3], MORGAN R. FRANK[4], ESTEBAN MORO[3,5], AND TAKAHIRO YABE[1,2,*]

[1] Department of Technology Management and Innovation, Tandon School of Engineering, New York University, Brooklyn, NY, USA
[2] Center for Urban Science and Progress, Tandon School of Engineering, New York University, Brooklyn, NY, USA
[3] Network Science Institute, Northeastern University, Boston, MA, USA
[4] Department of Design & Innovation, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH, USA
[5] Department of Physics, Northeastern University, Boston, MA, USA
[*] Corresponding author: Takahiro Yabe

1. INTRODUCTION

Human mobility links places into complex webs of economic interdependence, yet whether these dependencies arise from universal behavioral processes or locally specific activity patterns remains poorly understood. Human mobility itself is known to exhibit robust regularities across individuals, populations, and spatial contexts [1, 2]. Recent work introduced behavior-based dependency networks, showing that mobility creates invisible, often long-distance economic linkages between places that cannot be explained by geographic proximity alone [3]. However, whether these higher-order dependency structures inherit the universality of human mobility, or instead remain fundamentally local, remains unknown. This gap is also practical: because dependency networks require large-scale mobility data to observe, it remains unclear whether they can be inferred for cities and time periods where such data are unavailable. Resolving this question is essential for understanding the fundamental organization of urban dependency systems and their generalizability across metropolitan areas.

We address this question through cross-MSA transferability. If dependency networks are fundamentally local, models trained in one metropolitan area should fail when applied elsewhere. Conversely, successful transfer would indicate that urban dependency networks share common behavioral mechanisms across Metropolitan Statistical Areas (MSAs), while systematic variation in transferability would reveal how these mechanisms are modulated by metropolitan context. Using a nationwide dependency network and graph neural models trained separately within metropolitan areas, we find that dependency networks are substantially transferable across MSAs. Transferability is structured rather than uniform: metropolitan areas organize into behavioral regimes within which predictive performance is strongest. At the same time, substantial transferability persists across regimes, suggesting that urban dependency networks reflect shared behavioral mechanisms shaped by regional and socioeconomic variation.

2. DATA AND FRAMEWORK

We construct a nationwide behavior-based dependency network for the continental United States using one year of anonymized Cuebiq mobility data from 2021. The network contains 1.46 million building-level nodes aggregated from 5.28 million SafeGraph POIs, connected by 7.59 million inferred dependency relationships based on sequential visitation patterns [3]. Each node is enriched with POI attributes and neighborhood socioeconomic features from the American Community Survey (ACS). We then partition the nationwide network into Metropolitan Statistical Area (MSA) subnetworks using U.S. Census Bureau boundaries, retaining 6.55 million within-MSA dependency edges for analysis. Figure 1 presents the resulting MSA-level dependency networks.

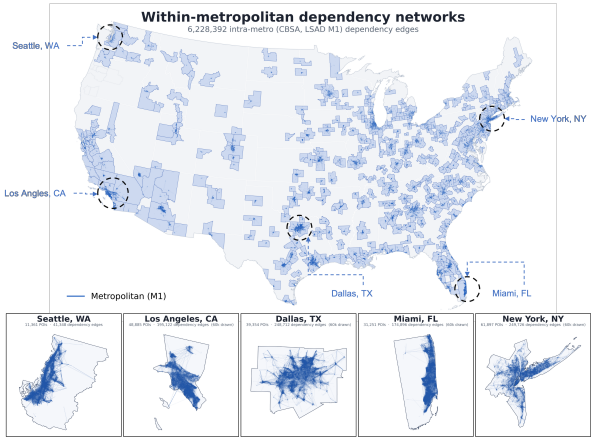


Fig. 1. Behavior-based dependency networks within Metropolitan Statistical Areas (MSAs). Edges denote inferred dependencies based on sequential human mobility patterns.

To quantify cross-MSA transferability, we train a single-layer GraphSAGE model on each source MSA to predict residual $\log(1+n_{ij})$ over a gravity baseline, where n_{ij} denotes the observed number of sequential visits between places i and j. We then apply each source-MSA model to every target MSA and measure transferability as the Pearson correlation between predicted and observed $\log(1+n_{ij})$ on the target test set. Repeating this procedure produces a directed cross-MSA transferability matrix, where each entry r_{ij} measures how well a model trained on MSA i generalizes to MSA j.

We analyze the resulting transferability matrix from two complementary perspectives. First, we symmetrize transferability as $(r_{ij}+r_{ji})/2$ and apply hierarchical clustering, with $k=8$ selected by the largest relative merge-height gap. Second, we use pairwise regression to explain variation in directed transferability r_{ij} as a function of differences in MSA size, socioeconomic composition, functional land-use structure, geographic separation, and network characteristics. Together, these analyses allow us to distinguish locally specific urban conditions from broadly shared behavioral regularities.

3. RESULTS

Transferability is structured rather than random. Figure 2 shows the cross-MSA transferability matrix after hierarchical reordering. Each entry is the Pearson correlation obtained by training

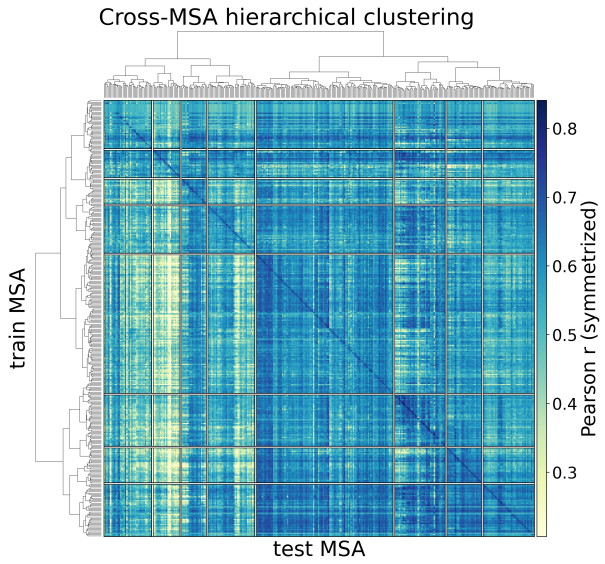


Fig. 2. Directed cross-MSA transferability matrix reordered by hierarchical clustering of the symmetrized matrix. Rows denote training MSAs and columns denote testing MSAs, and darker colors indicate higher Pearson correlation. Black lines indicate inferred clusters.

in one MSA and evaluating in another. We exclude 30 MSAs with in-domain $r_{ii} < 0.5$, where models fail to reliably recover dependency patterns even within their source MSA, leaving a 346×346 transferability matrix. Among the retained source models, cross-MSA transferability is substantial: even the lowest mean out-of-domain Pearson correlation exceeds 0.4. The matrix also exhibits a clear but non-uniform block structure, suggesting that dependency networks are neither purely local nor fully homogeneous. Instead, metropolitan areas share transferable dependency-formation mechanisms organized into partially distinct behavioral regimes.

Behavioral regimes emerge across metropolitan areas. Figure 3 maps the inferred transferability clusters onto the U.S. metropolitan system. The regimes are geographically structured but not geographically determined. C2 includes large established metropolitan economies such as New York, Chicago, San Francisco, and Miami. C6 is dominated by smaller college-town and university-oriented MSAs, while C8 captures older industrial or post-industrial metropolitan economies such as Detroit. Nearby MSAs often share transferability patterns, but distant MSAs with similar urban roles can also belong to the same cluster. This suggests that dependency formation reflects a common behavioral backbone modulated by metropolitan scale, functional specialization, and regional context.

What explains cross-MSA transferability? A baseline model using structural features explains 23.1% of the variation in r_{ij} (Table 1). Adding ACS-based socioeconomic features substantially increases adjusted R^2 to 0.437, corresponding to a gain of 20.6 percentage points. In contrast, POI composition and urban-form features add only modest explanatory power, increasing adjusted R^2 by 0.4 and 0.3 percentage points, respectively. This pattern suggests that transferability depends less on the detailed functional mix of places than on broader metropolitan context: models transfer best between MSAs with similar size, spatial proximity, and socioeconomic profiles.

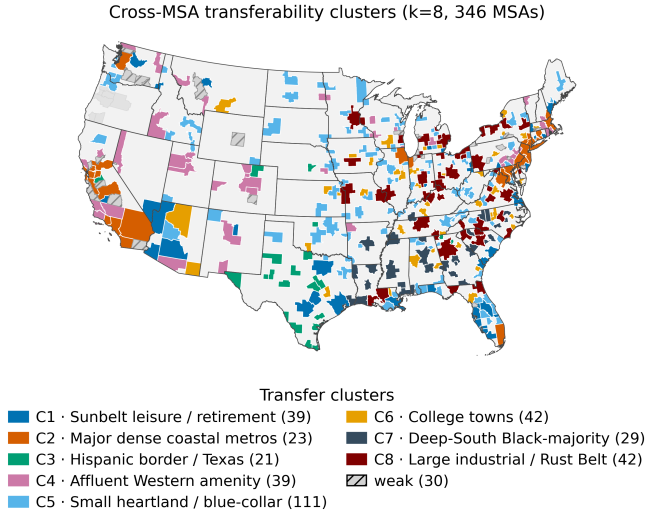


Fig. 3. Spatial distribution of cross-MSA transferability clusters across the United States. Colors indicate clusters inferred from the transferability matrix. Hatched gray areas denote MSAs excluded due to unstable in-domain model performance.

Table 1. Model comparison for explaining directed cross-MSA transferability.

Model	Added predictors	Adj. R^2	Gain
M1	Structural features	0.231	–
M2	+ ACS socioeconomic features	0.437	+0.206
M3	+ POI composition features	0.440	+0.004
M4	+ urban-form features	0.443	+0.003

4. DISCUSSION

These results suggest that the transferability of urban dependency networks is rooted in shared behavioral mechanisms underlying activity sequences. The strong role of ACS-based similarity is consistent with evidence that urban mobility reflects socioeconomic exposure beyond residential context [4]. ACS characteristics may therefore capture the populations, constraints, and activity needs that shape sequential connections between places. Rather than reflecting a universal network pattern, the shared organization of dependency networks emerges from a recurring logic by which similar populations connect places through similar visit sequences. This helps explain why models transfer across MSAs while still producing regimes shaped by local socioeconomic context and network organization.

REFERENCES

1. M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," nature **453**, 779 (2008).
2. O. Cabanas-Tirapu, L. Danús, E. Moro, *et al.*, "Human mobility is well described by closed-form gravity-like models learned automatically from data," Nat. Commun. **16**, 1336 (2025).
3. T. Yabe, B. García Bulle Bueno, M. R. Frank, *et al.*, "Behaviour-based dependency networks between places shape urban economic resilience," Nat. human behaviour **9**, 496 (2025).
4. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. communications **12**, 4633 (2021).

Do tsunami evacuation drills reproduce real evacuation response? Evidence from paired events in coastal Chile

KYRIAKI KALIMERI[1,2] AND LEO FERRES[1,3]

[1] ISI Foundation, Turin, Italy
[2] UNICEF, NYC, USA
[3] Faculty of Engineering, Universidad del Desarrollo, Chile
[*] kkalimeri@gmail.com

1. INTRODUCTION

Evacuation drills are widely used to assess tsunami preparedness, but their effectiveness is typically evaluated using attendance, self-reports, or other process measures rather than by comparing them with real evacuations [1–3]. Because real events are rare and difficult to observe at scale, it remains unclear how well drills reproduce actual behavior [4, 5].

We address this gap using aggregate mobile network data (XDRs) from the Valparaíso coastal region of Chile (Valparaíso, Viña del Mar, and Concón), comparing a real tsunami alert (30–31 July 2025) with a full-scale announced drill (2 October 2025) using the same analysis pipeline. We ask two questions: **(RQ1)** How closely does an announced drill reproduce real evacuation behavior? **(RQ2)** How unequal is access to safety, and does evacuation burden influence response?

Our study contributes a rare paired comparison of the same population observed during both a real emergency and a drill, extending recent work using mobile phone data to study evacuations across hazards [6–8]. Rather than measuring evacuation directly, we quantify changes in hazard-zone exposure and verify that these changes reflect population movement rather than network artefacts. Because the drill was pre-announced and the real evacuation followed several hours of prior official warning, our results should be interpreted as a comparison between these two operational scenarios rather than as a controlled experiment.

2. DATA AND METHODS

We analyse anonymized, aggregated mobile network data (XDRs) reporting the number of devices connected to each cell every 15 minutes, together with time-of-day matched baseline counts from reference days. We compare a real tsunami alert and a full-scale announced drill in the Valparaíso coastal region using the same analysis pipeline. The government-issued evacuation alert (SAE) defines T_0 for each event, and all comparisons are made at the same elapsed time after T_0. Attendance corrections are estimated separately for each commune using pre-alert periods only.

To ensure that the observed changes reflect population movement rather than network artefacts, we performed a series of robustness checks, including placebo tests on non-event days, bootstrap confidence bands, analyses restricted to common cells, sensitivity to attendance corrections, conservation of total device counts, and a detectability analysis to estimate movement that may remain invisible when trips occur within the coverage area of the same cell tower.

To quantify evacuation accessibility, we define *structural burden* as the terrain-adjusted walking distance from each census block to its nearest assembly point, computed on the pedestrian network using a 30-m digital elevation model. We then test whether structural burden influences evacuation by modelling attendance-corrected residual presence during the first 1–3 hours after the alert as a function of event type, structural burden, their interaction, and demographic controls, with commune fixed effects and Conley spatial-HAC standard errors.

3. RESULTS

A. Strong real-event evacuation; limited drill response

During the real event, hazard-zone presence dropped to 41% of its baseline within one hour of the alert and stabilized at 37–38% (Fig. 1). At the same time, the adjacent safe ring increased to 105–116%, consistent with movement from the coast towards safer inland areas. In contrast, the drill reduced hazard-zone presence by at most 20% and had almost returned to baseline (93%) by T_0+2h, well before the official all-clear.

Using comparable time windows, the real event produced a threefold larger peak reduction in exposure (63% vs. 20%) and a twelvefold larger reduction three hours after the alert (63% vs. 5%). The radius of gyration showed the same pattern, increasing by 4–7% during the real event but remaining within 1% during the drill.

Could the weak drill response simply reflect movement that is invisible at tower resolution? Our detectability analysis suggests not. Even under generous assumptions, at most 40% of exposure-reducing drill movement (46% with a threefold larger coverage radius) could have gone undetected, well below the 63% reduction observed during the real event. This result is also robust to attendance-correction sensitivity analyses (53–69%).

B. A staged response to warning escalation

The eight-hour interval between the alert state and the evacuation alarm reveals a two-stage behavioral response. Before dawn, presence in the hazard zone remained close to baseline (±2–4%). During the morning commute, however, it fell by about 13%, indicating that people were avoiding travel into the coastal area. By 11:45, presence was 31% below baseline and the adjacent safe ring had increased to 1.19, reflecting active departures. After the evacuation alarm, the reduction almost doubled within one hour, from 31% to 59%. Overall, the alert state accounted for roughly one third of the total exposure reduction, whereas the alarm triggered most of the evacuation.

C. Limits of drill-based forecasting

A forecast calibrated on the drill would have substantially underestimated the real evacuation. At the same time after the alarm, it would have underestimated exposure reduction by 58 percentage points (calibration ratio 0.08), by 43 points at the drill's peak (0.32), and by 23 points even under the most generous detectability assumptions (0.64). Equivalently, it would have greatly overestimated the number of people remaining in the hazard zone (Fig. 2). These results suggest that drill-based calibration alone is insufficient for operational planning.

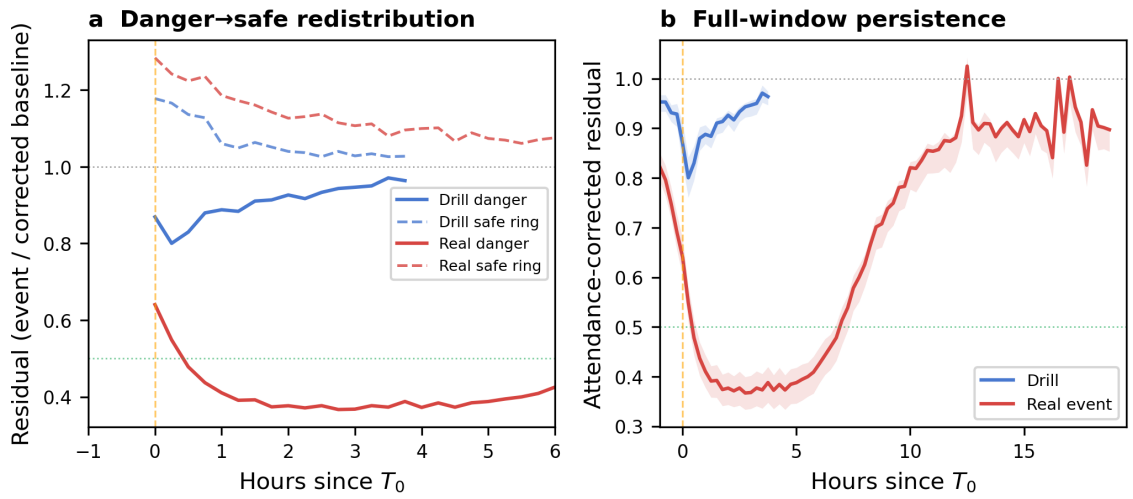


Fig. 1. Real alert vs. drill, hours since T_0. (a) Danger-to-safe redistribution over the first six hours: attendance-corrected residuals for the inundation zone (solid) and adjacent safe ring (dashed). (b) The danger-zone residual over each event's full official window (drill 4 h; real event 18.75 h) with 95% cell-bootstrap bands.

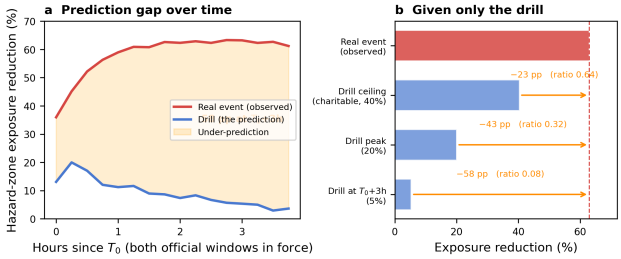


Fig. 2. The cost of drill-based calibration. (a) Exposure reduction for drill (prediction) and real event (observed) over the matched window; the shaded area is the planner's error. (b) Three readings of the drill against the observed 63%: under-prediction of 58, 43, and 23 percentage points.

D. Unequal accessibility does not explain evacuation response

Accessibility to assembly points is highly uneven. Overall, 52% of hazard-zone residents are more than 1 km from the nearest assembly point, and 40% are more than a 15-minute walk away (assuming 1.2 m s^{-1}). Terrain further increases travel effort, although only moderately (D_{eq}/D_{road} median 1.13; 90th percentile 1.24).

Despite these differences, we find no evidence that accessibility affected evacuation behavior. The real event reduced residual presence by 0.59 compared with the drill (Conley SE 0.09, $t = -6.6$), whereas both the main effect of D_{eq} and its interaction with event were small and not statistically significant (95% CI $[-0.08, 0.19]$ per km). Thus, while accessibility maps remain valuable for identifying where vulnerable populations are located, they do not predict differences in evacuation response over the 0.3–3 km range observed here.

4. DISCUSSION

Using the same population and analysis pipeline, we show that a real evacuation produced an order-of-magnitude larger response than a pre-announced drill. This is consistent with the Protective Action Decision Model [9], which predicts that protective behavior depends on perceived threat. A drill without environmental cues cannot fully reproduce real evacuation behavior.

Our findings have three implications. First, warning systems reduce exposure not only by triggering departures, but also by preventing people from entering hazardous areas before an evacuation order is issued. Second, mobile network data provide a scalable way to measure drill participation and compare drills with real events. Third, behavioral models calibrated solely on announced drills are likely to substantially underestimate real evacuation response and should instead be validated against observations from real emergencies whenever possible.

This study has limitations. The analysis relies on two-day baselines, cannot fully separate transit from stationary users, and is based on a single drill–event pair. However, the availability of recurring national evacuation drills provides a natural opportunity to replicate this analysis across different locations and hazards.

REFERENCES

1. C. Chen, A. Mostafizi, H. Wang, *et al.*, "Evacuation behaviors in tsunami drills," Nat. Hazards **112**, 845 (2022).
2. Y. Sun and K. Yamori, "Risk management and technology: Case studies of tsunami evacuation drills in Japan," Sustainability **10**, 2982 (2018).
3. N. Nakaya, H. Nemoto, C. Yi, *et al.*, "Effect of tsunami drill experience on evacuation behavior after the onset of the Great East Japan earthquake," Int. journal disaster risk reduction **28**, 206 (2018).
4. T. Takabatake and N. Hasegawa, "Tsunami evacuation behavior: a systematic review and proposal for a standardized research framework," Nat. Hazards **122**, 1 (2026).
5. F. Makinoshima, F. Imamura, and Y. Oishi, "Tsunami evacuation processes based on human behaviour in past earthquakes and tsunamis: A literature review," Prog. Disaster Sci. **7**, 100113 (2020).
6. S. D. Tirtha, D. A. Parvez, J. Lemp, *et al.*, "Understanding evacuation behavioral patterns using anonymized device data and high resolution parcel level data: application to hurricanes Irma and Matthew," Nat. Hazards **121**, 16857 (2025).
7. F. Makinoshima, S. Yotsui, S. Sato, and F. Imamura, "Large-scale geolocation data reveal evacuation behaviour during the 2024 Noto Peninsula earthquake and tsunami," Commun. Earth & Environ. **6**, 891 (2025).
8. E. Elejalde, T. Naushirvanov, K. Kalimeri, *et al.*, "Use of mobile phone data to measure behavioral response to SMS evacuation alerts," Int. J. Disaster Risk Reduct. p. 105919 (2025).
9. M. K. Lindell and R. W. Perry, "The protective action decision model: Theoretical modifications and additional evidence," Risk Analysis **32**, 616 (2012).

Comparing Gender-Labeled Urban Co-Visitation Networks Across U.S. Cities Using Google Places

JEMAL A JILO[1,*], MYRIAM DELGADO[1], AND THIAGO H SILVA[1,2]

[1] Universidade Tecnológica Federal do Paraná, Curitiba, Brazil
[2] University of Toronto, Toronto, Canada
[*] jemalabatejilo@alunos.utfpr.edu.br

1. INTRODUCTION

Digital traces from location-based platforms provide scalable observations of urban activity across neighborhoods and cities [1, 2]. However, large-scale gender-disaggregated mobility studies remain limited, partly because demographic information and comparable datasets are rarely available. Mobile-phone research has shown that women may visit fewer locations and concentrate their activity more strongly among them [3]. Recently, analyses of smartphone trajectories found women's mobility networks to be more clustered and home-anchored [4]. Studies using location-based social media have also identified gender differences, but have focused mainly on check-in distributions, temporal activity, and venue preferences [5, 6]. Consequently, less is known about whether gender-labeled groups produce different aggregate networks connecting urban areas, or whether they share the same strongest inter-area connections across cities.

We compare ZIP-code co-visitation networks generated by inferred male- and female-labeled Google Places reviewers across seven U.S. cities. Unlike continuous phone or GPS trajectories, reviews provide a consumption-oriented view of urban activity: two ZIP codes are connected when users review venues in both. Beyond individual-level mobility indicators, we examine how each group collectively connects urban areas, whether the resulting networks differ in global connectivity and local clustering, and whether the groups share the same strongest inter-area connections across cities.

Our contribution is a cross-city comparison of gender-labeled, aggregate urban co-visitation networks derived from venue-review activity. We find that male-labeled networks (MLN) are consistently more globally connected, whereas female-labeled networks (FLN) tend to be more locally clustered. We further show that the overlap and rank-sensitive agreement between the groups' strongest connections vary substantially across cities. Robustness checks evaluate whether these patterns remain stable under balanced group sizes and alternative network-construction choices. Although the findings capture only a subset of users' venue activity and gender inference may vary across cultural and linguistic groups, they nevertheless provide valuable insights into gendered mobility patterns.

2. DATASET

We use the Google Places Reviews dataset[1], whose observations are concentrated between 2015 and 2021. We retain reviewer identifiers, first names (used only to infer gender labels), venue identifiers, and venue coordinates. These fields associate users with reviewed venues and assign venues to ZIP codes. Table 1 summarizes the data. Male-labeled users outnumber female-labeled users in every city, motivating the imbalance checks in Section 3-C.

[1] "Uctopic: Unsupervised contrastive learning for phrase representations and topic mining," J. Li, J. Shang, and J. McAuley, Proc. ACL, 2022.

Table 1. Dataset statistics. Values are reported as male/female.

City	Users	Reviews	ZIPs
New York	4,648 / 2,660	409,104 / 212,616	213 / 213
Los Angeles	14,286 / 9,460	1,271,527 / 741,902	551 / 550
Chicago	7,404 / 4,833	633,592 / 373,108	300 / 299
Houston	6,106 / 4,760	515,901 / 349,193	208 / 207
Phoenix	7,100 / 5,193	578,695 / 375,808	194 / 193
Atlanta	2,533 / 1,888	187,592 / 126,024	130 / 129
Miami	6,795 / 4,546	542,532 / 312,574	218 / 216

3. METHODOLOGY

A. Gender-Label Inference

Because Google Places does not provide self-reported gender, we infer binary gender labels from reviewers' first names with the R `gender` package[2] through `rpy2`. It returns a predicted label and an associated confidence score. We retain only predictions with at least 95% confidence and use these male- and female-labeled groups solely for aggregate group-level comparisons. These labels are proxies and may not reflect users' self-identified gender or perform equally across cultural and linguistic groups.

B. Network Modeling and Evaluation

For each city c and inferred gender group g, we construct a weighted, undirected ZIP-code co-visitation network $G_{c,g} = (V_{c,g}, E_{c,g}, W_{c,g})$. Nodes represent ZIP codes, and an edge (i, j) connects two ZIP codes when a user from group g reviews venues in both. The weight w_{ij} is the number of shared users. Self-loops are removed, and we retain edges with $w_{ij} \geq 100$.

We report density, average degree, weighted local clustering, unweighted average shortest-path length on the largest connected component, and weighted degree assortativity. We rank edges by weight and compare groups using overlap@k and Rank-Biased Overlap (RBO) for $k \in \{20, 50, 100\}$. Overlap@k measures shared membership, whereas RBO measures rank-sensitive agreement. We use persistence parameter $p = 0.9$ and truncated RBO at depth k; edges with equal weights are ordered lexicographically by their ZIP-code pair.

To relate the aggregate network results to prior individual-level work, we also compute each user's review-based radius of gyration, defined as the root-mean-square distance of their reviewed venues from their geographic centroid, and compare group means within each city.

C. Robustness Checks

We assess sensitivity to representation imbalance and network construction. **Balanced-user resampling:** we downsample the larger group to the smaller group's size, reconstruct the networks, and repeat 100 times. **Edge retention:** we compare the baseline $w_{ij} \geq 100$ threshold with rules retaining the top 5%

[2] https://github.com/lmullen/gender.

and 10% of edges. **Null model:** for each city, we generate random weighted networks on the observed nodes, preserving total weight $t = \sum_{(i,j)\in E} w_{ij}$ by sampling t unordered node pairs with replacement. We repeat this 1,000 times and compare top-k overlap. **Imbalance diagnostic:** Across cities, we examine whether male–female differences in network metrics vary systematically with the corresponding ratios of users, reviews, and retained edges. We treat this analysis as a descriptive diagnostic rather than a formal test of association.

4. RESULTS

Figure 1 illustrates networks for New York City and Los Angeles.

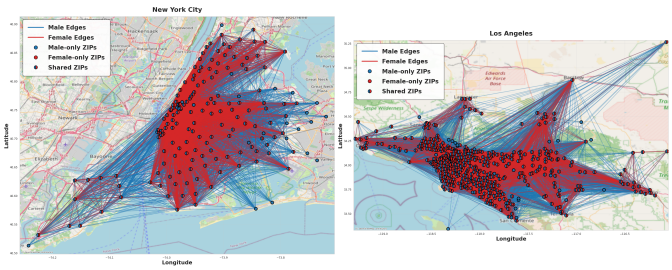


Fig. 1. Networks for New York City and Los Angeles

Across all seven cities, MLN are more globally connected than FLN: they have higher density and average degree, as well as shorter average shortest-path lengths. In contrast, FLN have higher clustering in all seven baseline comparisons, indicating more locally concentrated co-visitation patterns. Most networks are negatively assortative, suggesting that highly connected ZIP codes tend to connect with less-connected areas. MLN are more disassortative in most cities. Miami is the only city in which both networks are positively assortative (male: 0.028; female: 0.150). As an individual-level benchmark, female-labeled users have a lower mean review-based radius of gyration in all seven cities, consistent with prior work [3]. The male–female gap is largest in Miami (2.2 km) and Los Angeles (2.0 km), while the group means are nearly identical in Atlanta. This directional agreement complements, but does not explain, the aggregate network differences. Figure 2 summarizes the male–female differences in density, clustering, and average shortest-path length.

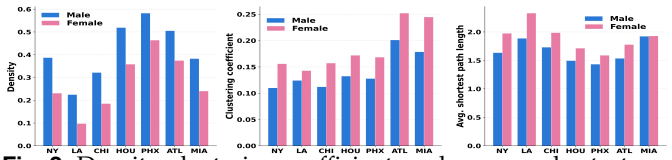


Fig. 2. Density, clustering coefficient, and average shortest-path length of male- and female-labeled networks across cities.

The similarity of the strongest connections varies substantially across cities. At $k = 100$, overlap is highest in New York and Houston (0.95 and 0.86, respectively) and lowest in Los Angeles and Chicago (0.50 and 0.61). RBO reveals a similar broad pattern while capturing differences in edge ordering. New York has the highest rank-sensitive agreement (RBO@100 = 0.82), followed by Atlanta (0.72), which ranks ahead of Houston (0.61) despite Houston's higher overlap@100. Miami and Los Angeles have the lowest RBO values (0.26 and 0.38, respectively). Phoenix and Chicago have quite similar RBO curves. Figure 3 presents both measures for $k \in \{20, 50, 100\}$.

Together, overlap and RBO distinguish whether the groups share the same strong connections from whether they prioritize those connections in a similar order. Moreover, observed overlap exceeds all 1,000 null-model realizations for every city and

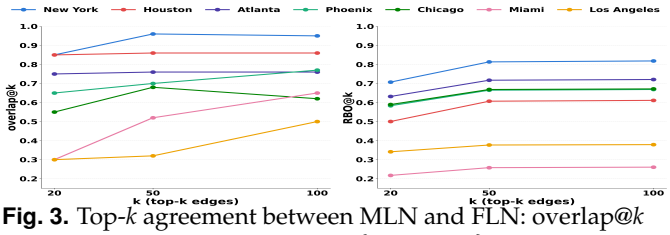


Fig. 3. Top-k agreement between MLN and FLN: overlap@k (left) and RBO@k (right), for $k \in \{20, 50, 100\}$. Higher values indicate greater agreement.

value of k ($p_{\mathrm{MC}} \leq 0.001$), indicating more shared top-ranked connections than expected under uniform random pairing.

The global-connectivity findings are robust to representation imbalance and network-construction choices. After balancing the number of users in each group, MLN remain denser in all seven cities. FLN retain higher clustering in six cities, with Los Angeles as the exception, indicating that the clustering tendency is broadly, but not universally, robust. The qualitative patterns are also largely preserved when retaining the top 5% or 10% of edges rather than applying the absolute threshold. E.g., Miami remains positively assortative under all filtering schemes: its male and female coefficients are 0.028 and 0.150 under the baseline threshold and 0.081 and 0.146 under top-10% retention. Finally, network-metric differences show no clear monotonic relationship with the ratios of users, reviews, or retained edges across cities. These analyses support the robustness of the main findings to group-size imbalance and edge-retention choices.

5. CONCLUSION

We compared gender-labeled ZIP-code co-visitation networks derived from Google Places reviews across seven U.S. cities. MLN were consistently more globally connected, whereas FLN tended to be more locally clustered. Female-labeled users also showed smaller mean review-based radius of gyration, consistent with previous mobility research. The groups shared more top-ranked connections than expected under uniform random pairing, while set overlap and rank-sensitive agreement varied across cities. The main network findings remained stable after balancing group sizes and changing edge-retention rules. These results show that aggregate metrics can conceal group specific differences in how urban areas are collectively connected.

ACKNOWLEDGMENTS

CNPq (processes 314603/2023-9, 441444/2023-7, 444724/2024-9, and INCT TILD-IAR 408490/2024-1).

REFERENCES

1. F. A. Santos, T. H. Silva, and L. A. Villas, "Real-up: Urban perceptions from lbsns helping moving real-estate market to the next level," in *Proc. of WWW*, (2024), p. 1071–1074.
2. G. Santos, F. Gubert, M. Delgado, and T. Silva, "Modeling interest networks in urban areas: A comparative study of google places and foursquare across countries," J. Internet Serv. Appl. **16**, 25–42 (2025).
3. L. Gauvin, M. Tizzoni, S. Piaggesi, *et al.*, "Gender gaps in urban mobility," Humanit. Soc. Sci. Commun. **7**, 11 (2020).
4. L. A. Sílvia de Sojo, Sune Lehmann, "Women mobility networks enable more efficient travel," in *Proc. of NetSci*, (2025).
5. M. Rizwan, W. Wan, O. Cervantes, and L. Gwiazdzinski, "Using location-based social media data to observe check-in behavior and gender difference: Bringing weibo data into play," ISPRS Int. J. Geo-Information **7**, 196 (2018).
6. W. Mueller, T. H. Silva, J. M. Almeida, and A. A. Loureiro, "Gender matters! analyzing global cultural gender preferences for venues using social sensing," EPJ Data Sci. **6**, 5 (2017).

Leveraging mobility behavior to re-identify users

LUCAS FÉLIX[1, 2,*], VINÍCIUS DA F. VIEIRA[3], ANNE J. KOUAM[1], ALEXANDRE G. EVSUKOFF[4], ALINE C. VIANA[1], NADJIB ACHIR[1], AND JUSSARA M. ALMEIDA[2]

[1] [1] *INRIA (France),* [2] *Universidade Federal de Minas Gerais (Brazil),* [3] *Universidade Federal de São João del Rei (Brazil),* [4] *Universidade Federal do Rio de Janeiro (Brazil),* [*] *lucasgsfelix@ufmg.com*

1. INTRODUCTION

Human mobility data have become an essential resource for understanding collective and individual movement patterns, enabling applications ranging from urban planning and traffic management to epidemic modeling and personalized location-based services [1]. However, the same fine-grained information that makes these data valuable also raises significant privacy concerns, as individual mobility patterns often contain distinctive characteristics that can reveal sensitive information about their owners. The privacy risks associated with mobility data have traditionally been studied from the perspective of raw spatio-temporal trajectories. In particular, a large body of work has demonstrated that a small number of visited locations or trajectory points is often sufficient to uniquely identify individuals or re-identify pseudonymized users across different datasets [2]. Consequently, most existing re-identification attacks rely directly on the unicity of raw mobility traces or on exact spatio-temporal matching between observations.

More recently, an alternative perspective has emerged by characterizing mobility through high-level behavioral descriptors rather than individual trajectory points. In this setting, the behavior is a set of processed information rather than extracted from the raw spatio-temporal trajectories. Thus, instead of modeling the user when and where, this approach summarizes the user characteristics by extracting mobility metrics that capture spatial, temporal, and structural behavioral patterns [3]. Although recent studies have proposed methods to quantify behavioral exposure and estimate users' vulnerability from these descriptors, it remains unclear whether such behavioral representations alone contain sufficient information to support an effective re-identification attack. In particular, to the best of our knowledge, no prior work has investigated whether a vector composed exclusively of mobility-derived behavioral metrics can serve as the sole basis for re-identification. Motivated by this gap, we investigate the following research question: **Can users be re-identified solely from their mobility behavior?** To answer this question, we propose a behavior-based re-identification attack that learns user-specific behavioral signatures from mobility feature vectors and evaluates whether these high-level representations alone are sufficient to recover users' identities.

2. METHODOLOGY

The proposed behavior-based re-identification attack assumes that the mobility dataset is divided into two disjoint subsets. The first contains trajectories identified by the users' pseudonymized identity and represents the adversary's background knowledge, while the second contains real-identity trajectories and constitutes the inference knowledge. Although both datasets originate from the same user population, no explicit correspondence between their identifiers is available. Consequently, the attack aims to recover these hidden associations by exploiting behavioral similarities extracted from mobility traces. We refer to these datasets as the **training set** and the **inference set**.

Each mobility trajectory is transformed into a behavioral feature vector by computing a set of mobility descriptors that summarize complementary aspects of human mobility. It should be noted that the number of behavioral feature vectors is equal to the number of trajectories that a user has. The metrics that compose the behavior are can be categorized in 3 groups: spatial (e.g., Radius of Gyration (RoG), 2kRoG, Avg. and Std. Jump Lengths, # of Visited Locations, # of Unique Locations, Max. Distance), temporal (e.g., Avg. Std. Waiting Time), and structural characteristics (e.g., Entropy, Regularity, Diversity, Stationarity), providing a compact representation of individual mobility behavior. Although other behavioral metrics could be incorporated, we selected these metrics because they have been shown to capture highly unique mobility behaviors [3], making them particularly suitable for user re-identification.

To learn user-specific behavioral profiles, a predictive model is trained and parameter-optimized **for each user in the training set**. The target variable is binary, indicating whether a behavioral vector belongs to the target user (positive class) or to any other individual (negative class). It is important to note that each user must have a sufficient number of behavioral vectors (i.e., training instances) to enable the model to learn their mobility behavior. To this end, we treat each day of a user's mobility history as an individual trajectory, thereby increasing the number of training instances available for model training. This definition can be naturally extended when longer observation periods are available (e.g., by considering each week as a single trajectory).

Instead of adopting a multi-class formulation, the proposed approach favors independent binary classifiers because they enable each model to specialize in recognizing a single behavioral profile while naturally supporting the one-vs-rest decision process required during inference. Moreover, in large datasets with an extensive number of users, such as the one held by mobile companies, multi-label classification algorithms may have an extreme classification problem, i.e., a scenario where the number of categories is massive.

During the **inference step**, each behavioral observations extracted from each trajectory in the inference dataset are predicted using the models trained for each the users in the **training set**, generating a matrix of probabilities $|P| = |U^I| \times |U^T|$, where $|U^I|$ and $|U^T|$ is the number of inference trajectories and training set respectively. Finally, the identity of each mobility trajectory is given by the model that yields the highest probability. Consequently, **the matrix P is composed of probabilities per trajectory**.

3. EXPERIMENTAL EVALUATION

Experimental setup: The mobility data used in this work are derived from anonymized Call Detail Records (CDRs) provided by a major mobile network operator for academic research. Because user locations are recorded only during communication events, such as phone calls or text messages, CDR data are inherently sparse. Although the original dataset comprises nearly two million users across the state of Rio de Janeiro in the last 6 months of 2012, most of the users have a very small number of records per day. Consequently, we select users with at least 10 records per day over one week. This strict filtering criterion ensures that the daily behavioral metrics extracted for each user are sufficiently informative and reliable for the subsequent analysis. We also restricted the observation period to August through the end of November, excluding vacation and holiday periods in Brazil. In total, the resulting dataset comprises 2125 users. However,

Scenario	total # users	# of trajectories	Avg. # trajectories per user	Dataset
Closed set	904	26630	29.47 ± 1.23	Train
	904	13109	14.50 ± 0.64	Inference
Open set	2125	44185	20.79 ± 0.67	Train
	1452	19108	13.15 ± 0.43	Inference

Table 1. Basic statistics of each evaluation scenario

	Scenario	
Metrics	Closed-set	Open-set
Accuracy	25%	14.88%
Recall	20%	8%
Precision	25%	9%
F1-Macro	20%	7%
Avg. Prob. Correct Prediction	87.07% ($\pm$ 0.71)	88.4% ($\pm$ 0.69)
Avg. Prob. Wrong Prediction	72.69% ($\pm$ 0.55)	72.67% ($\pm$ 0.54)
Coverage	62.61%	36.50%

Table 2. Overall statistics for each scenario

we train a model only for users with more than 10 trajectories, as this is a sufficient number of training instances required for the model to learn meaningful behavioral patterns. It is worth noting that users with fewer than 10 trajectories are not removed from the training set, as they may contribute behavioral variability that helps protect the identity of other users by acting as potential sources of noise. We temporally split the data into training and inference sets, using 70% of the behavioral observations for training and the remaining 30% for inference. Under this temporal split, there is no guarantee that the users present in the training set also appear in the inference set. Therefore, we evaluate the proposed methodology under two different settings: (i) a **closed-set** scenario, in which all users present during inference are also observed during training, and (ii) an **open-set** (realistic) scenario, in which no such assumption is made. The former evaluates whether the model can correctly identify users based solely on their behavioral patterns, whereas the latter assesses its performance under more realistic deployment conditions. The Table 1 presents the information regarding each experiment. For all average-based results, we also present the confidence interval with 95% of confidence. Finally, we use XGBoost model to learn the user behavioral patterns. We select XGBoost as it presents a very good performance for tabular tasks, and high computational performance. The hyperparameters of each model are optimized using the Optuna library. The performance of the attack is evaluated by observing matrix P.

Results: Table 2 presents the evaluation of the models in each scenario. Analyzing classic classification metrics such as accuracy, recall, precision, and F1-macro, it is noticeable that the difficulty is present for both closed-set and open-set scenarios. Such difficulty is justified by the several characteristics of the problem: (i) **behavioral ambiguity**, where users exhibit similar behavioral patterns, making them harder to distinguish and therefore naturally protected; (ii) **behavioral drift**, in which users change their behavior between the training and inference phases; and (iii) **out-of-distribution users**, i.e., users that are present in the training set but not in the inference set, and vice versa, which solely affects the **open-set scenario**. Nonetheless, we highlight that post-processing techniques may be applied during the user identity selection stage, as the average prediction probability is higher when the model is correct than when it is incorrect. This indicates that the model exhibits greater confidence in certain predictions, which could be exploited to improve the final identity assignment. Such a strategy could potentially avoid the out-of-distribution users and cases of ambiguity.

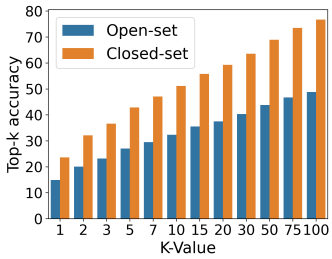


Fig. 1. Top-k accuracy for each of the evaluation scenarios

Analyzing the models' coverage, i.e., the percentage of users correctly identified at least once, we observe that nearly 62% of the users are re-identified at least once in the closed-set scenario, while approximately 36% are correctly re-identified in the open-set scenario. These results indicate that a substantial proportion of users can potentially be re-identified at least once. Although the coverage is lower in the open-set scenario, it remains considerable, especially given that a portion of the user identities is not available during the training phase, making their correct re-identification inherently impossible. This highlights that the behavior can be potentially used to re-identify the users.

Corroborating these findings, Figure 1 presents the top-k accuracy of the proposed methodology. The results show that the set of candidate identities can be substantially reduced while still achieving nearly 80% accuracy in the closed-set scenario and approximately 50% accuracy in the open-set scenario. We emphasize that, even for a relatively large value of k (e.g., $k = 100$), the number of candidate identities is substantially reduced during the inference phase. Specifically, this corresponds to a reduction of approximately 89% from the initial pool of 904 users in the open-set scenario and 93% from the initial pool of $1,452$ users in the closed-set scenario. These results indicate that the users' behavior can potentially be exploited to re-identify users, even when their identities or raw locations are obfuscated. Furthermore, the proposed behavioral attack could be combined with traditional spatio-temporal attacks, reducing the ambiguity in the candidate set and consequently increasing the likelihood of successful user re-identification.

4. CONCLUSIONS

In this paper, we propose a novel methodology for the re-identification of mobility datasets. The proposed approach leverages users' behavioral mobility patterns to perform the re-identification task. Within this context, we address the following research question: **Can users be re-identified solely based on their behavioral mobility?** Given our experimental setup, the answer is yes. However, we emphasize that these results are preliminary and should be interpreted within the scope of the evaluated scenarios. Nevertheless, they provide valuable insights into the inherent difficulty of the problem.

Acknowledgments: This work is part of the Mob Sci-Dat Factory project (ANR-23-PEMO-0004) under the France 2030 program. It was supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), by *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG), by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), and by the CAPES-STIC-AMSUD 22-STIC-07 LINT project.

REFERENCES

1. F. Asgari, V. Gauthier, and M. Becker, "A survey on human mobility and its applications," arXiv preprint arXiv:1307.0814 (2013).
2. Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," Sci. reports **3**, 1376 (2013).
3. L. G. Félix, A. J. Kouam, A. C. Viana, *et al.*, "How behavioral exposure reveals privacy risks beyond location uniqueness," in *ESORICS*, (2026).

Large Language Models Create an Uneven Informational Layer over Cities

LIN CHEN[1,*], GUANGYUAN WENG[1,2], AND ESTEBAN MORO[1]

[1] Network Science Institute, Northeastern University, Boston, MA 02148, USA
[2] Khoury College of Computer Sciences, Northeastern University, Boston, MA 02148, USA
[*] l.chen2@northeastern.edu

1. INTRODUCTION

Cities are organized not only by physical infrastructure, but also by information systems that guide how people perceive, navigate, and choose within urban space [1, 2]. With GPS routing [3] and review platforms effectively redistributing traffic and visits [4], large language models (LLMs) move this mediation to an even earlier stage of urban decision-making: rather than ranking a visible set of alternatives, they often construct the choice set itself when users ask where to eat, shop, or spend time [5]. This shift can be consequential because LLM responses do not expose the boundary between what has been included and what has been left out. Meanwhile, the underlying physical city is already deeply unequal, with neighborhoods differing in income, commercial density, and experienced mobility [6–9]. If LLM-generated visibility varies with these neighborhood conditions, it may add a representational layer to existing urban inequality. Prior work shows that LLM outputs can be socially and geographically uneven [10, 11], but it remains unclear whether such unevenness reflects correctable knowledge gaps or a more persistent form of selective attention. We address this question with a neighborhood-scale audit of restaurant recommendations across five U.S. cities, showing that LLMs fabricate venues, selectively ignore real establishments even when grounded in verified candidates, stratify recommendations by user description, and could redirect urban consumption flows.

2. AUDIT DESIGN

We audit LLM-generated urban information across 304 neighborhoods in Boston, New York, Chicago, Houston, and San Francisco. We focus on restaurants, because dining is a high-frequency urban decision that spans the full socioeconomic spectrum, with observable downstream consequences through large-scale mobility and transaction data. For each city-defined neighborhood [12], we generate factorial user profiles crossing household income (5 levels), age (8), sex (2), and residential status (4), yielding 320 profiles per neighborhood. For each profile, we prompt the LLM to provide ten restaurant recommendations near each neighborhood centroid. To disentangle the two failure hypotheses, we employ parallel experimental settings, one open-ended and one constrained to verified local venues. We experiment on models from three families (`GPT-4o-mini`, `Llama-3.3-70b-instruct`, and `Gemini-2.0-flash`). We measure (i) hallucination rate, the share of generated names that cannot be matched to a real establishment, and (ii) Neighborhood Algorithmic Invisibility Rate (NAIR), the share of real candidate venues receiving zero recommendations across all profiles.

