

Electronic Thesis and Dissertation Repository

---

6-25-2024 11:30 AM

## Integrating Environmental, Geographic and Network Context into Human Mobility Analysis

Milad Malekzadeh, *The University of Western Ontario*

Supervisor: Long, Jed A., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Geography and Environment

© Milad Malekzadeh 2024

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Geographic Information Sciences Commons](#), [Nature and Society Relations Commons](#), [Spatial Science Commons](#), and the [Urban Studies and Planning Commons](#)

---

### Recommended Citation

Malekzadeh, Milad, "Integrating Environmental, Geographic and Network Context into Human Mobility Analysis" (2024). *Electronic Thesis and Dissertation Repository*. 10172.  
<https://ir.lib.uwo.ca/etd/10172>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## Abstract

The study of human mobility provides a clear framework for examining and predicting a wide range of human activities. Irrespective of the approach, understanding mobility is not just about tracking movement but also about comprehending the context in which it occurs. This can be achieved by incorporating and integrating a variety of data sources into our mobility data analysis. This thesis uses three data sources—WORKANDHOME GPS tracking data, mobile network data, and NYC Yellow Taxi data—to thoroughly investigate how human mobility is related to environmental factors, geographical contexts, and community structures.

The research begins with an evaluation of the effects of environmental factors on stress and happiness levels during travel. By applying geographic ecological momentary assessment, it documents instantaneous emotional reactions to surroundings, underscoring the importance of green and blue spaces in boosting happiness and lowering stress. This study highlights the critical influence of the environment on the feelings individuals experience during daily travels, advocating for sensible urban planning and policies to improve emotional well-being. The thesis then introduces the *Mobility Deviation Index*, a new concept incorporating geographical context into the analysis of human mobility patterns. This approach contrasts observed mobility with expected levels based on amenity availability. The results advocate for including geographical context in mobility research directly when calculating indices of mobility, showing that patterns are significantly influenced by local characteristics. Next, it presents the *Local Mobility Index* (LMI), a novel metric for measuring local mobility behaviors. The LMI examines individual destination choices relative to amenity distribution. Finally, the thesis proposes the *Network Community Structure Similarity Index*, a new method for evaluating changes in community structures over time in mobility networks. This approach overcomes the limitations of previous methodologies, offering a clearer view of community dynamics. This thesis significantly advances human mobility research, offering a better view that underscores the critical role of context in determining mobility patterns. It not only deepens our understanding of human movement but also lays the groundwork for future research and policymaking aimed at fostering more liveable, sustainable, and equitable urban environments.

## Keywords

Human Mobility Analysis; Environmental Factors and Mobility; Geographic Ecological Momentary Assessment (GEMA); Urban Planning and Emotional Well-being; Sustainable Urban Development; Geographical Context in Human Mobility; Community Dynamics in Mobility Networks; Green and Blue Spaces in Urban Environments; Data Integration in Mobility Analysis

## Summary for Lay Audience

In today's fast-paced world, understanding how and why people move the way they do can tell us a lot about society. This research explores the study of human movement, not just by looking at where people go, but also by exploring the reasons behind their movements and the effects these journeys have on their emotions and well-being. Imagine walking through a park filled with green trees; doesn't it make you feel happier and more at peace? Our study confirms that natural spaces like these indeed boost happiness and reduce stress during travel.

We developed new tools and indices to better analyze human movement. One of these tools looks at how much people's movements differ from what we might expect based on the amenities available in their area. Another measures local travel behavior, considering the choices people make about where to go based on what is around them. We also introduced a way to understand how community connections change over time, reflecting on how people's movements can bring them together or apart.

Through our work, we emphasize the importance of considering the environment and local features when thinking about urban planning and policies. Cities should be designed with people's emotional well-being in mind, promoting access to nature and creating spaces that encourage positive community interactions.

This research sheds light on the complex relationship between human mobility, our surroundings, and our emotions. It suggests that by paying attention to the environment and how people interact with it, we can create more livable, sustainable, and happy communities. This is not just about moving from point A to point B; it's about making those journeys enrich our lives and the spaces we share.