3. FINDINGS

LLMs unevenly fabricate and overlook local places. In the open-ended setting, 36.8% of surfaced venues cannot be matched to real local establishments. Hallucination is higher in neighborhoods with fewer visits, fewer Yelp reviews, lower population density, and lower bachelor's-degree share. At the same time, most real venues remain absent from the generated answers: open-ended NAIR averages 96.8%, with strong neighborhood heterogeneity (Fig. 1). Factual reliability and visibility are therefore weakest where the city leaves thinner digital and mobility traces.

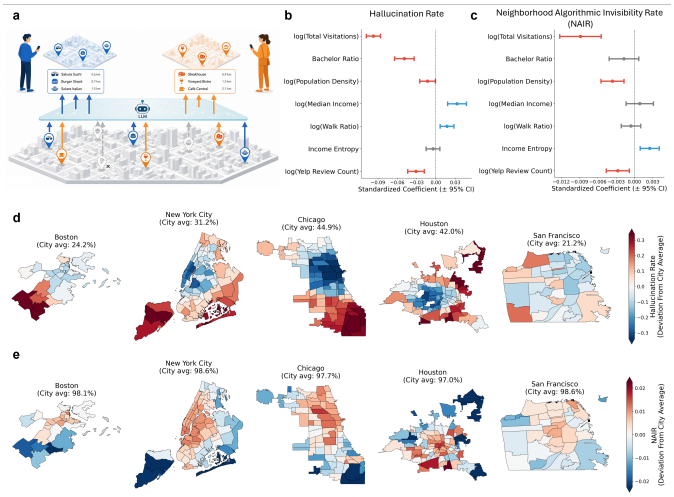


Fig. 1. Open-ended LLM recommendations are spatially uneven. Hallucination and algorithmic invisibility vary across neighborhoods and are associated with local digital, mobility, and socioeconomic footprints.

Grounding removes fabrication, not selective attention. Providing verified candidates eliminates hallucination, but the attention bottleneck remains. In the candidate-constrained setting, 47.5% of real venues are never recommended after all 320 profiles are exhausted (Fig. 2). Moreover, 31.9% of venues ignored by one model family are also ignored by the other two, far exceeding model-specific blind spots. These shared omissions indicate an allocative failure: the models repeatedly select from the same narrow slice of the city even when all options are known.

Demographic descriptions stratify recommendations. Holding the candidate pool fixed, user descriptions change which venues surface (Fig. 3). Higher-income profiles receive more expensive and less visited restaurants. Tourists are directed toward farther, costlier, and more socially diverse venues than local residents. These patterns mirror documented regularities in urban mobility and inequality [8, 9], suggesting that LLMs personalize from demographic expectations rather than observed behavior.

Visibility can redirect consumption. We project a counterfactual in which 10% of dining choices in each neighborhood shift from observed visitation shares to LLM recommendation shares, then convert visit changes into revenue using venue-level spending (Fig. 4). The shift moves attention away from chain and quick-service restaurants toward independent and full-service establishments. Conversational recommendations can therefore become an infrastructure for reallocating economic opportunity across urban places.

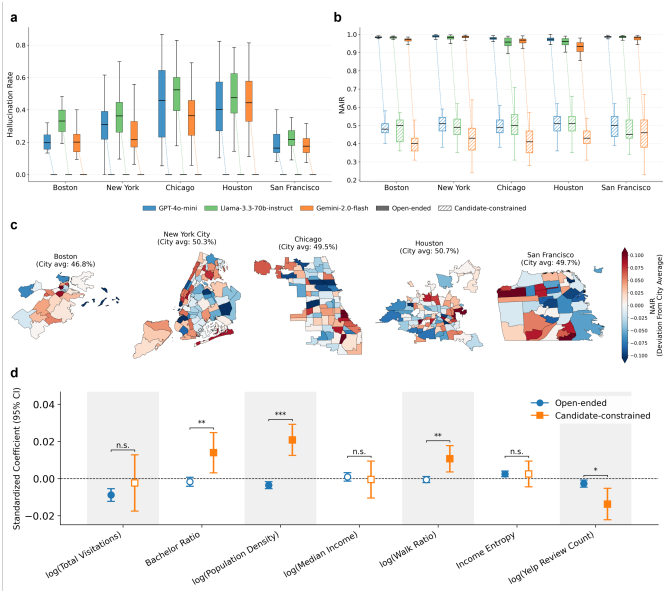


Fig. 2. Grounding removes hallucination but not invisibility. Candidate-constrained prompts eliminate false venues while leaving many real venues unseen.

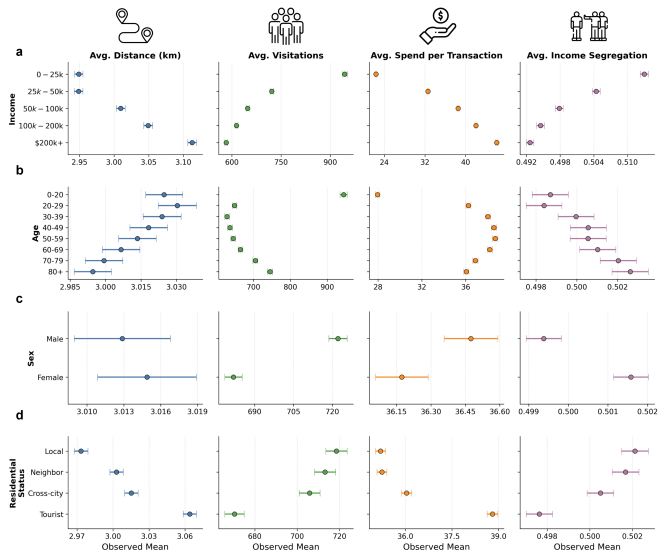


Fig. 3. LLM recommendations vary by user description. Within the same candidate pools, demographic attributes shift the price, popularity, distance, and social-mixing profile of recommended venues.

4. DISCUSSION

Our results suggest that LLMs are emerging as an urban information infrastructure that allocates visibility. They do not merely summarize local knowledge; they decide which places enter consideration before mobility occurs. The audit reveals both epistemic and allocative failure. Supplying verified local candidates removes fabricated venues, but algorithmic invisibility persists and many blind spots are shared across model families. Thus, better retrieval can improve factual grounding, but it does not by itself ensure that known places are represented broadly or equitably.

This matters because LLM-generated visibility can feed back into the city. Venues that are recommended may attract visits, reviews, and transaction traces, while omitted venues may become even less represented in the data on which future systems depend. A similar mechanism may operate across user groups: if income or residential status consistently narrows the slice of the city shown to a user, subsequent behavior can reinforce those associations. Future urban AI audits should therefore report not

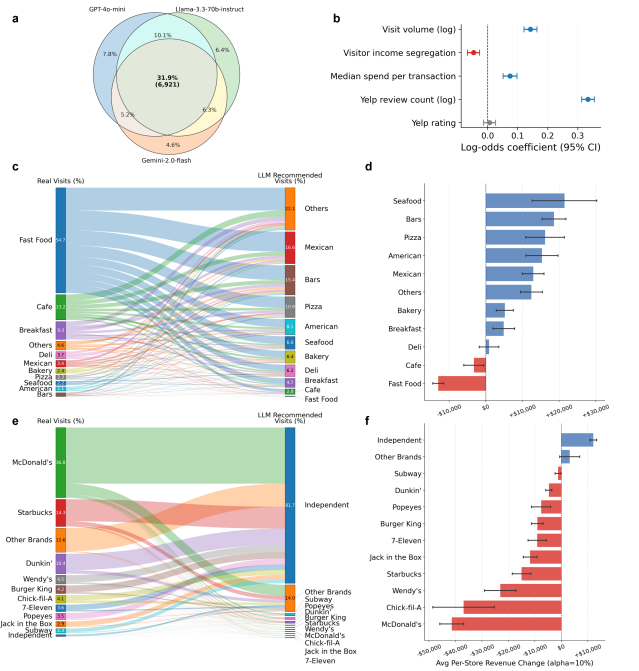


Fig. 4. LLM visibility redirects projected visits and revenue. Counterfactual demand shifts favor independent and full-service venues over chains and quick-service formats.

only factual accuracy, but also coverage, shared blind spots, and demographic stratification across the places that systems choose to make visible.

REFERENCES

1. S. Barns, *Platform urbanism: Negotiating platform ecosystems in connected cities* (Springer Nature, 2019).
2. M. Graham, "Regulate, replicate, and resist–the conjunctural geographies of platform urbanism," Urban geography **41**, 453 (2020).
3. T. Cabannes, M. A. S. Vincentelli, A. Sundt, *et al.*, "The impact of gps-enabled shortest path routing on mobility: a game theoretic approach," in *Transportation Research Board 97th Annual Meeting. Washington DC, USA,* , vol. 7 (2018).
4. M. Luca, "Reviews, reputation, and revenue: The case of yelp. com," Com (March 15, 2016). Harv. Bus. Sch. NOM Unit Work. Pap. (2016).
5. P. Li, M. de Rijke, H. Xue, *et al.*, "Large language models for next point-of-interest recommendation," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval,* (2024), pp. 1463–1472.
6. R. J. Sampson, *Great American city: Chicago and the enduring neighborhood effect* (University of Chicago press, 2012).
7. W. J. Wilson, *The truly disadvantaged: The inner city, the underclass, and public policy* (University of Chicago press, 2012).
8. E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," Nat. communications **12**, 4633 (2021).
9. F. Xu, Q. Wang, E. Moro, *et al.*, "Using human mobility data to quantify experienced urban inequalities," Nat. Hum. Behav. pp. 1–11 (2025).
10. M. Kamruzzaman, H. M. Nguyen, and G. L. Kim, ""global is good, local is bad?": Understanding brand bias in llms," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing,* (2024), pp. 12695–12702.
11. S. Dudy, T. Tholeti, R. Ramachandranpillai, *et al.*, "Unequal opportunities: Examining the bias in geographical recommendations by large language models," in *Proceedings of the 30th International Conference on Intelligent User Interfaces,* (2025), pp. 1499–1516.
12. S. Ansolabehere, J. R. Brown, R. D. Enos, *et al.*, "City-defined neighborhood boundaries in the united states," Sci. Data **12**, 1031 (2025).

Do LLM Agents Reproduce Human Mobility Laws?

GUSTAVO H. SANTOS[1,2,*], ALINE CARNEIRO VIANA[2], AND THIAGO H. SILVA[1,3]

[1] UTFPR, Curitiba, Brazil
[2] INRIA, Paris, France
[3] University of Toronto, Toronto, Canada
[*] gustavohenriquesantos@alunos.utfpr.edu.br

1. INTRODUCTION

Large Language Models (LLMs) are increasingly used to simulate urban behavior and human mobility. Frameworks like *AgentSociety* [1] and *CitySim* [2] deploy generative agents capable of context-aware routines. However, this emerging paradigm raises a central question: *Do synthetic agents reproduce empirical human mobility laws, or merely generate plausible trajectories?*

Decades of research demonstrate that human mobility follows complex but predictable patterns, including truncated power laws in travel distance [3], visitation dynamics [4], and recurrent mobility motifs [5]. While traditional simulators are rigorously validated against these empirical metrics [6], LLM-driven models are typically evaluated only on "face validity" [1, 2]—assessing whether schedules appear narratively coherent.

Plausibility does not guarantee objective mobility realism. An agent with a believable daily routine may still exhibit unrealistic spatial exploration or random visitation frequencies. We define *mobility realism* as the joint reproduction of empirical spatial visitation dynamics and temporal activity organization. Systematic evaluation of these coupled spatio-temporal properties remains largely absent in current LLM simulation literature [6].

To address this gap, we introduce *the first comprehensive validation framework for LLM-based urban simulators grounded in real-world mobility data and established empirical laws [7].* Our findings emphasize that rigorous empirical validation must become standard practice alongside plausibility-oriented assessments.

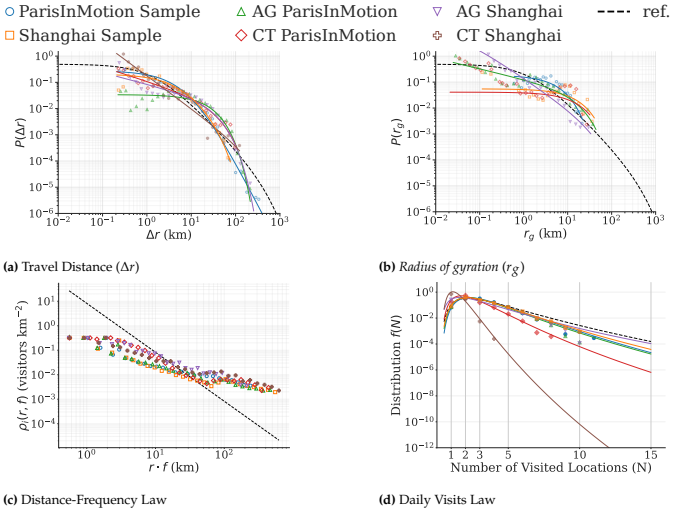
We evaluate the simulators using two real-world datasets: **GreaterParis dataset** [8], an anonymized 7-day GNSS dataset of Île-de-France residents featuring over 80,000 semantically annotated trips (purpose, mode) mapped to H3 level-10 cells; and **Shanghai** [9], a 10-day Cellular Data Record (CDR) dataset capturing hourly mobility for 58,502 users across 10,396 grid cells, which offers broader population coverage but lacks semantic labels. For our analysis, we extract representative cohorts from each: the GreaterParis dataset sample comprises 504 individuals with complete 7-day records within the simulated area, while the Shanghai sample contains 500 users selected to preserve empirical mobility profiles. A second, disjoint 500-user sample from Shanghai is utilized as a baseline reference for comparisons.

2. MOBILITY REALISM EMPIRICAL RESULTS

This section reports a brief comparison of results between simulated and real-world mobility datasets, while the full report is here [7]. We use the two real-world datasets and compare them against 3 simulation runs each on the *AgentSociety* and *CitySim* simulators.

A. Spatial and Temporal Mobility Realism

While LLM-based agents capture overarching macro-level mobility laws (Figure 1), they struggle to accurately reproduce individual-level spatial behavior. As Table 1 shows, simulator errors significantly exceed empirical sampling variability. In AgentSociety, unstable POI selection produces trajectories driven by local map-density artifacts. Although CitySim halves individual spatial errors (Δr and r_g), it still fails to yield realistic large-scale mobility dynamics, evidenced by weak Origin-Destination matrix agreement.



(a) Travel Distance (Δr) **(b)** *Radius of gyration* (r_g) **(c)** Distance-Frequency Law **(d)** Daily Visits Law

Fig. 1. Comparison of empirical mobility-law distributions between real and simulated trajectories across GreaterParis and Shanghai. AG refers to *AgentSociety* and CT to *CitySim*.

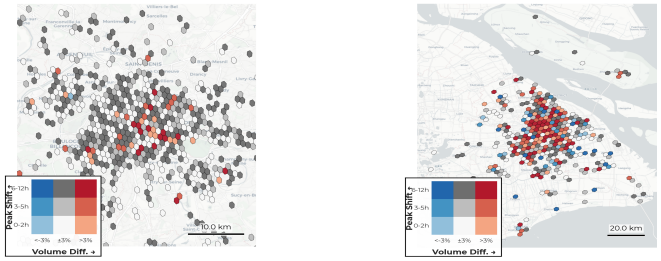
Table 1. Comparison of real vs simulated distribution of spatial mobility and spatio-temporal realism metrics. *STVD* and *OD matrix* results are shown for the H3 resolution 8.

Dataset	Source	Δr (W_1)	r_g (W_1)	*STVD* (W_1)	*OD Matrix* (*CPC*)
GreaterParis	*AgentSociety*	14.83 ± 0.35	7.29 ± 0.28	607.69 ± 5.88	0.088 ± 0.002
	CitySim	7.53 ± 1.67	3.47 ± 0.23	748.71 ± 23.00	0.138 ± 0.018
Shanghai	*AgentSociety*	8.73 ± 0.48	4.30 ± 0.67	113.88 ± 39.04	0.046 ± 0.008
	CitySim	3.97 ± 0.11	4.86 ± 0.09	432.55 ± 17.11	0.042 ± 0.013
	Reference sample	0.48	1.01	6.29	0.092

These spatial flaws inevitably propagate into unrealistic spatiotemporal and temporal dynamics. Figure 2 reveals that AgentSociety's simpler, gravity-based destination selection actually yields lower Spatio-Temporal Visit Distribution (STVD) errors than CitySim's complex LLM-driven approach. Still, both models exhibit severe discrepancies: GreaterParis agents capture spatial distributions but shift peak visitation by 3 to 12 hours (predicting where better than when), while the Shanghai simulation diverges sharply in the urban core. Consequently, these spatial deviations severely disrupt temporal realism (Table 2). AgentSociety systematically underestimates daily visits, and while CitySim improves individual spatial metrics, its complex goal-oriented POI selection causes excessive decision times that actively degrade temporal consistency. Ultimately, true mobility realism requires the joint reproduction of spatial and temporal dynamics; optimizing one in isolation undermines the other.

B. Mobility Motifs

Simulated trajectories successfully reproduce dominant routine motifs but fail to capture the full topological diversity of empirical daily mobility structures. Across both datasets, synthetic agents overproduce



(a) GreaterParis **(b)** Shanghai

Fig. 2. *Spatio-Temporal Visit Distribution* comparing real and simulated mobility patterns.

Table 2. Temporal mobility realism metrics comparing real and simulated trajectories.

Dataset	Source	*TD.* (min) (W_1)	*DT.* (h) (W_1)	*Vf.* (W_1)
GreaterParis	*AgentSociety*	24.41 ± 0.24	3.96 ± 0.12	9.63 ± 0.19
GreaterParis	*CitySim*	24.32 ± 18.93	30.24 ± 29.63	19.46 ± 3.62
Shanghai	*AgentSociety*	9.30 ± 0.24	4.79 ± 0.37	12.31 ± 4.19
Shanghai	*CitySim*	19.30 ± 0.27	36.39 ± 4.91	24.95 ± 0.65
Shanghai	Ref. sample	0.29	0.26	2.07

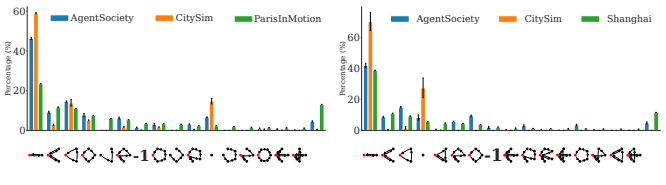
simple two-node motifs, such as alternating between home and work, while missing the complexity of real-world routines (Fig. 3). In the GreaterParis, these mismatches largely stem from open-ended trajectories that produce unclosed loops, whereas in Shanghai, the higher spatial granularity of real data generates complex motifs exceeding six nodes that simulators struggle to replicate. Although LLM-based agents capture coarse, mid-tier regularities, their motif distribution errors exceed baseline sampling variability, highlighting their current inability to reflect the nuanced topological complexity of real human mobility.

C. Behavioral Mobility Profiles

Simulated trajectories partially capture the exploration-return dynamics of human mobility, broadly aligning with established behavioral profiles: Routiners exhibit high regularity and stationarity, whereas Scouters display higher diversity and entropy. In the GreaterParis dataset, the profile-distribution Jensen-Shannon Divergence (JSD) remains low for both AgentSociety (0.13 ± 0.08) and CitySim (0.09 ± 0.01). However, generation artifacts frequently distort these dynamics, particularly in Shanghai. Instability in POI selection often causes agents to erratically search for unavailable destinations, artificially inflating exploratory behavior and creating extreme Scouter outliers. Conversely, a subset of agents classified as Routiners exhibit a degenerate pattern rather than true routine behavior, remaining at home for the entire simulation with maximum stationarity but near-zero regularity.

D. Semantic Mobility Dynamics

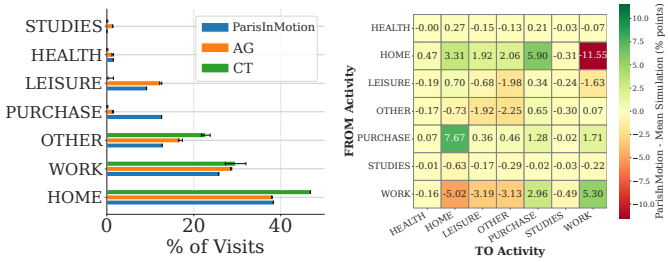
LLM-based agents reproduce first-order semantic activity distributions more reliably than sequential semantic dynamics. As illustrated in Figure 4 for the GreaterParis dataset, AgentSociety broadly matches empirical Visit Purpose Distributions, achieving a low divergence (0.0296 ± 0.0010) despite an underrepresentation of



(a) GreaterParis Motifs **(b)** Shanghai Motifs

Fig. 3. Comparison of the most frequent daily *mobility motifs* observed in real and simulated trajectories.

purchase-related activities likely caused by POI classification limits. In contrast, CitySim's more complex destination-selection mechanisms actually reduce semantic robustness when compatible POIs are sparse, yielding higher errors across all metrics. Crucially, the comparatively large divergence in the Activity Transition Matrix (ATM) across both simulators demonstrates that generating plausible aggregate activity proportions is significantly easier than accurately reproducing the ordered, sequential transitions that characterize real-world human mobility.



(a) *Visit Purpose Distribution* **(b)** *Activity Transition Matrix*

Fig. 4. Semantic mobility patterns for the GreaterParis.

3. CONCLUSION

This study introduces an empirical evaluation framework for LLM-driven urban simulators, assessing AgentSociety and CitySim against real-world mobility data from Greater Paris and Shanghai. While current models capture coarse semantic distributions and simple repetitive routines, they exhibit substantial discrepancies in large-scale spatial flows, temporal rhythms, sequential activity transitions, and heterogeneous exploratory behavior. These findings emphasize that generating narratively plausible trajectories fundamentally differs from reproducing empirically grounded urban mobility dynamics. Fine-tuning LLMs to handle this kind of context and focusing on these issues may help future simulations. Although our evaluation is constrained by the use of a CitySim reimplementation and the lack of semantic annotations in the Shanghai dataset, this work underscores the critical need for future simulators to incorporate richer empirical data and stronger spatial constraints to achieve behaviorally grounded urban mobility.

REFERENCES

1. J. Piao *et al.*, "Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society," (2025).
2. N. Bougie and N. Watanabe, "CitySim: Modeling urban behaviors and city dynamics with large-scale LLM-driven agent simulation," in *Proceedings of EMNLP: Industry Track*, (2025).
3. M. C. González *et al.*, "Understanding individual human mobility patterns," Nature **453**, 779–782 (2008).
4. D. Teixeira *et al.*, "On estimating the predictability of human mobility: the role of routine," EPJ Data Sci. **10** (2021).
5. C. M. Schneider *et al.*, "Unravelling daily human mobility motifs," J. The Royal Soc. Interface **10**, 20130246 (2013).
6. M. Larooij and P. Törnberg, "Do large language models solve the problems of agent-based modeling? a critical review of generative social simulations," (2025).
7. G. H. Santos, A. C. Viana, and T. H. Silva, "When plausible is not realistic: Evaluating human mobility in llm-based urban simulation," in *SIGSPATIAL*, (2026).
8. A. Chasse *et al.*, "The NetMob25 Dataset: A High-resolution Multi-layered View of Individual Mobility in Greater Paris Region," (2025).
9. J. a. P. Esper *et al.*, "Beauty or beast: Human behavioral insights and learning power of federated mobility prediction," in *Proceedings of the 32nd ACM SIGSPATIAL*, (2024).

Predictability of Urban Park Visitation Behavior Using Context-Aware Fine-Tuned LLM Agents

DongHak Lee[1,2], Xiyu Pan[1,2], Yixuan Hong[1], and Takahiro Yabe[1,2,*]

[1] Center for Urban Science and Progress, Tandon School of Engineering, New York University, 370 Jay St. Brooklyn, NY 11201
[2] Department of Technology Management and Innovation, Tandon School of Engineering, New York University, 2 Metrotech, Brooklyn, NY 11201
[*] Corresponding author: takahiroyabe@nyu.edu

1. INTRODUCTION

Human mobility prediction is a fundamental task for urban decision-making, including transportation planning, public health, environmental exposure assessment, and urban service provision [1, 2]. Recent advances in Large Language Models (LLMs) have created new opportunities for modeling human mobility through behavioral agents [3, 4]. However, it remains unclear how contextual mechanisms should be structured and transferred across heterogeneous urban environments.

Urban mobility behavior reflects two complementary drivers. Some regularities arise from broadly shared behavioral mechanisms that operate consistently across places, while others emerge from geographically localized contexts that vary across communities [5, 6]. Most existing LLM-based mobility prediction approaches treat contextual information as a single training distribution, obscuring this distinction between shared mobility mechanisms and localized contextual heterogeneity.

To address this limitation, we develop a context-aware hierarchical fine-tuning framework that first estimates cross-regional transferability of mobility mechanisms and then separates universal behavioral mechanisms from localized contextual heterogeneity for hierarchical LLM adaptation. Using park visitation data from New York City, Boston, and Atlanta, we examine when mobility knowledge should be shared across places and when it should be locally specialized.

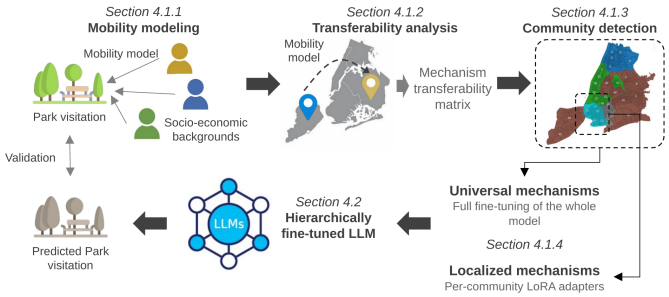


Fig. 1. Overview of the context-aware hierarchical fine-tuning framework.

2. METHODOLOGY

We first estimate neighborhood-specific Poisson pseudo-maximum likelihood (PPML) models of tract-to-park visitation flows using contextual features describing park characteristics, home and park neighborhoods, and travel accessibility [7]. A separate PPML model is fitted for each neighborhood cluster to estimate geographically varying effects of contextual factors on park visitation.

We hold out 20% of tract–park visitation observations before PPML estimation. The estimated mobility models are then cross-applied across neighborhoods to measure the transferability of mobility mechanisms using correlations between predicted and observed visitation. Pairwise transferability scores are used to construct a weighted network, where nodes represent neighborhood clusters and edge weights represent their predictive transferability. Louvain community detection [8] is then applied to identify groups of neighborhoods with mutually transferable mobility mechanisms.

Using these estimated mechanisms, contextual features with consistent effects across neighborhoods are identified as universal behavioral mechanisms, whereas those consistently important only within specific mobility communities are treated as localized contextual information. We use Llama-3.1-8B-Instruct and train it hierarchically: first through full fine-tuning with universal mechanisms to learn transferable behavioral patterns, followed by community-specific LoRA adaptation [9] ($r = 16$) using localized contextual information.

3. RESULTS

Figure 2 shows the transferability structure in New York City as a representative example. Rather than generalizing uniformly across all neighborhoods, mobility models exhibited substantially higher transferability within groups of neighborhoods. Louvain community detection identified coherent mobility communities, suggesting that park visitation is organized neither by a single city-wide mechanism nor by completely idiosyncratic neighborhood behavior.

Similar transferability structures were observed in Boston and Atlanta, yielding five, six, and four mobility communities in New York City, Atlanta, and Boston, respectively. These results indicate that mobility mechanisms are organized around a small number of coherent communities with shared behavioral structure.

We next examined which mobility mechanisms were shared across communities and which remained geographically localized. Park size consistently exhibited a positive effect across all three cities, while travel time showed a consistent negative effect in New York City and Boston. These predictors corre-

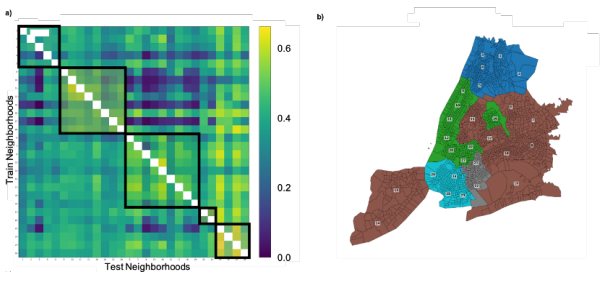


Fig. 2. Transferability matrix and Louvain mobility communities in New York City. Neighborhood-specific mobility models exhibit stronger transferability within identified mobility communities than across neighborhoods.

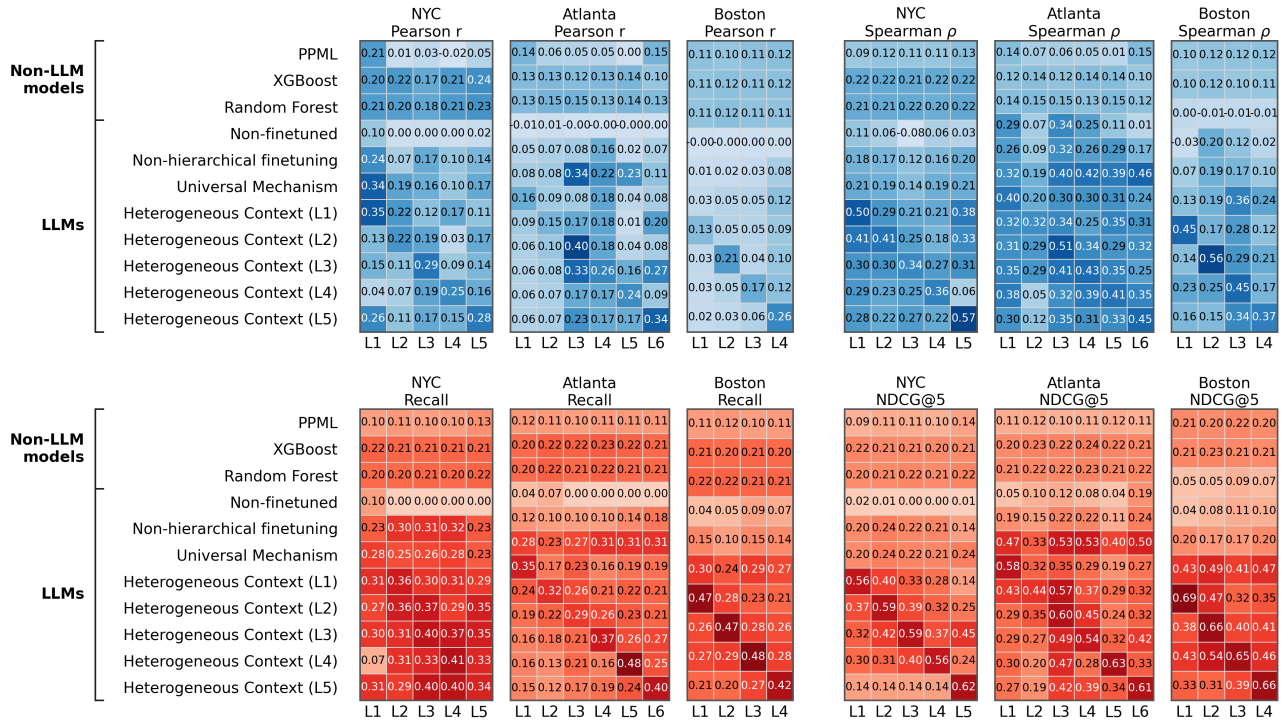


Fig. 3. Predictive performance across New York City, Atlanta, and Boston, evaluated using Pearson correlation (r), Spearman rank correlation (ρ), Recall@5, and NDCG@5. Community-specific heterogeneous context models perform best when evaluated within their corresponding mobility communities.

spond to classical destination attractiveness and distance decay mechanisms [10, 11], indicating that they represent broadly transferable mobility mechanisms. In contrast, recreational amenities, park-neighborhood characteristics, home-neighborhood characteristics, and socio-demographic variables remained significant only within particular mobility communities, suggesting that most contextual effects are geographically localized [5, 12, 13].

Finally, we evaluated whether this decomposition improves LLM-based mobility prediction. As shown in Figure 3, incorporating universal mobility mechanisms improved prediction over the non-fine-tuned LLM baseline and conventional non-hierarchical fine-tuning across all three cities. The largest gains were generally achieved by the community-specific heterogeneous context models when evaluated within their corresponding mobility communities. In contrast, localized models did not consistently outperform the universal mechanism model when evaluated on other communities. This result suggests that localized contextual knowledge is geographically bounded rather than universally transferable. Together, these findings demonstrate that separating transferable mobility mechanisms from localized contextual heterogeneity provides a more effective LLM adaptation strategy than treating all contextual information as a single training distribution.

4. CONCLUSION

We present a context-aware hierarchical fine-tuning framework for LLM-based human mobility prediction. By estimating cross-neighborhood transferability, identifying mobility communities, and separating universal mobility mechanisms from localized contextual heterogeneity, the framework provides a structured approach for adapting LLM agents to urban mobility tasks. Empirical results from New York City, Boston, and Atlanta show that mobility behavior is organized around transferable communities and that community-specific adaptation improves prediction when applied within the appropriate geographic context. These findings highlight the importance of explicitly modeling contextual transferability in LLM-based mobility prediction.

REFERENCES

1. M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," nature **453**, 779 (2008).
2. C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," Science **327**, 1018 (2010).
3. X. Liao, Q. Jiang, B. Y. He, *et al.*, "Deep activity model: A generative approach for human mobility pattern synthesis," arXiv preprint arXiv:2405.17468 (2024).
4. C. Han, Y. Yuan, Y. Liu, *et al.*, "Unimove: A unified model for multi-city human mobility prediction," arXiv preprint arXiv:2508.06986 (2025).
5. H. Barbosa, S. Hazarie, B. Dickinson, *et al.*, "Uncovering the socioeconomic facets of human mobility," Sci. reports **11**, 8616 (2021).
6. L. Alessandretti, U. Aslak, and S. Lehmann, "The scales of human mobility," Nature **587**, 402 (2020).
7. J. S. Silva and S. Tenreyro, "The log of gravity," The Rev. Econ. statistics pp. 641–658 (2006).
8. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. statistical mechanics: theory experiment **2008**, P10008 (2008).
9. E. J. Hu, Y. Shen, P. Wallis, *et al.*, "Lora: Low-rank adaptation of large language models." ICLR **1**, 3 (2022).
10. A. G. Wilson, "A family of spatial interaction models, and associated developments," Environ. Plan. A **3**, 1 (1971).
11. D. McFadden, "Conditional logit analysis of qualitative choice behavior," (1972).
12. R. M. Hilman, G. Iñiguez, and M. Karsai, "Socioeconomic biases in urban mixing patterns of us metropolitan areas," EPJ data science **11**, 32 (2022).
13. H. Nilforoshan, W. Looi, E. Pierson, *et al.*, "Human mobility networks reveal increased segregation in large cities," Nature **624**, 586 (2023).

Some Users Are Easier to Re-identify: Evidence from Mobility Profiles and Embedding Geometry

TIZIANO MARTÍN BERNAL[1], ANNE JOSIANE KOUAM[2,*], AND MARIANO G. BEIRÓ[1,3]

[1] Universidad de San Andrés, Argentina
[2] Inria, France & TU Berlin, Germany
[3] CONICET, Argentina
[*] anne-josiane.kouam@inria.fr

1. INTRODUCTION

Human mobility exhibits strong regularities: individuals repeatedly visit a limited set of significant locations and display predictable movement patterns over time [1, 2]. These regularities make mobility traces highly distinctive, enabling trajectory-user linking (TUL) attacks that associate anonymous trajectories with the users who generated them [3, 4]. Recent deep learning approaches such as TULER [5], TULVAE [6], and graph-based models [7, 8] have considerably improved re-identification performance by learning discriminative latent representations of user mobility. Despite these advances, existing work almost exclusively reports aggregate metrics such as Top-1 or Top-k accuracy. These averages hide an important source of heterogeneity: while some users are consistently re-identified with high confidence, others remain considerably harder to recognize. In this work, we investigate whether user identifiability can be explained jointly by mobility behavior and by the geometry of the learned embedding space. Using TULER on a large-scale mobile phone dataset, we first group users into three mobility profiles—scouters, regulars, and routiners—according to their mobility characteristics. We then characterize each user's embedding through two geometric properties, *cluster dispersion* and *cluster eccentricity*, computed on the training trajectories, and relate them to re-identification performance measured on unseen trajectories. Our results reveal a consistent relationship between mobility behavior, embedding geometry, and user identifiability. Routiners exhibit significantly higher re-identification accuracy than regulars and scouters, while also forming less disperse and less eccentric latent representations. Furthermore, we show that lower cluster dispersion and greater eccentricity user are associated with higher re-identification accuracy. These results suggest that embedding geometry provides complementary information for understanding why some users are easier to re-identify.

2. DATA AND METHOD

A. Dataset and preprocessing

We conduct our study on the Shanghai Call Detail Records (CDR) dataset [9]. The raw dataset contains approximately 18.1 million records from 9,618 users collected over a six-month period. Locations correspond to the dataset's cellular base stations, while timestamps are aggregated into 30-minute intervals. We define an active day as one containing at least 32 records, corresponding to approximately 66% of the maximum daily observations, and retain users with at least 7 active days. Home and work locations are inferred from nighttime and daytime activity patterns, respectively, and at most 100 active days with the largest number of records are retained per user. The resulting dataset contains 2,216 users, 12.7 million records, 116,578 daily trajectories and more than 3,000 unique locations.

Each user's mobility history is represented as a collection of daily trajectories, which are chronologically split into disjoint training (60%), validation (10%), and test (30%) sets. The training trajectories are used to learn user-specific mobility representations, the validation trajectories are used for model selection and hyperparameter tuning, and the held-out test trajectories are used to evaluate re-identification performance on previously unseen mobility traces.

B. User identifiability and representation analysis

Re-identification. To quantify user identifiability, we employ TULER [5], a deep trajectory-user linking model based on recurrent neural networks. Given an ordered sequence of visited locations, TULER learns a latent representation and predicts its originator through supervised multi-class classification. We represent each visited location by its latitude and longitude coordinates. For each user u, re-identification performance Top1_u is measured as the proportion of their held-out test trajectories correctly attributed by TULER.

Latent geometry. We exploit the learned embedding space to characterize the geometry of each user's trajectory cluster. Specifically, we extract the embedding of each training trajectory from the penultimate layer of the network, where trajectories belonging to the same user form a user-specific cluster. To characterize these clusters, we introduce two geometric descriptors. Let $\mathcal{Z}_u = \{\mathbf{z}_1, \ldots, \mathbf{z}_{n_u}\}$ denote the embeddings of the n_u training trajectories of user u, and let $\mathbf{c}_u = \frac{1}{n_u}\sum_i \mathbf{z}_i$ be the corresponding cluster centroid. We define *cluster dispersion* as

$$D_u = \frac{1}{n_u}\sum_{i=1}^{n_u} \|\mathbf{z}_i - \mathbf{c}_u\|_2,$$

where lower values indicate tighter and more coherent user representations. We further define *cluster eccentricity* as the distance to the nearest competing user cluster,

$$E_u = \min_{v \neq u} \|\mathbf{c}_u - \mathbf{c}_v\|_2,$$

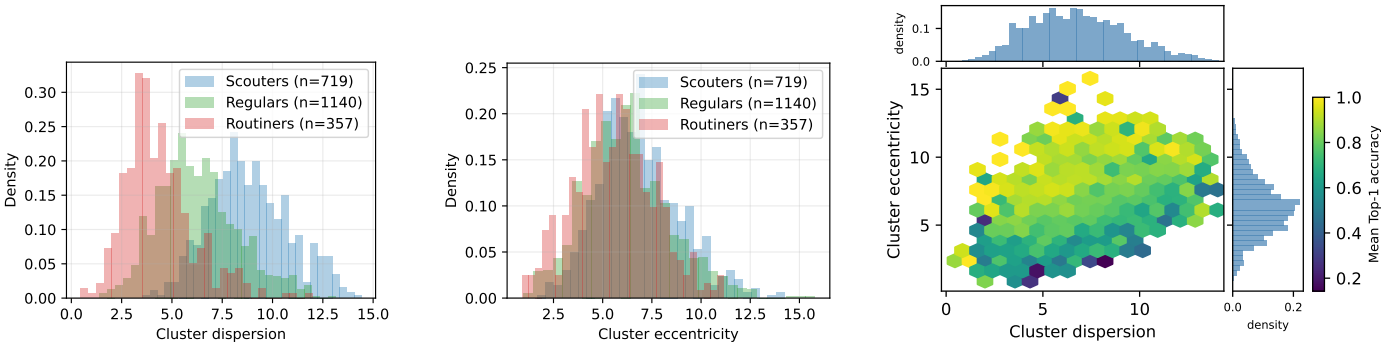
where larger values indicate greater separation from the closest competing user.

Mobility profiles. Following [10], users are assigned to three mobility profiles—scouters, regulars (Reg), and routiners (Rou)—from their mobility characteristics. We examine how these profiles differ in re-identification performance and latent geometry.

Association with re-identification. Since profiles also differ in the number of available training trajectories, we use linear regression to separate the effects of trajectory availability, mobility profile, and latent geometry on re-identification performance:

$$\text{Top1}_u = \beta_0 + \beta_1 T_u + \beta_2 I(\text{Reg}_u) + \beta_3 I(\text{Rou}_u) + \beta_4 D_u + \beta_5 E_u + \epsilon_u,$$

where T_u is the number of training trajectories and scouters are the reference profile. We compare a model containing trajectory count and profile with a second model additionally including dispersion and eccentricity.



(a) Distribution of cluster dispersion across the three mobility profiles.

(b) Distribution of cluster eccentricity across the three mobility profiles.

(c) Average Top-1 re-identification accuracy as a function of cluster dispersion and cluster eccentricity.

Fig. 1. Relationship between mobility profiles, embedding geometry, and re-identification performance. Routiners exhibit more compact latent representations, while higher re-identification accuracy is associated with lower cluster dispersion and greater separation from the nearest competing user.

Table 1. Mobility characteristics and TULER re-identification performance across mobility profiles. Values are reported as mean ± standard deviation. Rg = Radius of gyration.

Metric	Scouters	Regulars	Routiners
Users	719	1140	357
Rg (km)	25.08 ± 64.87	22.18 ± 76.83	12.50 ± 70.12
Repetitiveness	0.857 ± 0.052	0.911 ± 0.033	0.948 ± 0.030
Stationarity	0.806 ± 0.073	0.860 ± 0.071	0.913 ± 0.067
Unique locations	5.81 ± 2.24	3.74 ± 1.26	2.47 ± 1.00
Train trajectories/user	50.53 ± 20.48	53.36 ± 20.50	54.19 ± 20.16
Top-1 Acc.	0.765 ± 0.224	0.792 ± 0.222	0.804 ± 0.249

3. RESULTS

Mobility profiles exhibit distinct re-identification risks. Table 1 summarizes the mobility characteristics of the three user profiles together with their re-identification performance. As in previous work, routiners display the most regular mobility behavior, with higher repetitiveness and stationarity, and fewer unique locations. They have slightly more training trajectories per user. These behavioral differences are accompanied by distinct re-identification performance: routiners achieve the highest TULER Top-1 accuracy (80.4%), followed by regulars (79.2%) and scouters (76.5%). Hence, the performance differs significantly across profiles (Kruskal–Wallis, $p < 0.001$), with all pairwise comparisons remaining significant after Holm correction.

Embedding geometry reflects mobility behavior. To better understand these differences, we analyze the latent space learned by TULER. Figure 1a–b reports cluster dispersion and eccentricity across mobility profiles. Routiners exhibit substantially lower dispersion, indicating tighter and more coherent trajectory representations. A similar, although less pronounced, ordering is observed for eccentricity: routiners tend to lie closer to their nearest competing user, followed by regulars and scouters. This suggests that mobility regularity is primarily reflected in the consistency of a user's latent representation, rather than in greater separation from other users.

Compact and distinctive representations are associated with higher re-identification accuracy. Figure 1c relates cluster dispersion and eccentricity to user-level Top-1 accuracy. Users with lower dispersion and greater eccentricity tend to achieve higher re-identification accuracy. In user-level regressions controlling for mobility profile and the number of training trajectories, adding these geometric descriptors increases the explained variance from $R^2 = 0.153$ to 0.253. Both lower dispersion ($\beta = -0.020$, $p < 0.001$) and greater eccentricity ($\beta = 0.036$, $p < 0.001$) remain significantly associated with higher accuracy, while profile coefficients are no longer significant. This suggests that individual identifiability is better characterized by the coherence and distinctiveness of users' latent representations.

4. CONCLUSION

We investigated how re-identification performance varies across mobility users. Routiners exhibit significantly higher observed accuracy than regulars and scouters, but mobility profile alone does not fully explain individual identifiability. Lower cluster dispersion and greater cluster eccentricity are associated with higher re-identification accuracy, even after controlling for training trajectory availability. These findings show that user-level embedding geometry provides complementary information beyond mobility profiles for characterizing heterogeneous re-identification risk.

Acknowledgment. This work was supported by the Mob Sci-Dat Factory project (ANR-23-PEMO-0004) under the France 2030 program.

REFERENCES

1. M. C. González, C. A. Hidalgo, and A. Barabási, "Understanding individual human mobility patterns," Nature (2008).
2. L. Alessandretti, P. Sapiezynski, V. Sekara, *et al.*, "Evidence for a conserved quantity in human mobility," Nat. Hum. Behav. **2**, 485 (2018).
3. Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," Sci. Reports **3**, 1376 (2013).
4. L. Rossi, J. Walker, and M. Musolesi, "Spatio-temporal techniques for user identification by means of gps mobility data," EPJ Data Sci. (2015).
5. Q. Gao, F. Zhou, K. Zhang, *et al.*, "Identifying human mobility via trajectory embeddings," in *IJCAI-17,* (2017).
6. F. Zhou, Q. Gao, G. Trajcevski, *et al.*, "Trajectory-user linking via variational autoencoder," in *IJCAI-18,* (2018).
7. F. Zhou, S. Chen, J. Wu, *et al.*, "Trajectory-user linking via graph neural network," in *ICC,* (2021).
8. L. Deng, H. Sun, Y. Zhao, *et al.*, "S2tul: A semi-supervised framework for trajectory-user linking," in *WSDM '23,* (2023).
9. Mexwell, "Telecom shanghai dataset," https://www.kaggle.com/datasets/mexwell/telecom-shanghai-dataset (2020).
10. L. Amichi, A. C. Viana, M. Crovella, and A. A. Loureiro, "Understanding individuals' proclivity for novelty seeking," in *ACM SIGSPATIAL,* (2020).

BUBBLE-BLUE: private group messaging over an open Bluetooth mesh, measured on six boards

PAUL-MAXENCE BARATON[1,*], NADJIB ACHIR[1], PHILIPPE JACQUET[1], AND WILFRED DORÉ[2]

[1] *Inria, France*
[2] *Qualcomm, France*
[*] *paul-maxence.baraton@inria.fr; nadjib.achir@inria.fr; philippe.jacquet@inria.fr; wdore@qti.qualcomm.com*

1. THE IDEA

Local communication without infrastructure still collapses onto a Wi-Fi access point or a phone operating system that will not host a third-party mesh. Yet the settings that need it most are those where infrastructure is absent, saturated or distrusted: a crowd at an event, a site without coverage, an emergency response. Bluetooth is on every device and needs no operator.

BUBBLE-BLUE [1] separates two things that existing Bluetooth meshes merge. The *topological mesh* is open: any node may join it and relay traffic, including traffic it cannot read. The *bubbles* are private: a group exists only for the holders of its key. A stranger's phone can therefore extend the reach of a conversation it will never see. To keep the cost of flooding low, only an elected backbone, a connected dominating set (CDS), retransmits. The original note gave the algorithms and simulations. This work contributes the BLE transport and frame format, a revocation protocol that reaches members several hops away, and measurements on real boards, in a room and across a building.

2. STATE OF THE ART

Bluetooth meshes and messengers. The SIG Mesh profile [2] is a managed-flooding stack for lighting and sensors; nodes are provisioned, and a relay must hold the network key. BitChat [3] shows that a BLE-only chat can reach consumer scale; it floods over GATT without a backbone. Bridgefy flooded messages among protesters, and its cryptanalysis [4] shows what a mesh used in a crowd must withstand: impersonation, tracking and denial of service. Briar [5] synchronises between *pre-paired* contacts over Classic Bluetooth, Wi-Fi and Tor. Reticulum [6] routes to cryptographic identities over any link.

Backbones for broadcast. Electing a subset of forwarders is classic in ad hoc networks: Wu and Li's marking rule [7] yields a CDS from two-hop knowledge, and multipoint relays [8], the basis of OLSR, reduce flooding from each node's view. BLE makes this matter more than Wi-Fi, because it has no native broadcast on a connection: every local flood is a sequence of unicasts.

The gap. Table 1 places BUBBLE-BLUE: none of the other systems lets a node sit in an elected backbone and relay for a group it does not belong to.

3. DESIGN

Links without pairing. A node advertises a compact presence record in a connectable BLE advertisement: an anonymous identifier, a relay flag, and the address at which it accepts connections (manufacturer data under company identifier `0xFFFF`, reserved by the Bluetooth SIG for testing, marking the prototype as experimental). Neighbours then open an L2CAP connection-oriented channel at security level *low*: no pairing, no bond, no allow-list. Membership is a 32-byte group key, not a Bluetooth relationship.

Table 1. Where BUBBLE-BLUE sits. "Open relay": a node may forward a frame it cannot decrypt.

System	Data plane	Backbone	Open relay
SIG Mesh [2]	managed flood	provisioned	no
BitChat [3]	GATT flood	none	n/a
Briar [5]	RFCOMM, Wi-Fi, Tor	contacts	no
Reticulum [6]	multi-transport	identities	no
BUBBLE-BLUE	LE CoC	elected CDS	yes

Periodic HELLO messages fill a neighbour table; a neighbour silent for three periods (15 s) is dropped.

One frame, two planes. Every frame starts with a 26-byte clear header: type, source, destination, TTL, sequence and timestamp. The payload is sealed with AES-256-GCM under a per-packet key, HKDF-SHA256 of the group key with source, sequence and timestamp. All header fields but TTL are authenticated. A relay that does not hold the key can thus decrement TTL and forward the frame; it cannot read the text or forge its sender. A message of up to 470 bytes fits one frame (Fig. 1).

Who relays. Each node applies the Wu–Li rule [7] to its two-hop view and relays if it is marked, or if it bridges two links the backbone has not covered. Duplicates are suppressed on (source, sequence), so a flood crosses each relay once. The receiving screen shows the hop count and the neighbour a message came through.

Removing a member. The coordinator, the member with the smallest anonymous identifier, revokes by rotating the group key and flooding it wrapped under pairwise X25519 keys, skipping the victim. The victim keeps relaying but can no longer read. Keys of members beyond one hop arrive by a signed ANNOUNCE every 30 s; revoking before then is refused, which defines a join window.

What an adversary can do. A passive radio in range learns who transmits, toward which address, and how far a frame may still go; it does not learn the text. The Bluetooth address is stable and in clear, because that is how a neighbour is redialled without a bond: payload privacy is not location privacy. An active non-member can make the backbone carry junk, reset TTL, which is not authenticated, or pre-empt a (source, sequence) pair at a relay to cause selective loss. A member who controls its address can grind it to become coordinator. These are accepted limits of the prototype.

On real boards. A privileged daemon owns the Bluetooth adapter; touch screens are unprivileged clients. Screens are three Thundercomm RUBIK Pi 3 (Qualcomm QCS6490, Ubuntu 24.04, kernel 6.8, BlueZ 5.72); relays are three Arduino UNO Q (Qualcomm QRB2210, Debian 13, kernel 7.0, BlueZ 5.82). In the evaluated setup, connections were initiated by the relays and accepted by the screens, so every path alternates screen and

relay; on a table the graph is $K_{3,3}$ (Fig. 2).

Joining, step by step. A new node goes through four steps with different clocks. Its presence is seen within one advertising interval; the channel opens in one L2CAP handshake; the first HELLO puts it in its neighbours' tables; and the next ANNOUNCE, within 30 s, spreads its public key beyond one hop. Only after that last step can it be revoked safely, so the screens show link state and group state separately.

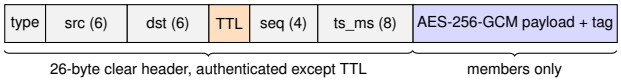


Fig. 1. One BUBBLE-BLUE frame. Relays read and rewrite only the header; TTL is left out of the authenticated data so they can decrement it.

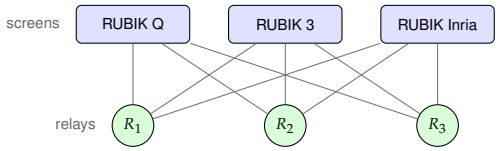


Fig. 2. The demonstration stand: every message between two screens crosses one of three relays.



Fig. 3. The line the boards formed on their own; median delays.

Table 2. Measurements on the six boards.

Stand ($K_{3,3}$, one room)	
Messages received by the five other boards	75/75
Rebooted screen reads the group, after daemon start	4.4 s
Recovery after an induced relay-controller interruption	130 s
Revoked board reads the next message (2 trials)	no
Four-hop line (building, battery, ∼15 m spacing)	
Messages received four hops away	76/79
Median delay, 1 / 2 / 3 / 4 hops (ms)	85 / 101 / 143 / 162
Last link carried away: held to	∼30 m
Endurance (stand, battery, one message every 30 s)	
Run on one charge; boards still up at the end	8.2 h; 6/6
Received, screens / relays (983 messages)	100 % / 99.3 %
Delay, median / 99th percentile (ms)	80 / 188

4. RESULTS

On the stand. Every message reached the five other boards (Table 2). Joining is dominated by the boot: a rebooted screen read the group 4.4 s after its daemon started. AES-GCM (30k–173k packets/s) is far above the link rate.

Across a building. With the boards about 15 m apart and one relay off, the mesh formed a four-hop line by itself (Fig. 3). From on-board logs (clocks within 50 ms), the far end received 76 of 79 messages, missing only the three sent while the line formed; each hop adds a few tens of milliseconds. Carried away, the far screen's link held to about 30 m, dropped, and returned within seconds of the board coming back: mobility appears as ordinary link churn. When the sender's only relay lost power, the line was dark for 95 s: a line has no detour.

A working day on battery. From one charge, the six boards ran the stand for 8.2 h, one message every 30 s, and all were still up when we stopped them. The screens received all 983 messages; the relays missed 9 and 7 (loss rates visible in Table 2). Over the 4 915 deliveries, the median delay was 80 ms and the 99th percentile 188 ms.

In simulation and next. The backbone adds no path stretch, and TTL 10 spans about twenty nodes (simulation). The link logs form a contact graph for replaying the protocol on mobility traces.

Acknowledgment. Boards were provided by the Qualcomm University Platforms program. AI-assisted tools were used for software, analysis and editing; the authors checked all results.

REFERENCES

1. N. Achir and P. Jacquet, "BUBBLE-BLUE a multihop private network based on Bluetooth," arXiv preprint arXiv:2509.00967 (2025).
2. Bluetooth SIG, *Mesh Profile Specification, Version 1.0* (2017).
3. J. Dorsey *et al.*, "BitChat: Bluetooth mesh chat," https://github.com/permissionlesstech/bitchat (2025).
4. M. R. Albrecht, J. Blasco, R. B. Jensen, and L. Marečková, "Mesh messaging in large-scale protests: Breaking Bridgefy," in *Topics in Cryptology – CT-RSA 2021*, , vol. 12704 of *LNCS* (Springer, 2021), pp. 375–398.
5. The Briar Project, "Briar," https://briarproject.org/ (2024).
6. M. Qvist, "Reticulum / LXMF," https://github.com/markqvist/Reticulum (2024).
7. J. Wu and H. Li, "On calculating connected dominating set for efficient routing in ad hoc wireless networks," in *Proc. 3rd Int. Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIAL-M)*, (1999), pp. 7–14.
8. A. Qayyum, L. Viennot, and A. Laouiti, "Multipoint relaying for flooding broadcast messages in mobile wireless networks," in *Proc. 35th Hawaii Int. Conf. on System Sciences (HICSS)*, (IEEE, 2002), pp. 3866–3875.

Privacy and anonymization of stochastic mobility trajectories from bus payment logs

VINCENZO RICCIARDELLI[1], NIKHIL JHA[1], AND LUCA VASSIO[1,*]

[1] *Politecnico di Torino, Italy*
[*] *luca.vassio@polito.it*

1. INTRODUCTION

Sharing transportation data comes with considerable opportunities, but also risks. While the amount of information to be shared with relevant stakeholders may play a decisive role in improving users' experience on the roads, the same information may contain sensitive features, whose publication might potentially violate the privacy of the users. Anonymizing information in spatio-temporal data, or in data derived from it, has been an ongoing problem for years, in a setting that presents unique challenges, such as sparsity of the individual traces, strong time-space correlation, and recurring patterns in users' activity [1–4]. The current literature often focuses on the scenario where each user is linked to a single spatio-temporal trajectory, and the threat model considers an attacker trying to re-establish such a link. In this work, using the NetMob26 challenge dataset, we extend the attack surface by considering a setting in which each user is represented not only by one trace, but by a set of *probabilistic* traces produced through stochastic reconstruction. Although this setting may initially appear more favorable to the user, since the true trace is hidden among the others, in reality it opens up new attack scenarios where the fingerprints left by a user are multiplied as a consequence of the stochasticity. In summary, protecting the user means protecting not only the true trajectory but also all non-zero-probability trajectories associated with the same profile.

In this work, we build the foundational concepts around our new threat model, show how it can endanger users, and test existing anonymization and protection strategies. We propose the framework illustrated in Fig. 1. Ticketing records, fleet telemetry, and GTFS stops are fused to infer probabilistic boarding locations and reconstruct stochastic user trajectories. Privacy risks are then evaluated under increasingly informative attack scenarios and mitigated through anonymization strategies.

2. RECONSTRUCTING TRAJECTORIES

The NetMob26 dataset [5] documents road-based public transport of Niterói (Brazil) over March 2026. It combines two sources: (i) *ticketing records*, comprising about 7.2 million boarding events from more than 370 000 users, with each event containing an anonymized user id, line id, and timestamp; (ii) *GPS fleet telemetry*, sampled every 15–30 s, with vehicle id, line id, GTFS trip/route ids and coordinates. Moreover, official stop locations are provided in an auxiliary `stops.json`. The data publisher declares irreversible user pseudonymization: identifiers are replaced, but the spatio-temporal quasi-identifiers persist.

We reconstruct each user's most plausible sequences of boarding locations from their payments, propagating uncertainty at every step. First, every GPS ping of the fleet is snapped to the nearest stop. Then, for each ticket purchase at time t we select the same-line GPS observations falling within the asymmetric window $[t - 10\,\text{min},\ t + 5\,\text{min}]$; the window is asymmetric because ticket purchase typically occurs at boarding, slightly after the vehicle reaches the stop, so the relevant GPS evidence lies in the past relative to the tap. Each observation is scored by combining its temporal offset from the tap with its snapping distance,

$$s(d, \Delta t) = \exp\left(-\frac{|\Delta t|}{\sigma_t}\right) \cdot \exp\left(-\frac{d}{\sigma_d}\right),$$

where d is the Haversine distance from the stop and Δt is the time between the tap and the observation, with $\sigma_t = 3\,\text{min}$ and $\sigma_d = 100\,\text{m}$. Temporal and spatial proximity act as two independent multiplicative decay factors, each down-weighting an observation that is far from the stop in time or in space. This scoring adapts the emission/transition structure of the HMM map-matching of [6], with an exponential kernel in place of the Gaussian and an added temporal term. The per-observation scores are then summed and normalized across all candidate stops in the window to form a probability distribution $P_t(\text{stop}_{i(t)})$; the five highest-probability stops are then retained as the candidate set for that payment.

We then combine the plausible stops for each payment to create stochastic trajectories for a user. The probability of a trajectory is the product of its per-payment stop probabilities,

$$P(\text{traj}) = \prod_t P_t(\text{stop}_{i(t)}),$$

For computational reasons, we only maintain the top-50 trajectories for each user.

3. PRIVACY VULNERABILITIES AND PROTECTION

We propose a threat model where an attacker has precise information about the (sub)trajectory of a user, e.g., by following them in real life. The goal of the attacker is twofold: i) to identify, among all the published probabilistic trajectories, the one that refers to their target, and ii) to improve their knowledge about the victim by being able to infer new points of the trajectory beyond what was already in their possession. Please note that this would allow the attacker to gain information about future stops the user will visit in subsequent data releases.[1]

In light of this threat model, every user's trajectory must be protected by taking into consideration the risk of an attacker learning the user's identifier based on a (sub)trajectory at their disposal. In the field of Privacy-Preserving Data Publishing, this protection is usually obtained by enforcing the k-anonymity property, which mandates that every record is indistinguishable from $k - 1$ other records. In our stochastic scenario, we propose a revised version of k-anonymity which we call (k, p)-anonymity. We claim that a dataset satisfies (k, p)-anonymity if every trajectory in the dataset is shared by at least k users, and none of the users has a probability whose ratio over all the probabilities for that trajectory exceeds p. For instance, in a (3,0.5)-anonymous dataset, every trajectory must appear in at least 3 users, and none of them can have a probability accounting for more than

[1] User ID rotation could complicate the task for the attacker. However, cross-time matching using recurring patterns would still be a viable threat. We leave this as promising future work.

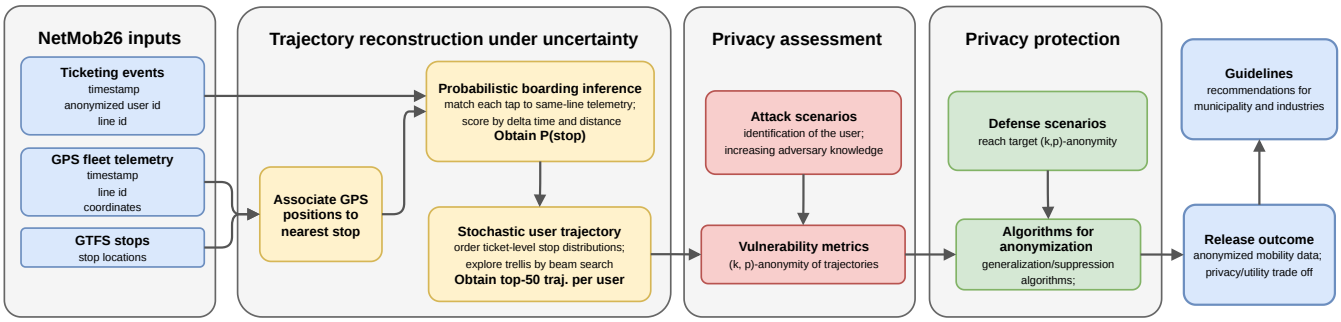


Fig. 1. Overview of the proposed pipeline to reconstruct probabilistic trajectories and to protect users' privacy.

50% of the sum of the probabilities. So, in this scenario, if 3 users exist that expose the same trajectory with probabilities $[0.1, 0.3, 0.8]$, then (3,0.5)-anonymity is not satisfied, as the trajectory is substantially more likely to belong to the third user $(0.8/(0.1+0.3+0.8) \approx 0.67 > 0.5)$, leaking relevant information to the attacker.

4. PRELIMINARY RESULTS AND DISCUSSION

Fig. 2 shows the distribution of the achieved level of k-anonymity for trajectories and *stop fingerprints*. By stop fingerprint, we mean the user's set of stops, without knowing in which order these locations have been visited. For simplicity, in this analysis both trajectories and stop fingerprints were constructed by only considering the most likely stop for every payment event. As expected, the number of distinct trajectories is larger than the number of distinct fingerprints, since multiple trajectories may collapse into the same fingerprint. Fig. 2 shows that 95.3% (92.9%) of the trajectories (fingerprints) are unique, i.e., $k = 1$.

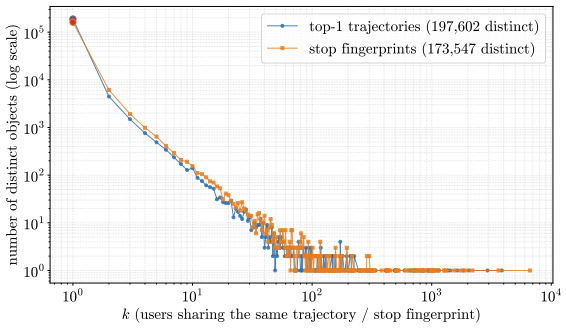


Fig. 2. k-anonymity distribution for trajectories and stop fingerprints.

Figure 3 shows how p, as defined in (k, p)-anonymity, is distributed within the dataset. Over 70% of user-trajectory pairs have $p = 1$, indicating that most trajectories are unique. This percentage is computed over all the stochastic trajectories and is therefore not directly comparable with the distinct-object analysis of Fig. 2. Taken together, the two plots highlight the need for appropriate protection both in terms of the number of users sharing a trajectory and in terms of the relative probability of each trajectory. This is the twofold protection that (k, p)-anonymity can provide.

To understand how existing geographical data anonymization algorithms behave against (k, p)-anonymity, we test the W4M algorithm [1] and measure how it impacts the (k, p)-anonymity property of the dataset. W4M groups trajectories into clusters of size k_{W4M} and then adaptively transforms them so that they lie within a radius of δ metres of a pivot trajectory of each cluster. We apply W4M using $k_{\text{W4M}} = 50$ trajectories and $\delta = 0$, needed to produce clusters of indistinguishable trajectories, as (k, p)-anonymity mandates. The resulting dataset satisfies (12, 0.113)-anonymity (recall that here $k = 12$ refers to users, not anymore to trajectories), showing that (k, p)-anonymity is a restrictive notion of privacy that needs the development of specific techniques.

Satisfying (k, p)-anonymity can offer a viable guideline to municipalities and industry stakeholders releasing similar data. More broadly, we showed that uncertainty-aware trajectory reconstruction must be accompanied by equally uncertainty-aware privacy guarantees before mobility data can be released responsibly.

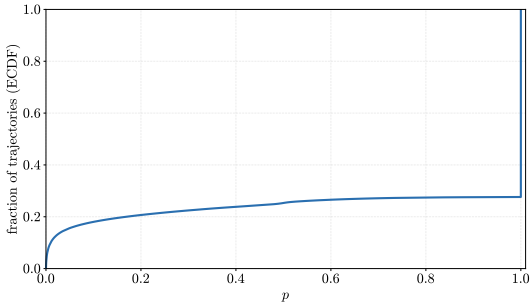


Fig. 3. Cumulative distribution of p for trajectories.

Acknowledgment. We thank Atman Systems https://atmansystems.com/ for their support of our research that led to this contribution.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [5].

REFERENCES

1. O. Abul, F. Bonchi, and M. Nanni, "Anonymization of moving objects databases by clustering and perturbation," Inf. Syst. **35** (2010).
2. P. Armenante, K. Huang, N. Jha, and L. Vassio, "Protecting participants or population? Comparison of k-anonymous Origin-Destination matrices," in *NetMob 2025 Data Challenge,* (2025).
3. M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *2013 ACM SIGSAC CCS,* (2013).
4. M. Maouche, S. Ben Mokhtar, and S. Bouchenak, "HMC: Robust Privacy Protection of Mobility Data against Multiple Re-Identification Attacks," ACM Interact. Mob. Wearable Ubiquitous Technol. **2** (2018).
5. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 Dataset: A High-Resolution Multi-Source View of Public Bus Mobility in Niterói," (2026).
6. P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *ACM SIGSPATIAL 2009,* (2009).

Unequal by Design: Bus Network Structure, Social Vulnerability, and Favela Access in Niterói

ISSA MOUSSA DIOP[1], ALI YASSIN[2], RANDY BRANDON GALLEGOS-RODRÍGUEZ[3], TONATIUH SUÁREZ-MEANEY[4], AND MARIBEL HERNÁNDEZ-ROSALES[5,*]

[1] Université de Pau et des Pays de l'Adour, LIUPPA, Pau, France
[2] ICB UMR 6303 CNRS, Université Bourgogne Europe, Dijon, France
[3] Instituto Politécnico Nacional, Centro de Investigación en Computación, UPALM-Zacatenco, Mexico City, Mexico
[4] Mexico's Ministry of Agrarian, Territorial, and Urban Development (SEDATU), Mexico City, Mexico
[5] Center for Research and Advanced Studies, CINVESTAV Irapuato Unit, Guanajuato, Mexico
[*] maribel.hr@cinvestav.mx

1. INTRODUCTION

Public transport plays a fundamental role in ensuring equitable access to employment, education, healthcare, and other essential urban opportunities. In Niterói, Brazil, the bus network is the primary mode of public transportation and constitutes a key component of daily mobility. Although its bus network provides the main public transport service across the municipality, the distribution of routes and service frequencies is spatially heterogeneous. The city also contains numerous favela-associated census sectors, where populations are generally more socially vulnerable and may face reduced transport opportunities. At the same time, access to key points of interest (POIs), such as hospitals, schools, universities, and pharmacies, depends not only on geographical proximity but also on the structure and performance of the public transport network[1].

This study examines whether public transport in Niterói serves socially vulnerable areas equitably. We focus on three related questions: whether vulnerable census sectors receive adequate bus service, whether favela-associated areas face additional disadvantages, and whether the routes serving vulnerable areas differ in their passenger profiles. By linking social vulnerability, bus supply, access to essential services, and route structure, the study identifies where transport disadvantages are concentrated and where improvements could most improve access. We further evaluate the impact of adding new bus routes, new healthcare facilities, or both, considering route network distance and network-based accessibility[2].

2. METHODOLOGY

This study examines socio-spatial inequalities in public transport supply and accessibility at three interconnected levels: census sectors, favela-associated areas, and bus routes. We construct a social vulnerability index from four standardized 2022 Brazilian Census (IBGE) indicators such as income, illiteracy, race/ethnicity, and environmental conditions, selected with support from PCA, whose first component explained 67.5% of the variance. Indicators are oriented so that higher values consistently represent greater vulnerability and combined with equal weights as

$$\tilde{V}_i = \frac{1}{K} \sum_{k=1}^{K} s_k z_{i,k},$$

where $z_{i,k}$ denotes the standardized value of indicator k for sector i, and $s_k \in \{-1, +1\}$ controls its direction: $s_k = -1$ when higher values indicate lower vulnerability, and $s_k = +1$ otherwise. The resulting score is rescaled to the 0–100 range, with higher values indicating greater relative vulnerability.

The empirical analysis comprises five complementary steps. We first examine how bus supply and proximity vary across vulnerability groups, then assess direct bus access to essential services. We compare favela-associated and non-favela-associated sectors and evaluate optimization scenarios involving new routes, healthcare facilities, or both, using direct connectivity in the observed bus-route network. At the route level, we cluster bus lines according to the vulnerability of the sectors they serve and compare their passenger profiles. Finally, bipartite networks linking sectors, favelas, and routes are used to identify transit-based communities and assess whether vulnerability is concentrated within particular network groups.

3. RESULTS AND DISCUSSION

We begin by examining the spatial distribution of the vulnerability index, since it provides the baseline for interpreting the remaining transport and accessibility results. As shown by the vulnerability heatmap in Panel D of Figure 1, vulnerability is not evenly distributed across census sectors. Instead, higher vulnerability values are concentrated in specific areas of the municipality, indicating clear socio-spatial heterogeneity. Many favela-associated areas also overlap with sectors displaying higher vulnerability, suggesting that favela association is closely linked to broader patterns of socio-spatial disadvantage. This spatial pattern motivates the subsequent analysis of whether bus service supply and access to essential services are aligned with areas of greater need. After establishing this spatial pattern, we examine whether observed bus access corresponds to sector-level vulnerability. As shown in Panel 1.A, bus service generally decreases toward the highest-vulnerability deciles, although the pattern is not strictly monotonic. In particular, deciles 1 and 2 show slightly lower service than deciles 3–5. This may reflect lower public transport dependence and greater private-vehicle use in more affluent areas. In this case, lower bus supply may reflect lower transit demand rather than a structural service deficit. By contrast, the lower service observed in the most vulnerable deciles is more indicative of transport disadvantage.

These disparities extend beyond service availability. Panel 1.B shows that residents of favela-associated sectors can directly reach substantially fewer hospitals, universities, schools, and pharmacies than residents of other sectors. For example, direct bus access to hospitals drops from 54% in non-favela sectors to 21% in favela-associated sectors. Similar gaps are observed for all categories of essential services, indicating that transport inequalities directly affect access to education, healthcare, and everyday opportunities. The spatial distribution of these inequalities is illustrated in Panel 1.C. Direct hospital accessibility is highest around the central and north-western bus corridors, while many eastern and south-eastern sectors remain poorly connected. These differences are not explained by route allocation alone. Niterói's hilly terrain and coastal geometry can also

constrain street-network continuity and limit feasible bus connections, especially in peripheral areas. The resulting accessibility pattern therefore reflects both network structure and the city's physical geography. Panel 1.D shows the priority locations identified using distances measured along the street network. These locations represent modeled potential improvements rather than observed or prescriptive infrastructure decisions, and are intended to support the identification of equity-oriented accessibility interventions.

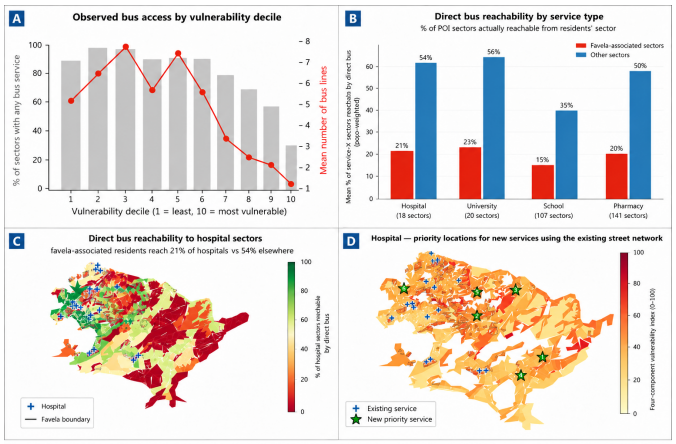


Fig. 1. Public transport equity patterns in Niterói: (A) bus service by vulnerability decile, (B) direct access to essential services, (C) hospital accessibility, and (D) vulnerability gradient and priority interventions for improving healthcare accessibility.

Figure 2 summarizes the route vulnerability analysis results. Panel 2.A shows the clustering of bus routes based on the vulnerability index composition of the census sectors they traverse. K-means clustering with $k = 2$ yielded the highest Silhouette score ($s = 0.512$), producing two distinct groups: Cluster C_1, composed predominantly of routes serving lower-vulnerability sectors, and Cluster C_2, whose routes concentrate in higher-vulnerability areas. Panel 2.B reveals that both student and senior ridership shares are systematically higher on C_2 routes, suggesting that concessionary-fare passengers disproportionately depend on the network segments serving the most vulnerable communities. Panel 2.C shows that direct bus access to essential services (hospitals and universities) decreases with sector vulnerability, indicating that residents of the most deprived areas face a compounded disadvantage: indicators of greater transit dependence coincide with poorer service reachability.

Panels 3.A and 3.B show that favela access to the observed bus network is both unequal and structurally fragile. Among 83 favela-associated areas, only 53 are served, through 46 bus lines and 232 favela–line links. A key result is that 21 of the 53 served favelas, around 40%, depend on a single observed bus line, while the three most connected routes account for 18% of all favela–line links. This concentration is reinforced by the nested structure of the network, indicating that poorly served favelas tend to rely on the same pivotal routes as better-connected ones rather than on alternative connections. The last two panels, 3.C and 3.D, show that the route-based network communities are not only topological structures, but also socio-spatially meaningful. The Louvain map reveals geographically coherent communities, with denser and more segmented groups around the centre and north-west, and more spatially extended communities in peripheral eastern and south-eastern areas. These community labels can then be related to social indicators in the scatter plot. The positive

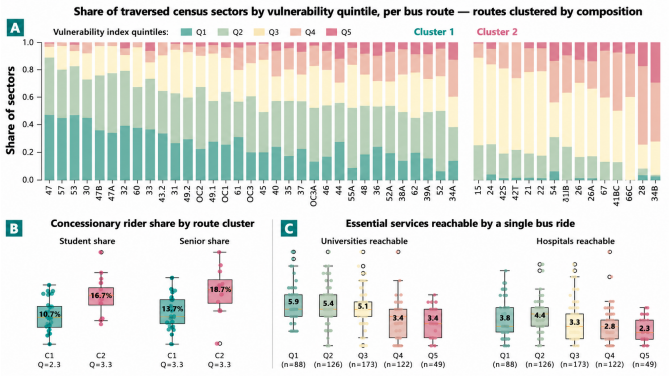


Fig. 2. Route-level transport inequality: vulnerability composition of bus routes, concessionary ridership by route cluster, and direct access to hospitals and universities by vulnerability quintile.

Spearman correlation shows that communities with higher mean vulnerability tend to contain a larger share of favela-associated sectors.

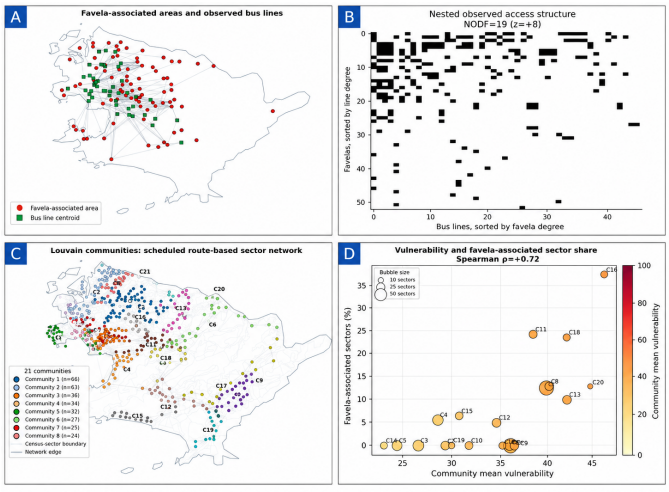


Fig. 3. Network structure of favela-associated access: spatial favela–bus links, nested access structure, Louvain route-based communities, and their relationship with social vulnerability.

Overall, our findings reveal persistent transport inequalities in Niterói, where socially vulnerable and favela-associated areas tend to receive lower bus service and have weaker direct access to essential services. The network analysis further highlights the structural fragility of favela access, with many areas depending on a limited number of bus routes. By integrating social vulnerability, accessibility, route profiles, and network structure, this study identifies priority areas where targeted transport interventions could improve urban accessibility and equity. Future work will further assess the role of terrain, urban form, and travel conditions in shaping these inequalities.

REFERENCES

1. A. Karner, R. H. Pereira, and S. Farber, "Advances and pitfalls in measuring transportation equity," Transportation **52**, 1399 (2025).
2. Y. Xu, L. E. Olmos, S. Abbar, and M. C. González, "Deconstructing laws of accessibility and facility distribution in cities," Sci. advances **6**, eabb4112 (2020).

Planned and experienced accessibility reveal contrasting diagnoses of transit poverty and inequality

EGOR KOTOV[1,2,3,*], YUAN LIAO[4], LEONARDO AKINAGA MOREIRA[5], AND RAFAEL H. M. PEREIRA[6]

[1] Department of Spatial Planning, TU Dortmund University, Dortmund, Germany
[2] Sociodemography Research Group, Political and Social Sciences Department, Universitat Pompeu Fabra (UPF), Barcelona, Spain
[3] Department of Digital and Computational Demography, Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany
[4] Department of Human Geography, Lund University, Lund, Sweden
[5] University of São Paulo, São Paulo, Brazil
[6] Institute for Applied Economic Research (Ipea), Data Science Division, Brasília, Brazil
[*] egor.kotov@tu-dortmund.de

Introduction. Public transport equity raises two questions [1]. Can communities reach enough essential activities? Do socioeconomic groups face systematic inequalities in access? The first concerns *accessibility poverty*; the second, *accessibility inequality*. Most studies answer them from scheduled GTFS timetables. Yet observed operations and monetary costs reshape accessibility [2, 3]. Using GPS telemetry and ticketing records from the NetMob26 dataset [4] for Niterói, Brazil, we compare planned and experienced accessibility and ask whether the resulting equity diagnoses agree.

Data. We combine March 2026 bus GPS traces and almost 6.8 million ticket validations (13.6% recorded as cash or unregistered) with 2022 population and income for 1,274 census tracts, the provided GTFS, geolocated hospitals and state schools, and 2019 formal-employment counts on an H3 hexagonal grid [5].

Methodology. For health, education, and employment, we compute accessibility poverty and inequality under three network representations (Scheduled, Typical, and High-delay) crossed with three payment conditions (Zero fare, Standard card, and Cash) — nine scenarios per domain. Every scenario holds the city, destinations, and population fixed and varies only how the bus network and its monetary cost are represented, so any change in the equity diagnoses is attributable to that representation. To evaluate these scenarios, we create three GTFS networks that are based on operator GTFS data and share all components except for per-route and per-window headways and running time: *Scheduled* is the operator's weekday timetable, *Typical* uses GPS-reconstructed median travel times and headways, and *High-delay* uses their 95th percentiles. To build these GTFS snapshots we developed two new open-source R packages: `gtfsrt2static` (https://github.com/e-kotov/gtfsrt2static) writes the headways, stop patterns, and feed tables, and `gps2gtfs` (https://github.com/e-kotov/gps2gtfs) supplies the terminals that cut the GPS traces into trips. Combining each network with OpenStreetMap, we route using `r5r` from every census-tract centroid to every destination over a 07:00–09:00 weekday window (16.7% of validations, 13.3% of GPS pings). For each origin–destination pair, travel time is the median over 120 departure minutes, with five routing simulations per minute for the main results. Sensitivity analyses average four independent runs, each using 20 routing simulations per minute, across four additional two-hour windows and an 85th-percentile delay scenario. Journeys combine walking (up to 30 minutes) and up to three transit rides, lasting at most 120 minutes. *Zero fare* is the no-cost baseline, including zero-debit student and senior populations. *Standard card* uses itinerary fares estimated from reconstructed linked-card journeys. Unlinkable *Cash* pays the full fare on each boarding, with no transfer integration, and therefore faces the highest out-of-pocket cost for multi-leg trips. Adapting Vale's effective-accessibility concept [3], we convert fares into an equivalent travel-time penalty using tract-average income before recounting destinations. Income is a sensitivity parameter, not an observed wage, and because tickets do not identify residence, each payment condition is applied hypothetically to every tract.

We use domain-specific sufficiency lines: one hospital within 45 minutes for health [6], three state schools within 30 minutes for education [7]. For jobs, the line is the fixed population-weighted median number reachable within 45 minutes [1] under Scheduled zero fare (98,776 of 129,350 jobs). The jobs poverty rate is therefore near 50% by construction; its contrasts are what matter. We measure accessibility poverty with population-weighted Foster–Greer–Thorbecke indices [8]. FGT_0 is the share of residents below the relevant sufficiency line. FGT_1 also captures how far their access falls below it. For inequality, the Erreygers Corrected Concentration Index [9] (CCI) ranks residents by tract income and summarises whether accessibility is concentrated among poorer or richer tracts. Positive values indicate a pro-rich distribution. We use the same normalization bounds for the Erreygers CCI in every scenario—0–20 hospitals, 0–41 schools, 0–129,350 jobs—so changes in the index reflect changes in accessibility rather than changes in scale. All comparisons use the same routes and time windows; Scheduled approximates the timetable rather than reproducing every departure exactly.

Findings. Figure 1a shows that fares and severe delays, rather than typical operations, drive accessibility poverty. At Zero fare, poverty rates under Typical morning-peak operations remain close to Scheduled levels (health 3.6% to 3.4%, education 20.5% to 21.0%, jobs 50.0% to 51.4%). Typical operations nevertheless lower health inequality (0.20 to 0.19). In contrast, Standard card and Cash raise health poverty from 3.4% to 26.1% and 27.9% and education poverty from 21.0% to 55.0% and 59.3%. Health CCI rises from 0.19 to 0.48 and 0.40, whereas education CCI falls from 0.22 to 0.20 and 0.13. Because we apply each payment condition hypothetically to every tract, a lower CCI does not mean that fares burden poorer riders less.

In Figure 1b, tract-level zero-fare access under Typical operations straddles the diagonal in all three domains: some tracts lose access while others gain. As a local diagnostic, at least one facility (for jobs, 10% of reachable jobs) is lost in 39.2%, 33.5%, and 14.9% of the 1,195 populated tracts. Typical job losses show no income gradient: the poorest and richest quintiles lose 1.1 and 1.2 percentage points of citywide job access, respectively. Under High-delay operations, however, losses become near-universal and regressive: at least one facility is lost in 88.8%, 83.0%, and 87.9% of populated tracts (89.3%, 84.2%, 89.3% of residents; none gain). Against Scheduled service (Fig. 1a), High-delay zero-fare poverty jumps to 8.7%, 31.2%, and 77.2%, and job losses concentrate in poorer areas (poorest quintile loses 22.1 points, richest 9.9). Figure 1c maps this spatial contraction toward the employment core. High-delay lowers inequality in health (CCI 0.20 to

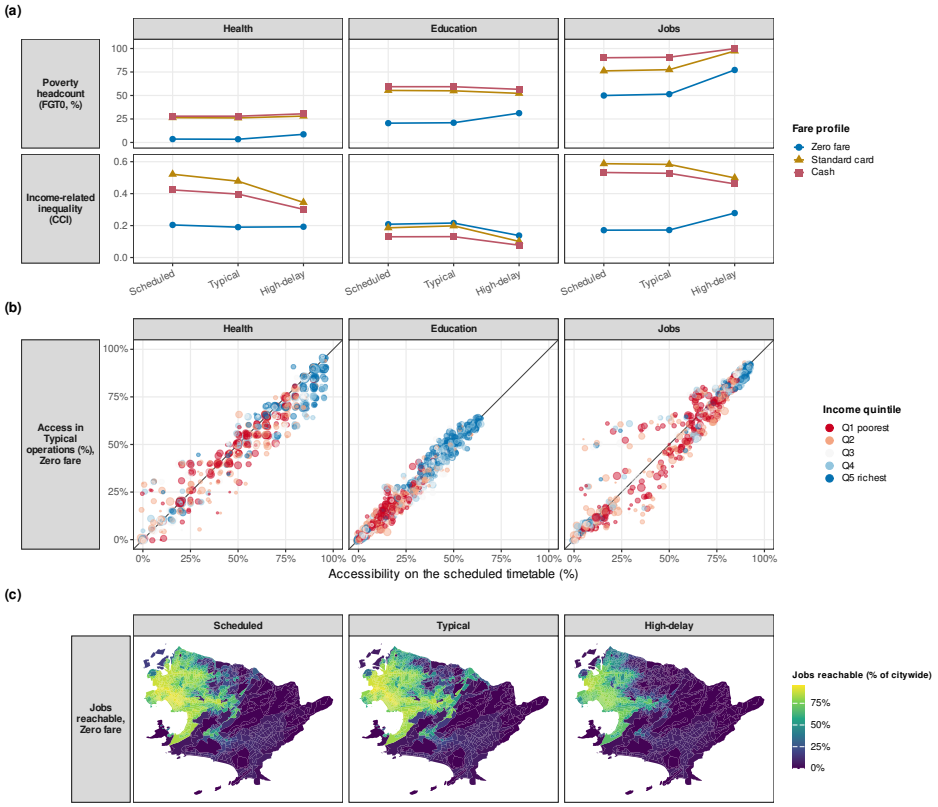


Fig. 1. (a) Accessibility poverty (FGT_0) and income-related inequality (CCI) by network and payment condition, expanded routing matrix; the fixed median jobs line puts the Scheduled zero-fare number near 50% by construction. (b) Tract-level zero-fare accessibility under Typical against Scheduled service (color = income quintile; area = population); points below the diagonal lose access. (c) Share of citywide jobs reachable at zero fare per network; color scale capped at the observed maximum, rounded up to 5%.

0.19) and education (0.21 to 0.14), but increases it in jobs (0.17 to 0.28; under Cash the jobs poverty rate nears 100%).

Employment access is especially vulnerable to fare penalties. Under Scheduled service the Standard-card penalty raises the jobs poverty rate from 50.0% to 76.2% and the CCI from 0.17 to 0.59, and similarly under Typical operations (0.17 to 0.58). Under Standard card, the poorest quintile retains only 4.3% of its zero-fare job access, whereas the richest retains 91.8% (ratios of population-weighted means). Alternative employment lines and 60/90-minute budgets are in the extended report; no jobs contrast reverses under any line.

Conclusions. Accessibility poverty and inequality do not always move together, and the diagnosis depends on whether Scheduled service is compared with Typical or High-delay operations. Fares affect health and education poverty more than changes in service, but their effects on inequality differ by domain. At the reported morning peak, zero-fare poverty under Typical operations remains close to Scheduled levels in every domain, while health inequality falls. Across five two-hour departure windows and an 85th-percentile delay scenario, High-delay operations raise poverty relative to both Scheduled and Typical service in all three domains. They consistently lower education inequality and increase jobs inequality, whereas their effect on health inequality varies across departure times. The comparisons are descriptive, cover one month, and rest on tract-average income and hypothetical payment conditions; within those limits, representing a transit system as scheduled, delivered, or paid for changes its equity diagnosis. Missing São Gonçalo, DETRO-RJ, and ferry GTFS leave cross-bay accessibility at zero, so the comparison is municipal, not regional.

Acknowledgment. This work used data made available by UFF within the NetMob 2026 Data Challenge [4]. **Disclosures.** The authors declare no conflicts of interest. **Data availability.** NetMob26 data [4] are restricted to challenge participants. Both packages introduced above are open source: https://github.com/e-kotov/gtfsrt2static and https://github.com/e-kotov/gps2gtfs.

REFERENCES

1. A. Karner, R. H. M. Pereira, and S. Farber, "Advances and pitfalls in measuring transportation equity," Transportation (2024).
2. C. K. V. Braga, C. F. G. Loureiro, and R. H. M. Pereira, "Evaluating the impact of public transport travel time inaccuracy and variability on socio-spatial inequalities in accessibility," J. Transp. Geogr. **109**, 103590 (2023).
3. D. Vale, "Effective accessibility: Using effective speed to measure accessibility by cost," Transp. Res. Part D: Transp. Environ. **80**, 102263 (2020).
4. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," (2026).
5. R. H. M. Pereira, D. Herszenhut, C. K. V. Braga, *et al.*, "Distribuição espacial de características sociodemográficas e localização de empregos e serviços públicos das vinte maiores cidades do brasil," Texto para Discussão IPEA **2772** (2022).
6. B. Cui, G. Boisjoly, B. Serra, and A. El-Geneidy, "Modal equity of accessibility to healthcare in Recife, Brazil," J. Transp. Land Use **15**, 1–15 (2022).
7. M. Saraiva, L. P. Silva, C. K. V. Braga, and R. H. M. Pereira, "Transporte urbano e insuficiência de acesso a escolas no Brasil," Texto para Discussão IPEA **2854** (2023).
8. J. Foster, J. Greer, and E. Thorbecke, "A class of decomposable poverty measures," Econometrica **52**, 761–766 (1984).
9. G. Erreygers, "Correcting the concentration index," J. Health Econ. **28**, 504–515 (2009).

Edge on Wheels: Forecasting Bus Fleet Compute with Calibrated Uncertainty

WELLINGTON LOBATO[1,*], NADJIB ACHIR[1], AND ALINE CARNEIRO VIANA[1]

[1] *Inria Saclay, TRiBE Team, Palaiseau, France*
[*] *viana-lobato-junior.wellington@inria.fr*

1. MOTIVATION

Multi-access Edge Computing (MEC) deploys computing resources at the network edge to satisfy the low-latency requirements of 5G and 6G applications. In the ETSI reference architecture, a MEC host is provisioned with a *fixed* amount of compute co-located with the access network [1], while the demand it must handle varies with vehicle density. Meanwhile, modern vehicles carry AI accelerators delivering tens to hundreds of TOPS, largely idle during normal driving, which collectively can approach the capability of a nearby MEC host [2]. Public buses are a natural carrier: they have fixed routes and timetables, consistent power, and dwell at terminals.

This supplemental capacity, however, is volatile: it rises and falls as buses enter and leave a cell, so an orchestrator admitting workloads onto it must allocate capacity *before* it is observed. A point forecast is not enough, allocating against a single predicted value ignores the risk that the actual value is lower, and cannot separate reducible (*epistemic*) from irreducible (*aleatoric*) uncertainty [3]. The decision needs a *calibrated predictive distribution*, whose nominal coverage agrees with its observed coverage, so capacity can be committed against a chosen lower bound.

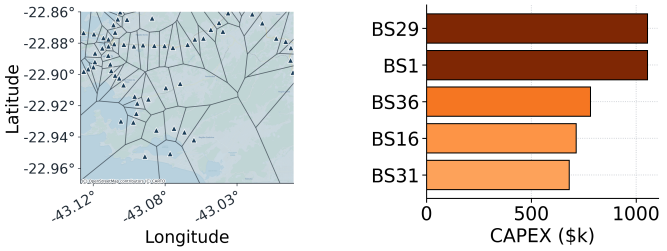
Our contribution is not the Bayesian machinery, which is standard, but the *construction* and the *evaluation*. From one operational transit dataset we build a per-host model of both sides of the ledger, the compute the fleet delivers to each MEC host, the task load its passengers offer to that same host, and evaluate the forecast on what an admission decision needs (coverage, sharpness, proper interval scores) across horizons from one minute to one hour. Two things follow that the literature does not report: how much of the widely quoted per-vehicle compute survives a realistic fleet and task length, and how a passenger workload modeled as a *device population* rather than a boarding count distributes over hosts.

2. SCENARIO AND DATA

We use the NetMob 2026 dataset, covering the Niterói (RJ, Brazil) bus system for March 2026: (i) GPS telemetry at 15-second intervals, 13.85 million records from 527 vehicles (321 active per day on average), and (ii) ticketing, 6 790 681 recorded boardings.[1] GPS is reliable for 19 of the 31 days, so supply-side results cover those, ticket-only results use the month.