## Co-Authorship Statement

This thesis was prepared according to the integrated-article layout designed by the Faculty of Graduate Studies at Western University, London, Ontario, Canada. All the work stated in this thesis including methodology and algorithm development, modelling, experimental testing, results interpretation, and writing draft manuscripts for publication was carried out by the author under the supervision of Dr. Jed A. Long. Dr. Long contributed to the development of methodology ideas and provided valuable comments, editing and revision on the manuscripts. He also provided financial support, software, hardware, and data etc.

Dr. Darja Reuschke contributed to the development of Chapters 2 and 4, offering invaluable insights and feedback. As a co-principal investigator for the project from which the dataset was derived, alongside Dr. Long, her role was crucial in both the acquisition and the use of the data.

Here is the list of all manuscripts:

**Malekzadeh, M.,** Reuschke, D., & Long, J. A. (submitted). How does travel environment affect urban stress and happiness: A study using geographic ecological momentary assessment in the England.

**Malekzadeh, M.,** & Long, J. A. Mobility deviation index: incorporating geographical context into analysis of human mobility. *J Geogr Syst* (2024). <https://doi.org/10.1007/s10109-024-00444-1>

**Malekzadeh, M.,** Long, J. A., & Reuschke, D. (submitted). Quantifying local mobility patterns in human mobility data.

**Malekzadeh, M.,** & Long, J. A. (2023). A network community structure similarity index for weighted networks. *Plos one*, 18(11), e0292018. <https://doi.org/10.1371/journal.pone.0292018>

## Dedication

To my mom, this is for you. Your unwavering support and love have been my foundation throughout this journey. You are my rock, and everything I have achieved is a reflection of the sacrifices you've made and the lessons you've taught me. Thank you for everything.

To Mahdi, my brother, thank you for being the best big brother anyone could ask for. You are one of the greatest blessings in my life. Your support has been invaluable. I will not forget when you gave me the only room we had, so I can study and work. Thank you for everything you have done for me.

To Jed, my supervisor, words fall short of expressing my gratitude. From the moment you gave me the opportunity to work with you, to this moment, I have learned a lot from you. From the moment you brought me a SIM card because I was in COVID-19 quarantine, to this day that I am leaving this campus, I cannot thank you enough for everything you've done for me. You were and are more than a supervisor; you are a friend and a mentor to me. Thank you for allowing me the freedom to explore various ideas, teaching me to be both creative and a problem solver. Thank you for believing in me and giving me a shot.

To Aranya, Kat, Nat, Michelle, Ethan, Matheus, Nate, Steffen, Ben, David, Mat, Daniel, Reyhaneh, Chelsea, Mohammad, Claire, and Hailyee, living in another country far from family could have been much more challenging without your support. Thank you for being there for me.

To Mohammad, Emad, Erfan, Mahyar, Arash, Ghazal, Helia, Yassi, Elham, Negin, Carlota, and Jonathan, although distance separates us, your support has always been a beacon of strength for me. Your encouragement means the world to me. Thank you for everything you have done for me.

## Acknowledgments

I would like to extend my deepest gratitude to Dr. Jed Long for his invaluable support throughout my PhD journey. His patience, guidance, and expertise in teaching me the nuances of conducting exemplary research have been fundamental to my success. Without his assistance and encouragement, my achievements would not have been possible.

Special thanks are also due to Dr. Darja Reuschke, whose insightful guidance significantly enhanced the quality of my research and contributed to my development as a researcher.

I am grateful to Dr. Jinhyung Lee for his readiness to always assist with my research and career questions, offering advice and support when it was most needed. Similarly, I appreciate Dr. Agnieszka Leszczynski for her encouragement and role as a supportive PhD committee member. To Dr. Grant Mackenzie, another esteemed member of my PhD committee, I am thankful for your constant support and the inspiration you have provided me.

I am deeply grateful to Dr. Ehab Diab, Dr. Kelly Yili Tang, Dr. Micha Pazner, and Dr. Jinhyung Lee for graciously accepting to be the examiners of my thesis and for their invaluable time and effort in reviewing my work. Their willingness to contribute to this significant phase of my academic journey is sincerely appreciated.

Furthermore, I wish to express my gratitude to the staff at our department, with a special mention of Lori Johnson. Her exceptional support has significantly eased each step of my journey, making challenges more manageable. I also owe a debt of gratitude to the former and current chairs, Dr. James Voogt and Dr. Katrina Moser, and graduate chair, Dr. Godwin Arku, whose support has been invaluable.