We obtain MEC-host locations from OpenCellID.[2] Within the mobility area, 203 LTE cells include groups that, when within 500 m of each other, share one physical mast; agglomerative clustering merges them and places each site at the *centre of mass* of its members, giving 65 hosts (Fig. 1a). Each host owns the Voronoi cell of points closer to it than to any other, and a bus is served by the host whose cell contains it.

A host's supplemental compute is its bus count times a per-vehicle contribution. We take 275 TOPS, the upper end reported for on-board accelerators [2], but treat it as the *upper bound of a measured envelope*, not a capacity (Sec. A).



(a) MEC host with its own Voronoi region.

(b) Edge-GPU CAPEX for the five highest-traffic MEC hosts.

Fig. 1. The fleet is a per-host resource with a concrete cost.

A boarding cannot be geolocated directly, the ticket has no stop identifier and its vehicle number does not overlap the GPS id, but it gives (route, minute), and the GPS says where that route's buses then are. We weight each host by its share of that route's distinct buses and split the minute's boardings in proportion, so the allocation is mass-conserving. 75% of a weekday's boardings are allocatable and 54 of the 65 hosts receive load. A boarding is an *arrival*, not a load, with mean ride time T_{ride} the device population is the rolling sum of the last T_{ride} minutes (Little's law), and two independent occupancy anchors, the measured IPK and a per-trip load profile, bracket it at 12–20 min.

3. CHARACTERISING THE RESOURCE

Both targets are strongly diurnal and out of phase. Supply is bimodal (peak 232 buses/min, 369 in the busiest minute) with a broad midday trough; weekday load is trimodal (450, 345, 480 boardings/min) against a flat weekend plateau near 100. At midday the load *per bus* therefore exceeds even the evening rush, so a rule parameterised on commute peaks alone would mischaracterise the resource exactly when it is scarcest per unit of demand.

Supply is also concentrated. The 10 busiest of 65 hosts hold 61% of hosted bus-time (Gini 0.58) and ten are never visited. Priced in Fig. 1b, the five busiest imply \$0.68–1.05 M in A100 CAPEX at one edge GPU per vehicle, and they are the hardest to forecast.

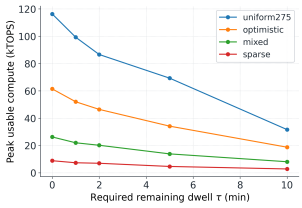
A. How much compute actually survives

Counting every bus at 275 TOPS assumes each carries such an accelerator, lends all of it, and is usable on entering the cell. We relax all three, a fleet mix over an Orin-class ladder $\{0, 40, 100, 275\}$ TOPS, an offerable fraction β, and a *dwell gate* τ counting a bus only while its remaining dwell is at least τ. The gate is measured, not assumed: applied to 577 820 confirmed cell visits (median dwell 196 s) with a 60 s hysteresis removing boundary flapping.

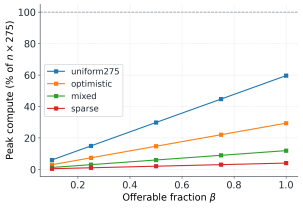
The resource is far more fragile than the headline suggests

[1] The dataset paper quotes ≈7.2 M transactions; the released files contain 6 790 681 rows, reported here unfiltered.

[2] Crowdsourced cell-tower database by Unwired Labs, https://www.opencellid.org, CC BY-SA 4.0.

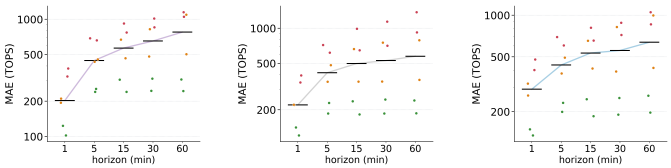


(a) Usable compute vs. required remaining dwell τ.



(b) Envelope at $\tau = 5$ min, as % of the $n \times 275$ TOPS assumption.

Fig. 2. The headline compute is an upper bound, reached only when every bus is equipped.



(a) Persistence. **(b)** Sequence LSTM. **(c)** BNN.

Fig. 3. Only the naive baseline decays steeply with the horizon. Per-host compute-forecast MAE vs. horizon by volatility group (• Stable • Moderate • Volatile).

(Fig. 2). Requiring merely that a bus stay 5 minutes longer leaves 38% of the usable bus-minutes and 60% of the peak compute. Under a plausible retrofit mix at $\beta = 0.5$ and the same τ, 5.9%. We therefore report 275 TOPS as an upper bound and state the configuration behind every downstream number.

4. FORECASTING SUPPLY AND DEMAND

We forecast each target from a 300-minute history plus seven calendar features, at horizons $h \in \{1, 5, 15, 30, 60\}$ minutes. The Bayesian Neural Network (BNN) places a distribution over the weights. Posterior samples give the *epistemic* component and a learned noise term the *aleatoric* one. Because a raw variational interval is rarely calibrated we rescale it with a single temperature fitted on validation, and apply the *same* rescaling to every competing model emitting a distribution, so none is advantaged by being the only calibrated one.[3]

The comparison spans nine models: persistence and seasonal-naïve, a sequence LSTM, N-BEATS and PatchTST as point forecasters, and the BNN, MC Dropout, a Deep Ensemble with a heteroscedastic head, a DeepAR-style Gaussian model and a quantile head, as distributional ones. Ranking is based on proper scores: CRPS and the Winkler interval score, which charges width plus a miss penalty and thus cannot be gamed by idening, alongside coverage (PICP, ACE), sharpness (MPIW), and calibration (ECE, PIT–KS).

The horizon decides whether learning is worth anything. At $h = 1$ the point target is nearly saturated: persistence reaches 325 TOPS MAE at the busiest host against 343 for the sequence LSTM, and beats every learned model at all six panel hosts. By $h = 5$ it is beaten at four of six, and from $h = 15$ at *all* six; at $h = 60$ the LSTM is 26% better than persistence (575 vs 776 TOPS median). Fig. 3 shows why this is a property of the baseline rather than of the task: persistence degrades by 284% from $h = 1$ to $h = 60$, the LSTM by 161% and the BNN by 119%. It also hides a finding: *which* learned model wins depends on volatility. The LSTM leads on stable and moderate hosts (22% and 28% over persistence at $h = 60$), but on the volatile ones it falls behind persistence while the BNN leads, by 17% over the LSTM and 13% over persistence — the model carrying an explicit noise term earns its place exactly where the resource is hardest to predict. A seasonal-naïve baseline fails *per host* ($R^2 < 0$ everywhere) yet reaches $R^2 \approx 0.48$ on the city-wide load: the daily cycle is exploitable in aggregate and vanishes at per-cell one-minute resolution. The crossover comes earlier on the passenger load, where learned models win four of six hosts already at $h = 1$.

Which uncertainty method, not whether. Against the LSTM and persistence the BNN is the only model offering a distribution at all; against the alternatives a reader would reach for, that framing is too generous. Over nine (target, host, horizon) cells the variational BNN with a global noise term ranks *last* of six distributional models on CRPS and Winkler, while a Deep Ensemble with a per-input $(\mu, \log \sigma^2)$ head ranks first, wins 6 of 9 cells, and costs 198× less at inference. N-BEATS and PatchTST, ranked separately because they emit points, trail it and win nothing. A full Temporal Fusion Transformer, run on 5 of 9 cells, never beats its own quantile head and falls further behind as h grows. DeepAR gives the narrowest intervals but under-covers badly (PICP@95 of 79.5% on the passenger target at $h = 15$), which is why we rank on Winkler, which charges width *and* misses.

Two targeted fixes isolate the cause. A heteroscedastic head improves the BNN's ECE by 23–89% at $h \geq 15$ but is counterproductive at $h = 1$, where a learned $\sigma(x)$ merely licenses the model to stop fitting a nearly deterministic mean. Split-conformal calibration, one multiplier per level instead of one temperature, halves ECE on the miscalibrated models (11.2→4.5 pp for MC Dropout) and repairs the 50% interval (66.0→53.8%) while leaving the 95% one untouched. This problem requires a *deep ensemble with a per-input noise head, calibrated conformally*.

5. CONCLUSION AND OUTLOOK

We forecast, with calibrated uncertainty, both the compute a real bus fleet supplies to each MEC host and the task load its passengers generate there, and bound how much of that compute survives realistic assumptions. Three limits are worth stating plainly. A single temperature is *level-local*: it hits 95% coverage but leaves the 50% interval at 58–65%, which conformal calibration repairs. The BNN's point advantage is confined to the volatile hosts. Elsewhere, a plain LSTM is better and cheaper. Admission control needs a task model, deadlines, and an SLA this dataset does not observe, so we report what such a policy consumes, calibrated quantiles, and leave it to a testbed.

ACKNOWLEDGMENT

This work was funded by the French government under the France 2030 ANR program "PEPR Networks of the Future" (ref. ANR-22-PEFT-0003)

REFERENCES

1. ETSI, "Multi-access edge computing (MEC); framework and reference architecture," Tech. Rep. ETSI GS MEC 003 V4.1.1, European Telecommunications Standards Institute (2025).
2. S. Lu and W. Shi, "Vehicle computing: Vision and challenges," J. Inf. Intell. **1**, 23 (2023).
3. J. Gawlikowski, C. R. N. Tassi, M. Ali *et al.*, "A survey of uncertainty in deep neural networks," Artif. Intell. Rev. **56**, 1513 (2023).

[3] Architecture, training settings, measured inference cost and the full metric tables are in the companion report. Splits are by whole day, the normaliser fitted on training days only, the temperature on validation.

Transit Opportunity Landscapes Under Climate Stress: Spatial Equity, Vulnerability Evaluation, and the Usage–Accessibility Gap

CHENCHEN SUN[1,2,*], YUXI ZHANG[1], YICHUAN ZHANG[1], SHUTA OKUAKI[3], AND YUYA SHIBUYA[1]

[1] The University of Tokyo, Hongo 7-3-1, Bunkyo-Ku, Tokyo, Japan
[2] Renmin University of China, Haidian District, Beijing, China
[3] Pitzer College, 1050 North Mills Avenue, Claremont, CA 91711, USA
[*] sun_chenchen@csis.u-tokyo.ac.jp

1. INTRODUCTION

In Latin American cities, public transport is not merely a service but a critical determinant of social inclusion and spatial justice[1, 2]. It has been acknowledged that inequalities in access to opportunities are deeply entrenched along sociodemographic lines, particularly income and race[1–5]. However, several gaps persist when considering the dynamic interaction between transit systems, environmental stressors, and actual user behavior in the Global South. First, many indicators remain rooted in "stop-coverage" logic, focusing on proximity to stations rather than the quality and quantity of opportunities reachable via the network [6]. Second, research has yet to fully explore how extreme weather events, such as floods, which are increasing in frequency due to climate change, spatially disrupt transit networks and exacerbate existing vulnerabilities[7]. Third, the disconnect between "potential" accessibility (what the network offers) and "realized" mobility (how people actually use the system). Identifying where high potential access coexists with low usage (or vice-versa) is vital for diagnosing systemic failures, yet this coupling remains underexplored in dynamic, climate-stressed environments [8].

To address these gaps, this study conducts a vulnerability evaluation of Niterói's transit system under extreme rainfall-induced flooding, landslides, and their compound scenarios to identify spatially uneven risks and stress-test the resilience of opportunity landscapes across sociodemographic groups. We further introduce a "Usage–Accessibility-Vulnerability" framework to examine how observed transit use, normal facility accessibility, and hazard-induced accessibility loss intersect spatially, thereby identifying distinct types of transit service areas and supporting differentiated planning and resilience interventions. This approach contributes to a nuanced diagnostic tool for equitable and climate-adaptive transit planning in the Global South.

2. METHOD

A. Study area

This study focuses on Niterói, RJ, Brazil, with a dense, service-rich center juxtaposed against steep, socio-economically segregated favelas and peripheral low-income neighborhoods[5]. The bus network serves as the primary motorized mode, exhibiting stark spatial imbalances in service frequency and last-mile walkability. Frequent extreme pluvial flooding further compounds transit vulnerabilities, making Niterói a pertinent case for climate-sensitive transit equity analysis[2, 5].

B. Accessibility Modeling

Transit accessibility was modeled as a door-to-facility journey. The travel time calculation formula was constructed shown below. It includes the walking time from home to the nearest bus stop, the travel time on the bus, the transfer waiting time, and walking time from bus stop to destination facility.

$$T_{ij} = T^{\text{walk}}_{i\to\text{so}} + T^{\text{bus}}_{\text{so}\to\text{sd}} + T^{\text{transfer}}_{\text{st}} + T^{\text{walk}}_{\text{sd}\to j} \quad (1)$$

Based on the GTFS and POIs dataset provided by netmob data challenge [9] and OSM, two types of accessibility indicators were calculated. 1) *Reachable facility counts*: the number of different types of facilities reachable within 15, 30, or 45 minutes from census centroid; 2) *Essential-facility visiting time*: the shortest travel time to the first- and second-nearest reachable essential facilities (include education, medical care, basic food supply, and community support).

C. Scenario settings

We simulate two major rainfall induced hazards in Niterói, flooding (using a 5 m 2D hydrodynamic model based on DEM, land cover, and IDF curves) and landslides (based on slope, susceptibility, source areas, and runout, incorporating susceptibility zones and LiDAR DEM). By varying rainfall intensity and hazard type, seven scenarios of differing severity are generated. Hazard impacts are assessed across three components including the delay or closure of walking paths, bus stops, and bus routes. Then the accessibility will be recalculated under updated transit network conditions accordingly.

D. Usage identification

Bus ridership was estimated from ticket transactions by linking each record to GPS-tracked vehicles based on Ticketing and Mobility datasets [9]. Transactions were firstly grouped into boarding bursts to identify stop-level swiping behavior. Each burst was matched to the nearest GPS observation within a ±90-second window and assigned to the nearest GTFS stop served by the corresponding route when the stop was within 60 m of the vehicle location. Matched transactions were counted as inferred boardings. Mean daily boardings were calculated across 19 analysis days, and population-normalized use was expressed as mean daily boardings per 1,000 residents within the stop's 500 m census catchment. Validation checks showed no significant bias in identification rates across income groups.

3. RESULTS

This study revealed profound spatial injustices across Niterói's transit opportunity landscapes under climate stress. First, low-income communities exhibit a lower count of accessible facilities in a given period compared with higher-income communities (Figure 1), especially for commercial, healthcare, educational, and recreational facilities within 15 minutes.

Moreover, the climate-stress vulnerability evaluation shows that flood- and landslide-induced disruptions fragment transit connectivity and disproportionately reduce essential facility

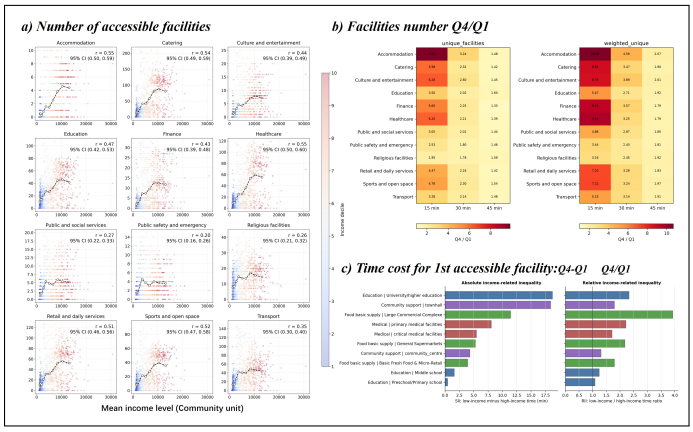


Fig. 1. Income disparities of community accessibility via transit system.

access for lower-income communities, supporting a "double disadvantage" interpretation. Figure 2 decomposes disaster-induced accessibility loss into four components: normal path interruption, rerouting recovery, final unreachability, and additional delay among journeys that remain connected. The results indicate that low-income communities experience greater disruption accessibility losses, as evidenced by higher rates of unreachability and additional delays, alongside lower rates of alternative route availability and rerouting recovery compared with their higher-income counterparts across scenarios.

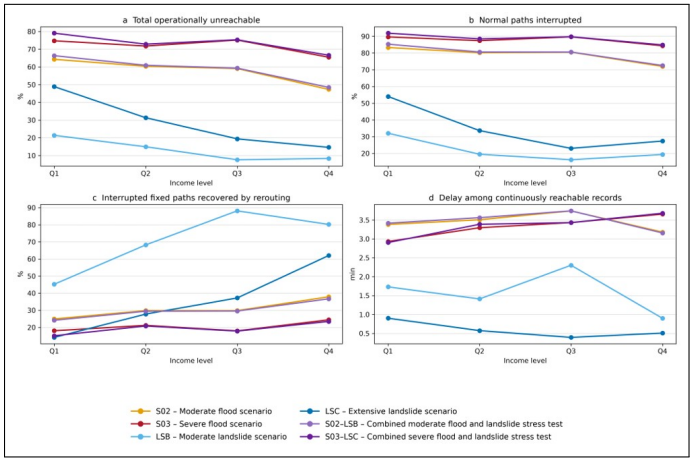


Fig. 2. Income disparities of transit vulnerability across various hazard scenarios.

Ticket transactions add an important nuance.Although low-income areas face poorer baseline accessibility and higher disaster-related unreachability, their observed bus use and boardings per 1000 resident are lower than those of middle-income areas (Figure 3). This may indicate constrained low use: potential transit demand is not fully realized because of limited service supply, fewer reachable opportunities, and weaker network resilience.

Finally, under a usage-accessibility-vulnerability framework, we integrate high-ridership pressure, low-accessibility vulnerability, and disaster-risk exposure to produce an intervention priority map (Figure 4). This map identifies stations requiring capacity and reliability protection, areas needing better route coverage and essential-service access, and high-risk locations requiring resilience investment and emergency alternatives. Equitable transit resilience planning should therefore consider re-

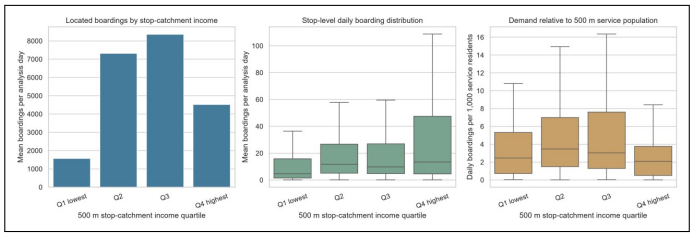


Fig. 3. Identified observed bus use and population-normalized use.

alized demand, unmet demand, essential-facility accessibility, and disaster-induced disruption risk together. Priority interventions should target both high-use, high-risk stops and hidden-vulnerability areas characterized by low income, low accessibility, and weak disaster recovery capacity.

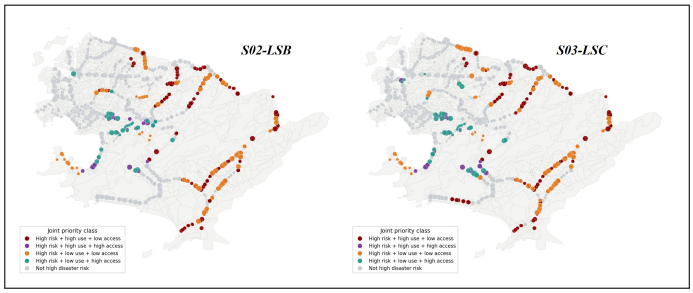


Fig. 4. Intervention priority map.

REFERENCES

1. J. P. Bocarejo S. and D. R. Oviedo H., "Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments," J. Transp. Geogr. **24**, 142 (2012).
2. R. H. M. Pereira, D. Banister, T. Schwanen, and N. Wessel, "Distributional effects of transport policies on inequalities in access to opportunities in rio de janeiro," J. Transp. Land Use **12** (2019).
3. C. Jaramillo, C. Lizárraga, and A. L. Grindlay, "Spatial disparity in transport social needs and public transport provision in santiago de cali (colombia)," J. Transp. Geogr. **24**, 340 (2012).
4. E. C. Delmelle and I. Casas, "Evaluating the spatial equity of bus rapid transit-based accessibility patterns in a developing country: The case of cali, colombia," Transp. Policy **20**, 36 (2012).
5. D. Oviedo Hernandez and H. Titheridge, "Mobilities of the periphery: Informality, access and social exclusion in the urban fringe in colombia," J. Transp. Geogr. **55**, 152 (2016).
6. M. Brussel, M. Zuidgeest, K. Pfeffer, and M. van Maarseveen, "Access or accessibility? a critique of the urban transport sdg indicator," ISPRS Int. J. Geo-Information **8** (2019).
7. G. Vecchio, I. Tiznado-Aitken, and R. Hurtubia, "Transport and equity in latin america: a critical review of socially oriented accessibility assessments*," Transp. Rev. **40**, 354 (2020).
8. K. Lucas, "Transport and social exclusion: Where are we now?" Transp. Policy **20**, 105 (2012).
9. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," (2026).

Impact of provided services and socio-demographic factors on bus accessibility: a Niterói case study

LUCA FANTIN[1,3,*], CHIARA PUGLIESE[2,*], FABIO PINELLI[1,*], GIULIO LODDI[1,*], CHIARA RENSO[4,*], ALESSANDRO BETTI[1,*], BRUNO ZAMENGO[3,*], AND ANDREA FRASSON[3,*]

[1] *IMT School for Advanced Studies Lucca, Lucca, Italy*
[2] *IIT-CNR, Pisa, Italy*
[3] *Motion Analytica s.r.l., Udine, Italy*
[4] *ISTI-CNR, Pisa, Italy*
[*] *luca.fantin@imtlucca.it, chiara.pugliese@iit.cnr.it, fabio.pinelli@imtlucca.it, giulio.loddi@imtlucca.it, chiara.renso@isti.cnr.it, alessandro.betti@imtlucca.it, bruno.zamengo@motionanalytica.com, andrea.frasson@motionanalytica.com*

1. INTRODUCTION

Public transportation services should guarantee equally accessible and efficient coverage throughout the entire urban area for all citizens, regardless of their neighbourhood, socio-economic status, or other factors. Our contribution to the NetMob26 Data Challenge aims to establish the influence of such factors on the *actual* bus accessibility in the Niterói area, taking full advantage of the multi-source, integrated nature of the NetMob26 dataset [1].

Measuring transit accessibility from observed operations rather than planned schedules is the defining methodological challenge in this literature. Previous works [2–5] explore the application of real-time transit data to compute accessibility metrics across diverse urban contexts, while departing from the previous approach based on static schedules. Several of these works [2, 6–9] examine Brazilian and Latin American urban areas, strengthening the value of our contribution regarding the city of Niterói. Several metrics have emerged as adequate summary instruments for transit analysis, such as Lorenz curves, Gini [10] and Suits coefficients [11], and Palma ratios [6, 12]. Despite these converging threads, the simultaneous combination of transit telemetry, payment transaction records, and census sociodemographics within a single framework remains sparse.

Our work aims to compute accessibility equality and examine how it is influenced by spatio-temporal and socio-demographic factors, while exploiting the dataset's integrated, multi-source nature. In particular, the presence of both real-time bus trips and ticket transactions enables fine estimation of demand and supply. Furthermore, the extensive availability of socio-demographic and contextual information enables segmentation of accessibility and equality metrics to identify particularly disadvantaged groups or areas. The same information also supports a complementary, service-oriented perspective: how travel times to essential public facilities, such as schools and hospitals, vary with socio-economic, demographic, and geographic factors. Together, these perspectives let us move from a purely operational description of accessibility to an equity assessment: do disadvantaged areas systematically experience slower, less frequent, less reliable, or more crowded service, and can they reach essential services on comparable terms?

2. RESEARCH CHALLENGE

Our study assesses bus accessibility in Niterói and examines correlations with the city's socio-economic indicators using three complementary analyses. Each is run identically on the published schedule and on a GTFS feed generated from GPS telemetry, in which scheduled stop times are replaced with an aggregation of real-time timestamps matched to each stop. Reporting the same measures on both feeds allows comparison between the accessibility riders actually experience from what the timetable assumes.

Accessibility. The core measure is the cumulative-opportunity accessibility [13]: for an origin area i, a departure time t, and a travel-time budget T,

$$A_i(t,T) \;=\; \sum_j O_j \, f\big(c_{ij}(t,T)\big), \tag{1}$$

where O_j is the number of opportunities in destination area j (e.g. population, essential-service points of interest split by type), $c_{ij}(t)$ is the door-to-door travel time estimated with the R5 routing engine [14] over the GTFS schedule, and $f(\cdot)$ is an impedance function of T (e.g., a cumulative cutoff $\mathbb{1}[c_{ij}(t) \le T]$ or a smooth decay). A distinguishing feature of our work is the time-dependent nature of the measure: whereas accessibility is static in most works, the availability of real-time bus data allows us to compute A_i across different time-of-day bands and analyse the influence of the hour on the equality metrics.

Equality and its drivers. On top of A_i, we compare accessibility-inequality measures such as Lorenz curves, Gini and Palma [12] indices, and Corrected Concentrated Indices (CCIs) [15–17], stratified by census-tract attributes such as income, ethnicity, age, literacy, and household and neighbourhood status. We then quantify how accessibility correlates with these attributes, whether its distribution across the city is more unequal than expected by chance, and whether the association survives controlling for obvious geographic confounders such as distance from the city centre.

Segregation induced by the network. Finally, we assess the influence of the transport network on the computed inequalities. We consider a transport-independent baseline for residential economic segregation, computed as a dissimilarity index of census-tract incomes, assuming that the distributions of these values are lognormal. Afterwards, each bus line is profiled on a segregation scale through a normalised income-entropy computed over its served tracts, to distinguish lines that predominantly serve a single income group from those that bridge socio-economically diverse areas.

3. PRELIMINARY ANALYSES

Figure 1 pairs the spatial pattern of morning-peak accessibility, computed from the original GTFS schedule, with the socio-economic composition of the same census tracts. The accessibility surface (left) is strongly polarised: travel times grow almost monotonically from the well-served western tip of the peninsula, where the ferry terminal, city centre and adjacent neighbourhoods are located, towards the eastern and oceanic periphery.

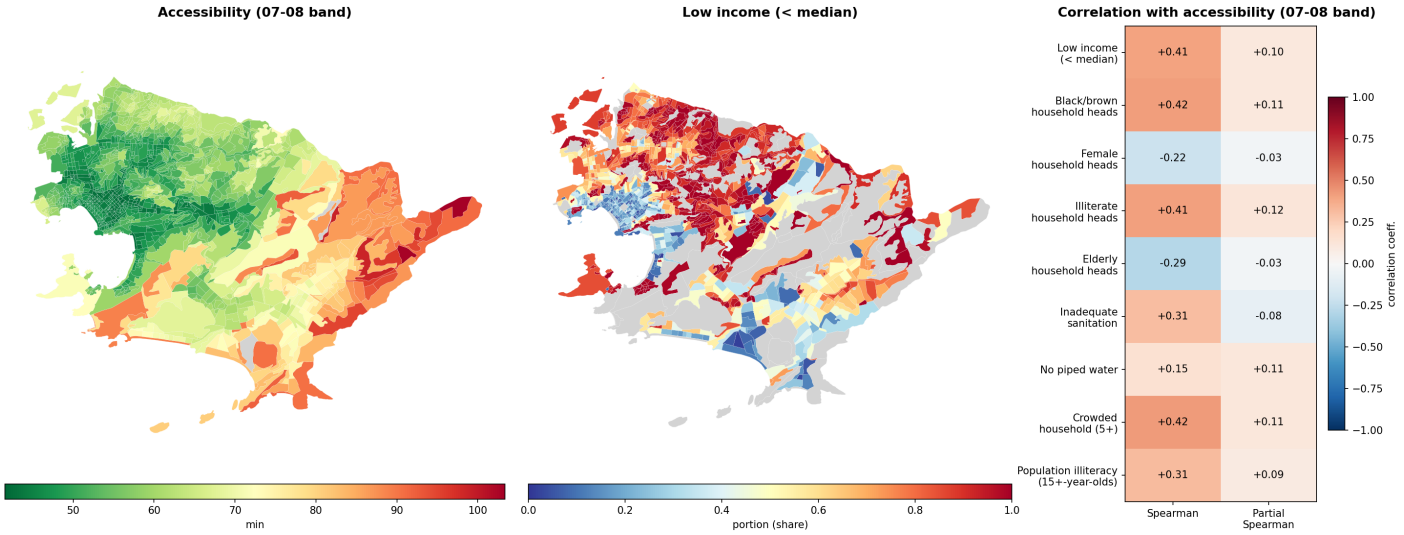


Fig. 1. Accessibility vs. socio-economic indicators (07–08 morning band). *Left*: per-tract accessibility, expressed through the mean R5 door-to-door transit travel time (minutes) needed to reach each census tract from the rest of the city, computed from the original GTFS schedule. *Centre*: estimated share of low-income (below the city median) household heads per tract. *Right*: correlation between different tract-level disadvantage shares and accessibility, expressed through Spearman coefficients and partial Spearman coefficients, the latter removing the distance to the GTFS-derived city centre.

Moving between these two regions, the mean time to reach other tracts increases from 45-60 minutes to over 90. This gradient is already a service-quality signal, as it shows a large fraction of the territory sits well above the network median.

The correlation matrix (right) turns this spatial reading into an equity statement. Most of the displayed indicators of socio-economic disadvantage correlate positively with travel time, indicating that the most disadvantaged tracts tend to be the least accessible. The visual overlap between the low-income map (centre) and the red eastern periphery of the accessibility map reflects these findings for this specific indicator. The two negative indicators, female- and elderly-headed households, instead reflect their concentration in the central, better-served neighbourhoods. Crucially, once we exclude the correlation with the distance to the city centre, the values of the Spearman coefficients, which are now called *partial Spearman coefficients*, collapse towards zero from both directions, showing that most of the measured inequality is actually geographic: disadvantaged groups are penalised mainly because they live far from the centre. Residuals with the same directions persist across all indicators, suggesting a genuine, if modest, accessibility penalty beyond mere location.

The full report will expand on these preliminary analyses by extensively examining correlations between accessibility and all socio-economic indicators, computing inequality and segregation measures, and comparing the results across the original and recreated GTFS schedules. A comprehensive discussion of the full results will yield an exhaustive view on the service quality and how evenly this quality is distributed across socio-economic and geographic population segments.

REFERENCES

1. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, “The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói,” (2026).
2. C. K. V. Braga, C. F. G. Loureiro, and R. H. Pereira, “Evaluating the impact of public transport travel time inaccuracy and variability on socio-spatial inequalities in accessibility,” J. Transp. Geogr. **109** (2023).
3. A. Nalin, N. Fulman, E. C. Wilke, *et al.*, “Evaluation of accessibility disparities in urban areas during disruptive events based on transit real data,” Commun. Transp. Res. **5**, 100160 (2025).
4. L. Liu, A. Porr, and H. J. Miller, “Measuring the impacts of disruptions on public transit accessibility and reliability,” J. Transp. Geogr. **114** (2024).
5. F. Pinelli, A. Hou, F. Calabrese, *et al.*, “Space and time-dependant bus accessibility: A case study in Rome,” in *2009 12th International IEEE Conference on Intelligent Transportation Systems*, (IEEE, St. Louis, 2009), pp. 1–6.
6. D. Herszenhut, R. H. Pereira, L. d. S. Portugal, and M. H. d. S. Oliveira, “The impact of transit monetary costs on transport inequality,” J. Transp. Geogr. **99**, 103309 (2022).
7. F. Xu, Q. Wang, E. Moro, *et al.*, “Using human mobility data to quantify experienced urban inequalities,” Nat. Hum. Behav. **9**, 654 (2025).
8. G. Luz, M. H. Barboza, L. Portugal, *et al.*, “Does better accessibility help to reduce social exclusion? Evidence from the city of São Paulo, Brazil,” Transp. Res. Part A: Policy Pract. **166**, 186 (2022).
9. N. Valenzuela-Levi, “Income inequality and rule-systems within public transport: A study of Medellín (Colombia) and Santiago (Chile),” J. Transp. Geogr. **112**, 103700 (2023).
10. A. Raza, M. Zhong, R. Akuh, and M. Safdar, “Public transport equity with the concept of time-dependent accessibility using Geostatistics methods, Lorenz curves, and Gini coefficients,” Case Stud. on Transp. Policy **11**, 100956 (2023).
11. G. Sharma and G. R. Patil, “Public transit accessibility approach to understand the equity for public healthcare services: A case study of Greater Mumbai,” J. Transp. Geogr. **94**, 103123 (2021).
12. K. Silver, A. Lopes, D. Vale, and N. M. da Costa, “The inequality effects of public transport fare: The case of Lisbon’s fare reform,” J. Transp. Geogr. **112**, 103685 (2023).
13. K. T. Geurs and B. Van Wee, “Accessibility evaluation of land-use and transport strategies: review and research directions,” J. Transp. Geogr. **12**, 127 (2004).
14. M. W. Conway, A. Byrd, and M. van der Linden, “Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks,” Transp. Res. Rec. **2653**, 45 (2017).
15. N. C. Kakwani, “Measurement of Tax Progressivity: An International Comparison,” The Econ. J. **87**, 71 (1977).
16. D. B. Suits, “Measurement of tax progressivity,” Am. Econ. Rev. **67**, 747 (1977).
17. G. Erreygers, “Correcting the Concentration Index,” J. Health Econ. **28**, 504 (2009).

Reaching the Stop, Reaching the City: Socio-Spatial Inequality in Public-Transport Access in Niterói (SDG 11.2.1)

KININ KONÉ[1,*] AND ABDOULAYE DOUMBIA[1]

[1] *Agence Nationale de la Statistique (ANStat), Abidjan, Côte d'Ivoire*
[*] *Corresponding author: kinin.kone@stat.plan.gouv.ci*

1. INTRODUCTION

SDG indicator 11.2.1 measures the share of the population with convenient access to public transport. Its stop-proximity component commonly uses a 500 m network-walking threshold for low-capacity public-transport stops [1]. Two equity questions remain: does reaching a nearby stop provide satisfactory access to essential destinations, and is the service specified in the timetable actually delivered?

We study Niterói (metropolitan Rio de Janeiro; 481,749 inhabitants, 52 neighbourhoods and 1,274 census tracts). We compare the service scheduled in GTFS—which specifies planned stops, routes and timetables—with bus passages observed in GPS data. We combine these measures with a study-specific socio-spatial vulnerability index and network travel times to essential services to identify priority neighbourhoods. Our central finding is that inequalities in stop proximity and destination travel times are substantially larger than the gap between schedule-based and GPS-observed coverage. Service regularity and upper-tail waiting times remain to be assessed.

2. DATA AND METHODS

We use the NetMob 2026 Niterói dataset [2]: bus GPS telemetry, smart-card ticketing (6.79 M retained transactions, March 2026), and two operator GTFS feeds (TransNit, TransOceânico), with the 2022 IBGE census [3], the OpenStreetMap walk network, and municipal facility layers. All spatial layers were reprojected to SIRGAS 2000 / UTM zone 23S (EPSG:31983). The GTFS files from the two operators were combined to cover the complete network, and mismatched identifiers between sources were reconciled using explicit crosswalks, after a data-quality audit (decimal-format, time-zone and identifier issues). Favela/FCU membership was taken directly from the official IBGE delimitation (the *Favelas e Comunidades Urbanas* layer) as a binary per-tract flag; hereafter "favela/FCU" denotes tracts in this official category and "non-favela" the rest. We constructed a study-specific *vulnerability index* as the equal-weighted mean of seven standardised components, all oriented so that higher values indicate less favourable conditions—favela/FCU membership, illiteracy, household-infrastructure deficits, female- and Black/Brown-headed household shares, density, and inverted walkability—all kept after a collinearity check ($|r| < 0.8$). Smart-card data were used only to characterise overall demand volume. Because vehicle identifiers could not be matched between the smart-card and GPS datasets, transactions could not be assigned reliably to individual stops and therefore do not enter the coverage or accessibility indicators.

Stop proximity was measured as the share of residents able to reach a served stop within 300, 500 or 800 m along the pedestrian network, within the Niterói boundary. The 500 m threshold is the UN-Habitat reference distance for low-capacity public-transport stops [1]. Population was allocated uniformly to a 50 m grid within each census tract, with Euclidean distance used as a fallback for 6.3% of grid points. Stop-proximity coverage (Fig. 1) counted stops with at least one scheduled passage. Separately, planned and GPS-map-matched service were compared *symmetrically* at a ≤ 30 min mean-headway threshold, by stop, day-type and time band. An exploratory bivariate Local Indicators of Spatial Association (LISA) analysis identified tracts combining high vulnerability and high access deficit [4]. Service accessibility was measured as the median scheduled travel time combining walking and bus travel from each tract's geometric centroid to the nearest facility in each of four categories: hospital, health unit, school and university. Computed with r5py [5] for departures between 07:00 and 08:00 on 18 March 2026, this measure includes walking, scheduled waiting before the first bus, in-vehicle time and transfers. Finally, to synthesise these dimensions, we constructed a study-specific composite equity index. It is the equal-weighted mean of four standardised components oriented in the same adverse direction: coverage, effective frequency, service access and vulnerability. A weight-sensitivity analysis assessed the influence of the aggregation weights.

3. PRELIMINARY RESULTS

The results highlight four complementary dimensions of access equity: stop proximity, travel time to essential services, the gap between schedule-based and GPS-observed coverage, and the spatial concentration of disadvantage.

Stop proximity: a last-mile gap. City-wide, 83.9% of residents reach a stop within 500 m, but favela/FCU residents trail by 11.0 points (74.8% vs. 85.8%). The gap *rises to 25.6 points at 300 m*—just over half the favela/FCU population then lacks a stop within the stricter 300 m threshold—and closes to 1.8 points at 800 m (Fig. 1).

Destination access: unequal travel times. Beyond proximity, network travel time to essential services is longer from favelas for three of four services (Table 1): about +2 min to a hospital, +3 min to a health unit, and +6 min (+26%) to a university (29 vs. 23 min), while school access is equal. In Várzea das Moças, the median time to the nearest university reaches 102 minutes.

Scheduled and observed coverage: a small gap. Under the same 30-minute mean-headway threshold, GPS-observed coverage was only 0.7 percentage points below schedule-based coverage city-wide. The corresponding gap was approximately 1.1 points in favela/FCU tracts and 0.5 points elsewhere. This threshold-based comparison does not capture frequency deterioration that remains within the 30-minute threshold, punctuality, or upper-tail waiting times. Within the scope of this indicator, however, the gap between scheduled and observed service is approximately one order of magnitude smaller than the stop-proximity gap.

Spatial concentration: localised disadvantage. Access is strongly spatially clustered (global Moran's $I = 0.61$). Because 1,274 local tests were run simultaneously, we applied the Benjamini–

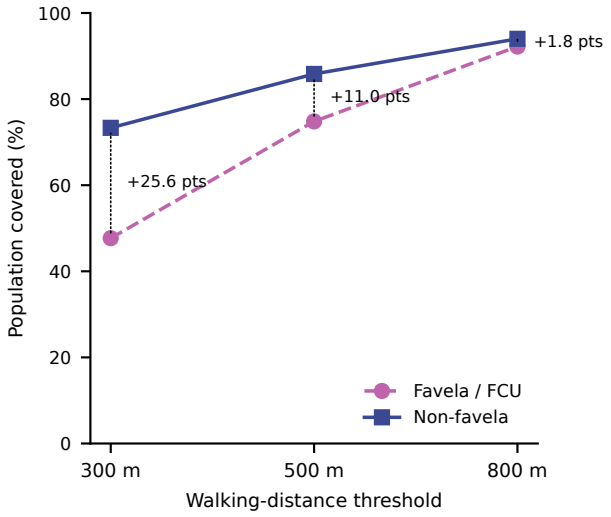


Fig. 1. Stop-proximity component of SDG 11.2.1 (stops with ≥ 1 scheduled passage) by walking threshold, for favela/FCU (dashed, circles) and non-favela (solid, squares) populations. The gap widens at short distances and closes at 800 m—the last-mile inequality. Curves are distinguished by line style and marker, not colour alone.

Table 1. Unweighted tract-level median scheduled travel times combining walking and bus travel to the nearest essential service (minutes), by residential group

Service	Favela/FCU	Non-favela
Hospital	16	14
Health unit (UBS)	20	17
School	6	6
University	29	23

Hochberg false-discovery-rate correction [6] to limit the expected share of false positives: 59 High-High tracts (combining high vulnerability and high access deficit) remained significant, compared with 85 at an unadjusted directed permutation pseudo-$p < 0.05$. The LISA-based and composite-index rankings partially overlap: Várzea das Moças, Muriqui and Jacaré appear among the top five neighbourhoods under both rankings. Sapê ranks first under the LISA-based ranking and sixth under the composite index (Fig. 2). Formal aggregation of the two rankings is left to the full report.

4. DISCUSSION AND CONCLUSION

Among the dimensions measured here, proximity and destination access show substantially larger inequalities than the gap between schedule-based and GPS-observed coverage: vulnerable neighbourhoods have poorer stop proximity (the last mile) and longer network travel times to hospitals and universities. Where service is spatially available, planned coverage is largely delivered, so the results point first to fine-grained redistribution (stop densification, feeder services, pedestrian access) in the identified neighbourhoods. Reliability interventions or new trunk capacity require further study.

Limits: the effective-service comparison rests on a 19-day GPS window; disjoint vehicle identifiers preclude individual-

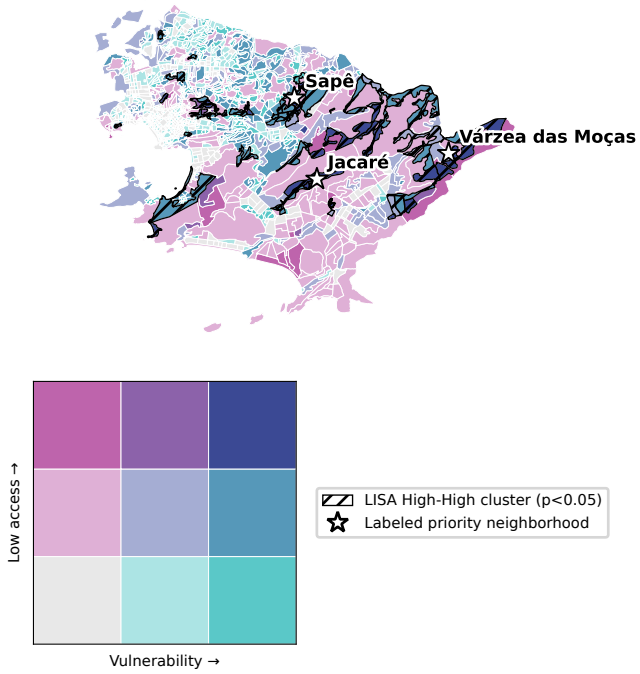


Fig. 2. Bivariate map of socio-spatial vulnerability and bus-stop access (500 m). High-High denotes tracts combining high vulnerability and high access deficit; the map displays 85 tracts at an unadjusted directed permutation pseudo-$p < 0.05$, of which 59 remain after Benjamini–Hochberg FDR correction across all 1,274 local tests. Stars mark selected illustrative high-ranked neighbourhoods; the labels are non-exhaustive. Categories are shown by pattern and label, not colour alone.

trip demand reconstruction; most health-unit locations (9 of 11) rely on neighbourhood-centroid proxies; and OpenStreetMap pedestrian-network completeness may differ between formal and informal areas, which could affect the last-mile gap. Travel times computed with r5py already include scheduled waiting before the first bus and scheduled transfer waiting; future work will derive route- and destination-specific GPS headway distributions and implied upper-tail waiting times, refine the LISA (two-sided, more permutations), and test the hypothesised north–south gradient. The reusable pipeline is transferable to other statistical offices—here, to Abidjan (2021 census, SOTRA network).

REFERENCES

1. UN-Habitat, "SDG indicator 11.2.1 training module: Public transport system," Tech. rep., United Nations Human Settlements Programme, Nairobi (2018).
2. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 dataset: A high-resolution multi-source view of public bus mobility in Niterói," (2026). ArXiv:2605.20263 [physics.soc-ph].
3. IBGE, "Censo demográfico 2022," Tech. rep., Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro (2023).
4. L. Anselin, "Local indicators of spatial association—LISA," Geogr. Anal. **27**, 93 (1995).
5. C. Fink, W. Klumpenhouwer, M. Saraiva, *et al.*, "r5py: Rapid realistic routing with R5 in Python," (2022).
6. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. Royal Stat. Soc. Ser. B (Methodological) **57**, 289 (1995).

Urban Mobility and Demand Dynamics: A Telemetry and Ticketing Baseline for Niterói's Bus Network

EVALDO ALVES DE BRITO JÚNIOR, LUIZ SERGIO POMPEU ALVES FILHO, AND CLAUDIO E. C. CAMPELO

Systems and Computing Department, Federal University of Campina Grande (UFCG), Campina Grande, PB, Brazil

1. INTRODUCTION AND MOTIVATION

The public transportation networks of rapidly expanding urban centers confronts a complex task in balancing operational efficiency with dynamic commuter needs, which in face of it data-driven approaches providing multi-perspective datasets offer critical insights into the city's systemic demand.

In this work, we utilize the NetMob 2026 Data Challenge dataset for the city of Niterói in Brazil, to investigate the structural dynamics of the municipal bus network. Specifically, our research framework is driven by three interconnected urban mobility objectives:

1. **Urban Movement Patterns:** To determine whether passenger boarding logs trace periodic, routine, and predictable movement patterns across distinct urban zones in Niterói.
2. **Operational Anomalies:** To analyze vehicle trajectory profiles via GPS telemetry and detect both scheduled and unscheduled stops, besides possible bottlenecks.
3. **Predictive Modeling:** To develop and evaluate a supervised learning framework capable of predicting continuous bus trip durations.

2. DATASET AND PREPROCESSING

Our analytical framework relies on the parsing and synchronization of two primary data repositories provided by the NetMob 2026 Data Challenge: the mobility telemetry streams and the smart card electronic ticketing registers.

A. Mobility Telemetry and Network Mapping

The vehicle tracking dataset consists of high-frequency GPS ping sequences aggregated from multiple CSV source files. To construct a consistent timeline, raw string timestamps are cast into standardized `datetime` instances, enabling the easy extraction of trip informations like calendar day, operational hour, and precise temporal indices, while spatial validation is performed by enforcing a geometric boundary matrix restricting geographical data points to Niterói's municipal limits.

B. Ticketing Validation and Identity Cross-Matching

The passenger boarding dataset captures individual transaction events categorized by four discrete structural day periods (21:00–02:00, 03:00–08:00, 09:00–14:00, and 15:00–20:00). To bridge individual passenger demand with the vehicles routes, we evaluate the cross-table entity overlap between the transaction logs and mobility traces. In that process, an identity mismatch anomaly was identified during the vehicle registration process: checking raw identifier indices reveals distinct overlap limits between unique vehicle numbers (`vehicle_number` $\cap$ `id`). Furthermore, we also found that route descriptions contained parsing discrepancies due to inconsistent string formatting and fractional suffix definitions between the tables(e.g., trailing decimal structures), problem which we solved by executing a regular expression normalization function to remove the aforementioned irrelevant fractional parts.