To each individual mentioned and to all who have contributed to my academic and personal growth during this PhD journey, I extend my sincere thanks.

# Table of Contents

Abstract.....	ii
Summary for Lay Audience.....	iv
Co-Authorship Statement.....	v
Dedication.....	vi
Acknowledgments.....	vii
Table of Contents.....	viii
List of Tables.....	xii
List of Figures.....	xiii
List of Appendices.....	xvii
List of Acronyms.....	xviii
Chapter 1.....	1
1 Introduction.....	1
1.1 Background.....	1
1.2 Movement data: Eulerian and Lagrangian approaches.....	3
1.3 Mobility data analysis.....	4
1.4 Research questions.....	6
1.5 Research objectives.....	7
1.6 Structure of the dissertation.....	8
References.....	9
Chapter 2.....	16
2 How does travel environment affect urban stress and happiness? A study using geographic ecological momentary assessment in England.....	16
2.1 Introduction.....	16
2.2 Background.....	18



2.3	Methods.....	19
2.3.1	Data.....	19
2.3.2	Happiness and stress .....	21
2.3.3	Natural amenities .....	21
2.3.4	Other travel and environmental variables .....	23
2.3.5	Statistical analysis.....	25
2.3.6	Sample description.....	26
2.4	Results.....	28
2.5	Discussion.....	31
2.6	Conclusion .....	33
	References .....	34
Chapter 3	.....	45
3	Mobility deviation index: Incorporating geographical context into analysis of human mobility .....	45
3.1	Introduction.....	45
3.2	Materials and methods .....	47
3.2.1	Human mobility data.....	47
3.2.2	Mobility Deviation Index.....	48
3.2.3	Observed mobility.....	49
3.2.4	Expected mobility .....	50
3.2.5	Statistical analysis of mobility deviation index and Socio-Economic Factors.....	55
3.3	Results.....	57
3.4	Discussion.....	65
3.5	Conclusion .....	70
	References .....	71
Chapter 4	.....	80

4	Quantifying local mobility patterns in human mobility data .....	80
4.1	Introduction.....	80
4.2	Methodology.....	82
4.2.1	Local Mobility Index (LMI) .....	82
4.2.2	Incorporating multi-centric activity patterns into measures of local mobility.....	87
4.3	Case study .....	89
4.3.1	GPS tracking data and study area .....	89
4.3.2	Data analysis .....	92
4.3.3	Results.....	94
4.4	Discussion .....	102
4.5	Conclusion .....	107
	References .....	108
	Chapter 5.....	117
5	A Network Community Structure Similarity Index for Weighted Networks .....	117
5.1	Introduction.....	117
5.2	Background/Literature Review.....	118
5.3	Network Community Structure Similarity Index (NCSSI).....	121
5.3.1	Step 1 – Pairing communities .....	122
5.3.2	Step 2 – Calculating edit costs of nodes .....	123
5.3.3	Step 3 – NCSSI calculation.....	126
5.4	Simulation Tests.....	127
5.4.1	Benchmark .....	127
5.5	Different community detection methods .....	129
5.5.1	Manually changing nodes' community memberships and their weights .....	130
5.6	Case study: NYC Taxi Flow dataset.....	132

5.7 Discussion and conclusion.....	137
References .....	140
Chapter 6.....	147
6 Conclusion .....	147
6.1 Summary.....	147
6.2 Contributions.....	150
6.2.1 Empirical findings.....	150
6.2.2 Conceptual and methodological contributions.....	151
6.2.3 Open-source code and reproducibility .....	152
6.3 Future research.....	153
References .....	155
Appendices.....	158
Curriculum Vitae .....	171

## List of Tables

Table 2-1 Summary of individual-level variables of n = 606 individuals tracked by a mobile app.....	26
Table 2-2 Summary statistics of environmental and control trip-level variables associated with n = 8083 trips by n = 606 individual tracked by a mobile app to study daily travel patterns in England. ....	27
Table 2-3 Linear mixed-effect regressions of happiness and stress level with and without control variables.....	30
Table 3-1 Six socio-economic variables chosen to compare with observed mobility (ROG) and mobility deviation index based on different models of expected mobility. All descriptions are based on Statistics Canada, (2017a). ....	55
Table 3-2 Summary statistics of observed and computed values for ROG and MDI resulted from different models. ....	62
Table 3-3 Summary of the model selection procedure using Moran’s I test of spatial autocorrelation on ordinary least squares regression model residuals, and the Lagrangian multiplier (LM) and robust Lagrangian multiplier (RLM) tests for the spatial error and spatial lag models. ....	62
Table 3-4 Spatial error models summaries using radius of gyration (ROG) and mobility deviation index (MDI) values computed by different models as dependent variables. ....	64
Table 4-1 Summary of currently available mobility metrics often used to capture relative local mobility patterns and a selected reference. ....	92
Table 4-2 Results ( $\beta$ coefficients) from eight linear regression models with the eight mobility metrics as the dependent variable. Bold relationships indicate $p > 0.05$ .....	98

# List of Figures

Figure 2-1 Trapezoidal fuzzy memberships for 6 different modes for 4 variables. The bottom plots feature black lines illustrating the overlapping membership functions of modes other than the bus on the left and the train on the right..... 24

Figure 3-1 Example showing a single simulation of a mobility pattern (EROG). The flowchart demonstrates the process of selecting a set of POIs for the computation of expected mobility. The numbers in the parentheses are simply an example and vary based on the local data. We ran the simulation  $s=100$  times for each ADA to simulate expected mobility..... 53

Figure 3-2 Implementation flowchart for how activity patterns were simulated..... 54

Figure 3-3 A) Map of types of areas based on Statistics Canada’s definition of population centers: urban areas, medium population centers (PC), small population centers (PC), and rural areas Statistics Canada, (2017a). These population centers are defined based on a specific range of population and population density per square kilometer. B) Map of POI Density defined as the ratio of the number of POIs over the area of ADA in Ontario, Canada using on OSM POI data (OpenStreetMap contributors, 2017). C) ROG values for the week of February 02, 2020, to February 08, 2020 in Ontario, Canada. Mobility data sourced from a large, de-identified network mobility dataset provided by TELUS Communications Inc. .... 58

Figure 3-4 A) Boxplots of observed ROG in different types of population centers. The red dotted line is the provincial median ROG. B) Scatter plot of the log of ROG values against the log of POI density for each ADA. The different colors show different types of population centers. The definition of population centers is based on Statistics Canada. .... 59

Figure 3-5 A) box plots of ROG values: observed and expected. The red dashed line is the provincial median of observed ROG values. B) box plots of MDI values. The red dashed line is associated with the value of MDI=1 (indicating an OE ratio of 1). C) Scatter plot of ROG values against EROG values. We colored each node based on their area type. D) Scatter plot of the natural logarithm of ROG values against MDI values. The black horizontal dashed line is the provincial median of observed ROG values, and the black vertical dashed line is where MDI=1. .... 61

Figure 4-1 An example scenario of the components of LMI. The stop point is surrounded by several nearby POIs. Among these, the restaurant (a), the nearest POI to the stop point (i), is identified as the destination at the stop point. Accordingly, the analysis identifies the closest POI of the same type as i (a) from the anchor location.  $d^a$  is the distance from the anchor location (e.g., home) to the nearest activity POI of type a, and  $d_i^a$  is the distance from the anchor location (e.g., home) to the chosen activity stop i. All other POI types depicted in gray, are not considered for this particular stop point in the calculation. .... 85

Figure 4-2 An example scenario of the components of multi-centric LMI with two anchor locations. The stop point is surrounded by several nearby POIs. Among these, the restaurant (a), the nearest POI to the stop point, is identified as the destination at the stop point. Accordingly, the analysis identifies the closest POI of the same type as a (a restaurant) from the two anchor locations. **dia1** and **dia2** are the distances between the first and second anchor locations and the nearest POI to the stop point. **da1** and **da2** are the distances between the first and second anchor locations and the closest POI of the type a. All other POIs, depicted in gray, are not considered for this particular stop point in the calculation. .. 88

Figure 4-3 LMI1: local mobility index to home; and LMI2: local mobility index to home and second-place. a) Scatter plot of LMI1 and LMI2, showing that almost all LMI2 values (with the exception of 26 data points) are equal to or greater than LMI1 values. b) Histograms of LMI1 and LMI2, where a positive skewness is observed in LMI1 and a negative skewness in LMI2. .... 95