3. METHODOLOGY

A. Spatiotemporal Demand Profiling and Geometry Mapping

To characterize routine commuter patterns (Objective 1), we isolated the shared intersection of validated routes, $R_{\text{common}} = L_{\text{mobility}} \cap R_{\text{ticket}}$. Electronic boarding logs are aggregated into a line-by-period matrix to isolate localized spatial configurations. To map this transaction index onto physical transit infrastructure, the aggregated volume for each 5-hour period is joined to the corresponding route's mobility coordinates. Crucially, this join associates the period's cumulative demand with the entire physical route layout to enable a geographic rasterization. To mitigate telemetry noise, the captured coordinates are rounded to a fixed decimal threshold, establishing a uniform spatial coordinate rasterization layer with an approximate topological resolution of $\approx$ 11 meters, while the coordinates lacking active transaction pairs are dynamically truncated prior to network plotting.

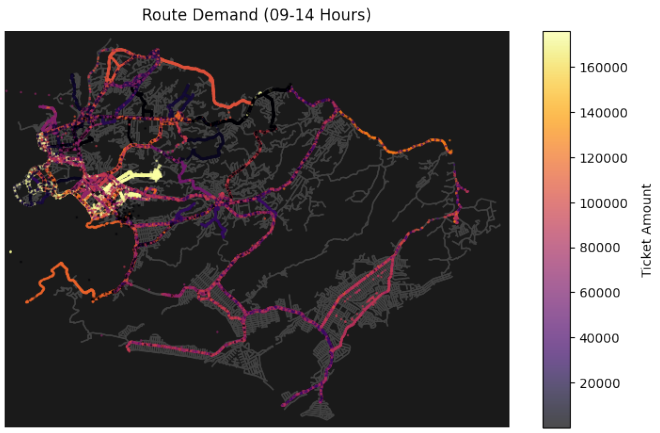


Fig. 1. Geographical rasterization ($\approx$ 11 m resolution) of ticket volume variations across Niterói's public bus network during the mid-day period (09:00–14:00 hours).

B. Operational Stop Validation and Spatial Anomalies

The extraction model isolates recurring vehicle dwell events across Niterói's public transport network. To define a valid dwell event, one is triggered when the spatial displacement between sequential records remains within a decimal threshold of 10^{-5} degrees ($\approx$ 1.1 meters) for a continuous duration of at least 3 consecutive GPS pings ($\Delta t \geq 30$ s on average). Although using only 2 consecutive GPS pings would substantially increase the number of detected events, accounting for approximately 76% of all detected stops, this criterion was not adopted due to the high probability of false detections arising from GPS positioning errors. The chosen spatial-temporal constraint mathematically implies a velocity threshold of $v \leq 0.04$ m/s (≈ 0.13 km/h), effectively filtering out minor pauses and transient positioning inaccuracies. As illustrated in Fig. 2, the duration of detected stops exhibits high spatial variance. Moreover, the detected dwell

events were compared with relevant OpenStreetMap (OSM) features along the bus routes in Niterói, including Traffic Signals, Bus Stops, Signalized Crossings, Bus Stations, Public Transport Stations, Crossings, and Roundabouts. This analysis aimed to assess the spatial relationship between detected dwell events and nearby transportation-related infrastructure. Furthermore, the consistency and distribution patterns of the detected dwell events across different days of the week were analyzed.

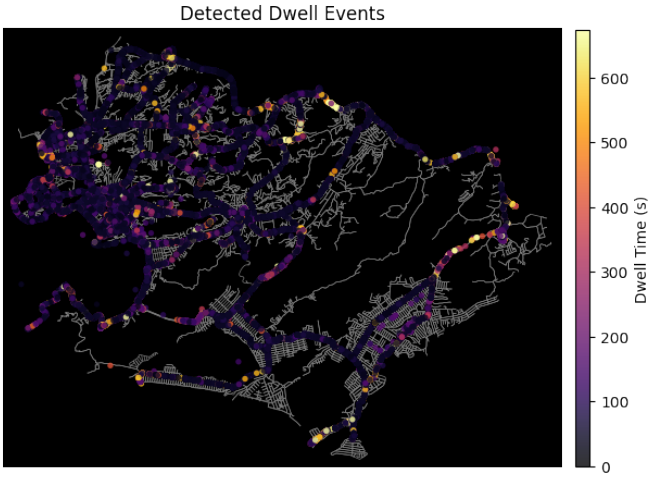


Fig. 2. Spatial distribution and duration (seconds) of algorithmically detected public transit vehicle dwell events.

C. Continuous Trip Segmentation and Predictive Modeling

To construct the predictive framework (Objective 3), telemetry streams are processed sequentially to partition continuous tracking data into individual, distinct operational trips. Because the dataset's `tripId` actually represents a persistent route itinerary rather than an isolated run, a custom multi-criteria segmentation algorithm is applied. A new trip segment initiation is triggered whenever a change is detected in: (i) the vehicle registration identifier (`id`), or (ii) the active `headsign`. Outliers are filtered using an Interquartile Range (IQR) mask bounding trip durations between 3.5 and 107.24 minutes. The resulting trips exhibited an average duration of 46.9 minutes and a median duration of 44.8 minutes.

For feature engineering, temporal attributes including *start hour*, *start minute*, and *weekday* (ranging from 0 to 6) were incorporated. In addition, the vehicle's *direction* and *headsign* were included as categorical features. The *headsign* variable was numerically encoded by assigning a unique integer identifier to each distinct headsign. Finally, the geodesic distance between the geographic coordinates of the trip's starting and ending points was included as a feature. This measure represents the shortest distance along the Earth's surface between the two coordinates, independently of the actual road network or route followed by the vehicle. The processed feature matrix X is split (train=80%; test=20%) to train and benchmark three regression architectures: Linear Regression, Random Forest Regressor, and eXtreme Gradient Boosting (XGBoost).

4. RESULTS AND DISCUSSION

The spatial integration of smart card transactions onto vehicle telemetry layers reveals clear operational structures and demand variations. As shown in Fig. 1, midday demand (09:00–14:00) exhibits extreme monocentric polarization, where a high-volume nucleus exceeding 140,000 transactions converges tightly around the central terminal zone. Beyond this commercial core, the network displays a steep demand gradient; radial arterial corridors branching north and south maintain stable, moderate passenger volumes (60,000–100,000 tickets), while peripheral feeder lines drop off significantly. This clear segmentation confirms that electronic ticketing data captures highly routine, structured displacement flows rather than stochastic passenger movements.

It was also made clear that the demand structure directly influences the vehicle dwell events mapped in Fig. 2. The analysis revealed substantial spatial and temporal variation in vehicle dwell events. Expected dwell times of 30–120 seconds were concentrated in central areas, whereas isolated delays exceeding 600 seconds occurred along major highways and peripheral junctions, indicating potential traffic bottlenecks. A strong spatial correspondence with OSM infrastructure was also observed: 64.28% of dwell events were within 50 m of at least one OSM-tagged feature, while 99% of the OSM points had a nearby dwell event. Temporally, weekdays recorded approximately 60,000–75,000 events, nearly twice the 30,000–35,000 observed on weekends, with counts increasing from Sunday to Wednesday, remaining stable through Friday, and decreasing toward Sunday. This pattern is consistent with the higher operational activity associated with weekday transit demand and further demonstrates that the detected dwell events capture meaningful spatial and temporal variations in bus operations.

Finally, the supervised learning framework demonstrated strong performance in predicting transit trip durations. The baseline Linear Regression model achieved a Mean Absolute Error (MAE) of 11.65 minutes, a Root Mean Squared Error (RMSE) of 14.71 minutes, and a coefficient of determination (R^2) of 0.45. The ensemble-based models substantially improved predictive performance, indicating their ability to capture the non-linear relationships and complex spatiotemporal interactions underlying transit travel times. The Random Forest Regressor reduced the MAE to 5.23 minutes and the RMSE to 8.06 minutes, while achieving an R^2 of 0.835. XGBoost yielded the best overall results, with an MAE of 5.22 minutes, an RMSE of 7.90 minutes, and an R^2 of 0.84. Overall, these results indicate that integrating route-related characteristics with temporal features can substantially improve the prediction of transit trip durations, providing a promising foundation for applications such as real-time fleet management, operational planning, and passenger travel-time estimation.

5. CONCLUSION AND FUTURE WORK

This work established a robust data fusion and predictive framework by synchronizing heterogeneous passenger boarding transactions with high-frequency GPS telemetry across Niterói's bus network. Our spatial analyses successfully mapped highly monocentric demand structures (Objective 1) and isolated critical roadway friction points by parameterizing structural dwell events (Objective 2).

Addressing the core predictive task (Objective 3), we developed a multi-criteria trip segmentation algorithm to overcome the temporal inconsistencies associated with the native itinerary indices. By integrating these engineered temporal, directional, and route-specific features into supervised learning models, we achieved highly precise travel duration forecasts.

Future developments may focus on deploying these trained architectures within dynamic, real-time prediction pipelines to feed passenger-facing arrival systems. Additionally, we aim to incorporate external environmental streams, such as real-time weather logs and sudden road event notifications, to enhance model robustness against unpredicted roadway disruptions.

The Predictability of Bus Bunching and Who Pays for It: Evidence from Niterói, Brazil

HUGO BARBOSA[1], FEDERICO BOTTA[1,*], ZIHAO CHEN[1], LUIZ CORREIA[1], YIFAN HE[1], SOROUSH KARIMI[1], RONALDO MENEZES[1], DIOGO PACHECO[1], AND YUCHEN ZHOU[1]

[1] *Department of Computer Science, University of Exeter, UK*
[*] *f.botta@exeter.ac.uk*

1. INTRODUCTION

Bus bunching refers to the tendency of buses that should be evenly spaced to arrive in tight clusters separated by long gaps, and it is a familiar failure mode of high-frequency transit, driven by an unstable feedback in which a late vehicle meets more waiting passengers, dwells longer, and falls further behind [1, 2]. Because mean waiting time grows with the *variability* of headways rather than the average headway alone [3], a route can dispatch its scheduled number of buses per hour and still leave passengers waiting far longer than the timetable implies. These costs are not borne equally: unpredictable service compounds disadvantage for the transit-dependent riders least able to absorb it, and perceived waiting time is worse for women in unsafe settings, older passengers, and stops lacking basic amenities [4–6]. Identifying where bunching concentrates is, in this sense, also a way of asking who pays for it. Prior work has identified bunching's principal drivers and a range of metrics to quantify it [4, 7], yet most services still get little or no real-time control [4].

The NetMob26 Data Challenge dataset, which includes GPS telemetry every 15-30 seconds paired with 7.2 million March 2026 ticketing transactions, route geometries, and socio-demographic layers for Niterói, Brazil [8], allows us to study this directly. Here, we reconstruct headways from the GPS traces with our own matching pipeline and compute four reliability metrics from them. We identify hotspots of access burden, and try to link burden to socioeconomic context.

2. METHODOLOGY

A. Matching and headway calculation

We calculate scheduled and observed bus headways using GTFS timetable and mobility data, using our own pipeline (via the Python package `rt2gtfs`[9]) to match real-time vehicle positions to scheduled GTFS trips and the nearest stops along each route, after which we estimated the actual arrival time at each stop. For each route-direction pair, we identify each trip's previous trip and take the headway as the time between them at each shared stop. The first trip of the service day is excluded for lack of a predecessor. This ran separately for every date in March 2026 with available mobility data, over both the original scheduled GTFS and a corrected GTFS.

To answer the questions of where the network is bunched and who it affects most, we require measures of both how badly service deviates from plan and what it costs passengers. We calculate a series of metrics by groups defined by route, direction, stop, and day type (weekday / Saturday / Sunday), with a minimum-sample threshold (typically $n \geq 30$ pairs, $n \geq 300$ observations for the predictability estimator) below which a group is marked low-confidence and excluded from reported figures. Groups are further stratified by scheduled frequency band (≤ 10, 10–20, 20–40, > 40 minute mean scheduled headway) where the metric's interpretation depends on it.

B. Metric definitions

Within a given route/direction/stop/day-type group, i indexes matched pairs of an observed trip and its immediately preceding trip. We then calculate the scheduled headway S_i and observed headway O_i as the time differences between them.

Bunched share refers to the share of pairs with $O_i < 0.25 \cdot S_i$ (*severe*, p25) and, separately, $O_i < 0.50 \cdot S_i$ (*moderate*, p50), reported side by side as a consistency check.

Headway adherence measures the spread of how far each bus's actual headway deviates from its scheduled headway. Let $D_i = O_i - S_i$, the trip-level deviation from the group's own plan. The metric itself, $\mathrm{SD}(\{D_i\})$, is how consistently that deviation holds across the trips in the group.

Coefficient of variation (CV+) measures whether running is more irregular than the timetable, independent of a route's own average headway. Let $\mathrm{CV}_{\mathrm{sched}} = \mathrm{SD}(S)/\mathrm{mean}(S)$ and $\mathrm{CV}_{\mathrm{obs}} = \mathrm{SD}(O)/\mathrm{mean}(O)$, each computed over every headway of that type in the group. The metric itself is $\mathrm{CV}+ = \mathrm{CV}_{\mathrm{obs}} - \mathrm{CV}_{\mathrm{sched}}$, unitless and scale-free.

Excess Waiting Time measures the passenger cost of headway irregularity in minutes of waiting, rather than in the headway units the preceding metrics use. The Osuna–Newell approximation [3]

$$\mathrm{AWT} = \frac{E[H^2]}{2 \cdot E[H]} \tag{1}$$

gives a random passenger's expected wait, where H is a headway and $E[\cdot]$ averages over all headways in the group. Applied to the group's scheduled and observed headways in turn, $\mathrm{EWT} = \mathrm{AWT}(\mathrm{observed}) - \mathrm{AWT}(\mathrm{scheduled})$ gives the extra passenger-minutes attributable to running not matching plan. Because AWT depends on headway spread, not just the mean, EWT is positive whenever running is more variable than planned, even on an apparently punctual route.

3. RESULTS

Figure 1 depicts all four metrics described above divided by scheduled frequency band.

Bunched share is the clearest result: of 1.08 million matched pairs, 3.64% were severely bunched (p25) and 8.96% moderately (p50). It is the only metric falling monotonically with frequency, from 14.4% moderately bunched on $\leq$10 min routes to zero on $>$40 min routes, since a headway must already be short for a further short gap to register. Both thresholds pick out the same stops (Fig. 2), concentrated along corridors rather than scattered.

Adherence has a median of 7.14 min against a raw observed SD of 8.54 min, and peaks mid-range (20–40 min) rather than varying monotonically. Seven minutes is a large fraction of most scheduled headways here, so the timetable is a weak guide to the gap a passenger will face. In 8.8% of groups adherence exceeds raw spread: there the plan itself is unachievable, and retiming rather than real-time control is the remedy.

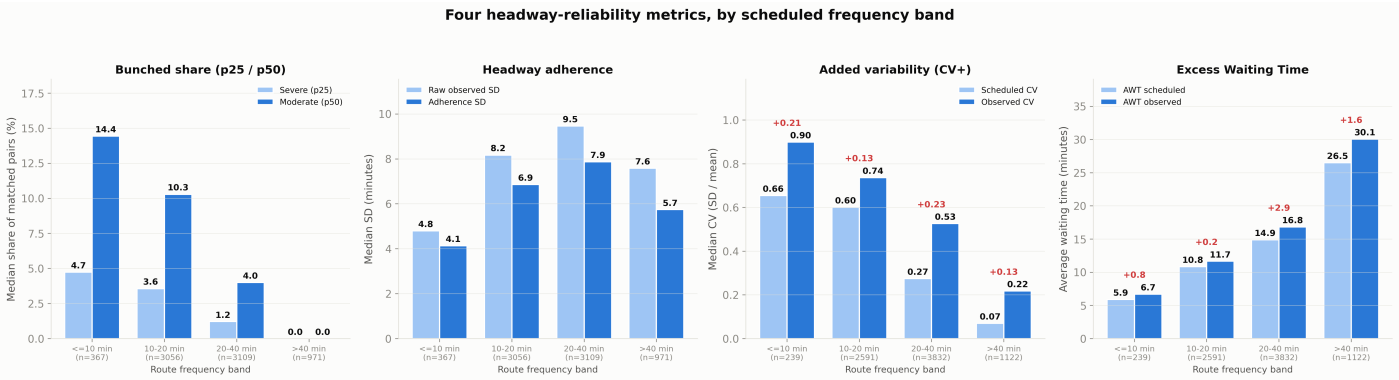


Fig. 1. Bunched share, headway adherence, added variability (CV+), and Excess Waiting Time, each by scheduled frequency band. Bars show the group-median of the two distributions named in the legend; the red annotation on the third and fourth panels is the median of each group's own gap (observed minus scheduled), not the bar-height difference – medians do not subtract cleanly. Group counts (n) differ between panels because each metric applies its own minimum-sample threshold independently.

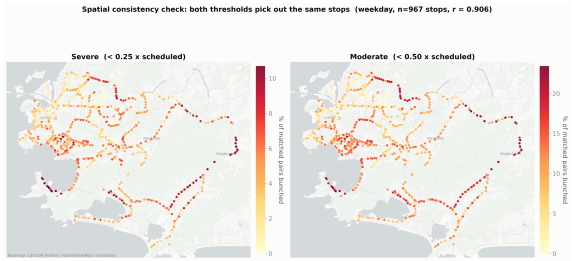


Fig. 2. Severe (p25) and moderate (p50) bunched share, per physical stop, weekday only. The two thresholds agree almost perfectly on which stops are worst ($r = 0.906$, $n = 967$ stops), and high-bunching stops cluster spatially rather than appearing at random.

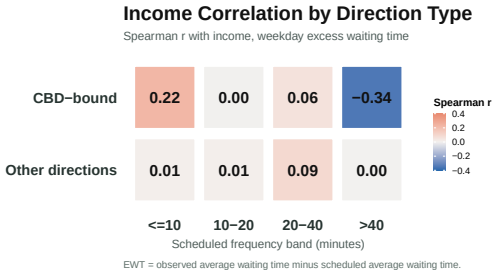


Fig. 3. Income-reliability correlations by service direction and scheduled frequency band. Each cell shows the Spearman correlation between local census income and weekday excess waiting time or reliability burden for route-stop groups, split between CBD-bound and other services.

CV+ has a network median of +0.175, positive in 77.6% of groups, and is likewise non-monotonic: largest on 20–40 min (+0.23) and ≤10 min routes (+0.21). Since CV is scale-free, this adds about a sixth of the mean headway to gap-to-gap variation – under two minutes on a ten-minute route, over five on a thirty-minute one – and most of it is generated in operation rather than inherited from the timetable.

EWT has a network median of 1.98 min across 1.83 million pairs. The excess is smallest at 10–20 min (+0.2), rises to +2.9 at 20–40 min, then falls to +1.6 at >40 min; 28.6% of groups sit at or below zero, chiefly low-frequency ones where few observations ran smoother than plan.

Taken together, where bunching is most visible is not where it costs most. Clustering concentrates on the most frequent routes, but there a shortened gap costs a passenger little: the ≤10 min band has the highest bunched share and among the lowest excess wait. The largest penalties fall instead on the 20–40 min band, which bunches least yet leads on both added variability and excess waiting time. Bunching alone would therefore direct attention to the wrong part of the network.

To test whether reliability disadvantages are concentrated among lower-income users travelling toward the city centre, we stratified route-stop observations by scheduled frequency and inferred service direction. Correlations between local income and reliability burden are generally weak (Fig. 3), suggesting no strong system-wide income gradient in excess waiting time. The main exception is the least frequent CBD-bound services, where the negative correlation indicates higher burden around lower-income stops, although this group contains fewer observations and should be interpreted cautiously. Overall, the results point to localised and frequency-specific reliability inequalities rather than a uniform network-wide pattern.

REFERENCES

1. G. F. Newell and R. B. Potts, "Maintaining a bus schedule," in *Proc. 2nd Australian Road Research Board Conf.*, (1964), pp. 388–393.
2. C. F. Daganzo, "A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons," Transp. Res. Part B: Methodol. **43**, 913 (2009).
3. E. E. Osuna and G. F. Newell, "Control strategies for an idealized public transportation system," Transp. Sci. **6**, 52 (1972).
4. A. Tirachini, J. Godachevich, O. Cats, *et al.*, "Headway variability in public transport: A review of metrics, determinants, effects for quality of service and control strategies," Transp. Rev. **42**, 337 (2022).
5. I. Psarros, K. Kepaptsoglou, and M. G. Karlaftis, "An empirical investigation of passenger wait time perceptions using hazard-based duration models," J. Public Transp. **14**, 109 (2011).
6. Y. Fan, A. Guthrie, and D. Levinson, "Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security," Transp. Res. Part A: Policy Pract. **88**, 251 (2016).
7. M. Trompet, X. Liu, and D. J. Graham, "Development of key performance indicator to compare regularity of service between urban bus operators," Transp. research record **2216**, 33 (2011).
8. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 dataset: A high-resolution multi-source view of public bus mobility in Niterói," arXiv preprint arXiv:2605.20263 (2026).
9. Z. Chen and F. Botta, "rt2gtfs: A scalable framework for correcting public transport timetables using real-time data for accessibility analysis," arXiv preprint arXiv:2603.11477 (2026).

Finite-size-corrected synchronization in a city-scale bus network: crossover, not demonstrated criticality

PO-HSUN LAI[1,*]

[1] Physics, National Taiwan University, Taipei, Taiwan
[*] b11202025@ntu.edu.tw

1. INTRODUCTION

Buses on a loop can be treated as self-oscillators with discrete, local, unidirectional coupling specific to the bus ahead, and that model produces critical coupling thresholds [1]. City-scale telemetry poses a harder identification problem. Fleets are small and change membership from hour to hour, motion is nonuniform, and real routes are open rather than closed. A rising circular order parameter is therefore a candidate measurement of bunching, not by itself evidence that the fleet obeys mean-field Kuramoto dynamics.

The NetMob26 record [4] covers 11–31 March 2026: 13,850,467 GPS fixes from 527 hardware IDs and 6,790,681 tickets from 556 operator vehicle numbers. The two identifier sets do not overlap and no crosswalk exists, so a ticket cannot be assigned to a specific GPS bus. We join the sources by normalized route and local hour alone. Twenty routes and 40 directions give 11,641 line-direction-date-hour cells, of which 10,135 are eligible. Boardings per active bus, $\lambda = B_{\ell t}/N_{\ell dt}$, divides route-hour demand by an endogenous supply; it is a realized load proxy, not exogenous demand and not a vehicle-level load.

2. FLEET SIZE AND OPEN GEOMETRY

For route position s and route length L the conventional phase is $\phi_{\text{arc}} = 2\pi s/L$, and for N active buses

$$r_1 = \left| \frac{1}{N} \sum_{j=1}^{N} e^{i\phi_j} \right|, \qquad r_{\text{exc}} = \frac{r_1 - \mu_N}{1 - \mu_N}, \tag{1}$$

where μ_N is the same-N null mean. Under independent uniform phases Nr_1 is the endpoint length of an N-step planar random walk, and Kluyver's exact distribution gives $E[r_1^2] = 1/N$, while $E[r_1] \simeq 0.8862/\sqrt{N}$ holds only asymptotically [3]. The median cell contains $N = 6$ buses and the lower quartile 4, so an uncorrected r_1 compared across cells largely reports fleet size. In Fig. 1 the observed medians track the same-N null closely enough that only the vertical offset from it carries information.

Arc phase also periodicizes an open coordinate. The two ends of a route are far apart on the ground but land at nearly the same phase, so concentration at the boundary lifts r_1 with no dynamical phase locking. Travel-time phase corrects unequal typical speed but leaves terminal occupancy in place.

3. LOAD RESPONSE AFTER CORRECTION

The simulated same-N null reproduces the exact distribution. We repeated the analysis with PPC $= (Nr_1^2 - 1)/(N - 1)$, a pairwise estimator with lower sample-size bias [2], under five normalizations, both phase definitions and four null models; seven executable phase-null specifications yield 35 estimator-specification contrasts.

Fleet size shows why the estimand must be stated. Raw arc-phase r_1 is $+0.045$ higher at weekends while mean N falls from 7.99 to 4.47, and exact matching within N leaves $+0.014$. Under

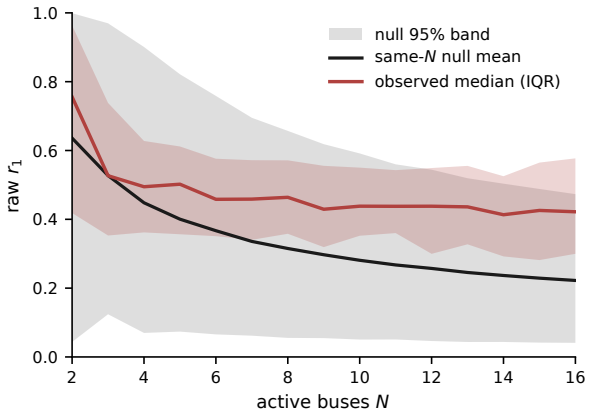


Fig. 1. Raw order has a fleet-size-dependent baseline. Observed medians must be read against the same-N null rather than compared directly across N.

the arc-uniform correction the marginal excess-order contrast is -0.068, against $+0.012$ when matched exactly on N. The marginal number carries the weekend supply pathway through N; the matched number asks about day type at fixed fleet size. Neither is a causal weekend effect.

Corrected order rises with the load proxy in a random-intercept model with line-direction effects and operational controls (standardized coefficient 0.04256, $p < 10^{-15}$). On the same fixed-effect frame the piecewise, linear and sigmoid AIC values are 10308.21, 10334.61 and 10396.36, and the piecewise slope change sits at $\lambda_c = 135.74$ boardings per active bus (Fig. 2).

The transition diagnostics do not support a stronger reading. Susceptibility has no interior peak, and the free-exponent collapse $r = N^{-b}F((\lambda - \lambda_c)N^a)$ is start sensitive and reaches optimizer boundaries for whole-route and interior fits alike, so we report its exponents as not identified. That failure does not rule out a transition here, but it blocks any reading of λ_c as a critical point.

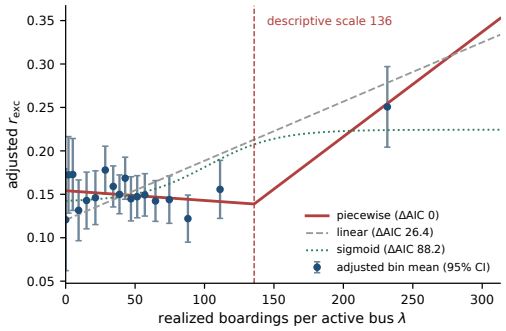


Fig. 2. Adjusted corrected order changes slope near 136 boardings per active bus. The fitted scale is descriptive and is not a demonstrated critical threshold.

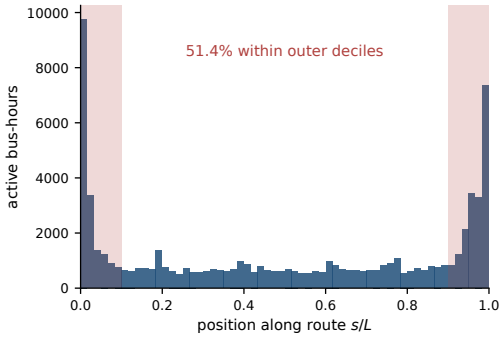


Fig. 3. After layover exclusion, 51.37% of representative active bus-hours remain in the outer route deciles.

4. WHERE THE BUSES ACTUALLY ARE

After layover exclusion, 27.19% of active bus-hours lie within 2% of a route end and a further 24.18% between 2% and 10%, so 51.37% of the representative sample sits in the outer deciles (Fig. 3). Median matching residuals are comparable across terminal, near-terminal and interior zones, which rules out projection error as the source of the pile-up.

Recomputing N and the order parameter separately inside each zone reverses the picture (Fig. 4): mean r_{exc} is 0.9974 at terminals and 0.9186 near them, against -0.1887 in the 10–90% interior and 0.1574 whole route. The terminal value records concentration in the circular coordinate. The negative interior value means spacing more even than the specified same-N null, consistent with a splay-like state but not proof that dispatch control succeeds.

The load association splits the same way (Fig. 5). Whole route, the standardized coefficient in a line-direction fixed-effect model is 0.03758 (95% CI $[0.01097, 0.06420]$, $p = 0.00565$); in the interior it is 0.00111 ($[-0.02460, 0.02683]$, $p = 0.9325$), and no zone shows a detected within-zone coefficient. Adding terminal fraction to the whole-route model cuts the coefficient to 0.02106, an attenuation of 44%. Spatial restriction changes both the estimand and the within-zone N, so this is substantial attenuation rather than complete mediation.

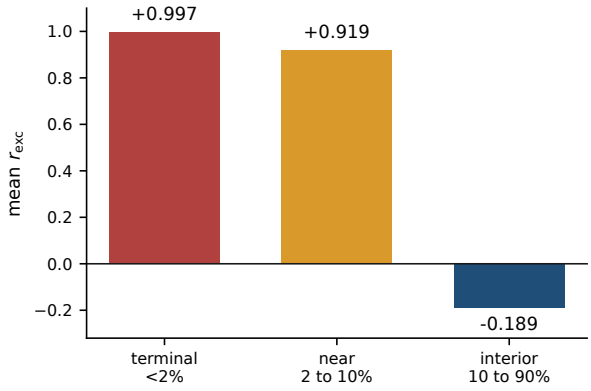


Fig. 4. Corrected order reverses sign between terminal and interior zones once fleet size and order are recomputed separately within each zone.

5. ROBUSTNESS AND LIMITS

GPS points are projected onto continuous route polylines whose members are globally ordered and oriented before arc position is computed. Each bus-hour is represented by the eligible fix nearest HH:30; we drop representations more than 20 minutes from the midpoint or with median route-matching residual above 150

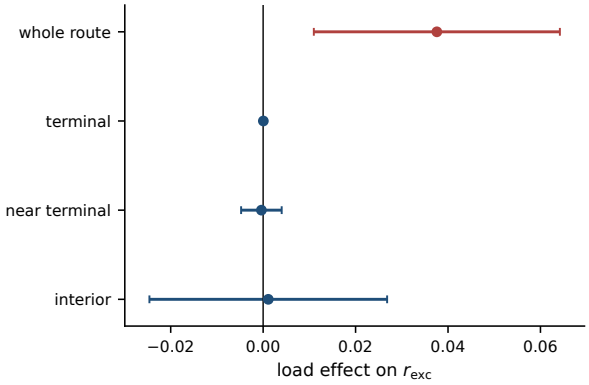


Fig. 5. The whole-route load association is detected, but the corresponding within-zone coefficients are not. Intervals are clustered 95% confidence intervals.

m, and remove terminal layovers by a rule on duration, boundary mass and motion span. These choices define cell membership and none of them filters on order.

The spatial result does not hinge on one cutoff. Trimming 2% from both ends moves mean arc-phase r_{exc} from 0.155 to -0.120, and 5% trimming gives -0.223. Heavier trimming deletes genuine route and drives the median within-cell fleet down to three, so each cutoff defines its own estimand.

Finite-N correction is necessary and not sufficient, since corrected order stays conditional on the phase coordinate, the occupancy null and the zone. The opposite signs of terminal and interior order, the absence of a detected within-zone load association, and the 44% attenuation under terminal-fraction control together make whole-route coherence largely a spatial mixture, though they do not show that terminals cause every residual association.

The remaining limits are structural. Tickets and GPS cannot be linked at vehicle level, so λ is a route-hour ratio with an endogenous denominator; midpoint positions compress within-hour trajectories; null models cannot reproduce every directed interaction or dispatch decision; and zone restriction changes occupancy and fleet size at once. The record covers 19 days in one city, so the magnitude of the terminal concentration needs replication elsewhere.

6. CONCLUSION

A conventional circular order parameter makes this fleet look Kuramoto-like, but the appearance depends on fleet size and on periodicizing an open coordinate. After same-N correction a positive load association and a descriptive crossover near 136 boardings per active bus survive, while susceptibility and collapse diagnostics identify no critical point and about half of whole-route coherence is spatial composition. The supported conclusion is crossover, not demonstrated criticality.

REFERENCES

1. V.-L. Saw, N. N. Chung, W. L. Quek, Y. E. I. Pang, and L. Y. Chew, "Bus bunching as a synchronisation phenomenon," Sci. Rep. **9**, 6887 (2019).
2. M. Vinck, M. van Wingerden, T. Womelsdorf, P. Fries, and C. M. A. Pennartz, "The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization," NeuroImage **51**, 112–122 (2010).
3. J. C. Kluyver, "A local probability problem," Proc. Sect. Sci. K. Akad. Wet. **8**, 341–350 (1905).
4. F. Domingos, H. T. Marques-Neto, B. Pereira, C. Celes, S. Knoblauch, and V. F. S. Mota, "The NetMob26 dataset: a high-resolution multi-source view of public bus mobility in Niterói," arXiv:2605.20263 (2026).

GraniteMob: An AI Framework for Mobility Intelligence in the City of Niterói, Rio de Janeiro, Brazil

KRISHNA MOHAN BATHULA[1,2], JAIDEN ROMAN[1,3], AMIR ABDULLOZODA[1,4], AND CHRISTELLE SCHARFF[1,5]

[1] Seidenberg School of Computer Science and Information Systems, Pace University, Department of Computer Science, New York, 10038, USA
[2] kbathula@pace.edu
[3] jr79854n@pace.edu
[4] aa60864n@pace.edu
[5] cscharff@pace.edu

1. INTRODUCTION

Recent advances in artificial intelligence (AI), machine learning (ML), and large language models (LLM) have created new opportunities for developing intelligent transportation systems capable of not only predicting mobility patterns but also generating explainable and actionable insights [1, 2]. While existing studies have focused on mobility prediction, traffic forecasting, and route optimization, relatively little attention has been given to the integration of multimodal urban datasets with foundation models to provide explainable decision support for transportation planning [3]. In particular, the combination of mobility data, ticketing data, and social infrastructure information remains an underexplored area within urban mobility research. In this work, we propose a lightweight AI framework that generates context-aware, grounded explanations and decision-support insights for urban transportation planning in the city of Niterói.

2. BACKGROUND

Earlier work in urban mobility research has demonstrated the value of integrating heterogeneous data sources, including GPS trajectories, smart-card transactions, GTFS data, mobile phone records, weather information, and urban infrastructure, to gain a deeper understanding of the urban transportation systems [4]. Prior studies have applied ML and deep learning (DL) techniques for passenger demand forecasting [5–7], travel time prediction, bus bunching detection [8], anomaly detection, traffic flow estimation, and mobility pattern analysis, while more recent work has explored foundation models, LLMs, and retrieval-augmented generation (RAG) for transportation intelligence and urban decision support [1–3]. The emergence of urban foundation models and small language models (SLM) provides new opportunities to combine predictive analytics with natural language reasoning, enabling more explainable and interactive mobility analysis [9]. Building upon these developments and leveraging the NetMob26 multimodal dataset [10], this work proposes a simplified framework that integrates mobility, ticketing, and social data, employs a time-series transformer for route-level passenger demand forecasting, and utilizes RAG with the IBM Granite SLM [11] to provide insight for urban mobility planning.

3. DATA INTEGRATION AND EXPLORATION

The Ticket, Mobility, and Social datasets are merged into a unified analytical representation. Spatial integration associates transportation routes with nearby social infrastructure, while temporal integration aligns events based on common temporal intervals, such as route by hour or route by day combinations. The integrated data is then passed to the data preparation Layer, which consists of data preprocessing and feature engineering.

Mobility and ticketing records were integrated using shared route identifiers and aggregated into hourly intervals. The exploratory data analysis focused on smart-card categories defined in the NetMob26 dataset, enabling passenger counts to be computed for each card type at every hourly interval. Figure 1 illustrates the ridership based on the smart-card type.

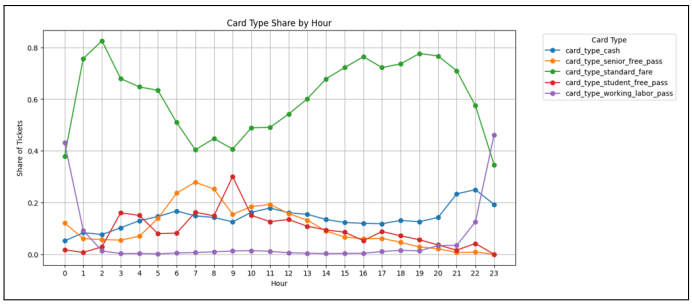


Fig. 1. Smart Card Ridership

To better identify the mobility patterns of smaller subsets of the transporting community, the card types were separated, and the student free-pass category was analyzed as a proof of concept. From the analysis, it was discovered that certain routes are more skewed towards student passes, meaning that these routes saw the most student activity compared to others, showcasing student dependence that simply observing total ridership alone would not reveal. This analysis demonstrates how multimodal integration can reveal route-specific travel behavior that is not observable from total ridership alone. We also take note that the time frame that students in particular spike, which are around 6:00 AM and 12:00 PM, accounting for the BRT (UTC-3) timezone, which helps to focus in on where demand is happening.

4. METHODOLOGY

The proposed methodology integrates multimodal urban data with ML models, and implements RAG, SLM to generate data-driven grounded explanations and planning insights. The core research goal is to combine spatio-temporal prediction with contextual mobility data for more accurate, interpretable intelligence rather than approaches based on isolated data sources or prediction models.

A. Feature Engineering

Feature engineering generated meaningful variables that characterize the transportation system, such as the number of active buses, trips, passenger counts, transfer rates, passenger-to-bus ratios, peak-hour indicators, scheduled travel time, service frequency, route length, number of stops, and social accessibility measures. The output of this stage is a high-quality feature dataset suitable for machine learning analysis.

B. System Architecture

Figure 2 represents the pipeline of the system architecture. The **ML Model layer** uses a time series Transformer for passenger-demand forecasting and a classification model for predicting service-frequency levels. The **Knowledge Base** integrates transportation data, model outputs, and contextual information for retrieval. The **RAG Component** retrieves relevant evidence from the knowledge base, enabling IBM Granite to generate grounded explanations and planning insights. Finally, the **Dashboard** presents forecasts, classifications, and AI-generated insights through interactive visualizations and natural language queries, helping planners identify mobility patterns, accessibility gaps, and transit disparities.

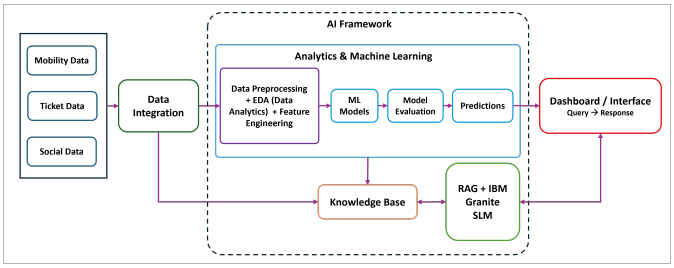


Fig. 2. System Architecture

5. RESULTS

The experiments demonstrate the feasibility of multimodal transit data for route-level mobility intelligence. As shown in Figure 3, the Time Series Transformer captured passenger-demand patterns with an MAE of **39.93 passengers**, while quantile forecasts represented prediction uncertainty. As shown in Table 1, for service-frequency classification, **XGBoost achieved the highest recall for the high-frequency class 0.891**, for service-frequency classification, outperforming Random Forest (0.850) and the neural network (0.857). Exploratory analysis further revealed route-specific passenger patterns, including disproportionately high student-pass activity on certain routes, demonstrating the value of multimodal integration. The forecasts and classifications are integrated into the knowledge base and retrieved through RAG-enabled IBM Granite for explainable mobility insights as shown in Figure 4, while multimodal integration reveals route-specific patterns beyond aggregated ridership.

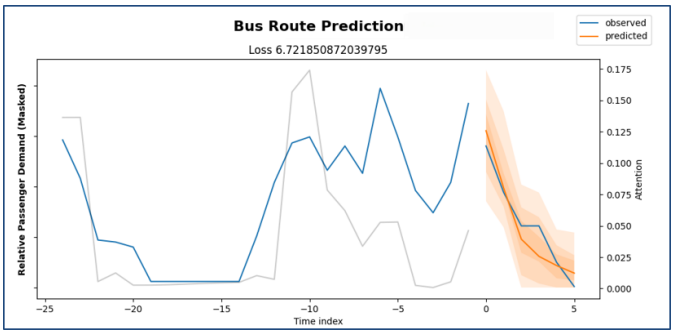


Fig. 3. Passenger Demand Forecasting

6. CONCLUSION AND FUTURE WORK

Preliminary results show that combining predictive outputs with route-level context enables interpretable mobility intelligence.

Table 1. Model Evaluation

MODEL	METRIC	VALUE
Transformer	MAE	**39.93**
XGBoost	Recall	**.891**
Random Forest	Recall	.850
Neural Networks	Recall	.857

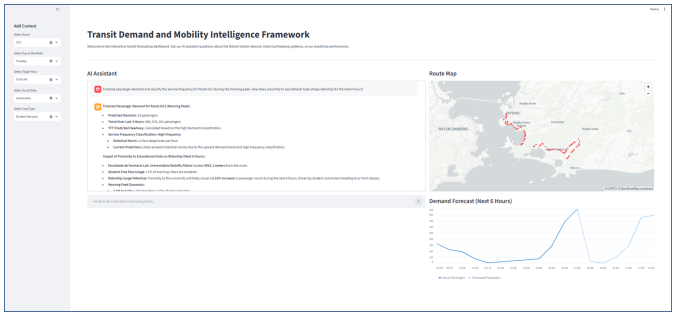


Fig. 4. GraniteMob Interface

Future work will incorporate weather and seasonal effects and enhance GTFS integration for richer transit-network analysis.

ACKNOWLEDGMENT

The authors gratefully acknowledge the CRA UR2PhD Program and Seidenberg Summer Undergraduate Research Initiative for their support and IBM for GPU access through Heidi AI.

REFERENCES

1. P. Bhandari, A. Anastasopoulos, and D. Pfoser, "Urban mobility assessment using llms," (2024).
2. W. Zhang, J. Han, Z. Xu, *et al.*, "Urban foundation models: A survey," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* (Association for Computing Machinery, New York, NY, USA, 2024), KDD '24, p. 6633–6643.
3. C. Zhao, X. Dai, Y. Lv, *et al.*, "Foundation models for transportation intelligence: Its convergence in transverse," IEEE Intell. Syst. **37**, 77 (2023).
4. R. Teeuwen and J. Gil, "Estimating traffic flows from vehicle trajectories based on sparse mobile phone geolocation data," NetMob Book Abstr. pp. 129–130 (2025).
5. Y. Ou, "Ai for real-time bus travel time prediction in traffic congestion management," in *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership,* (Springer, 2021), pp. 63–84.
6. R. Cousin and T.-M.-T. Nguyen, "Machine Learning-Based Netmob25 Mobility Model for Network Simulators," in *NetMob 2025 Data Challenge,* (INRIA, Paris, France, 2025).
7. S. Liyanage, R. Abduljabbar, H. Dia, and P.-W. Tsai, "Ai-based neural network models for bus passenger demand forecasting using smart card data," J. Urban Manag. **11**, 365 (2022).
8. M. Yan, B. Xie, and G. Xu, "Identifying spatial–temporal characteristics and significant factors of bus bunching based on an ega and dt model," Appl. Sci. **12**, 11778 (2022).
9. F. Wang, Z. Zhang, X. Zhang, *et al.*, "A comprehensive survey of small language models in the era of large language models: Techniques, enhancements, applications, collaboration with llms, and trustworthiness," ACM Trans. on Intell. Syst. Technol. **16**, 1 (2025).
10. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," (2026).
11. I. Granite Team, "Granite 3.0 language models," URL: https://github.com/ibm-granite/granite-3.0-language-models (2024).

Beyond Geographic Proximity: A Methodological Framework Informed by Mobility Data for Assessing Access to Specialized Hospital Services

DARCIO LEITE MARIO[1,*], RODRIGO PAULA CAMPOS[2,†], ALUIZIO NUNES DA SILVA JUNIOR[3,†] AND MARCO DE OLIVEIRA DA SILVA[3,†]

[1]Hospital Universitário Antônio Pedro, Empresa Brasileira de Serviços Hospitalares, Niterói, Rio de Janeiro, Brazil.

[2]Hospital Universitário Antônio Pedro, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil.

[3]Instituto de Computação, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil.

†These authors contributed equally to this work.

**darcio.mario@ebserh.gov.br*

1. PROBLEM AND MOTIVATION

Access to specialized hospital services depends on the spatial distribution of populations and facilities, but also on actual travel conditions. In metropolitan regions, a nearby hospital may still require long travel times, transfers, or journeys incompatible with service hours [1,2]. Because hospitals differ in the services they provide and in their care capacities, the nearest facility may not meet the health need under consideration [2].

Accessibility assessments commonly use residence, geographic distance, or estimated travel times as proxies for access. However, studies based on mobility data show that journeys to healthcare services may differ from the theoretically nearest or fastest alternatives [1], and that access to everyday activities is also organized around recurrent locations beyond the home [3]. This distinction is especially relevant when specialized services are concentrated in a limited number of facilities that serve populations from different municipalities [2].

We propose a methodological framework informed by mobility data to investigate how territory, transport, and hospital supply interact in the assessment of access to specialized services in metropolitan regions. Rather than treating hospitals as equivalent spatial points, the framework differentiates facilities by services and care capacities and incorporates the temporal and modal conditions of travel. It therefore enables the identification of barriers and inequalities that remain poorly captured by assessments based solely on proximity.

2. RESEARCH GAP AND QUESTIONS

Conventional accessibility approaches often represent populations by their place of residence and services as destinations reachable by the shortest distance or estimated travel time [1,3]. Although useful, these measures may fail to capture temporal variation in travel and constraints associated with different transport modes [1,2], as well as recurrent locations within everyday routines [3]. Consequently, spatial proximity and effective accessibility may lead to different interpretations of which populations are reached by services.

The gap lies in the limited integration of observed or estimated mobility, transport conditions, the territorial distribution of the population, and specialized hospital supply. The proposed framework addresses this gap by bringing these dimensions together within a single methodological design. The main research question is: how do actual mobility conditions influence access to specialized hospital services in metropolitan regions?

This question is divided into three subquestions:

Q1) To what extent does geographic proximity differ from effective accessibility to specialized hospital services?

Q2) Which areas face greater disadvantage when travel time, waiting, transfers, and temporal variation in the transport network are considered?

Q3) How do differences in hospital services and care capacities alter catchment areas defined solely by proximity?

3. METHODOLOGICAL FRAMEWORK

The framework integrates four dimensions: population and territory, mobility and transport, specialized hospital supply, and accessibility and inequality indicators. The first characterizes origin areas using demographic and socioeconomic data, including vulnerability and dependence on public transport. The second represents travel to hospitals through walking, waiting, in-vehicle travel, transfers, transport mode, and temporal variation [1,2]. When available, aggregated mobility data can complement formal transport networks and timetables, bringing the analysis closer to observed travel patterns [1,3].

The third dimension characterizes hospital supply by service, care capacity, and operating hours. This extends a D4D Challenge Senegal study presented at NetMob 2015, which combined mobility data, population distribution, and hospital locations but assumed equivalent hospital capacities and acknowledged the need for finer temporal granularity and

differentiation of treatment capacities [4]. Our framework therefore avoids treating physical proximity as equivalent to availability of the required care [2,4].

The fourth dimension operationalizes this integration: for each origin area and specialized service, the framework identifies hospitals able to provide the required care and evaluates access to them under time-, mode-, and network-specific travel conditions. Indicators include total access time, transfers, population reachable within time thresholds, the difference between the nearest and most accessible hospital, and territorial inequality. This supports time-, mode-, and service-specific catchment areas and comparisons among geographic proximity, transport-network accessibility, and mobility-informed accessibility.

Figure 1 summarizes the four dimensions and their interactions.

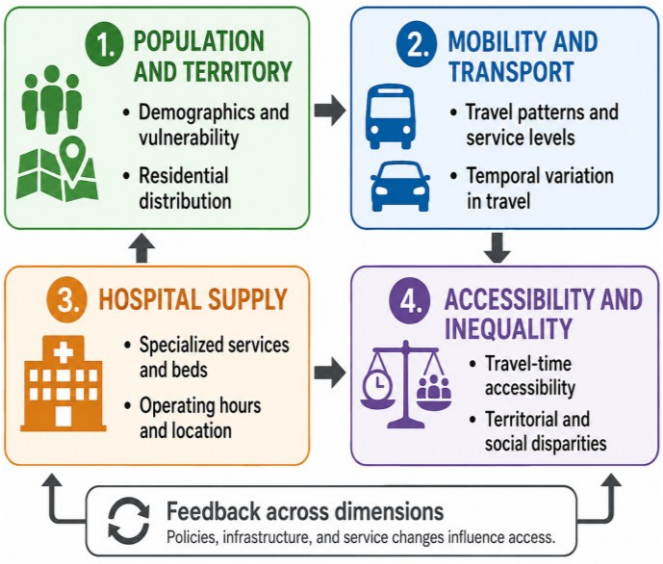


Fig. 1. Methodological framework.

4. OPERATIONALIZATION AND EVALUATION

The framework is designed to analyze Brazilian metropolitan regions by integrating demographic data, transport networks, aggregated mobility, and specialized hospital supply. GTFS feeds, stop data, route geometries, and census indicators available in the official NetMob 2026 repository provide concrete support for the initial operationalization of the territorial and transport dimensions [5]. The evaluation plan compares two baselines—geographic proximity and transport-network accessibility—with a mobility-informed scenario based, when available, on aggregated patterns of observed mobility.

The framework enables the examination of how timetables, transport modes, and hospital capacities affect catchment areas and the identification of the most disadvantaged territories. The evaluation considers indicator consistency, sensitivity to different time periods, modes, and services, and the stability of the resulting patterns. At later stages, validation may use aggregated administrative data on patients' areas of origin, origin–destination surveys, or mobility records, while preserving privacy. This strategy distinguishes potential accessibility, observed mobility, and actual service use [1].

5. CONTRIBUTIONS AND LIMITATIONS

The proposal makes three contributions: (i) conceptual, by distinguishing geographic proximity from accessibility to the specialized healthcare service required by the health need under consideration; (ii) methodological, by integrating territory, mobility, transport, and specialized hospital supply; and (iii) application-oriented, by enabling the delineation of catchment areas specific to time period, travel mode, and specialized service. The framework provides a basis for identifying inequalities that are poorly captured by conventional measures and may support healthcare and transport planning.

The proposal also has limitations. Estimated accessibility represents the possibility of access, not necessarily actual service use [1]. Mobility data may be affected by coverage biases and may underrepresent vulnerable groups [1], while transport timetables and hospital records may not fully reflect operational conditions [1,2,4]. Access regulation, service availability, installed capacity, and users' choices also influence care delivery. Empirical application therefore requires the integration of multiple data sources, validation of the indicators, and procedures consistent with privacy and data protection.

Funding. This work received no specific funding.
Acknowledgment. The authors thank Hospital Universitário Antônio Pedro (HUAP), Universidade Federal Fluminense (UFF), and Empresa Brasileira de Serviços Hospitalares (Ebserh/HU Brasil) for the institutional support for research, development, and innovation that contributed to the preparation of this work.

Disclosures. The authors declare no conflicts of interest.

Data Availability. No datasets were generated or analyzed in this study. The methodological framework is based on the literature and public sources cited herein.

References

1. K. Gligorić, C. Kamath, D. J. Weiss, et al., Commun. Med. 3, 157 (2023), doi: 10.1038/s43856-023-00384-9.
2. P. B. M. Garcia and A. A. Raia Jr., urbe Rev. Bras. Gest. Urbana 7, 21 (2015), doi: 10.1590/2175-3369.007.001.AO02.
3. S. Kulkarni and E. Moro, "Beyond proximity: rethinking urban accessibility through human mobility patterns in United States," in NetMob 2025 Book of Abstracts (2025), pp. 67–68.
4. E. Mutafungwa, F. Thiessard, M. P. Diallo, et al., "Mobile data as public health decision enabler: a case study of cardiac and neurological emergencies," in Book of Abstracts: Scientific Papers—Data for Development Challenge Senegal (2015), pp. 97–102.
5. LPRM-UFES, "NetMob 2026: Public Transport & Passenger Demand in Niterói, RJ, Brazil," GitHub (2026), https://github.com/lprm-ufes/Netmob2026.

Mobile Data for Public Good: Sustainable Pathways in a Changing Mobile Ecosystem

Anna Colquhoun,1*

1 GSMA Mobile for Humanitarian Innovation, London, United Kingdom, * *acolquhoun@gsma.com*

1. INTRODUCTION

Since NetMob began in 2010, mobile network data has moved from a novel research resource to an established input for decision-making across public-interest contexts. Its value has been demonstrated in research, official statistics, public health, disaster response and humanitarian action [1,2]. Over the same period, the mobile ecosystem underpinning these applications has expanded: mobile networks now cover 96% of the global population and connect 5.8 billion unique subscribers, with network-generated data more embedded in digital systems [3].

The tenth edition of NetMob offers an opportunity to take stock of these developments. The field has made significant advances in methodological development and real-world applications of mobile data for good initiatives. Meanwhile, questions around data access, governance and responsibility have become increasingly important. How mobile data is made available is directly shaped by the sector that generates and governs it. This paper therefore examines how changes within the mobile industry are influencing the conditions for responsible and sustained public-interest data use.

Drawing on published evidence, documented use cases and stakeholder consultations, four shifts are examined: the growing strategic and regulatory significance of mobile data; expanded analytical capabilities within mobile operators; more controlled approaches to data access and analysis; and the institutionalisation of mobile data within public-interest systems. Its contribution is to show how these changes are reshaping which mobile data initiatives become viable and endure.

2. EVIDENCE BASE AND APPROACH

The paper draws on GSMA's 2026 analysis of mobile data for humanitarian action, bringing together evidence from more than 100 publications, over 70 documented use cases and consultations with mobile network operators (MNOs), researchers, members of the NetMob community, humanitarian organisations, public institutions and intermediaries [4]. The review focused primarily on low- and middle-income countries (LMICs) and examined mobile data applications, responsible use, MNO perspectives and partnership models.

Humanitarian action is used as a lens for examining broader questions about public-good access to mobile data. The sector brings several recurring issues into focus: the demand for timely data-driven insights, strong requirements around privacy and data responsibility, uneven institutional capacity and interest in access to regulated, private sector datasets. Case studies from Chile, Ghana and the Philippines were selected to capture different regions, partnership models and operational settings, as well as factors contributing to their longevity [4].

3. FOUR SHIFTS SHAPING RESPONSIBLE AND SUSTAINED MOBILE DATA USE

3.1 Mobile data as a strategic asset

First, mobile network-generated data now carries greater commercial significance and is managed through more formal governance and risk structures. Data that was once treated primarily as an operational by-product can reveal near real-time customer behaviour, network performance and commercially sensitive market information.

This has changed how mobile data is treated within operators. What may have previously been considered a technical data-sharing decision is now reviewed through legal, privacy, security, commercial and regulatory lenses. Public-interest initiatives must therefore align with operators' legal obligations, internal approval processes and responsibilities to customers. Establishing these agreements and governance arrangements can take 12 to 18 months, reflecting the breadth of considerations for an MNO engaging in such partnerships [4].

3.2 Expansion of operator analytics capabilities

Analytical capabilities within the mobile industry have expanded, although unevenly across markets. Earlier collaborations often assumed a relatively clear division between operators holding data and external actors conducting analysis. MNOs can now support a wider range of approaches, including provider-side analysis, operator-hosted processing and managed analytical services. This creates opportunities to generate insights without transferring sensitive data externally. However, this varies considerably by geography, with ongoing capacity strengthening needed to enable national ownership in many low- and middle-income countries [4].

3.3 More established access and sharing models

Access models have advanced. While several early projects relied on bespoke data-sharing arrangements, the review identified five broad models in use: direct data release, controlled remote access, query-only access, provider-side analysis and ad hoc emergency access [4]. The direction of change is towards controlled approaches that keep sensitive data within MNO environments while sharing aggregated outputs. These approaches reduce MNO data exposure while enabling repeated analytical use.

At the same time, regulators and data protection authorities are playing a more influential role in determining what forms of mobile data access are appropriate. Across markets, approaches differ depending on national laws, policy priorities and institutional arrangements. GDPR and the ePrivacy Directive have influenced operator practice well beyond Europe [5], while newer instruments such as the EU Data Act and EU AI Act point towards a more formalised landscape for private-sector data access and processing [6,7]. For mobile data initiatives, this places greater emphasis on whether access models can withstand regulatory scrutiny and operate at scale.

3.4 Institutionalisation of public-interest data ecosystems

Mobile data initiatives are moving away from short-term or one-off pilots, towards established systems for national statistics, public policy and disaster response. In Chile, Telefónica and Universidad del Desarrollo maintained a research partnership for a decade, supporting mobile data analysis across public health, disaster response and urban research [8,9,10]. In Ghana, a partnership established in 2018 between Telecel, Flowminder and the Ghana Statistical Service produces recurring indicators of population distribution and mobility within the national statistical system [11,12]. In the Philippines, Globe and UN DISHA's always-on platform provides mobility insights that are

refreshed daily and available for immediate use during crises [13].

These examples show that sustainability is strengthened when mobile data capabilities are embedded within institutions that can apply them repeatedly to evolving priorities and decision-making needs. For operators, this can also create a clearer basis for longer term engagement, with the same underlying infrastructure supporting multiple public-interest applications.

Table 1. Mobile industry shifts influencing public-good access

Shift	Observed change	Implications
Data as a more regulated asset	Network data has more defined commercial and regulatory significance.	Public-good use must fit operator risk and authorisation models.
Analytics capability	Operators can increasingly process data within their own environments.	Partnerships may involve shared analytical responsibilities.
Access models	Controlled systems mediate access to outputs.	Research design must address transparency and scrutiny.
Institutionalisation	Public-good use is moving into official and multi-country systems.	Sustained impact depends on governance and local capacity.

4. EMERGING SIGNALS

Alongside these observed shifts, several developments are likely to shape the next phase of the ecosystem. One is the ongoing evolution of network-derived data. While the NetMob community has long worked with datasets beyond Call Detail Records (CDRs), changing technologies, communication behaviours and digital services continue to change what mobile networks capture and how those signals can be interpreted – particularly in LMICs. The growing use of over-the-top services, alongside richer network and signalling data, is likely to influence both the design of indicators and the assumptions that underpin established analytical approaches [4].

A second signal is the growing integration of data and analytics within the mobile industry itself. Mobile operators and hardware providers are investing in analytical platforms, location intelligence products and AI-enabled services, embedding data more deeply within commercial strategies. This may expand the infrastructure available for public-interest applications, but it could also strengthen expectations that such applications align with commercial priorities and more clearly defined value propositions [4].

AI is expected to play a growing role in this landscape, although its wider contribution is still taking shape [14]. The DISHA platform in the Philippines illustrates an emerging role for AI, combining mobile usage patterns with census data and other anonymised signals to produce poverty estimates alongside population movement insights.

5. IMPLICATIONS FOR MOBILE DATA FOR GOOD INITIATIVES

The four shifts identified in this paper are changing the basis on which mobile data initiatives become viable. Earlier collaborations often focused on whether researchers or public institutions could access mobile data for analysis. The more relevant question now is whether an initiative can generate useful insight within the commercial, regulatory and institutional conditions surrounding the data. What is possible will vary by context, the use case being supported, the risks attached to the data and the capacity available within the institutions involved.

There are now more proven pathways for sustained use than a decade ago. Longevity of these models will depend on stronger governance, clearer use cases and compatibility with operator risk frameworks. This should not be read as a retreat from public-good use. It reflects a field that has become more formalised and structured [4].

These conditions also have methodological implications. As the underlying data and processing environments evolve, established assumptions will need periodic reassessment. Where analysis takes place within controlled or automated systems, researchers will require sufficient transparency to validate outputs and understand their limitations. Questions of data and algorithmic sovereignty will also become more important, with growing expectations that analytical capabilities should be more deeply rooted in national ownership. [14]

For the mobility data community, these shifts create an important agenda. Methodological innovation remains central, but the next phase of public-good use will also depend on the systems through which mobile data is accessed and applied. The field has an opportunity to shape those systems so that mobile-derived insights can continue to generate public value within the industry and governance realities that define their use.

BACK MATTER

Funding. This paper is based on research funded by UK International Development from the UK government and supported by the GSMA and its members.

Acknowledgements. The underlying review was informed by perspectives from mobile operators, researchers, humanitarian organisations, public institutions and intermediaries.

Disclosures. The authors declare no conflicts of interest.

Data availability. No new datasets were generated for this abstract. The analysis draws on a published evidence review.

REFERENCES

1. L. Bengtsson, X. Lu, A. Thorson, et al., “Improved response to disasters and outbreaks by tracking population movements with mobile phone network data,” PLoS Medicine 8(8), e1001083 (2011).
2. N. Oliver, B. Lepri, H. Sterly, et al., “Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle,” Science Advances 6(23), eabc0764 (2020).
3. GSMA, The Mobile Economy 2025 (2025).
4. A. Colquhoun, A. Lozano, T. Etulain, et al., Mobile Data for Humanitarian Action: Pathways to Sustained and Responsible Data Use, GSMA (2026).
5. R. Murphy, “Mapping the Brussels Effect: The GDPR Goes Global,” Centre for European Policy Analysis (2025).
6. European Commission, “Data Act explained” (2025).
7. European Parliament, “EU AI Act: first regulation on artificial intelligence” (2025).
8. Pappalardo, L. et al. (2023). “A dataset to assess mobility changes in Chile following local quarantines”. Scientific Data, 10.
9. De Alarcon, P. et al. (2021). “The contribution of telco data to fight the COVID-19 pandemic: Experience of Telefonica throughout its footprint”. Data &
Policy.
10. Elejalde, E. et al. (2025). Behavioral response to mobile phone evacuation alerts.
11. Flowminder, Urban-Rural Population Mobility Patterns in Ghana: 2022–2024 (2025).
12. Flowminder. (2 March 2026). Conflict-derived displacements in Ghana: Focus on Sawla-Tuna-Kalba and Bawku Municipal districts (October 2025).
13. Globe. “Globe is First Telco to Join UN-Led DISHA Initiative for Disaster Response” (2025).
14. GSMA Mobile World Live, AI Survey Report 2026 (2026).

Using UAVs To Improve The Logistics Of Medical Supply Transport In Urban Areas

PEDRO HENRIQUE CARDOSO MAIA[1,*], GABRIEL FERREIRA PEREIRA[1,*], FELIPE DE CASTRO BELÉM[1,**], AND HUMBERTO T. MARQUES-NETO[1,**]

[1] *Department of Computer Science – PUC Minas – Belo Horizonte, MG, Brazil.*
[*] *{pedro.maia1337,gabrielferreirape502}@gmail.com*
[**] *{felipebelem,humberto}@pucminas.br*

1. INTRODUCTION

The Brazilian Popular Pharmacy Program (*Farmácia Popular do Brasil*) is the largest initiative in the country for drug access, currently covering 86% of the national territory [1]. Despite this extensive reach, the program faces severe logistical challenges in the Belo Horizonte Metropolitan Region (RMBH, for its Portuguese acronym). The delivery of services in the region is frequently threatened by climatic and geographic factors. The severity of these events is amplified by the rugged terrain of the RMBH, characterized by high elevations that favor landslides capable of blocking highways and disrupting conventional ground transport [2]. This isolation directly compromises the continuity of chronic disease treatments – such as diabetes and hypertension – which can increase hospitalization rates and costs for Brazil's Unified Health System (SUS).

To mitigate this scenario, Unmanned Aerial Vehicles (UAVs) have emerged as a promising solution. However, the region's own geography imposes complex challenges related to signal stability and battery consumption, turning route optimization into a critical vehicle routing and logistics engineering problem. In the literature, the effectiveness of aerial logistics in extreme conditions has been demonstrated by Chen et al. [3], who proposed prioritizing deliveries in flood-affected regions based on patients' clinical vulnerability. From an algorithmic standpoint, Kitjacharoenchai and Lee [4] modeled the Vehicle Routing Problem with Drones (VRPD) through exact formulations, providing the mathematical foundation for capacity and flight-autonomy constraints. To address the region's unique logistical challenges, this study innovates by hybridizing a constructive heuristic with a Genetic Algorithm, proposing a multi-objective formulation that optimizes multi-point delivery routes by weighing time, energy, distance, and critical priority costs.

To address this gap, this paper proposes a route optimization model for UAVs tailored to the region under study. The methodology is first grounded in the collection of geospatial data and the mapping of partner pharmacies' distribution. A predictive analysis of these data is then performed to identify supply bottlenecks and map areas of higher vulnerability to isolation. This information feeds a routing algorithm guided by a multi-objective function that integrates variables such as energy consumption at different altitudes, payload capacity, and prioritization during climatic emergencies – thereby overcoming the limitations of conventional algorithms that disregard topographic and environmental variables. The proposed model aims to reduce response time and dependence on conventional transport. In doing so, this study contributes to protecting the constitutional right to health and reducing avoidable hospitalization costs within the public healthcare system.

2. METHODOLOGY

This research adopts an applied methodological approach with an exploratory and experimental character, structured into three stages: definition of the experimental scenario, construction of the initial solution, and route optimization.

The logistics problem is formulated as an extension of the classical Vehicle Routing Problem (VRP). To accommodate the specific demands of medical supply delivery, the model is classified in the literature as VRP-D-TW-C, encompassing drone routing (D) subject to operational capacity constraints (C) and strict time-window compliance (TW).

In the first stage, a geospatial survey was conducted of Popular Pharmacy Program units located within the Belo Horizonte Metropolitan Region (RMBH). The dataset was obtained from the public registry of establishments accredited to the Popular Pharmacy Program, made available by the Brazilian Federal Government. From the registered addresses, the geographic coordinates (latitude and longitude) of each unit were extracted, enabling the construction of a geospatial database representing pharmacy distribution across the study region. The collected data were processed and standardized to ensure consistency across the analyses, for more information [5].

Thus, an experimental scenario was designed comprising 20 orders distributed across different municipalities within the RMBH. Order selection aimed to represent realistic supply conditions, encompassing locations with varying geographic and logistic characteristics. These data constitute the input to the routing system and provide the information required for distance calculation, route generation, and algorithm performance evaluation. Figure 1 presents the spatial distribution of the orders used in the experiments.

In the second stage, a feasible initial solution is generated using the Clarke-Wright algorithm [6]. This constructive heuristic uses the distance matrix to identify savings between pairs of orders and cluster geographically proximate points into more efficient routes. The output of this stage constitutes the first feasible solution to the distribution problem.

In the third stage, the solution produced by the Clarke-Wright algorithm is used as the initial population of a Genetic Algorithm. This method applies selection, crossover, and mutation operators to explore new route combinations, seeking to minimize both distance traveled and delivery time while satisfying the operational constraints defined for the system. Additional details about the mathematical formulas and system settings can be found in [7].

Finally, the resulting routes are compared against equivalent routes performed via conventional urban transport. The analysis considers metrics such as distance traveled, estimated delivery time, and logistical efficiency, enabling an assessment of the proposed optimization's impact on medication distribution in hard-to-reach regions.

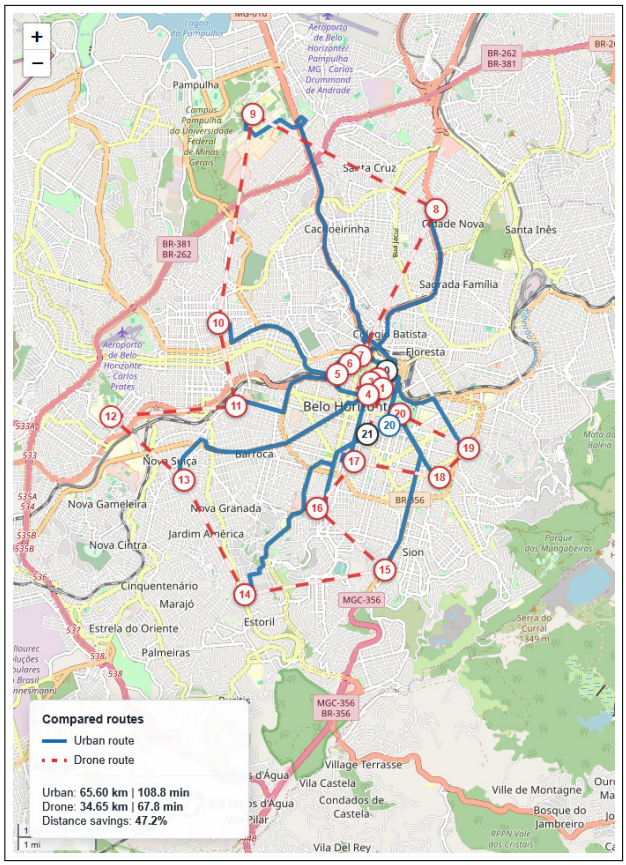


Fig. 1. Comparison map of the best routes between drones and urban transport.

3. RESULTS & DISCUSSION

The problem is modeled as a complete directed graph $G = (V, A)$, where $V = \{0\} \cup \{V_1, \ldots, V_n\}$ represents the central depot, located at coordinates $(-19.924057, -43.935237)$ and the n delivery points. Service time at each waypoint accounts for 30 s to cover landing and take-off maneuvers. Each aircraft has a maximum payload capacity of 4.0 kg. The theoretical maximum range of 10.0 km is derived from a baseline consumption of 15.0 Wh/km, with consumption dynamically penalized and operations constrained by wind speed. Additionally, the aircraft operates at a constant speed of 36.0 km/h.

To enable an efficiency comparison, the model also incorporates ground transport. The conventional vehicle is parameterized with a substantial payload capacity of 10.0 kg, a range of 400.0 km and a baseline fuel consumption of 10.0 km/l. Nominal operating speed is set at 60.0 km/h, while urban congestion reduces operational speed to 40.0 km/h, with actual road-network distances (d'_{ij}) used to compute travel distances.

The simulations revealed the substantial operational superiority of the aerial mode. By leveraging direct flight paths, drones shortened travel distances by 47.2% and reduced total mission time by 65.8%, confirming that the observed gains in temporal efficiency stem from spatial route optimization. Beyond the advantages in time and distance, drone-based delivery yielded a financial saving of 86.6% (reducing costs from R$ 40.67 to R$ 5.44) and produced zero direct carbon emissions, in sharp contrast to the 15.15 kg of CO_2 emitted by ground transport.

These results carry important logistical and regulatory implications. The feasibility of drone operations depends on clear airspace regulations, decentralized charging networks, and fiscal incentives tied to emissions reduction. Conversely, ground transport remains superior under strict operational constraints, representing the necessary choice for no-fly zones (such as areas surrounding airports), severe weather conditions, or situations in which demand exceeds drone payload capacity.

4. CONCLUSION

This work presented a route-optimization approach for drones applied to the distribution of medication in the Belo Horizonte Metropolitan Region. The proposed approach combined distance computation via the Haversine formula, initial solution generation through the Clarke-Wright algorithm, and route refinement using a Genetic Algorithm.

Drones demonstrated significant time and distance reductions compared to urban transport, which proved effective for deliveries in vulnerable or hard-to-reach regions. Consequently, the proposed solution shows potential to improve the efficiency of medication distribution and to broaden access to essential health services.

5. FUTURE WORKS

As future work, we intend to scale up the experiments by incorporating a larger number of orders, multiple depots, and different drone models. We also plan to integrate real-time meteorological data and a telemetry-based monitoring module, enabling dynamic route replanning during mission execution.

6. FINAL REMARKS

Acknowledgment. The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas, its Computer Science Student Council, FAPEMIG, CNPq, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - INFORMÁTICA and Finance Code 001).

Disclosures. The authors declare no conflicts of interest.

Data Availability Statement. The dataset used in this study was obtained from the Brazilian Ministry of Health's official database. [8].

Supplemental document. Results and the code presented in this paper are available in our GitHub repository [5].

REFERENCES

1. Brazil, Ministry of Health, "Programa Farmácia Popular amplia cobertura para 93% do território nacional e alcança mais de 5.000 municípios," (2025). (In Portuguese - PTBR). Accessed: May 01, 2026.
2. R. Dalagnol, C. B. Gramcianinov, N. M. Crespo, *et al.*, "Extreme rainfall and its impacts in the brazilian minas gerais state in january 2020: Can we blame climate change?" Clim. Resil. Sustain. **1**, e15 (2022).
3. Y.-H. Chen, A. M. El-Adle, K. J. O'Brien, *et al.*, "Drone-based medication delivery for rural, flood-prone coastal communities," PLOS ONE **20**, e0333696 (2025).
4. P. Kitjacharoenchai and S. Lee, "Vehicle routing problem with drones for last mile delivery," ScienceDirect (2023).
5. "Dronepharm," GitHub (2026). Retrieved on July 9, 2026], https://github.com/HubSeAILab/DronePharm.git.
6. G. Clarke and J. W. Wright, "Scheduling of vehicles from a central depot to a number of delivery points," Oper. Res. **12**, 568 (1964).
7. "Dronepharm," GitHub (2026). Retrieved on July 9, 2026], https://github.com/HubSeAILab/DronePharm/tree/main/docs.
8. Brazil, Ministry of Health, "Farmácia Popular program (PFPB) pharmacy address locator," (2025). (In Portuguese - PTBR). Accessed: May 01, 2026], https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_PFPB_ENDERECOS/index.html.

Scalable multilayer urban social networks using Call Detail Records and census data

PAULO GABRIEL NUNES CANÇADO[1], CAROLINA RIBEIRO XAVIER[1], ALEXANDRE GONÇALVES EVSUKOFF[2], AND VINÍCIUS DA FONSECA VIEIRA[1,*]

[1] *Departamento de Ciência da Computação, Universidade Federal de São João del-Rei (UFSJ), São João del-Rei, MG, Brazil*
[2] *COPPE – Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil*
[*] *vinicius@ufsj.edu.br*

1. INTRODUCTION

As cities grow and become more complex, understanding how individuals distribute themselves across space, interact with one another and with urban infrastructure becomes a fundamental challenge for urban planning [1, 2].

Call Detail Records (CDR), stored by telecommunication companies, can introduce biases stemming from the representativeness of the city's population, social strata, geographical regions, data extraction and storage methods, and the duration of the database capture, among other factors [1]. Incorporating data from diverse sources, especially those that consider the demographic aspects of a population, can significantly enhance the quality of urban social models, thereby improving their applicability. In this context, census information can be used to characterize the demographic composition of the population and its interaction across different perspectives, enabling the construction of models that reflect social relations and urban structures with greater realism [3]. Furthermore, a single element can assume different roles in each of its relationships [4, 5]. For instance, relationships between people in a social network occur within different groups, which can be understood through multiple layers or levels and considering only a single layer can lead to interpretations that overlook crucial details of the system [4].

Given this context, the primary objective of this work is to propose a methodology for constructing multilayer urban social networks by integrating *Call Detail Records* (CDR) and census data, aiming to represent relationships between people that emerge from their social interactions and the mobility patterns they determine in urban environments. Additionally, we intend to adopt a data sampling strategy to ensure the resulting network-based model is scalable, allowing for the representation of different dimensions of population organization—including social, demographic, and spatial aspects—in more realistic scenarios. To guide this study, two Research Questions (RQs) are proposed: *RQ1: How can data from different sources be integrated to construct a social network-based model that characterizes a population's mobility and its interaction across different dimensions?*; *RQ2: How can this model be scaled, enabling its application in useful practical scenarios, even in medium and large-sized cities?*

2. MATERIALS AND METHODS

The proposed methodological steps aim to reduce biases in social networks constructed from the integration of census data to define different perspectives of interaction among individuals. These perspectives are represented by different layers in the network, built based on the demographic distribution of the population across different dimensions [3]. We build upon the formalism introduced by Battiston *et al.* [6] and define, for each city, a multilayer network, which can be represented by a graph $G^{[\alpha]}$, composed of a set of vertices $v_i \in V$, such that $|V| = n$, representing the individuals of a given city, and l layers. Each layer α ($\alpha = [1, ..., l]$) represents the relationship between individuals v_i under a perspective α (e.g., communication, mobility, family) and can thus be represented by $G^{[\alpha]} = (V, E^{[\alpha]})$. $E^{[\alpha]}$ is the set of edges (v_i, v_j) that represents the relationships between individuals v_i and v_j under a perspective α.

We consider CDRs generated over a 30-day period between 03/15/2013 and 04/14/2013 and, here, we illustrate the application of the methodology in three municipalities in Brazil: Uberlândia, Betim, and Belo Horizonte. The following layers are defined:

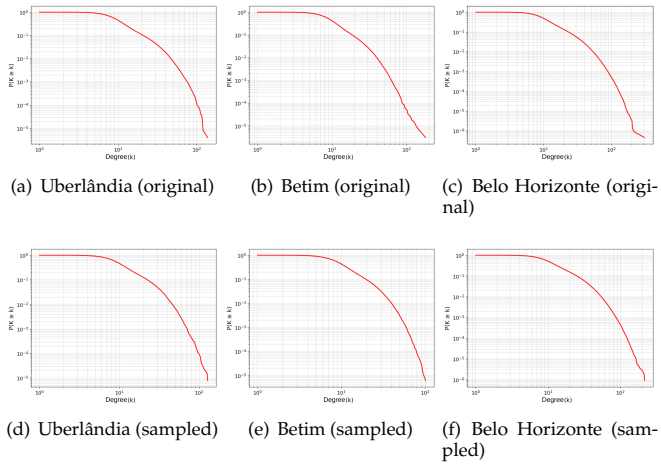
- **Comunication relationships layer ($G^{[1]} = (V, E^{[1]})$):** A relationship between i and j are represented as an edge $(v_i, v_j) \in E^{[1]}$ when phone calls are observed between them.
- **Mobility layer ($G^{[2]} = (V, E^{[2]})$):** This layer is based on the estimated mobility of individuals with a presumed residence derived from communication activity, combined with an Origin-Destination (OD) matrix constructed from the aggregated movement of individuals between the regions. Edges $(v_i, v_j) \in E^{[2]}$ are created between a proportion of individuals v_i and v_j residing in each region of the municipality, controlled by a parameter $\beta^{[2]}$.
- **Familiar layer ($G^{[3]} = (V, E^{[3]})$):** Based on the age pyramid observed in census data, the proportion of adults and children in the population is identified. Edges $(v_i, v_j) \in E^{[3]}$ are created between all members i of the family, generating small cliques that represent the families.
- **Work/school layer ($G^{[4]} = (V, E^{[4]})$):** Considering the previously defined categorization into Adults" and "Children", individuals are aggregated into work and school institutions of size $\beta^{[4]}$ for 79% of adults and 100% of children, respectively, accounting for census information.
- **Religious layer ($G^{[5]} = (V, E^{[5]})$):** Models interactions arising from religious practices, organizing individuals into small groups of size $\beta^{[5]}$ and connecting each pair of vertices v_i and v_j by a parameter $\mu^{[5]}$ for the edge $(v_i, v_j) \in E^{[5]}$.
- **Casual relationships layer ($G^{[6]} = (V, E^{[6]})$):** Incorporates random or sporadic interactions among individuals, increasing the network's diversity. For each pair of vertices v_i and v_j, edges are added randomly based on a probability parameter $\mu^{[6]}$, creating connections $(v_i, v_j) \in E^{[6]}$ in an unstructured manner.

The resulting aggregated network model associates, for each edge $(v_i, v_j) \in E$, a weight $w(i, j) = \sum_1^l w^{[\alpha]}$, $\alpha = [1, ..., l]$. Consequently, individuals who share multiple social contexts have their interactions represented by stronger connections.

The developed model frequently results in large-scale graphs, often exceeding millions of vertices and edges, which limits its applicability—especially when dealing with medium and large-sized urban centers. This issue is particularly challenging in contexts where the network model is applied to simulate phenomena such as disease propagation and the diffusion of ideas. To address this issue, a sampling technique is employed to create a reduced version of the original network while maintaining a topology that preserves certain characteristics, ensuring that its capacity to represent interactions among individuals is not severely compromised. We propose the use of a strategy based on Respondent-Driven Sampling (RDS) [7], a sampling method that preserves several structural properties of the original network. These include, for instance, the vertex degree distribution, not only globally across the network but also within its local topological organization through the identification of vertex communities.

3. RESULTS

After defining the parameters for the multilayer model, networks are generated for the studied cities in their original form. The results of a preliminary analysis of some topological properties of these networks can be observed in Figure 1 and Table 1, which present, respectively, their Complementary Cumulative Degree Distribution (CCDF) curves, and their main structural measures. The results for the networks sampled through the strategy based on Respondent-Driven Sampling (RDS) are also presented. Ideally, the sampled network should act as a *proxy* for the original network, preserving its fundamental topological properties. Such structural equivalence is important so that reduced-scale simulations can reproduce, with statistical consistency, the trends and dynamics observed in the full network. Indeed, it is possible to notice that the degree distribution curves of the sampled networks globally preserve the characteristics of the original networks, both in their shape and in the presence of a scale-free distribution pattern.



(a) Uberlândia (original) (b) Betim (original) (c) Belo Horizonte (original)

(d) Uberlândia (sampled) (e) Betim (sampled) (f) Belo Horizonte (sampled)

Fig. 1. Degree distribution (CCDF) of the multilayer networks.

In general, the observed values for the properties displayed in Table 1 are quite consistent across the networks for all cities (≈ 11). A moderate assortativity is observed, indicating that the architecture of connections may be predominantly influenced by exogenous factors, such as geographical and socioeconomic variables, rather than degree homophily. For the sampled networks, it is observed that the properties underwent very modest variations, suggesting the preservation of the original topological structure in the sampled networks, thus validating the methodology proposed here.

Table 1. Topological properties of the networks: number of nodes (n), number of edges (m), average degree ($\hat{k}$), average global clustering coefficient ($\hat{cc}$), average assortativity index (ρ), number of communities w/ Louvain method (nc), modularity w/ Louvain method (Q).

Rede	n	m	$\hat{k}$	$\hat{cc}$	ρ	nc	Q
Uberlândia (original)	254417	1443005	11.34	0.09	0.282	99	0.45
Uberlândia (sampled)	127208	740133	11.64	0.10	0.326	95	0.48
Betim (original)	329924	1735944	10.52	0.09	0.255	62	0.44
Betim (sampled)	164964	884632	10.72	0.10	0.311	66	0.46
Belo Horizonte (original)	2228226	13726879	12.32	0.09	0.305	161	0.46
Belo Horizonte (sampled)	1114117	7136255	12.81	0.09	0.332	160	0.49

4. CONCLUSIONS AND FUTURE WORK

This work presented a methodology for constructing multilayer urban social networks, integrating Call Detail Records (CDR) and census data. The conducted topological analyses address the proposed research questions: The combination of CDR and demographic data proved effective in characterizing population interaction, and the generated networks exhibited a scale-free distribution pattern and pronounced heterogeneity, with the presence of *hubs* and a modular community structure, which is consistent with social network literature (RQ1). The application of the Respondent-Driven Sampling (RDS) technique allowed for a 50% reduction in the networks' scale without compromising their fundamental properties, and a structural convergence was observed between the original and sampled networks, as evidenced by the stability of several topological properties (RQ2). The methodology proposed in this work proved to be robust, enabling the use of complex network models in large urban centers for simulations of dynamic processes, such as disease propagation, with high statistical representativeness. In future works, we intend to explore the impact of these different layers on epidemiological simulations and investigate how variations in layer weights ($\mu^{[\alpha]}$) can reflect different intervention policies for containing the spread of diseases.

ACKNOWLEDGMENTS

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

1. M. Lenormand and J. J. Ramasco, "Towards a better understanding of cities using mobility data," Built Environ. **42**, 356 (2016).
2. P. Sapiezynski, A. Stopczynski, R. Gatej, and S. Lehmann, "Tracking human mobility using wifi signals," PLOS ONE **10**, 1 (2015).
3. L. F. Scabini, L. C. Ribas, M. B. Neiva, *et al.*, "Social interaction layers in complex networks for the dynamical epidemic modeling of covid-19 in brazil," Phys. A **564**, 125498 (2021).
4. M. Kivela, A. Arenas, M. Barthelemy, *et al.*, "Multilayer networks," J. Complex Networks **2**, 203–271 (2014).
5. A. Cardillo, J. Gómez-Gardeñes, M. Zanin, *et al.*, "Emergence of network features from multiplexity," Sci. reports **3**, 1344 (2013).
6. F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," Phys. Rev. E **89**, 032804 (2014).
7. D. D. Heckathorn, "Respondent-driven sampling: a new approach to the study of hidden populations," Soc. problems **44**, 174 (1997).

Daily municipal bus demand forecasting in Niterói: combining weekly recurrence, calendar effects, and operational heterogeneity

MATHEUS DE SOUZA E SILVA[1,*], MARINA LEITE DE BARROS BALTAR[1], LUIS EDUARDO SANTOS FORTES[2], ANDERSON MARDEN DE SOUSA SILVA[1], GLAYDSTON MATTOS RIBEIRO[1], AND CINTIA MACHADO DE OLIVEIRA[2]

[1] *Transportation Engineering Program, COPPE, Federal University of Rio de Janeiro*
[2] *Federal Center for Technological Education Celso Suckow da Fonseca (CEFET/RJ)*
[*] *matheus.souza@poli.ufrj.br*

1. INTRODUCTION

Passenger-demand forecasting supports fleet allocation and service planning. Bus ridership varies across time, lines, service periods, holidays, and disruptions; machine-learning models can represent nonlinear interactions among these dimensions [1–3], but require comparison with strong historical baselines [4].

This study models daily municipal bus demand in Niterói at the date–line–period level. It asks how much weekly recurrence explains demand, whether calendar and historical information are complementary, and when forecast reliability deteriorates. The contribution is therefore not only an algorithm benchmark, but evidence about the structures that make demand predictable.

2. DATA AND METHODS

MobNit passenger records [8] were matched to municipal GTFS routes and aggregated by date, line, period, and day type. The selected source comprised 8,172,685 transaction records and 98,019,656 passengers across 47 lines and four periods from January 2, 2025, to May 31, 2026. Historical-feature construction yielded 77,534 complete observations.

Predictors represented calendar effects, operational categories, and historical demand through lags 1 and 7 and 7/30-observation moving averages. Rolling statistics were shifted before calculation to prevent target leakage. Training covered February 1–December 31, 2025 (n = 52,947), and testing covered January 1–May 31, 2026 (n = 24,587); the split was strictly chronological.

Random Forest [5], XGBoost [6], and LightGBM [7] were compared with Lag-1, Lag-7, and 7-observation moving-average baselines using MAE, RMSE, WAPE, and R^2. LightGBM ablation contrasted calendar-only, history-only, combined, and expanded specifications. The expanded model added daily observed weather and static GTFS descriptors. Feature importance and SHAP were treated as predictive, not causal, evidence.

3. RESULTS

Table 1 combines the benchmark and ablation results. All supervised models outperformed the historical baselines. LightGBM achieved the lowest RMSE (209.71), MAE (117.88), and highest R^2 (0.9794), reducing RMSE by 62.5% relative to Lag 7. Nevertheless, Lag 7 was the strongest baseline, confirming marked weekly recurrence. The supervised-model gain does not arise from omitting lags: these models explicitly combined historical demand with calendar and operational information.

Calendar and history alone reached RMSE 281.64 and 375.83, respectively, while their combination reached 209.71. Their information is therefore complementary. The expanded specification slightly reduced MAE but increased RMSE to 213.05, showing that appending observed weather and static network descriptors did not improve the primary metric.

Table 1. Predictive performance and LightGBM ablation.

Model/specification	MAE	RMSE	WAPE	R^2
Model comparison				
LightGBM	117.88	209.71	10.16	.9794
XGBoost	119.06	215.24	10.26	.9783
Random Forest	118.88	225.79	10.25	.9761
Lag 7	229.34	559.17	19.77	.8533
Moving average 7	417.67	729.04	36.00	.7506
Lag 1	382.44	809.95	32.96	.6921
LightGBM ablation				
Calendar	166.61	281.64	14.36	.9628
History	159.61	375.83	13.76	.9337
Calendar + history	117.88	209.71	10.16	.9794
Expanded	117.52	213.05	10.13	.9787

Historical demand dominated interpretation: Lag 7, Lag 1, and the 30/7-observation moving averages were among the most influential attributes. Calendar variables and operational categories also contributed (Fig. 1), consistent with the ablation result.

Reliability was heterogeneous. Low-demand observations reached WAPE 36.90% despite small absolute errors. Atypical days reached RMSE 258.35 and WAPE 15.18%, compared with 186.45 and 8.64% on regular days. Global performance thus conceals operationally relevant failure conditions.

The scale-dependent metrics clarify this contrast. High-demand observations produced the largest absolute deviations because more passengers were exposed to error, whereas the lowest-demand quartile produced the largest proportional error. A model can therefore appear accurate in system-wide RMSE while remaining unreliable for low-volume services. For operational use, aggregate metrics should be accompanied by segment-level diagnostics and thresholds tailored to the planning decision.

The atypical-day gap also indicates that calendar flags only

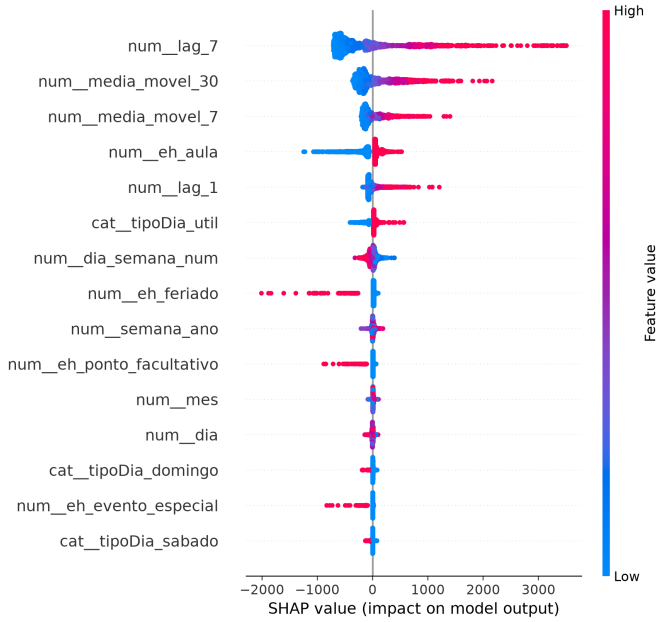


Fig. 1. SHAP summary for a LightGBM test sample.

partially represent departures from routine operation. Holidays and school vacations identify broad demand regimes, but they do not encode event location, event intensity, disruptions, detours, cancellations, or the service actually supplied. Because the target is observed boardings, it reflects both travel needs and available capacity; it should not be interpreted as latent demand independently of operations.

The spatial extension must be interpreted similarly. Static GTFS attributes summarize route size and connectivity, but lines still enter the predictive model largely as independent categories. Shared corridors, stops, terminals, and transfer hubs may create correlated changes that static route descriptors cannot express. The absence of improvement from the expanded specification is thus evidence against simple feature appending, not evidence that weather or network structure is irrelevant.

The chronological design strengthens the evaluation because the 2026 observations were never used for fitting. It nevertheless represents a single temporal holdout. Demand patterns can drift after fare changes, route restructuring, land-use change, or prolonged service disruptions. Repeated rolling-origin tests across longer periods are needed to distinguish stable weekly structure from gains that depend on this particular training and test window.

Reproducibility also requires preserving the complete forecasting pipeline rather than only the final estimator. The set of eligible municipal routes, temporal aggregation, categorical encoding, shifted lag construction, and missing-history removal jointly define the 77,534-observation sample. Any change to these stages changes the prediction task and requires recalculating every baseline and supervised result on the same test observations.

4. DISCUSSION AND CONCLUSION

Weekly recurrence was strong but insufficient: Lag 7 was the best baseline, while combining historical, calendar, and operational information substantially reduced RMSE. The expanded specification did not improve RMSE, indicating that static network descriptors and observed weather are not equivalent to representing the mechanisms through which network and contextual conditions affect demand.

Forecast reliability deteriorated on atypical days and proportionally on low-demand observations. Future models should explicitly represent dynamic interactions among lines, realized service supply, events, and forecast weather. Overall, predictability depends less on algorithm choice alone than on the complementary temporal and operational information available to the model.

These results also delimit the intended use of the forecasts. The estimates can support fleet allocation, headway review, and exception monitoring, but should be combined with scheduled capacity and managerial knowledge rather than used as automatic operating commands. Prospective validation should replace observed weather with forecast information, use rolling-origin evaluation, and quantify uncertainty through prediction intervals.

A natural spatial extension is a line–line or stop–stop graph weighted by shared stops, corridor overlap, or transfer intensity. Coupling this representation with time-varying GPS, scheduled and realized trips, and origin–destination flows would test whether disruptions and demand shocks propagate through the network. Such extensions preserve the central finding of the present study: predictive gains arise when complementary information represents distinct structures instead of merely increasing the number of available variables.

Finally, model comparison should remain tied to operational costs. RMSE emphasizes large misses that may affect fleet adequacy, whereas MAE summarizes typical absolute deviation and WAPE supports comparisons across demand scales. No single metric captures crowding risk, missed trips, or the asymmetric consequences of underprediction. A deployment study should therefore translate forecast errors into service-level outcomes and examine whether improvements persist for the lines and periods where planning decisions are most consequential.

REFERENCES

1. F. Toqué, N. Kühl, M. Jakob, and H. Darmer, “Short & long term forecasting of multimodal transport passenger flows with machine learning methods,” in *2017 IEEE International Conference on Big Data*, 2017.
2. F. Rowe, M. Mahony, and S. Tao, “Assessing machine learning algorithms for near-real-time bus ridership prediction during extreme weather,” *arXiv preprint arXiv:2204.09792*, 2022.
3. L. Zou *et al.*, “Passenger flow prediction using smart card data from connected bus system based on interpretable XGBoost,” *Wireless Communications and Mobile Computing*, 2022.
4. F. Rodrigues, “On the importance of stationarity, strong baselines and benchmarks in transport prediction problems,” in *2023 IEEE International Conference on Intelligent Transportation Systems*, 2023.
5. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
6. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD Conference*, pp. 785–794, 2016.
7. G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
8. Municipality of Niterói, “MobNit – Sistema de Gestão da Mobilidade Urbana de Niterói,” available at https://mobnit.niteroi.rj.gov.br/.