Figure 4-4 Correlations between two LMI metrics and six commonly used mobility. LMI1: local mobility index to home; LMI2: local mobility index to home and second-place; NDS: average number of daily stops; TDD: average total daily travel distance (km); FDH: farthest distance from home (km); ROG: radius of gyration (km); CHA: convex hull area (km<sup>2</sup>); LTI: local travel index. .... 97

Figure 4-5 a) Boxplots of employment characteristics against LMI models. b) Boxplots of worker type categories against LMI models. SE/BO: self-employed/business owner. .... 101

Figure 4-6 Scatter plots and best fit lines of distance to city center against LMI1 and LMI2 (n=765). To enhance clarity in visualization, we have applied a log10 transformation to the distance values before plotting..... 102

Figure 5-1 An example of edit cost calculation when node *i* has changed its community membership. Blue lines and red lines represent the edges that are considered in the calculation of the edit cost related to node *i*, as intra-community edges and edges to the paired community. Dashed black lines represent the inter-community edges and black lines represent the intra-community edges that were not considered in the calculation of the edit cost related to node *i*. The paired communities in different community sets are colored the same (dark green and light green). All the other communities that are not considered in the edit cost calculation for node *i*, are colored gray. .... 126

Figure 5-2 A) LFR benchmark (Lancichinetti et al., 2008) when  $\mu$  equals 0.05, and B) LFR benchmark when  $\mu$  equals 0.95. Each color is representative of the nodes' communities.... 129

Figure 5-3 Similarity scores deploying Louvain similarity measure on LFR benchmarks using two existing similarity measurement methods (the Jaccard index and mutual information) and the proposed method NCSSI..... 130

Figure 5-4 Similarity values for an LFR benchmark while changing the weights associated with 10 randomly chosen nodes using two existing similarity measurement methods (the Jaccard index and mutual information) and the proposed method NCSSI. .... 131

Figure 5-5 The seasonal level of similarity in community structures for pairs of consecutive years using different indices from 2011 to 2021. The vertical axis demonstrates the similarity score. The dotted red line highlights the three periods of time we discuss in this section (spring 2016-2017; summer 2016-2017; and summer 2020-2021). .... 134

Figure 5-6 Maps of New York communities computed based on NY Yellow Taxis flow data. Taxi zones colored red represent the areas that changed their communities, and gray zones, represent the areas that did not. Base map and its related data are from OpenStreetMap and OpenStreetMap Foundation (OpenStreetMap Wiki, 2022). .... 135

Figure 5-7 Intra community adjusted flows and adjusted flows to the paired community for taxi zone number 222 for the spring of 2016 and 2017, and the summer of 2020 and 2021. This zone is chosen since it changed its community in all the chosen periods. Intra-community adjusted flows are presented with blue and adjusted flows to the paired community are presented with orange. The width of the flows is based on their intensity; the thicker, the higher rate of adjusted flows..... 137



## List of Appendices

<u>Appendix A: Results of linear mixed-effect regression models of happiness and stress level</u> .....	158
<u>Appendix B: MDI POI categorization</u> .....	159
<u>Appendix C: MDI Sensitivity Analysis</u> .....	161
<u>Appendix D: ROG, EROG, and MDI maps</u> .....	162
<u>Appendix E: Socio-economic factors-MDI scatter plot</u> .....	163
<u>Appendix F: LMI search radius sensitivity analysis</u> .....	166
<u>Appendix G: Community detection algorithms' Similarity scores</u> .....	169
<u>Appendix H: Ethics Approval Letters</u> .....	169

## List of Acronyms

ADA: Aggregated Dissemination Areas

AEC: Adjusted Edit Cost

AF: Adjustment Factor

CDRs: Call Detail Records

CHA: Convex Hull Area

EC: Edit Cost

FDH: Farthest Distance from Home

GEMA: Geographic Ecological Momentary Assessment

GIS: Geographic Information Science/System

GPS: Global Positioning System

LFR: Lancichinetti-Fortunato-Radicchi Benchmark

LM: Lagrangian Multiplier

LMI: Local Mobility Index

LTI: Local Travel Index

MDI: Mobility Deviation Index

MIDAS: Meteorological Office Integrated Data Archive System

NCSSI: Network Community Structure Similarity Index

NDS: Average Number of Daily Stops

NO<sub>2</sub>: Nitrogen Dioxide

NYC: New York City

OE: Observed/Expected

O3: Ozone

OS: Overlap Score

OSM: Open Street Map

PM10: Particulate Matter less than 10 micrometers in diameter

PM2.5: Particulate Matter less than 2.5 micrometers in diameter

POI: Point of Interest

RLM: Robust Lagrangian Multiplier

ROG: Radius of Gyration

SO2: Sulfur Dioxide

TDD: Average Total Daily Travel Distance

TTR: Travel Time Ratio

UK: United Kingdom

UKCEH: UK Centre for Ecology & Hydrology

USA: United States of America

VGI: Volunteered Geographic Information

VIF: Variance Inflation Factor

# Chapter 1

## 1 Introduction

This doctoral thesis explores the integration of geographical context in measuring and analyzing human mobility. Its primary objective is to enhance our understanding of how environmental, geographical, and social contexts influence mobility patterns. By methodically examining various contextual factors and their interaction with human movement, the research aims to contribute modestly to the broader field of human mobility studies, focusing on developing and refining metrics and methodologies that consider a wide range of influencing factors within urban and environmental settings.

### 1.1 Background

Human mobility casts light on the examination and prediction of a wide range of human activities (Wang et al., 2019). Studying human mobility offers insights beyond just locational shifts, exploring the reasons, patterns, and impacts of everyday human travel and interactions (Feng et al., 2020). The details of our daily movements reveal a wealth of information about social and economic aspects of life (Wu et al., 2016). Every individual's movement, be it routine urban travel or a short local trip, represents a mix of personal choices, socio-economic influences, and environmental factors (Chen et al., 2023; Cuttone et al., 2018; Wu et al., 2018). These patterns of movement are fundamental to urban planning, influencing the development of transportation systems, urban layouts, and accessibility within cities (Wang et al., 2020). Studying these movement patterns helps us understand the mechanisms and interactions that drive modern societies.

One of the key areas where analyzing human mobility can be beneficial is in urban and transportation planning. As cities grow larger, understanding human mobility becomes vital for urban and transportation planners (Boyce and Williams, 2015; Wang et al., 2012). Every person's movement, whether within city limits or inter-regionally, provides valuable information that offers insights into individuals' needs, preferences, and

challenges. By analyzing these patterns, city planners can effectively strategize urban development (F Xu et al., 2021), ensuring efficient transportation networks and smooth traffic flows (Asgari et al., 2013; Ferreira et al., 2018). Moreover, as cities sprawl and densities vary, knowing how, when, and where people move can significantly assist in designing sustainable cities, reducing congestion, and enhancing the overall quality of urban life.

Beyond urban planning, human mobility analysis offers an invaluable perspective for health-related studies (Oliver et al., 2015; Tizzoni et al., 2014). The dynamics of disease transmission, identification of hotspots, and containment strategies are closely associated with human mobility patterns (Roy and Kar, 2020). This was particularly evident during the COVID-19 pandemic, where mobility data was crucial in tracking and predicting the spread of the virus (Bergman and Fishman, 2020; Borkowski et al., 2021; Elarde et al., 2021; Franch-Pardo et al., 2020; Guzman et al., 2021; Long and Ren, 2022; Noi et al., 2022), enabling health experts to implement timely interventions. By understanding these patterns, health experts can predict potential outbreaks, strategize interventions, and manage public health crises more effectively.

Additionally, in the emerging field of neuro-urbanism, the interaction between urban environments and human psychological and physiological health is gaining attention (Pykett et al., 2020). Diverse urban environments, such as green spaces in cities, can have varying impacts on cognitive functions, emotional states, and overall mental health (Krekel et al., 2016; Willberg, Poom, et al., 2023). These findings underscore the significance of mobility patterns in influencing the psychological well-being of urban populations, as they interact with diverse urban landscapes.