Disclosures. The authors declare no conflicts of interest.

Data availability. The data underlying the results are not publicly available but may be obtained from the authors upon reasonable request.

Detecting bus delays and service anomalies in Niterói with autoencoder latent feature representations

MATHEUS GERMANO AND CLAUDIO E. C. CAMPELO

Systems and Computing Department,Federal University of Campina Grande (UFCG), Campina Grande, PB, Brazil

1. INTRODUCTION

Riders judge a bus system by two failures: trips that run late, and days when service as a whole misbehaves; travellers value reliability on a par with travel time itself [1]. We ask whether a single unsupervised representation of raw GPS telemetry can support both diagnoses at once on the NetMob 2026 challenge data from Niterói [2]. Our approach transfers the latent feature representation learning pipeline of Duncan and Chen [3], originally built to detect internet censorship in network traffic: an autoencoder (AE) is trained to compress each trip into a low-dimensional latent vector, the frozen encoder then feeds lightweight supervised classifiers for *delay detection*, and the same AE's reconstruction error doubles as an unsupervised *service anomaly* score. Autoencoders are long established [4] and have been applied to anomalous bus trajectories [5]; here we reuse one label-free representation for both tasks. This is particularly useful when timetable-based comparisons are unreliable — for example, when schedule joins are missing or outdated, when service runs at variable intervals, or when trips are evaluated before they are complete.

2. DATA AND PREPROCESSING

We use the NetMob26 mobility feed (11–31 March 2026, ~ 950,000 GPS points per day), the accompanying General Transit Feed Specification (GTFS) schedule snapshots, and hourly weather [2]. Each GPS trip report is joined to its scheduled GTFS trip (matching identifiers per direction and dated snapshot, over 99% of trips), map-matched onto the route shape, and stop arrival times interpolated where along-route progress crosses each stop. The dataset carries no delay annotations, so we derive the label ourselves. The Transit Capacity and Quality of Service Manual defines a fixed-route service as on-time when it runs no more than 5 minutes behind schedule [6]; adopting this industry boundary, we mark a trip *delayed* when its median stop delay against the schedule is at least 5 minutes, additionally requiring at least five matched stops and 50% stop coverage for the median to be trustworthy. In this work, GTFS is used only to construct the delay labels and is excluded from the model inputs. We then retain trips with at least 10 points and durations between 8 minutes and 3 hours, split trips at internal gaps longer than 30 minutes, and remove points more than 200 m from the route. Instances with more than 40% of their points removed are discarded, as are trips failing the five-stop and 50%-coverage criteria. These filters yield 22,605 trip instances across 20 service dates, of which 20.9% are labeled delayed.

Each trip becomes a 64×12 sequence of 1-minute bins carrying only label-free signals: kinematics (speed, acceleration, heading change), route geometry (progress fraction and rate, cross-track error), dwell fraction, time-of-day encodings, a weekend flag, hourly rainfall, and headway to the preceding trip of the same line and direction. No schedule-derived quantity enters the features, so the delay label cannot leak into the representation. Trips longer than 64 min are truncated to their first 64 bins; shorter trips are zero-padded and carry an explicit length, so the pooled latent and the reconstruction error are both masked to real timesteps. Empty bins are filled from the interpolated along-route progress, with cross-track error carried forward and rainfall forward-filled. Each trip is scored from its complete trajectory: the task as posed is retrospective classification, not early prediction.

3. METHOD

Two AE variants are compared under one scaffold: a one-dimensional convolutional AE (CNN-AE) [7] and a sequence-to-sequence LSTM AE [8], both with a 64-dimensional latent space. For each cross-validation fold the AE is trained unsupervised on the training split only; the frozen encoder then produces a pooled latent vector z for four classifiers (MLP, XGBoost, random forest, logistic regression) and a latent sequence z_{seq} for two recurrent classifiers (LSTM [8], GRU [9]). Evaluation uses `StratifiedGroupKFold` with 5 folds grouped by *service date* [10]: day-level effects (weather, traffic, telemetry quality) are the dominant leakage channel, so every model is scored exclusively on unseen days. No hyperparameter search was performed; all models use fixed settings based on their respective literature.

For anomaly detection each trip is scored by the AE of the fold where it was held out, so no trip is scored by a model that saw it. Since every fold has its own AE, raw errors are not comparable across folds: the heavy-tailed MSE is log-transformed and standardized within each fold before pooling, and trips above the 95th percentile of the pooled score are flagged. That percentile sets an inspection budget rather than an anomaly rate, since no ground truth exists for the latter. At 5% of 22,605 trips, this corresponds to ~ 57 alerts per service day. Applied once to the pooled score, the daily alarm count is free to vary (3–9% of trips), but a global cut is dominated by between-line differences. On 22 March, the one day with known degraded telemetry, it flags 5.7% of trips against 29.3% when each score is re-standardized within its own line. We assess the score here against two external checks: three days with known degraded telemetry (20, 22, and 28 March) and rainfall. Because rainfall is also an AE input, we repeat the analysis with the CNN-AE retrained on the 11 non-weather channels.

4. RESULTS

A. Delay detection

Table 1 reports mean test metrics over the five day-grouped folds. CNN-AE latents beat LSTM-AE latents for every classifier, and tree ensembles on the pooled latent work best: **CNN-AE + XGBoost reaches** 0.838 ± 0.014 **ROC-AUC and** 0.581 ± 0.035 **F1**, with random forest indistinguishable from it (paired t over folds, $p = 0.66$ ROC-AUC). The CNN-AE advantage over the LSTM-AE, by contrast, holds in every fold ($p = 0.001$ ROC-AUC; Wilcoxon $p = 0.031$, the floor at $n = 5$). Fold spread is modest, so the representation generalizes to unseen days despite day-to-day shifts in weather and telemetry quality.

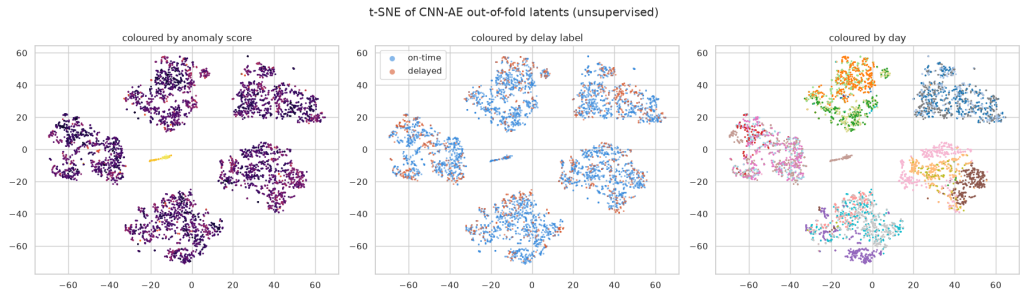


Fig. 1. t-SNE of 5,000 out-of-fold CNN-AE latents, coloured by anomaly score (left), delay label (centre; delayed trips marked in orange and concentrated at lobe margins) and service date (right). Each of the five lobes corresponds to one fold-specific encoder.

Table 1. Delay detection on unseen days (mean over 5 folds)[a]

Classifier	ROC-AUC	PR-AUC	Prec.	Rec.	F1[b]
CNN-AE					
AE+MLP	0.824	0.593	0.437	0.738	0.548
AE+XGBoost	0.838	**0.629**	0.539	0.637	**0.581**
AE+RF	**0.839**	0.621	**0.546**	0.612	0.574
AE+LR	0.811	0.564	0.446	0.694	0.541
AE+LSTM	0.826	0.601	0.448	0.738	0.557
AE+GRU	0.828	0.607	0.437	**0.752**	0.552
LSTM-AE					
AE+MLP	0.801	0.548	0.421	0.698	0.522
AE+XGBoost	0.814	0.575	0.471	0.650	0.544
AE+RF	0.819	0.576	0.476	0.645	0.545
AE+LR	0.785	0.512	0.403	0.703	0.512
AE+LSTM	0.809	0.559	0.418	0.717	0.526
AE+GRU	0.808	0.554	0.407	0.751	0.525

[a]Cross-fold SDs: 0.014–0.020 (ROC-AUC), 0.046–0.061 (PR-AUC), 0.028–0.051 (Prec.), 0.043–0.079 (Rec.), 0.033–0.043 (F1). Bold = best. Delayed prevalence 0.209.
[b]Binary F1 on the delayed class.

B. Service anomaly detection

The two AEs produce similar anomaly scores: their out-of-fold errors correlate at Pearson $r = 0.876$ on the log scale (Spearman $\rho = 0.896$), with 705 of the 1,131 trips flagged by both models (Jaccard 0.45). The degraded-telemetry signal is also present in the weather-free model, although only on two of the three affected days. On 20 and 22 March, scores shift upward under both variants: the median score is $+0.37$ versus -0.20 with rainfall as input and $+0.50$ versus -0.19 without it (one-sided Mann–Whitney $p < 10^{-114}$). The corresponding flag rates remain similar (4.6% vs. 5.1%), suggesting a broad shift in the fleet rather than a small number of extreme trips. On 28 March, which has only 449 trips, scores instead fall below the normal-day median under both variants. This shows that degraded telemetry does not necessarily produce higher reconstruction error. Because degraded telemetry reflects data quality rather than service disruption, we use it only as a positive control for sensitivity.

The rainfall association is not preserved when rainfall is excluded from the input. With rainfall included, 12.2% of rain-hour trips are flagged, compared with 4.4% of dry-hour trips; without rainfall, the rates are 5.4% and 5.0% ($p = 0.84$). Similarly, all top-20 scoring trips occur on 23 March between 15:30 and 17:00, during an 11.8 mm/h rain cell, but the median score for this window falls from $+5.43$ to -0.22 when rainfall is removed, and none of these trips remain in the top 20. Thus, the apparent anomaly is driven by rainfall as an input rather than by a broader operational signal. The day-level ordering is nevertheless stable: the two variants have Spearman $\rho = 0.961$ across day-median scores, with 22 March ranked first by both.

Figure 1 shows the CNN-AE latent space. The five visible lobes reflect the five fold-specific encoders used in the analysis. Within each lobe, delayed trips occupy distinct regions, while trips from the same service date tend to cluster together. This structure is consistent with the two patterns captured by the downstream tasks: trip-level delay and day-level variation.

5. CONCLUSION

Under identical folds, classifiers trained on the raw features perform similarly to those using the CNN-AE representation. Gradient boosting reaches 0.845 ROC-AUC and 0.632 PR-AUC on the raw features, compared with 0.838 and 0.629 for the latent features; per-fold differences remain below 0.04. A contrastive (SimCLR) encoder trained on the same folds does not improve on these results. The main benefit of the AE representation is therefore not predictive accuracy but compression and linear separability: logistic regression improves by 0.027 ROC-AUC and 0.064 PR-AUC over the raw features. For delay detection, the representation itself has little effect on performance. The anomaly results are more consistent: the same days receive the highest scores across four independently trained representations, including models without weather inputs and without a reconstruction objective. These results suggest that the day-level signal is less dependent on the specific representation or training objective.

6. DATA AVAILABILITY

The used dataset was obtained under the NetMob 2026 Data Challenge terms and is available from the organizers to admitted participants [2]. Weather data are from the public municipal feed. All analysis code is openly available. Code: github.com/mttgermano/Netmob2026. Dataset: github.com/lprm-ufes/Netmob2026.

REFERENCES

1. J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," Transp. Res. Part E: Logist. Transp. Rev. **37**, 191–229 (2001).
2. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 dataset: A high-resolution multi-source view of public bus mobility in Niterói," arXiv preprint arXiv:2605.20263 (2026).
3. S. P. Duncan and H. Chen, "Detecting network-based internet censorship via latent feature representation learning," Comput. & Secur. **128**, 103138 (2023).
4. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science **313**, 504–507 (2006).
5. X. Zhang, Y. Zheng, Z. Zhao, *et al.*, "Deep learning detection of anomalous patterns from bus trajectories for traffic insight analysis," Knowledge-Based Syst. **217**, 106833 (2021).
6. Transportation Research Board, "Transit capacity and quality of service manual," 3rd ed., TCRP Report 165 (2013).
7. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).
8. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**, 1735–1780 (1997).
9. K. Cho, B. van Merriënboer, C. Gulcehre *et al.*, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. EMNLP*, (2014), pp. 1724–1734.
10. F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011).

Demand-Weighted Reliability and Diagnostic Observability in Public Transport: A Multi-Source Alignment Approach

MANUEL MENDOZA-HURTADO[1*], JOSÉ MANUEL CUEVAS-MUÑOZ[1], AND DOMINGO ORTIZ-BOYER[1*]

[1] Department of Computer Science and Artificial Intelligence, Campus de Rabanales, Albert Einstein Building, University of Cordoba, 14071, Cordoba, Spain
[*] mmendoza@uco.es, dortiz@uco.es

1. INTRODUCTION

Urban mobility studies often infer behavior from passively collected records such as Call Detail Records (CDR), but these provide only indirect and partial observation. Public transport data offer a more structured setting—routes, schedules, fares, and vehicle trajectories—but are not automatically complete: Automated Fare Collection (AFC) systems may be "tap-on only" [1], while Automatic Vehicle Location (AVL/GPS) telemetry may contain dropouts or lack reliable vehicle-to-route assignments.

We address the resulting problem of unlinked AFC and AVL/GPS records through a multi-source alignment pipeline. By fusing 6.8 million ticket boardings, 13.8 million GPS pings, GTFS schedules, weather observations, and 2022 Census data from the NetMob26 dataset [2], we reconstruct a route-hour view of demand and supply for the Niterói bus network. A labeled subset of 40 routes supports the diagnostic multi-label experiments.

Our contribution is both engineering and diagnostic. In data-scarce settings, planners may have aggregate route, schedule, fare, and census/context information but not the aligned spatiotemporal evidence required to measure every route state directly. We therefore ask: *which route-level states are recoverable from aggregate features, and which require granular alignment or manual spatial inspection?* Building on a supervised multi-label framing [3], our hybrid 40-route diagnostic pilot uses three model families and 27 aggregate predictors to recover six semantic groups under nested repeated group cross-validation. The objective is not a deployable predictor, but an observability probe that quantifies what the available data can and cannot support.

2. DATA AND METHODOLOGY

A. Multi-Source Alignment Pipeline

The pipeline first normalizes route identifiers across sources. GTFS trips are expanded using service calendars, while GPS pings are segmented into vehicle-trip records, matched to GTFS by trip identifier and direction, and filtered to records with at least five pings and durations of 5–300 minutes. Observed and scheduled trips are then aggregated to route-date-hour cells. AFC boardings are aggregated at the same grain; missing and known low-coverage GPS days are excluded from supply comparisons.

For each valid date-hour cell i of route r with $S_{ri} > 0$, let $d_{ri} = \max(S_{ri} - O_{ri}, 0)/S_{ri}$ and $e_{ri} = \max(\bar{H}^o_{ri} - \bar{H}^s_{ri}, 0)/(60\,\text{min})$. The route score is $M_r = \sum_{i \in \mathcal{I}_r} B_{ri}(d_{ri} + e_{ri})$, where B is boardings, S/O are scheduled/observed trip starts, and $\bar{H}^s/\bar{H}^o$ are scheduled/observed mean headways. Thus, M_r prioritizes reconstructed service gaps by passenger exposure; it is not a confirmed-failure count.

B. Diagnostic Multi-Label Classification Setup

The hybrid reference set contains 40 routes and 16 binary columns forming six semantic groups: five three-class one-hot groups and binary urban-core exposure. Urban-core exposure is assessed by one annotator from spatial evidence. FCU exposure is the mean percentage of route length and stops within 300 m of official FCU polygons, discretized as low ($< 35\%$), medium (35–50%), or high ($> 50\%$). Operational states follow deterministic March 2026 rules: demand-intensity tertiles use 108.3 and 318.7 boardings per demand-positive route-hour; general-delay tertiles use 2.04 and 3.18 minutes of observed-trip-weighted positive excess over GTFS duration. Rainy cells are compared with dry baselines matched by route, hour, and weekday/weekend status. Rain-demand uses $\pm 5\%$ thresholds; rain-delay requires both $\pm 10\%$ and ± 0.5 minutes. Eligible routes have at least six matched rainy cells. This pilot is therefore a diagnostic reference set, not independently adjudicated ground truth.

Each route has 27 aggregate predictors: two schedule/service, two fare-composition, four corridor-scale, 15 population-weighted census/context, two GTFS multiplicity, and two geometric features. The demand, delay, rain-effect, and direct spatial-exposure metrics defining the labels are excluded. Predictors are median-imputed with missingness indicators and standardized within each training fold.

We apply nested repeated Group K-Fold cross-validation to Binary Relevance with kNN (BR-kNN), a Classifier Chain with SVC (CC-SVC), and Label Powerset with Random Forest (LP-RF), grouping suffix variants of the same route family. The outer evaluation uses 4 folds and 5 repeats; configurations and the global model selector are chosen only through 3-fold inner cross-validation using mean balanced accuracy across the six semantic groups. We report balanced accuracy, macro-F1, and 2,000-replicate route-bootstrap 95% confidence intervals, with raw accuracy and fold-local majority predictions as secondary diagnostics. Separate label-permutation, permutation-importance, and feature-group-ablation runs provide complementary checks but do not drive model selection. Exact commands, seed, environment versions, and input hashes are recorded in an accompanying reproducibility protocol. Models fitted on all 40 reference routes are applied to 20 unlabeled routes only as an exploratory inference pass. Interpretation is restricted to the 10 routes with 92.6–100% feature completeness; the other 10 have only 11.1% completeness.

3. DESCRIPTIVE RESULTS

Across the 6,790,681 March boardings, 6,333,872 (93.3%) use route identifiers represented in both GPS and GTFS and are therefore structurally alignable. On valid comparison days, after excluding missing or low-coverage GPS dates, 72,914 reconstructed versus 75,600 scheduled trip starts yield a 96.5% system-level ratio. This aggregate masks substantial route-level heterogeneity.

Figure 1 ranks the 60 records in the unified route-level table by M_r. On those valid days, route 33 combines 157,502 boardings with 1,447 reconstructed versus 4,922 scheduled starts (29.4%); 97.5% of its score comes from the relative-deficit component.

Routes 33 and 48 jointly account for 26.0% of the total index, while the top five account for 50.0%. Because route-specific telemetry or identifier-matching loss can depress reconstructed counts, the ranking identifies audit priorities rather than confirmed failure rates.

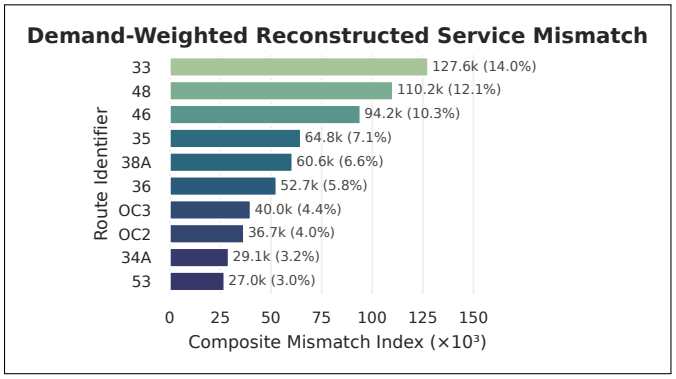


Fig. 1. Demand-weighted composite mismatch for the top ten of 60 records in the unified route-level table. Each bar is M_r; larger values reflect some combination of demand, relative trip-start deficit, and positive headway excess. Parentheses show each route's share of the total 60-record index. The ranking prioritizes audits and is not a confirmed failure rate.

4. DIAGNOSTIC OBSERVABILITY

Nested repeated cross-validation reveals uneven observability from aggregate route-level features (Figure 2). Zero denotes nominal balanced-accuracy chance, not the raw-accuracy majority baseline. Inner folds determine all configurations and one global selector across the six groups.

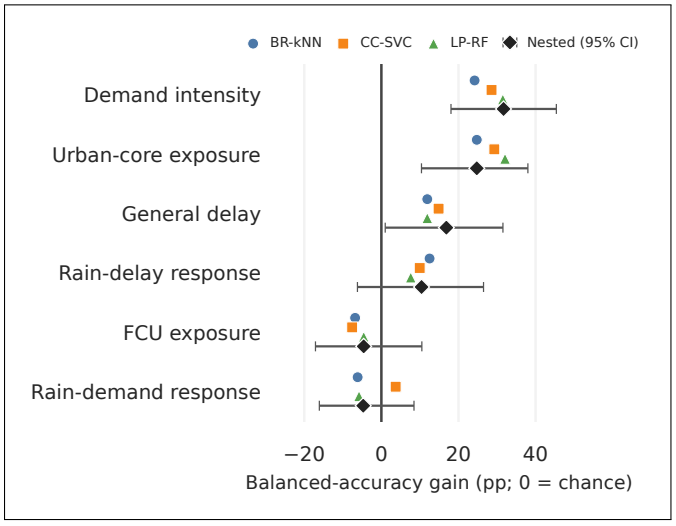


Fig. 2. Balanced-accuracy gain above nominal chance (1/3 for three-class states; 1/2 for binary urban-core exposure). Markers show internally tuned families; black diamonds show the global selector with route-bootstrap 95% confidence intervals. Zero is not the raw-accuracy majority baseline.

A. More Observable: Demand and Urban Structure

Aggregate features clearly recovered demand intensity and urban-core exposure, with weaker evidence for general delay. The global selector achieved balanced accuracies of 0.650 (95% CI: 0.514–0.787), 0.747 (0.604–0.880), and 0.502 (0.343–0.649), respectively; general delay remained tentative because its lower bound was only 0.010 above chance. No family dominated: LP-RF led demand/urban core, CC-SVC general delay/rain demand, and BR-kNN rain delay; all were weak on FCU. A separate full-sample control yielded micro-F1 0.564 on real labels versus 0.348 after label permutation. Exploratory importance highlighted scheduled frequency and spatial extent; small, model-dependent ablation drops (maximum 0.044) indicated redundancy.

B. Less Observable: FCU and Weather Response

The aggregate predictors were less informative for FCU exposure and weather response. Balanced accuracy was 0.287 for *FCU Exposure*, 0.286 for *Rain Demand Response*, and 0.438 for *Rain Delay Response*; all three intervals included chance. Route 32 missed all six groups and routes 35/60 five, despite complete predictors and diverse frequency (2.26–6.45 trips/route-hour) and length (15.4–32.0 km); overall, FCU and rain demand were wrong for 27/40 and 24/40 routes. This identifies predictor and reference-design limits, not intrinsic unpredictability: FCU requires direct spatial proximity, while rain response requires matched route-hour variation.

Among the 10 high-completeness unlabeled routes retained for interpretation, BR-kNN, CC-SVC, and LP-RF were unanimous most often for urban core (10/10) and general delay (9/10), but least often for FCU and rain-delay response (2/10 each). All three assigned *stable* rain demand to every route, suggesting majority-class tendency. Spatial rules agreed with consensus for urban core in 8/10 and FCU tier in 6/10; without adjudicated ground truth, these outputs support triage, not validation.

5. CONCLUSION

The pipeline enables demand-weighted reliability audits and diagnostic observability tests. A 96.5% reconstructed-to-scheduled trip-start ratio coexisted with concentrated route-level mismatch, identifying audit priorities rather than confirmed failures. Aggregate features recovered demand and urban-core exposure most clearly; general delay was tentative, and FCU/rain-response evidence inconclusive. The hybrid 40-route pilot, single-annotator urban-core labels, rule-defined FCU and operational states, and wide intervals preclude citywide generalization. Larger, adjudicated reference sets plus direct spatial and matched route-hour evidence are needed.

Acknowledgments. We thank the NetMob 2026 Data Challenge organizers and Fluminense Federal University (UFF) for providing access to the dataset [2].

REFERENCES

1. M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," Transp. Res. Part C: Emerg. Technol. **19**, 557 (2011).
2. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The NetMob26 dataset: A high-resolution multi-source view of public bus mobility in Niterói," arXiv preprint arXiv:2605.20263 (2026).
3. M. Mendoza-Hurtado, J. A. Romero-del Castillo, and D. Ortiz-Boyer, "MAPLID: A new multi-label approach for place identification using data supplied by mobile network operators," Int. J. Geogr. Inf. Sci. (2026).

Counterfactual Dispatch Regularization Reduces Passenger Waiting in Urban Bus Services

M. MAVIN DE SILVA[1,2*], SANTIAGO GARCIA GABILONDO[2], KETIAN CHEN[2], SHUSEI HAYASHI[4], JUE MA[2*], AND YOSHIHIDE SEKIMOTO[2]

[1] Extreme Energy-Density Research Institute, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan
[2] Center for Spatial Information Science, University of Tokyo, Komaba, Tokyo, Japan
[3] Graduate School of Information Sciences, Tohoku University, Tohoku, Japan
[*] s225063@stn.nagaokaut.ac.jp/majue@csis.u-tokyo.ac.jp

1. INTRODUCTION

Public transport reliability is often evaluated relative to scheduled service, but passengers experience reliability through the service that actually arrives. In high-frequency bus operations, irregular departures produce uneven headways and can contribute to bus bunching, long gaps, and elevated waiting time even when the total number of departures is not especially low. Many analytical workflows evaluate this problem using schedule-linked trips and stop-level timetables. However, those assumptions become difficult to sustain when scheduled GTFS trip records are incomplete or when GPS observations cannot be reliably matched to individual GTFS trip instances[1].

For this challenge, because no GTFS Realtime feed was available, we were provided access to high frequency GPS data that we could use to identify the actual operating patterns and level of service provided[2]. Through our initial analysis, we identified a lack of consistency in the stated static GTFS service provision and the supporting GPS data: many scheduled trip instances could not be matched to corresponding GPS trajectories, the dispatch times had deviations from their stated schedules, and more importantly, we observed significant irregularities in the departure headways.

To evaluate how this lack of service reliability affects passenger waiting times, we first matched the Autonomic Fare Collection(AFC) data to specific stations to quantify the empirical demand. We then propose an aggregate route-direction-hour headway reliability framework that uses observed GPS trip origins as the service backbone and AFC-derived boardings as passenger exposure[3]. We then evaluate how much passenger waiting time could be reduced under two fleet-neutral counterfactual regularization scenarios, in which the same observed departures are redistributed either evenly across the full hourly block or conservatively after the first observed departure.

2. METHODOLOGY

A. Spatiotemporal Demand Reconstruction at the Stop Level

We begin the analysis with a citywide preprocessing pipeline that integrates bus GPS, and AFC records. While the source data provided detailed ticketing records for each vehicle, the vehicle ids did not match to those of the GPS data. Therefore we attempted to conciliate the information from both data sources. To do so, we matched the vehicle progression along the route to identify the expected arrival time of a vehicle to a given station. Next, based on the assumption that all AFC ticketing times occurred at the exact location of a given station, we cross-reference the interpolated GPS station arrival times with the AFC ticketing times, applying a $/pm5$ second offset to account for delays from the bus stopping to the passenger validating it boarding. We do this cross-referencing comparing all vehicles from both data sources. Finally, we use a recursive algorithm greedily assign the best matches based on the number of successful matches between the ticketing times and our expected station stops. We enforce a strict 1-to-1 vehicle mapping and then assign the passenger volumes to specific stations by interpolating the boardings based on the transaction time (Figure. 1).

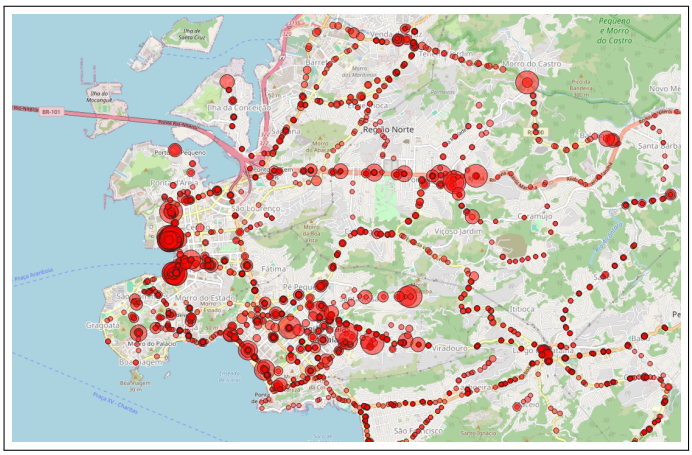


Fig. 1. Matched stop level demand

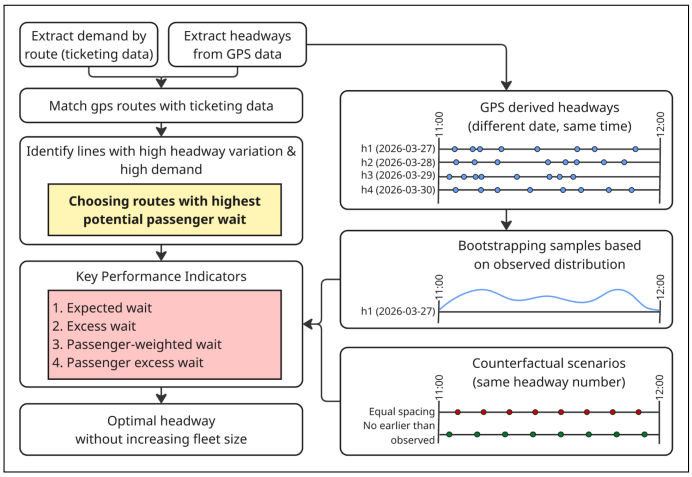


Fig. 2. Overall workflow summary.

B. Measuring Origin-Headway Reliability under Missing GTFS

For each hourly block, observed headways are calculated from GPS-derived trip origins. Let h_i denote the i-th observed headway within the hourly block. Expected waiting time is calculated using the squared-headway formulation: $E(W)_{r,d,t} = \frac{\sum_i h_i^2}{2\sum_i h_i}$. For a fixed 60-minute hourly cycle, this becomes $E(W)_{r,d,t} = \frac{\sum_i h_i^2}{2\times 60}$.

This produces a route-direction-hour origin-headway-based expected wait, rather than a stop-level waiting estimate. Observed passenger waiting burden is then calculated as $PW^{obs}_{r,d,t} = B_{r,d,t} \times E(W)^{obs}_{r,d,t}$. We compute baseline reliability indicators, including expected wait, excess wait, passenger-weighted wait, and passenger excess wait, and restrict the evaluation to hourly blocks that satisfy data-quality and service-reliability screening.

To examine whether GPS observations can reveal a stable latent dispatch pattern in the absence of usable GTFS trips, we bootstrap the empirical dispatch distribution. For each selected route-direction-time window, daily GPS-derived departure profiles are resampled with replacement to estimate the mean dispatch distribution and its 95% uncertainty interval. This assesses whether repeated GPS observations form a stable schedule-like pattern or instead show day-to-day variability. In addition, we apply a day-level bootstrap to the counterfactual results by resampling service days and recalculating passenger-hours saved: $PHS = \frac{\sum_{r,d,t} PW^{obs}_{r,d,t} - \sum_{r,d,t} PW^{cf}_{r,d,t}}{60}$, and wait reduction percentage: $WR = 100 \times \frac{\sum_{r,d,t} PW^{obs}_{r,d,t} - \sum_{r,d,t} PW^{cf}_{r,d,t}}{\sum_{r,d,t} PW^{obs}_{r,d,t}}$.

C. Fleet-Neutral Counterfactuals for Waiting-Time Reduction

Finally, we compare observed service with two explicit fleet-neutral counterfactual scenarios. In the Uniform Hourly Regularization, the same number of observed departures is evenly redistributed across the full hourly block. In the Constrained Regularization, the first observed departure is fixed and only the remaining departures are evenly redistributed until the end of the hour. The first scenario represents an upper-bound regularization benchmark, while the second provides a more operationally conservative alternative. In both scenarios, the number of buses and passenger demand in each hour are held constant; only the spacing of departures changes. This allows us to quantify how much passenger waiting time could be reduced by regularizing irregular GPS-observed dispatches without increasing fleet size. The overall workflow is shown in Figure 2.

3. PRELIMINARY RESULTS

Across the four analyzed route-directions of Routes 30 and 47, the final evaluation includes 995 reliable hourly blocks, 11,540 GPS-observed departures.

The counterfactual results show substantial waiting-time reductions when the same observed departures are regularized without increasing service volume. Aggregating across all four route-directions, passenger-weighted waiting time decreases from 4.61 to 2.64 minutes per passenger under the uniform hourly regularization scenario. Under the constrained regularization scenario, the aggregate passenger-weighted waiting time decreases to 3.28 minutes per passenger. Table 1 reports the corresponding route-direction-specific results. The largest improvement is observed for Route 30, Direction 0, where uniform hourly regularization saves 4,688.5 passenger-hours, equivalent to a 51.3% reduction. The constrained regularization scenario for the same route-direction still saves 2,717.3 passenger-hours, or 29.7%. Across the remaining route-directions, uniform hourly regularization reduces passenger waiting by 33.2% to 42.3%, while constrained regularization produces reductions of 25.4% to 33.3%. Day-level bootstrap resampling further confirms that waiting-time savings remain positive across all route-direction and scenario combinations.

These results indicate that the main source of passenger waiting burden is not only insufficient service volume, but also the irregular temporal distribution of the service already being operated. Because both counterfactual scenarios hold hourly passenger demand and the number of observed departures constant, the estimated savings represent the potential benefit of improving dispatch regularity rather than adding additional buses.

Table 1. Passenger waiting-time reduction under fleet-neutral counterfactual regularization.

Case	Bdgs.	Obs. (min)	CF (min)	Red. (%)
30–0 Uniform	100,591	5.46	2.66	51.3*
30–0 Constrained	100,591	5.46	3.83	29.7*
47–1 Uniform	92,787	4.78	2.96	38.1*
47–1 Constrained	92,787	4.78	3.57	25.4*
47–0 Uniform	71,029	4.09	2.36	42.3*
47–0 Constrained	71,029	4.09	2.73	33.3*
30–1 Uniform	69,602	3.66	2.45	33.2*
30–1 Constrained	69,602	3.66	2.65	27.6*

Notes: Bdgs. = adjusted boardings; Obs. = observed passenger-weighted waiting time per passenger; CF = counterfactual passenger-weighted waiting time per passenger; PHS = passenger-hours saved; Red. = waiting-time reduction. Uniform = uniform hourly regularization, where observed departures are evenly redistributed across the full hourly block. Constrained = constrained regularization, where the first observed departure is fixed and the remaining departures are redistributed until the end of the hour. * indicates that the day-level bootstrap 95% interval for waiting-time savings is above zero.

4. EXPECTED CONTRIBUTIONS

This paper makes three contributions. First, it introduces a robust method to match inconsistent GPS and ticketing data, allowing to explicitly map the actual demand to conduct more detailed analysis. Second, it proposes a schedule-free evaluation framework that combines GPS-derived trip origins, route-time AFC matching, and route-direction-hour expected waiting time. Third, it shows that substantial benefits can be achieved without increasing service volume.

The broader implication is both methodological and operational. When GPS observations do not reveal a stable latent dispatch template, reliability evaluation should not depend on an inferred schedule that is not clearly present in operations. Instead, explicit counterfactual regularization scenarios grounded in observed departures provide a transparent benchmark for service improvement. Since both counterfactuals hold demand and the number of departures constant, the estimated benefits reflect improved temporal regularity rather than additional buses.

The current results focus on route-direction-hour origin-headway reliability rather than stop-level waiting. We expect to expand this framework to stop-level analysis to examine where along the corridor passengers experience the largest waiting-time burden.

5. BACK MATTER

Funding. M.M.D.S. acknowledges support by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for JSPS Fellows [24KJ1180].

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. N. Wessel, J. Allen, and S. Farber, "Constructing a routable retrospective transit timetable from a real-time vehicle location feed and gtfs," J. Transp. Geogr. (2017).
2. B. P. Ashwini, R. Sumathi, and H. S. Sudhira, "Analyzing travel time reliability of a bus route in a limited data set scenario: A case study," Int. J. Intell. Syst. Appl. Eng. **11**, 30 (2023).
3. J. Wang and L. Sun, "Reducing bus bunching with asynchronous multi-agent reinforcement learning," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, (2021), pp. 426–433.

A GPS-Based Framework to Distinguish Road-Network Congestion from Bus Operation in Niterói, Brazil

ALEXANDER BOUSMAN[1,2,*]

[1] *Washington and Lee University, 204 W Washington St, Lexington, Virginia 24450, USA*
[2] *DIS Study Abroad in Scandinavia, Vestergade 7, 1456 Copenhagen, Denmark*
[*] *abousman@mail.wlu.edu*

1. INTRODUCTION

Public bus GPS data can act as a citywide traffic probe, but slow bus movement does not always imply road-network congestion. Buses stop for passengers, wait at terminals, queue behind other buses, and slow down for operational reasons. Prior work has shown that transit vehicles can support real-time traffic monitoring [1], but bus trajectories remain operationally biased because the vehicles belong to a scheduled transit system. This project addresses that ambiguity by separating likely road-network slowdowns from transit-operational causes while identifying route vulnerability to congestion. Using the NetMob26 mobility dataset for Niterói, Brazil [2], we develop a GPS-based framework combining map-matched trajectories, rolling speed, General Transit Feed Specification (GTFS) context, route information, and spatiotemporal clustering. The framework identifies candidates for road-network congestion and quantifies route exposure to congestion events.

2. DATA AND METHODOLOGY

Let each bus GPS observation be indexed as i, with projected coordinates $\mathbf{x}_i$, timestamp t_i, bus identifier b_i, and observed speed v_i. To align GPS observations with the street network, each point is assigned to its nearest road-network edge:

$$\epsilon_i^\star = \arg\min_{\epsilon \in E} d(\mathbf{x}_i, \epsilon), \tag{1}$$

where E is the set of OpenStreetMap (OSM) road-network edges and $d(\mathbf{x}_i, \epsilon)$ is the point-to-edge distance [3, 4].

Speeds are smoothed within each bus trajectory using a time-based rolling mean,

$$\tilde{v}_i = \frac{1}{|W_i|} \sum_{j \in W_i} v_j, \quad W_i = \{j : b_j = b_i, |t_j - t_i| \le \Delta t_{\text{roll}}\}. \tag{2}$$

For the NetMob26 mobility dataset [2], a 60 second window was experimentally determined for the dataset's ~15 second reporting interval, reducing GPS-speed noise without erasing minute-scale slowdowns. For each edge, a local reference free-flow speed is then defined as the 85th percentile of smoothed speeds above the stopped-vehicle threshold $v_{\min} = 1$ km/h [5]:

$$v_{\text{ff}} = Q_{0.85}(\{\tilde{v}_i : \tilde{v}_i > v_{\min}\}), \qquad s_i = \max\left(0, 1 - \frac{\tilde{v}_i}{v_{\text{ff}}}\right). \tag{3}$$

Values near 0 indicate free-flow movement, while values near 1 indicate stopped or near-stopped movement. To filter for slow-speed scores of interest, a minimum slow-speed score threshold, denoted by σ_s, is established experimentally and used as a baseline for later comparison.

To distinguish bus operation from road-network slowdowns, each slow observation is compared with GTFS stop and terminal locations. Let d_i^{stop} and d_i^{term} be the corresponding nearest distances. For $s_i \ge \sigma_s$, the point-level label is

$$L_i = \begin{cases} \text{road-network}, & d_i^{\text{stop}} > \delta_{\text{stop}} \wedge d_i^{\text{term}} > \delta_{\text{term}}, \\ \text{stop/terminal}, & d_i^{\text{stop}} \le \delta_{\text{stop}} \vee d_i^{\text{term}} \le \delta_{\text{term}}. \end{cases} \tag{4}$$

We set δ_{stop} and δ_{term} to represent curbside-stop and terminal footprints. These cutoffs indicate plausibility as opposed to causality, since queues can reach stops or delay can occur elsewhere. Slow observations are then grouped into ~15 second DBSCAN cluster-frames per NetMob26 dataset resolution [2]. The DBSCAN parameters include a radius $\varepsilon = 2\delta_{\text{stop}}$ meters and a three-point minimum, and it links adjacent-frame clusters within $2\delta_{\text{term}}$ meters to follow the cutoff distances δ_{stop} and δ_{term} for labeling. Mixed cases are retained as short-cluster counts and boundaries change rather than being forced into one class.

3. CLUSTERING ANALYSIS

We demonstrate the framework with a morning rush-hour case study in Centro, Niterói's downtown bairro, from 07:30–08:30 AM (See Fig. 1). After snapping observations to the Centro road network and adding GTFS context, 16,824 observations were retained and matched to GTFS routes. The rolling free-flow speed was estimated from these data at 19.70 km/h. Using an experimental threshold $\sigma_s = 0.48$, it identified 11,670 slow observations, including 3,380 stop/terminal dwell observations and 8,290 non-dwell road-network candidates.

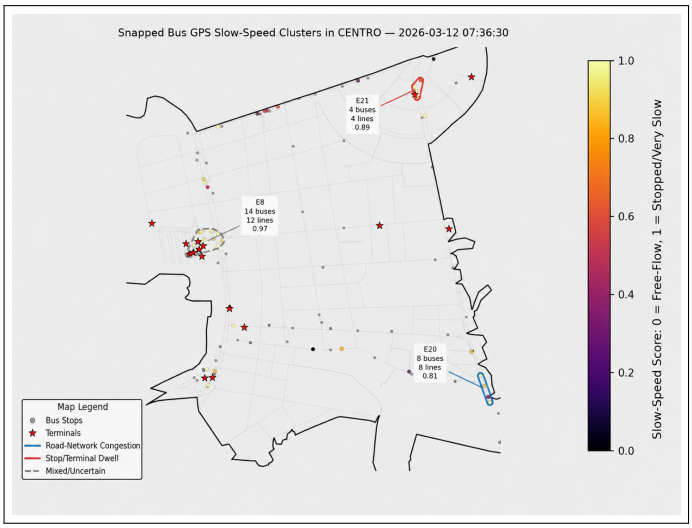


Fig. 1. A snapshot of morning rush-hour traffic in Centro. Cluster categories are identified in the road-network along with their event id and mean slow-speed score.

Slow observations formed 634 road-network clusters, 179 stop/terminal clusters, and 94 mixed/uncertain clusters. The mixed cases suggest that stop activity, terminal proximity, and

road congestion often overlap spatially in Centro, as expected of a dense, active downtown.

Table 1. Clustering Event Results for Centro, March 12, 2026

Event Id	Duration (Min)	Max Buses	Max Lines	Slow Score
31	33.50	19	15	0.934
73	25.25	23	19	0.917
0	6.75	10	8	0.921
95	5.75	4	4	0.975
20	3.75	9	9	0.917

The road-network clusters showed high slow-speed scores and involvement from multiple buses and lines, supporting a shared roadway explanation rather than isolated vehicle behavior. From there, Table 1 shows that clustering events also identify persistent multi-route disruptions involving many buses at once, with Event 31 lasting 33.5 minutes and Event 73 including up to 23 buses across 19 lines.

4. ROUTE VULNERABILITY

Route vulnerability, per our framework, is the extent to which a route is disproportionately represented in road-network clusters relative to its overall contribution to slow bus observations after filtering out stop and terminal activity. For route r, let C_r and S_r denote its road-cluster and all-slow observations, and let C and S denote the corresponding city totals. We compute

$$V_r = \frac{C_r/C}{S_r/S}, \tag{5}$$

so ratios above 1 are considered "over-represented." Normalization prevents frequently sampled routes from appearing vulnerable merely because they contribute more GPS points.

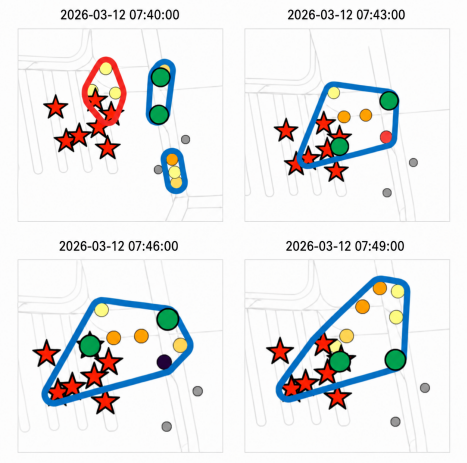


Fig. 2. Three-minute snapshots from Centro showing two OC3 buses, colored in green, with high route vulnerability since they remain within the same congestion event for an extended period. Map legend as in Fig. 1.

Several TransOceanico routes are strongly over-represented in road-network clusters, such as OC3 in Fig. 2. Across the hour-long case study, 78.6% of OC3 slow observations were located inside road-network clusters, producing a vulnerability ratio of 2.30: its road-cluster share was therefore 2.30 times its baseline share of all slow activity. Other highly exposed routes included lines 57, 46, 40, and 32, all with vulnerability ratios above 1.5 as shown in Fig. 3.

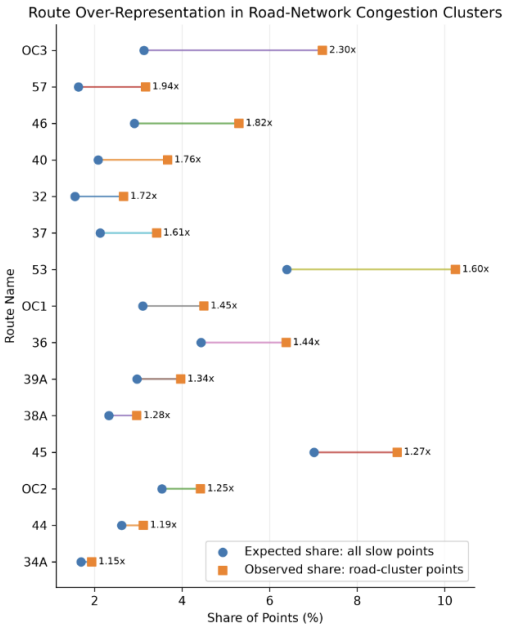


Fig. 3. A dumbbell chart depicting over-represented routes with vulnerability ratios above 1 in the city.

5. CONCLUSION

In this paper, we show that bus GPS data can identify plausible urban congestion while accounting for operational bias. In the Centro case study, map-matched trajectories, rolling speed, GTFS context, route information, and spatiotemporal clustering distinguish operational, road-network, and mixed slowdowns. The framework is transferable to other cities with GPS, GTFS, and OSM data, though parameter tuning for many metrics remains a key limitation. Edge matching may also fail on parallel streets because it ignores heading and trajectory continuity. Future work should use trajectory-aware matching and validate detected events against sensors, incident records, or external travel times.

Data Availability

Data underlying the results presented in this paper is provided by the NetMob26 Data Challenge, for which access can be requested per the Data Challenge organizers, Ref. [2].

REFERENCES

1. S. Jiang, Y. Sun, W. Wong, *et al.*, "Real-time urban traffic monitoring using transit buses as probes," Transp. Res. Rec. J. Transp. Res. Board pp. 219–236 (2024).
2. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," arXiv preprint pp. 1–39 (2026).
3. P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, (Association for Computing Machinery, New York, NY, USA, 2009), GIS '09, pp. 336–343.
4. G. Boeing, "Osmnx: A python package to work with graph-theoretic openstreetmap street networks," J. Open Source Softw. **2**, 215 (2017).
5. Federal Highway Administration, "The Urban Congestion Report (UCR): Documentation and Definitions," https://ops.fhwa.dot.gov/perf_measurement/ucr/documentation.htm.