To conduct such studies effectively, it's essential to access precise and relevant data that captures the mobility patterns of individuals, whether through detailed individual-level data or comprehensive aggregate data of populations. This data is foundational for in-depth analysis and understanding of human mobility. Since the year 2000, the field of

human mobility research has seen significant advancements, primarily due to the availability of diverse and comprehensive datasets (Noulas et al., 2012).

## 1.2 Movement data: Eulerian and Lagrangian approaches

In the study of human mobility, data can be categorized into two main types, each offering a distinct perspective on how people move and interact with their environment (Laube, 2014). The Eulerian perspective observes the world from fixed locations, similar to watching water flow at a specific point in a river. This approach captures data on entities as they interact with these points, providing a broad understanding of movement patterns (Demšar et al., 2021). Data sources that take an Eulerian approach include stationary sensors (H Xu et al., 2021), traffic cameras (Chen et al., 2021), and mobile network datasets like Call Detail Records (CDRs) (González et al., 2008; Song et al., 2010). CDRs are created each time a mobile phone is used for calls or texts, providing an Eulerian view of mobility by recording the user's location in relation to nearby cell towers (Blondel et al., 2015). These datasets are especially valuable in regions or demographics where continuous individual tracking might be challenging or invasive (Forghani et al., 2020). They provide a broad understanding of mobility patterns, helping to identify congested areas, popular transit nodes, and general movement trends.

On the other hand, the Lagrangian perspective focuses on following entities throughout their movement, capturing their entire trajectory (Demšar et al., 2021). GPS tracking is a key example of this approach, continuously recording an entity's location over time. This method provides a detailed view of individual movement patterns, revealing travelling habits, regular routes, or changes in mobility over time (Rhee et al., 2011; Zheng et al., 2008). However, while it offers detailed insights into individual movements, since it commonly features a smaller population, its granularity can sometimes overshadow broader, macroscopic trends evident in Eulerian datasets (Barbosa et al., 2018). Also, growing concerns about data privacy and ethics in continuous high-precision tracking are influencing how this method is used in human mobility studies (Leszczynski, 2018).

### 1.3 Mobility data analysis

In today's data-rich environment, the primary challenge is to derive meaningful information and insights from the vast amounts of raw data available (Fayyad et al., 1996; Laube, 2014). In human mobility studies, this transformation is facilitated using various mobility metrics, each offering a unique perspective on movement (Barbosa et al., 2018; Wang et al., 2019). Time-based metrics focus on the temporal aspects of movement, considering factors like how long an individual stays at a location (Demissie et al., 2019; Dijst and Vidakovic, 2000). Range-based metrics concentrate on the spatial and geometric aspects of movement. Metrics like the Radius of Gyration (ROG), which measures the extent of an individual's movement around a mean center of gravity (González et al., 2008), are examples of range-based metrics. Entropy-based metrics, such as Shannon Entropy, measure the unpredictability in a person's choice of places (Song et al., 2010). A high Shannon Entropy value implies a varied and spontaneous pattern, while a lower value suggests more predictable and routine movements. These metrics collectively facilitate a comprehensive analysis of data, creating an empirical basis for urban planning, policy development, and enhancing our understanding of spatial-temporal human behaviors. However, although we have significantly advanced in understanding human mobility, there is a noticeable gap here: the direct integration of the context in which movement takes place. Mobility is not an isolated phenomenon; it is closely connected with the surrounding environment, societal frameworks, and individual experiences (Siła-Nowicka et al., 2016).

Irrespective of the approach, understanding mobility is not just about tracking movement but also about comprehending the context in which it occurs. This can be achieved by incorporating and integrating a variety of data sources into our mobility data analysis. Satellite imagery, for instance, provides a bird's eye view of landscapes, helping researchers differentiate urban from rural, forested from barren, and residential from commercial (Jia et al., 2018). Volunteered Geographic Information (VGI) leverages crowd-sourced data, often offering real-time insights into road closures, construction activities, or even public events that might influence mobility (Antoniou and Skopeliti,

2015; Goodchild, 2007; Sangiambut and Sieber, 2016). Publicly available datasets, statistical reports from governmental bodies, and specialized questionnaires further enrich the context, providing socio-economic, demographic, and infrastructural details that influence movement patterns (Hanson and Johnston, 1985; Havet et al., 2021; Simini et al., 2012).