The Impact of Precipitation on Public Transportation in Niterói, Brazil

LUCAS TOSHIOKA TENÓRIO[1,*], FERNANDA YUMI ASSONUMA[2], PEDRO HENRIQUE CAMPOS RIBEIRO[3], ALAN UEMA TAVARES[4], AND MATHEUS TEIXEIRA PASSOS GIAROLA[5]

[1,2,3,4,5] *Institute of Mathematics and Statistics, University of São Paulo, Rua do Matão 1010, 05508-090 São Paulo, SP, Brazil*
[*] *lucastoshioka@usp.br*

1. INTRODUCTION

Located on the eastern shore of Guanabara Bay, Niterói is one of the most prominent municipalities in the Metropolitan Region of Rio de Janeiro. The spatial mismatch between housing supply and employment concentration in the metropolitan area generates intense daily commuting flows between peripheral municipalities and the capital [1, 2]. By leveraging heterogeneous data sources [3], our research centers on Data Fusion, integrating meteorological, mobility, and ticketing data to uncover their interrelations.

Research Question: Based on this multi-layered framework, this study addresses a critical urban mobility challenge: *How does heavy precipitation affect the travel times of bus routes in Niterói, and to what extent are specific bus routes and the areas they serve disproportionately penalized by these weather-induced disruptions?*

2. DATASET AND DATA FUSION METHODOLOGY

This research is structured around two main interconnected axes: a **Demand Study** and a **Fleet Study**. The Demand Study focuses on passenger behavior and the spatial distribution of transit users, while the Fleet Study evaluates vehicle trajectories and systemic bottlenecks. Finally, a machine learning model is constructed to predict future delays.

Since the meteorological data were recorded at an hourly resolution, the ticket transactions and mobility data were temporally aggregated and merged to align both datasets by the hour (converted to BRT, UTC-3).

For the Demand Study, given that the initial aggregate results were inconclusive, the dataset was further stratified first by day of the week, and subsequently by geographic location, allowing us to isolate spatial and temporal effects. Due to the substantial imbalance between weekday (5,928,502) and weekend (862,179) sample sizes, this study focuses strictly on weekdays to avoid potential biases introduced by this disparity.

3. EXPLORATORY DATA ANALYSIS

The correlation coefficient calculated for the aggregate passenger data was negligible (-0.05). Based on a threshold of 2.5 millimeters, recognized by the World Meteorological Organization [4] as the minimum rainfall required to classify the event as moderate, the p-value derived from the Welch's t-test comparing samples with and without rain was not statistically significant at the 95% confidence level.

Consequently, these aggregate findings necessitated a deeper analysis. Passenger counts were first normalized using each route's daily mean volume, after which a route-specific Welch's t-test [1] was conducted under the same 2.5 mm rainfall threshold. As anticipated, specific routes exhibited statistically significant impacts at the 95% confidence level.

[1]Applying Bonferroni's correction would have severely reduced statistical power, however the consistently high Cohen's d values across these results suggest that the effects found are unlikely to be purely artifacts of noise.

The spatial distribution of precipitation impacts across Niterói is presented in Figure 1. The routes are represented on the map alongside favela regions (historically classified by the IBGE as Subnormal Agglomerations [5]). Notably, all affected routes either converge in or traverse the northwestern sector of the city. Evidence indicates that this localized phenomenon is not coincidental and requires deeper investigation, as discussed in Section 4.

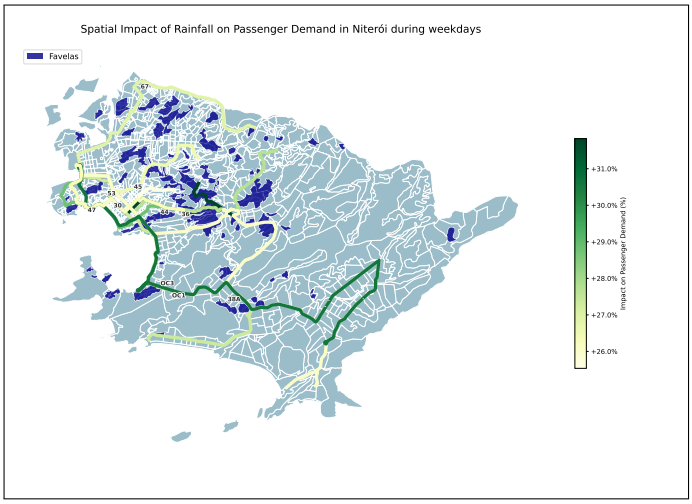


Fig. 1. Geospatial distribution of rainfall impact on public transit routes in Niterói during weekdays (95% confidence level).

To evaluate the impact of precipitation on public bus lines in Niterói, a Mann-Whitney U test[2] was conducted to assess differences in trip duration between rainy and non-rainy days. This analysis also applies the 2.5 mm threshold [4]. The results provide sufficient evidence to conclude that a statistically significant difference exists between these groups.

Furthering this analysis, we quantified the delay for each route and visualized the corresponding spatial densities in Figure 2. Consistent with earlier findings, the maps reveal a distinct concentration of delays in northwestern Niterói, accompanied by a substantial increase in accumulated delay density whenever the 2.5 mm precipitation threshold is exceeded.

4. DISCUSSION AND POTENTIAL IMPACT

The results in Section 3 reveal a distinct spatial concentration of rainfall impacts across the bay-facing areas of Niterói. Ferreira [6] identifies this area as densely populated and, in parts, characterized by lower per capita income, resulting in medium coastal vulnerability scores. This socio-economic and drainage-related vulnerability likely extends to urban mobility, providing a plausible explanation for the concentrated transit delays observed in

[2]The p-values of this test were corrected using the Bonferroni method.

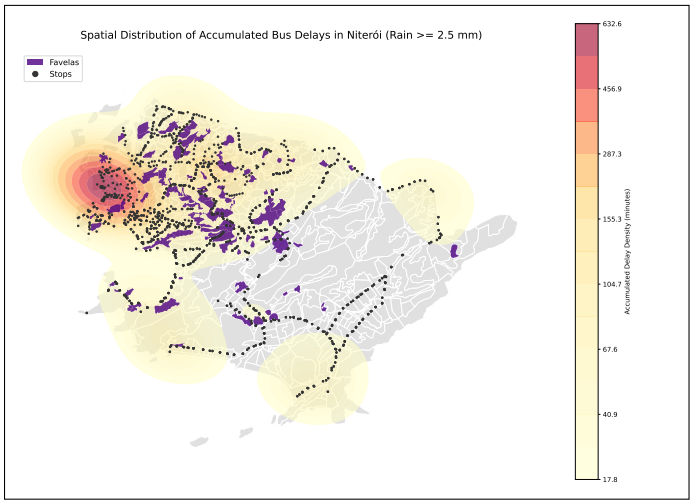


Fig. 2. Spatial density of accumulated delays for precipitation exceeding the 2.5 mm threshold.

our data. However, definitively verifying this hypothesis falls outside the current scope and requires dedicated future research.

Based on the observed correlation between rainfall and transit disruptions, a predictive model was developed and detailed in Section 5 to forecast expected durations for individual trips.

5. PREDICTIVE MODEL

The predictive model was configured with a multidimensional feature set designed to capture the complex dynamics of the transit system using XGBoost [7]:

- **Input Features:**
  - *Meteorological:* Accumulated precipitation (mm).
  - *Temporal:* Time of day and day of the week.
  - *Spatial and Operational:* Line ID, route direction, and geospatial coordinates (start and end latitude/longitude).
- **Target Variable:** Total trip duration (in minutes).

The model was trained using a histogram-based tree algorithm. The final model achieved a Mean Absolute Error (MAE) of approximately 5.5 minutes and an R^2 score of 0.84. When compared to the descriptive statistics of trip duration in the sample, this MAE corresponds to approximately 11% of the mean (~48 minutes) and 23% of the standard deviation (~23 minutes), indicating that the model's typical error is substantially smaller than the natural variability of the data — a result consistent with the high R^2 obtained.

Despite these robust overall performance metrics, the model still exhibits mispredictions, particularly for trips with exceptionally long durations. Figure 3 illustrates this limitation, showing an increase in outliers as the actual trip duration grows.

Finally, although earlier analyses demonstrated a correlation between rain and delays, meteorological variables yielded the lowest feature importance scores in the predictive model. This discrepancy is likely attributable to severe class imbalance, as the dataset contains remarkably few instances recorded under rainy conditions compared to dry ones. Conversely, spatial and operational factors (e.g., Line ID and direction) emerged as the primary determinants for the model's predictions. This aligns with the prior findings, which highlighted distinct delay concentrations depending on the geographic region of the city.

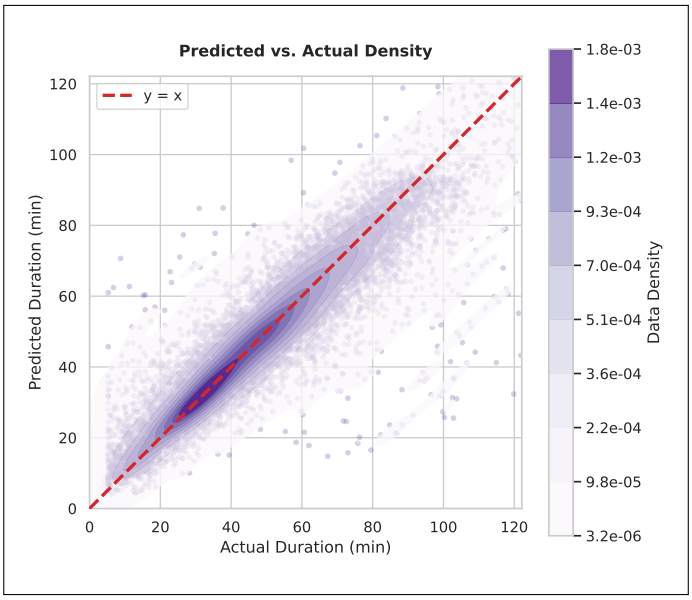


Fig. 3. Comparison between predicted and actual trip durations (in minutes).

6. CONCLUSION

In summary, the two main analyses in this article emphasize the impact of weather conditions on public transportation dynamics. Section 3 demonstrates how rainfall affects specific transit lines, altering passenger demand and primarily increasing delays. These lines might be particularly sensitive to weather changes due to population density and geographic conditions—especially those traversing northwestern Niterói, as discussed in Section 4. Finally, while the developed predictive model achieved satisfactory performance, it struggles to accurately estimate instances with extremely long durations. Furthermore, owing to data quality limitations and a scarcity of rainy samples in the dataset [3], weather variables demonstrated limited influence on the model's predictions. Addressing this data imbalance to enhance model robustness remains a key objective for future work, which will require extending the observation window to capture a higher frequency of precipitation events.

REFERENCES

1. L. C. d. Lago, "Trabalho, moradia e (i)mobilidade espacial na metrópole do rio de janeiro," São Paulo em Perspect. **19**, 9 (2005).
2. U. C. S. Ferreira, P. d. M. Jannuzzi, and L. d. C. Giannella, "Mover-se na metrópole: movimentos pendulares na região metropolitana do rio de janeiro," Cadernos Metrópole **25**, 617 (2023).
3. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," (2026).
4. World Meteorological Organization, *Guide to Instruments and Methods of Observation*, World Meteorological Organization (WMO), Geneva (2021). Annex: Criteria for Slight, Moderate and Heavy Precipitation Intensity.
5. Instituto Brasileiro de Geografia e Estatística (IBGE), "Censo demográfico 2022," https://censo2022.ibge.gov.br/ (2023). Accessed: 2026-08-24.
6. V. L. D. Ferreira, E. S. Pereira, L. P. Souza de Mello, *et al.*, "Analysis of coastal environmental vulnerabilities in the municipality of niterói, rio de janeiro, brazil, in the face of sea level rise projections," Coasts **5**, 11 (2025).
7. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), pp. 785–794.

Assessing the Effects of Precipitation on Public Transportation Demand Through Eigenvector Analysis

GEYMERSON S. RAMOS[1,*], SACHIT MISHRA[1], AND RAZVAN STANICA[1]

[1] Inria, INSA Lyon, CITI, EA3720, 69621 Villeurbanne, France
[*] geymerson.ramos@inria.fr

1. GENERAL PROBLEM AND MOTIVATION

Public transportation demand varies with the time of day, day of the week, socioeconomic conditions, and weather. In Brazil, intense rainfall frequently causes flooding and disrupts road traffic and public transportation services. This study uses an eigendecomposition-based method to characterize dominant spatiotemporal demand patterns and quantify precipitation-associated differences across locations and hours. We address three research questions: (1) How can dominant public transportation demand patterns be quantified and represented? (2) Which areas exhibit the largest precipitation-associated differences in typical demand behavior? (3) Which time periods exhibit the largest deviations from typical demand behavior? Using the NetMob 2026 dataset [1], complementary point-of-interest (POI) data from OpenStreetMap[1], and meteorological data from Brazil's National Institute of Meteorology (*Instituto Nacional de Meteorologia*, INMET)[2], we apply an eigenvector-based decomposition method to characterize the dominant spatiotemporal patterns of public transportation demand and analyze their association with precipitation in Niterói, Brazil. Compared with direct hourly comparisons based on individual days, the dominant eigenvector approach is less sensitive to random and day-specific variations. Single-day demand may be influenced by factors unrelated to weather, such as holidays or traffic disruptions. Dominant eigenvectors mitigate these effects by capturing the underlying demand patterns across multiple days. Although previous studies have established that precipitation is associated with reductions and spatially heterogeneous changes in public transportation use, they have not identified the dominant coupled spatiotemporal modes through which system-wide demand varies under precipitation. The main strength of our approach is its applicability to large-scale, fine-grained transaction data, enabling the joint analysis of spatial and temporal variations in transportation demand. The eigenvectors used in this analysis are obtained through Principal Component Analysis (PCA) [2], following the general pipeline illustrated in Fig. 1.

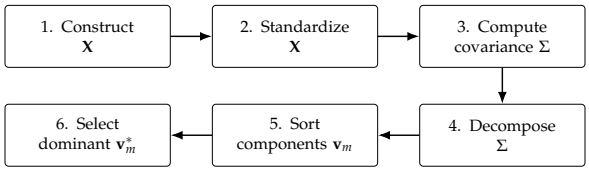


Fig. 1. Eigendecomposition pipeline for extracting the dominant component $\mathbf{v}_m^*$.

The data matrix $\mathbf{X} \in \mathbb{R}^{d\times h}$ contains passenger boarding records aggregated by hour. Each row, $\mathbf{x}^{(i)} = (x_{i,1}, x_{i,2}, \ldots, x_{i,h})$, represents the hourly boarding profile observed on day i, where d denotes the number of days (observations) and h denotes the number of hourly time points considered (features). After standardizing $\mathbf{X}$, we compute the covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{h\times h}$, which captures the dependence structure of boarding volumes across different hours. We then compute the eigenvalues and eigenvectors of $\mathbf{\Sigma}$ and identify the principal component associated with the largest eigenvalue. Its corresponding eigenvector, denoted by $\mathbf{v}_m^*$, represents the dominant direction of variation in the data. By labeling days according to precipitation conditions, we partition $\mathbf{X}$ into the submatrices $\mathbf{X}_\eta$ and $\mathbf{X}_\xi$, which contain daily passenger-boarding profiles under favorable and adverse weather conditions, respectively. We then perform an eigendecomposition separately on $\mathbf{X}_\eta$ and $\mathbf{X}_\xi$. This procedure yields the dominant eigenvectors $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$, which characterize the principal temporal modes of variation in passenger boardings under the two weather conditions. Because both submatrices represent the same type of daily temporal profile, $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$ are expected to be similar when precipitation is not associated with a substantial change in passenger boarding behavior.

The angle between $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$, defined as $\theta_{\eta,\xi}^* = \frac{180}{\pi}\arccos\left(\left|\frac{\mathbf{v}_\eta^*\cdot\mathbf{v}_\xi^*}{\|\mathbf{v}_\eta^*\|\,\|\mathbf{v}_\xi^*\|}\right|\right)$, quantifies the difference between the dominant temporal patterns of passenger boardings under favorable and adverse weather conditions by measuring the angular separation between their corresponding eigenvector directions. A change in the orientation of these vectors indicates a change in the dominant direction of variation in the data and, consequently, in the temporal structure of passenger boardings. The angle $\theta_{\eta,\xi}^*$ ranges from 0°, indicating identical dominant directions, to 90°, indicating orthogonal dominant directions. These two extremes correspond, respectively, to the minimum and maximum angular dissimilarity between the dominant boarding patterns associated with the weather conditions under consideration. Because $\mathbf{v}$ and $-\mathbf{v}$ represent the same principal direction, the absolute value of the inner product is used within the $\arccos(\cdot)$ function to ensure invariance to the arbitrary sign of the eigenvectors. To assess hour-specific differences in passenger-boarding patterns across weather conditions, we compute the element-wise absolute difference between the dominant components, defined as $\Delta\mathbf{v}^* = \left||\mathbf{v}_\eta^*| - |\mathbf{v}_\xi^*|\right|$., where $|\cdot|$ denotes the element-wise absolute value. Accordingly, for $\mathbf{v}_\eta^*, \mathbf{v}_\xi^* \in \mathbb{R}^h$, the j-th element of $\Delta\mathbf{v}^*$ is given by $\Delta v_j^* = \left|\left|v_{\eta,j}^*\right| - \left|v_{\xi,j}^*\right|\right|$, $j = 1, \ldots, h$. If $\Delta v_j^* = 0$ for all j, then $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$ have identical element-wise magnitudes, indicating that the magnitude structure of the dominant temporal mode remains unchanged between the two weather conditions. In this case, no hour-specific difference is detected in the magnitude of the dominant component.

2. RESULTS

Fig. 2 presents the hourly number of passenger-boarding transactions in Niterói for each day of the week. A clear distinction can be observed between weekday and weekend mobility patterns. From Monday to Friday, boarding volumes remain consistently

[1] https://overpass-turbo.eu/
[2] https://bdmep.inmet.gov.br/

higher throughout most of the day, which is consistent with greater demand associated with work, education, and access to services. Weekdays also exhibit a trimodal temporal pattern, with pronounced peaks highlighted by the shaded bands at 5:00, 9:00, and 14:00.

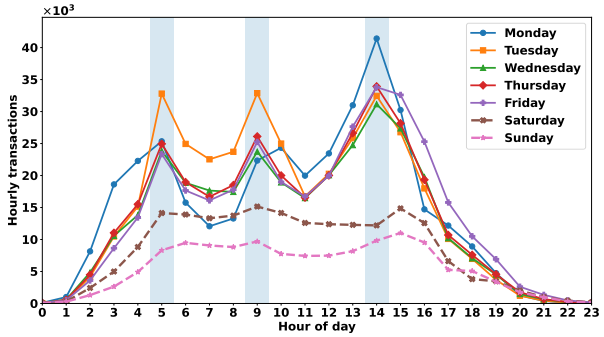


Fig. 2. Total hourly number of passenger-boarding transactions aggregated across the city of Niterói for each day of the week during the period from March 1 to March 31, 2026. The shaded bands indicate peak travel periods centered at 05:00, 09:00, and 14:00.

According to INMET, precipitation occurred on 14 days in March 2026 in Niterói. By applying eigendecomposition to the transaction data organized into the matrices $\mathbf{X}_\eta$ (boardings transactions under nominal weather) and $\mathbf{X}_\xi$ (precipitation conditions), we obtain and analyze the dominant eigenvectors $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$ for different areas of the city. Fig. 3 illustrates the dominant eigenvectors of the sector containing the Barcas Terminal, a key hub that presented the highest number of boarding events (189729) in March 2026. On the y-axis, the loading v_j associated with hour j indicates the contribution of that hour to the dominant principal direction. Because the sign of an eigenvector is arbitrary, larger absolute loading values, $|v_j|$, indicate a stronger contribution of the corresponding hour to the dominant temporal pattern of passenger boardings in that sector.

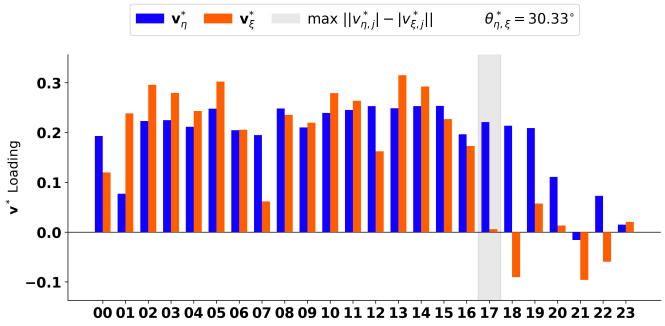


Fig. 3. Dominant eigenvectors representing the dominant components of passenger boardings under favorable-weather conditions, $\mathbf{v}_\eta^*$, and precipitation conditions, $\mathbf{v}_\xi^*$, at the sector containing the Barcas Terminal hub.

The largest difference in the relative contribution of individual hours occurs at 17:00, when the absolute loading under favorable-weather conditions is markedly higher than that observed under precipitation conditions. A clear divergence between the dominant eigenvectors $\mathbf{v}_\eta^*$ and $\mathbf{v}_\xi^*$ persists throughout the evening and nighttime period, from 17:00 to 23:00. This pattern indicates that the dominant temporal structures of passenger boardings under the two weather conditions differ substantially during these hours. In particular, the lower absolute loadings under precipitation conditions suggest that late-afternoon and evening boardings contribute less strongly to the dominant temporal mode on rainy days. This difference may reflect precipitation-related changes in travel behavior, including the cancellation or substitution of trips. The spatial distribution shown in Fig. 4 illustrates the angular dissimilarity between the dominant temporal patterns of passenger boardings under favorable and adverse weather conditions across sectors in Niterói. Sectors with larger values of $\theta_{\eta,\xi}^*$, represented by darker shades on the map, exhibit greater deviations between the corresponding dominant eigenvector directions. Conversely, sectors with smaller values of $\theta_{\eta,\xi}^*$, represented by lighter shades, exhibit more similar dominant temporal patterns across the two weather conditions. Gray areas indicate sectors for which no passenger-boarding transaction data are available.

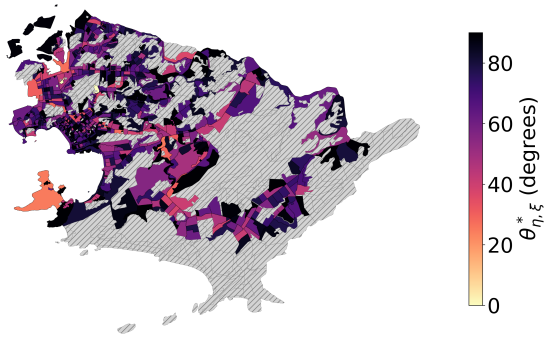


Fig. 4. Spatial distribution of $\theta_{\eta,\xi}^*$ values, illustrating the extent to which each area is affected by precipitation.

Using the proportion p_n of non-residential POIs in each sector, defined as $p_n = \frac{N_n}{N_r+N_n}$, where N_n and N_r denote the numbers of non-residential and residential POIs, respectively, we assess the association between land-use composition and precipitation-related changes using Pearson's correlation coefficient. The results show a weak negative correlation between the proportion of non-residential POIs and $\theta_{\eta,\xi}^*$ ($r = -0.202$, $p < 0.001$). Thus, sectors with higher proportions of non-residential POIs tend to exhibit smaller angular differences between their dominant temporal patterns under favorable and precipitation conditions.

3. CONCLUSION AND FUTURE WORK

Using public transportation demand data and the dominant eigenvectors of different areas of Niterói, we identified sectors exhibiting different degrees of variation in passenger-boarding patterns between rainy and non-rainy conditions. Sectors with higher proportions of non-residential points of interest appear less sensitive to precipitation-related changes in passenger boardings, a pattern that may be associated with the concentration of routine or essential activities that are less likely to be postponed or canceled. Future work will extend the spatial analysis to identify and classify areas with similar passenger-boarding demand patterns.

REFERENCES

1. F. Domingos, H. T. Marques-Neto, B. Pereira, *et al.*, "The netmob26 dataset: A high-resolution multi-source view of public bus mobility in niterói," (2026).
2. M. Greenacre, P. J. Groenen, T. Hastie, *et al.*, "Principal component analysis," Nat. Rev. Methods Primers **2**, 100 (2022).

Spatial Analysis of Operational Discrepancies in Niterói's Municipal Bus System

MATHEUS DE SOUZA E SILVA[1,*], MARINA LEITE DE BARROS BALTAR[1], LUIS EDUARDO SANTOS FORTES[1,2], ANDERSON MARDEN DE SOUSA SILVA[1], GLAYDSTON MATTOS RIBEIRO[1], CINTIA MACHADO DE OLIVEIRA[2], DANIEL KIM IMAI[1], LUCAS RIBEIRO SAMPAIO[1], AND BRUNO PEÇANHA SIMÕES[1]

[1] *Transportation Engineering Program, Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering (COPPE), Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil*
[2] *Department of Production Engineering, Federal Center for Technological Education Celso Suckow da Fonseca (CEFET/RJ), Itaguaí, RJ, Brazil*
[*] *matheus.souza@poli.ufrj.br*

1. INTRODUCTION

Travel time is one of the determining factors in a user's decision to use public transit. The disparity between scheduled travel times and those observed in real-world operations often exhibits mismatches that penalize passengers and reduce the competitiveness of buses relative to private transportation. The increasing availability of high-resolution operational data, such as electronic ticketing, vehicle telemetry (GPS), and planned service specifications (GTFS), enables a systematic comparison between the service provided and the planned one, identifying routes and corridors in which this mismatch is most pronounced. This study aims to identify, among the highest-demand bus routes, those with the greatest travel time discrepancies in Niterói, adopting deviations exceeding 10 minutes [1, 2] during peak hours as the evaluation threshold, while also assessing the spatial coverage of these routes and the potential exposure of the population to these delays, using data provided by the MobNit platform [3], consolidated in the NetMob26 dataset [4], alongside data from the IBGE census [5] and municipal land use [6].

2. METHODOLOGY

The methodology of this study is divided into four stages. The first stage consisted of a sequential data engineering pipeline to consolidate the sources. The electronic ticketing data were aggregated into route, date, and hour keys, deriving fare profile variables (paying passengers, fare-exempt riders, and integrated transfers). In parallel, the raw telemetry data underwent spatial filtering (via a *bounding box* of the municipality of Niterói) and the removal of duplicate readings, and were converted into indicators for trips and vehicles observed, aggregated by route and direction. The planned service was extracted from GTFS files, subjected to deduplication processes, and matched with the GPS data by truncating the direction suffix in the *trip_id* field. Finally, the three operational components were merged into a unified temporal key, preserving a traceability matrix for the manual equivalences applied to the original route numbers.

The second stage aimed to identify, through electronic ticketing analysis, the bus routes with the highest passenger volume during peak hours (from 6:00 a.m. to 10:00 a.m. and from 4:00 p.m. to 8:00 p.m.). It should be noted that, although the data processed in the first stage identified 55 bus routes in operation, only 47 of them had ticketing records available. Boarding records were grouped by day and one-hour intervals, allowing the calculation of the total passengers transported in each analyzed period. Subsequently, demand concentration was analyzed using a cumulative load curve. The resulting distribution revealed that a small number of routes concentrate most of the passengers, allowing the selection of the top 20 municipal bus routes.

The third stage was responsible for diagnosing the temporal performance of the operation based on the mismatch between scheduled (GTFS) and observed (GPS) travel times. The unit of analysis corresponded to a trip identified by the combination of date, vehicle, route, direction, and *trip_id*, enabling the mapping of the most critical times, routes, and regions of the network.

To reduce the influence of incomplete trips, only observations with at least ten GPS readings, a minimum duration of five minutes, and a maximum difference of 30 minutes between the observed and scheduled start times were considered valid. After applying these criteria, the analytical database included 25,734 valid trips during peak hours. For each trip, the absolute difference (Equation 1) and percentage difference (Equation 2) of the travel times were calculated, selecting routes with an average delay exceeding 10 minutes for spatial detailing, where positive values indicate operational delays:

$$\Delta T_i = T_i^{GPS} - T_i^{GTFS} \tag{1}$$

$$\Delta T_i^{\%} = \frac{T_i^{GPS} - T_i^{GTFS}}{T_i^{GTFS}} \times 100 \tag{2}$$

Finally, the fourth stage consisted of analyzing the spatial and social distribution of the impacts. The 2022 census dataset [5] was refined based on the municipality's land use classifications [6], removing uninhabitable areas and obtaining a more realistic population distribution across the territory. Considering a 300-meter coverage radius [7] for bus route accessibility enabled the analysis of the operational coverage efficiency of each route, assessing which population groups are potentially most exposed to delays and public transit unpredictability.

3. RESULTS

Figure 1 presents the demand distribution of the top 20 municipal bus routes obtained from the analyzed passenger volume. These routes represent a cumulative demand accounting for approximately 83% of the total demand obtained from the electronic ticketing database.

Based on this subset, Table 1 presents a comparison between scheduled (GTFS) and observed (GPS) average travel times for routes that registered average delays exceeding 10 minutes.

It is worth noting that certain high-demand services are not represented, namely routes 30 and 47. Although these routes represent the highest passenger volume in Niterói, they were excluded from the final selection due to the inconsistency of their GPS records. This limitation highlights a common challenge in public transit big data, where localized transmission failures, hardware anomalies, or signal dropouts can generate critical telemetry gaps in high-density corridors [2].

Nevertheless, spatial and demographic indicators can be utilized to infer their operational impact. Given that 12 of the 18

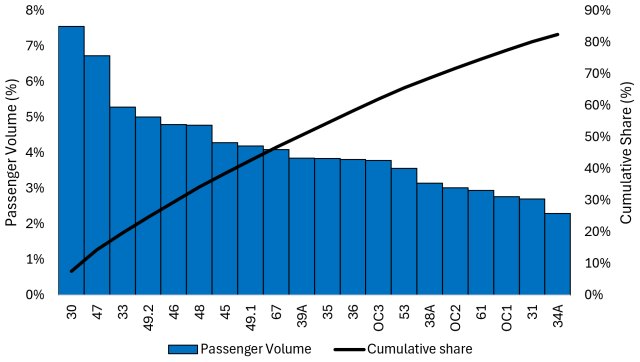


Fig. 1. Passenger volume by route during peak hours

Table 1. Travel time delay by route (GTFS vs. GPS)

Route	GTFS (min)	GPS (min)	ΔT (min)	ΔT (%)
31	48,0	77,7	29,7	62
53	35,9	58,9	23,0	64
46	79,4	100,5	21,1	27
OC3	77,2	96,8	19,6	25
33	39,8	56,2	16,4	41
OC1	58,7	72,3	13,6	23
67	44,9	57,9	13,1	29
35	56,9	69,7	12,8	22
61	63,5	76,2	12,6	20
OC2	67,8	80,4	12,6	19
39A	65,9	77,4	11,5	17
45	37,6	48,9	11,3	30

analyzed routes experienced delays exceeding 10 minutes, and considering that high-demand corridors are highly susceptible to dwell-time variability and traffic congestion [1], it is highly probable that routes 30 and 47 also suffer from operational mismatches. Consequently, the system's actual unpredictability is expected to be broader than what is shown by the sample.

The social dimension of operational impacts is observed in Figure 2, which overlays the 300-meter coverage of the 12 critical routes onto the income classes of the served sectors. The mapping enables the identification of how operational mismatch is distributed across the territory, affecting populations with different household income profiles, classified according to the adapted IBGE income criteria [8]. Although routes 30 and 47 were excluded from the delay calculations, their spatial corridors are plotted on the map to confirm that the lack of telemetry does not negate the socio-spatial profile of the affected areas. The predominance of low and middle-income sectors (classes D/E and C) within the coverage areas of the critical routes confirms that the failure to meet planned operations penalizes the social groups most vulnerable to service unpredictability.

4. CONCLUSIONS

This study identified, among the main high-demand bus routes in Niterói, 12 routes where the mismatch between scheduled and observed travel times resulted in delays exceeding 10 minutes (a value associated in the literature with losses in service reliability [1, 2]). The spatial distribution of these routes covers different regions of the municipality and coincides with the network's main structural axes, demonstrating that operational unpredictability directly impacts the corridors most essential to urban mobility. Although other criteria for evaluating the level of service are important to refine the diagnosis, complementary analyses require more robust or complete datasets. Therefore, the results reinforce the urgency of roadway prioritization policies targeted at these critical corridors and point to the need for expanding telemetry data coverage and consistency in future studies.

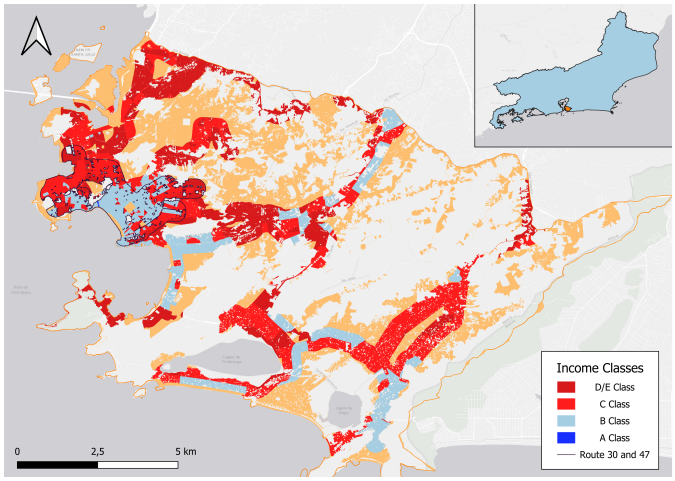


Fig. 2. Census tracts served by critical routes and income levels

REFERENCES

1. National Academies of Sciences, Engineering, and Medicine, *Transit Capacity and Quality of Service Manual, 4th Edition* (The National Academies Press, 2026). doi:10.17226/29449.
2. E. Graves, S. Zheng, L. Tarte, et al., "Customer Journey Time Metrics for New York City Bus Service using Big Data," *Transportation Research Record* **2673**(2), 1–10 (2019). doi:10.1177/0361198118821632.
3. Niterói Municipal Government, "MobNit – Urban Mobility Management System of Niterói," available at https://mobnit.niteroi.rj.gov.br/ (accessed July 1, 2026).
4. F. Domingos, H. T. Marques-Neto, B. Pereira, et al., "The NetMob26 Dataset: A High-Resolution Multi-Source View of Public Bus Mobility in Niterói," *arXiv preprint arXiv:2605.20263* (2026).
5. Niterói Municipal Government and Brazilian Institute of Geography and Statistics (IBGE), "Demographic Census Data – Restricted Database under Confidentiality Agreement," internal restricted-access document, accessed July 7, 2026.
6. Niterói Municipal Government, "Land Use – Geospatial Layer of the Geographic Management System of Niterói (SIGeo)," available at https://www.sigeo.niteroi.rj.gov.br/datasets/geoniteroi::uso-do-solo/about (accessed July 9, 2026).
7. Z. D. Ebrahimi, M. Momenitabar, A. A. Nasri, et al., "Using a GIS-based spatial approach to determine the optimal locations of bikeshare stations: The case of Washington D.C.," *Transport Policy* **127**, 48–60 (2022).
8. Rio de Janeiro Municipal Government, "Income Classes in Rio: Special Study," Municipal Secretariat of Economic Development (SMDE), Rio de Janeiro, RJ, 2025.

Diagnosing Transit Performance from Passive Data: Reliability, Vulnerability and Boarding-Pressure in Niterói, Brazil

Swetha Cheemakurthi

SHIV NADAR UNIVERSITY, CHENNAI, INDIA

swetha.chsnu@gmail.com

1. INTRODUCTION

Diagnosing service failures typically requires occupancy sensors or rider surveys — costly and often unavailable at scale. Agencies running General Transit Feed Specification (GTFS) schedules, Automatic Vehicle Location (AVL), and Automated Fare Collection (AFC) systems already collect GPS traces, schedules, and fare logs, yet report only aggregate metrics (average headway, on-time performance, ridership) that cannot localize where service breaks down.

Prior work treats these failure modes separately or in pairs. Deepa et al. (2023) link headway variability to ridership loss, but at the route level. Candelieri et al. (2019) and Abdelaty et al. (2020) expose network vulnerabilities invisible to aggregate statistics, but neither tests persistence across schedule revisions. Kim and Song (2018) integrate accessibility and reliability (ACCREL) without incorporating crowding. Luo et al. (2018) combine AFC, AVL, and GTFS for space-time occupancy graphs, without benchmarking structural vulnerability or reliability. Yap et al. (2020) show fare data can support crowding valuation, yet this remains uncommon.

We are not aware of prior work combining reliability, structural vulnerability, and crowding in one passive-data pipeline, or formally testing whether they converge on the same routes.

This study makes no pillar-level methodological contribution — coefficient-of-variation (CV) reliability scoring, betweenness-centrality node removal, and threshold-based anomaly flagging are each standard. Its contribution is (a) integrating three standard diagnostics into one pipeline; (b) grounding design choices in sensitivity analysis rather than convention; and (c) formally testing, rather than asserting, convergence — itself stress-tested for selection bias and fragility.

RQ1: Does headway irregularity track a real deficit against the published schedule?
RQ2: Does connectivity depend on a small number of structurally critical stops, beyond random failure?
RQ3: Does boarding-pressure concentrate at identifiable peaks, and do reliability, vulnerability, and crowding co-locate on the same routes?

2. DATA

The analysis uses one month (March 2026) of operational data from Niterói's transit system: daily GPS vehicle traces (~15-second resolution) and daily fare collection logs (6.7 million transactions monthly), one CSV per day. GPS traces cover March 11–31 (19 days), excluding March 18–19, due to a telemetry outage before data collection; fare data covers the full month. Data spans two operators across three schedule snapshots (TransOceanico: 2, TransNit: 1), capturing a mid-month timetable revision. An auxiliary feeder/van dataset with a separate stop-ID system was excluded to avoid conflating it with the main GTFS network.

3. METHODS

GPS pings are snapped to the nearest stop via k-d tree (60 m radius, swept 30–120 m for sensitivity) and grouped by (stop, line, direction, hour), ≥3 observations/day.

Reliability. For each group,

$$CV = \sigma/\mu \quad (1)$$

Mean and CV are min-max normalized and combined into:

$$R = 1 - (0.5 \times \mu norm + 0.5 \times CVnorm) \quad (2)$$

Equal weighting is supported by near-identical correlation of R with each component alone (r = −0.634, −0.622) and a five-point Spearman sweep (ρ ≥ 0.87 near equal weighting). Observed headways were matched to GTFS-scheduled headways via a 30 m spatial stop crosswalk (54.9% match) and a tripId-to-route_id join (100% of 63 lines mapped unambiguously), to test whether irregularity reflects a real scheduling deficit.

Structural vulnerability. A stop-level graph from stop_times.txt, weighted by median travel time. Stops shared by multiple lines are represented as a single node, so betweenness reflects cross-route dependency rather than route-specific structure. Betweenness is recomputed after each iterative removal of the highest-betweenness node, benchmarked against 50-trial random removal (mean ± 1 SD), with connectivity loss (CL) measured as

$$CL\,(\%) = 100 \times [1 - reachable(G')/reachable(G_0)] \quad (3)$$

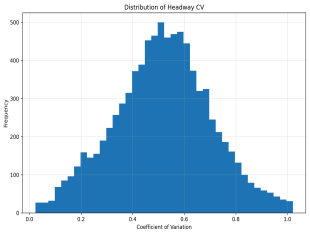
Boarding-pressure (not "overcrowding": no onboard load/capacity data exists). Boardings bin into 5-minute windows; a window is flagged if :

$$boardings > \mu + 2\sigma \quad (4)$$

(swept k ∈ {1.5, 2.0, 2.5, 3.0} for sensitivity). A route-specific threshold and a fleet-wide 95th-percentile threshold both reproduce the same diurnal pattern (Spearman ρ = 0.999, 0.997), and an independent IsolationForest confirms the same shape.

Cross-pillar convergence test (RQ3). Reliability, betweenness, and boarding-pressure were merged to route_id (route-name crosswalk: 82.6% match); 12 routes had coverage across all three pillars. Spearman correlations were computed among unreliability (1 − R), betweenness, and boarding-pressure rate, with two robustness checks: a Mann-Whitney U test on observation-volume (merged vs. excluded routes) and a full leave-one-out (LOO) sensitivity analysis.

4. RESULTS



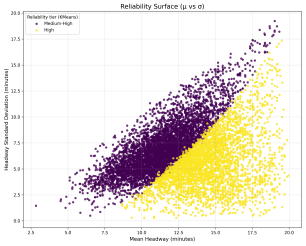


Fig. 1 (left) Headway CV distribution (n=8,636), clusters 0.3–0.7. Fig. 2 (right) Reliability surface (μ vs. σ) by KMeans tier; boundaries overlap (silhouette=0.359).

Headline finding (RQ3). No positive convergence. Betweenness is uncorrelated with unreliability (ρ = 0.007, p = 0.983) and with boarding-pressure (ρ = −0.364, p = 0.245 — underpowered, not clearly null, at n = 12). Unreliability and boarding-pressure are significantly, strongly *negatively* correlated (ρ = −0.699, p = 0.011): more reliable routes run more crowded — consistent with a trade-off, not the three failure modes reinforcing each other.

The result withstands three checks. Merged and excluded routes do not differ significantly in observation volume (Mann-Whitney U = 144.5, p = 0.835), ruling out a busiest-routes selection bias. The correlation stays negative and significant across all 12 leave-one-out deletions (ρ ∈ [−0.782, −0.609], all p < 0.05), including exclusion of the highest-volume route. It also survives Bonferroni correction for the three pairwise tests (p_adj = 0.033); the two null results stay non-significant under both Bonferroni and Holm. RQ3: not convergence — a trade-off, significant after correction

Reliability (RQ1). Headway CV clusters between 0.3 and 0.7 (mean 0.517, n = 8,636). On the 52.5% of the network with a matched schedule, service runs 2.7 -- 2.8× less frequently than GTFS implies (median ratio 2.72, IQR [2.28, 3.25]) despite weak point-level correlation (r = 0.164) — a fairly uniform under-frequency gap. Supported, with coverage caveat.

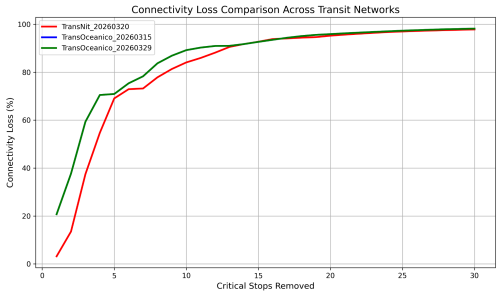


Figure 3. Connectivity loss under targeted removal, TransOceânico vs. TransNit

Structural vulnerability (RQ2). Targeted removal reaches ≥95% CL within 9 stops (TransOceânico, identical across both schedule snapshots) and 10 stops (TransNit), versus 46.7% and 29.4% under random removal — gaps of 49.0 and 67.2 points. Explained substantially by multi-route stop-sharing (82.0% / 48.5% of stops serve >1 route). Supported.

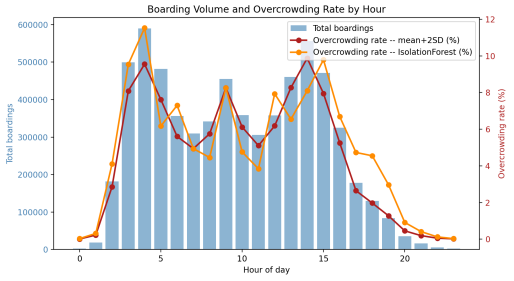


Figure 4. Boarding volume vs. overcrowding rate by hour: threshold method vs. IsolationForest.

Boarding-pressure. Peaks at 4:00 and 14:00 (9.5–9.9% flagged), near zero overnight; timezone and fleet-size artifacts were ruled out by re-deriving timestamps against each operator's UTC offset and normalizing flagged counts by concurrent fleet size, both of which left the two peaks unchanged. Tracking of volume is expected by construction — evidence against sparsity-driven crowding, not proof of demand-driven causation.

Parameter sensitivity. Mean CV moved only 0.523→0.535 across the 30–120 m snap-radius sweep — the reliability headline is not materially sensitive to this choice. The stop-crosswalk match rate, by contrast, moved 54.9%→64.2% over the same range: the 30 m tolerance trades coverage for match confidence and bounds RQ1's scheduling-deficit comparison to the matched network, rather than being inconsequential. For boarding-pressure, the flagged rate varies substantially with k (5.71% at k=1.5 to 2.36% at k=3.0, as expected), but the diurnal shape is stable (ρ ≥ 0.986 vs. the k=2.0 baseline; peak hours identical at every k tested). The paper's timing claim is robust; the rate claim was never made.

5. DISCUSSION AND LIMITATIONS

Each diagnostic is computable from data agencies already collect, and each surfaces a genuine finding — but the findings do not co-locate. A pipeline built to test unification instead shows the failure modes are statistically independent, or in the RQ3 case, in tension. This argues for route-specific interventions over a single composite risk score.

Machine-learning extensions (KMeans, two Random Forests, IsolationForest) were evaluated, not assumed: k = 2 was fixed a priori to test only a binary high/low reliability split relevant to intervention triage, not selected by silhouette optimization; KMeans tiering is coarse and overlapping (silhouette = 0.359); the Random Forest betweenness approximation matched exact removal for TransOceânico but failed for TransNit (52.7-point gap) and is not adopted.

The RQ3 result is drawn from only 12 of 63 lines and should be read as a well-supported signal in this sample, not a population-level claim.

6. CONCLUSION

This study offers a validated integration of three existing diagnostics into a passive-data-only workflow, not a new method. Findings are specific to Niterói, March 2026, and are planning implications rather than firm recommendations. The trade-off between reliability and boarding-pressure, if it holds elsewhere, argues against composite risk scores as a general diagnostic tool for transit agencies. Future work: extend the stop-crosswalk coverage to enable a stop-level convergence test with adequate power; validate boarding-pressure against occupancy data; and test whether the reliability–crowding trade-off replicates elsewhere. More broadly, the pipeline requires only data agencies already collect, making replication in other GTFS/AVL/AFC-equipped systems straightforward.

REFERENCES

1. Deepa, L. et al. (2023). The adverse impact of headway variability on bus transit ridership. *Transport Policy*, 141, 343–356.
2. Candelieri, A. et al. (2019). Vulnerability of public transportation networks against directed attacks and cascading failures. *Public Transport*, 11, 27–49.
3. Abdelaty, H., Mohamed, M., Ezzeldin, M., & El-Dakhakhni, W. (2020). Quantifying and classifying the robustness of bus transit networks. *Transportmetrica A*, 16(3), 1176–1216.
4. Yap, M., Cats, O., & van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A*, 16(1), 23–42.
5. Kim, H., & Song, Y. (2018). An integrated measure of accessibility and reliability of mass transit systems. *Transportation*, 45(4), 1075–1100.
6. Luo, D., Bonnetain, L., Cats, O., & van Lint, H. (2018). Constructing spatiotemporal load profiles of transit vehicles with multiple data sources. *Transportation Research Record*, 2672(8), 175–186.