To better define and understand context in human mobility studies, it is important to recognize that context is a complex and ever-changing concept, involving various factors that impact human movement (Brum-Bastos et al., 2021; Purves et al., 2014). Context can be related to the physical environment, including environmental factors like weather and the availability of green and blue spaces, and urban infrastructure design (Zhou et al., 2019). These factors can significantly impact mobility choices (König and Axhausen, 2002; Willberg, Tenkanen, et al., 2023). Additionally, context encompasses the socio-economic environment, including cultural norms and personal experiences, which play a role in shaping commuting habits, residential preferences, and leisure activities (Palmer et al., 2013; Wu et al., 2016). Therefore, understanding context in human mobility is not just an additional aspect but a fundamental component that shapes movement patterns, underscoring the need to consider these diverse elements to fully grasp mobility behaviors (Purves et al., 2014; Siła-Nowicka et al., 2016).

Integrating context with mobility data presents significant challenges due to the multifaceted nature of context and the intricacies of mobility patterns. This process requires a careful combination of varied datasets and metrics, and an interdisciplinary approach. Successfully combining context into mobility analyses involves using a variety of datasets, including geographical, socio-economic, and behavioral data, as well as developing and applying metrics that can capture the complex relationship between context and mobility. Integration of contextual factors into human mobility analyses, though challenging, allows for a more accurate interpretation of mobility patterns, ensuring that the patterns we observe, predict, and seek to influence are rooted in a deeper understanding of human behavior.



## 1.4 Research questions

When examining context in human mobility analysis, the impact of the environment on our well-being emerges as a pivotal consideration. Understanding the relationship between an individual's psychological well-being and their daily interactions with geographical elements is particularly crucial. Given the significant importance of well-being and the substantial time spent in transit, it is essential to explore how the environments we navigate during our travels influence our mental and emotional health.

As the focus of our analysis expands to encompass large-scale movement patterns, the importance of geographical context becomes increasingly apparent. At this macro level, geographical context is a critical element in meaningful analysis. People from different regions exhibit distinct mobility behaviors influenced by their specific geographical context, highlighting the need for a robust method to compare these behaviors across different contexts. This method should directly incorporate the context into its measurement approach, taking into account the varied impacts of geographical context on movement patterns.

A crucial aspect of studying mobility patterns is understanding the local nature of an individual's movements. However, analyzing just the geometric aspect of mobility falls short in capturing the full picture, as the context in which mobility occurs can significantly influence the degree of localness in a person's movements. Recognizing the importance of localness necessitates the development of a metric that not only quantifies it but also integrates the contextual factors influencing it. This approach allows for a more thorough examination of how various urban features and environmental contexts impact local mobility patterns, thereby providing a better understanding of individual movement behaviors within different geographical contexts.

Beyond geographical considerations, human movement is influenced by a variety of contexts, including community structures. Analyzing mobility data can uncover complex community patterns, offering deeper insights into the relationship between human

mobility and communities. However, there is a notable gap in methodologies for comprehensively comparing these community structures over time. Developing such methodologies is essential for gaining insights into how mobility patterns and community structures influence each other through time, enhancing our understanding of their mutual relationship.

Against this background, the research questions in this dissertation are:

1. How do environmental factors during daily travel activities impact individuals' happiness and stress levels?
2. How does incorporating geographical context into human mobility measurements alter the interpretation of human mobility analyses results?
3. How to measure local human mobility behavior considering the influence of geographical context?
4. How to compare two sets of communities derived from mobility flows?

## 1.5 Research objectives

The primary goal of this thesis is to explore the field of human mobility in depth, with a focus on integrating and intricately characterizing context in these studies. To answer the research questions proposed above, the following four specific objectives are defined:

1. Investigate the relationship between environmental variables and happiness and stress levels by conducting a geographic ecological momentary assessment study.
2. Introduce and define a measure that incorporates geographical context, facilitating researchers in the study and comparison of human mobility patterns among individuals residing in different contexts.
3. Define a metric for quantifying the degree of localness in human mobility behavior, unveiling nuanced insights into localized mobility preferences, contributing to a more comprehensive understanding of movement behavior in urban contexts.

4. Define a measure that effectively captures the impact of label and weight changes in community structure, offering a new approach for community similarity measurement in complex networks.

These research objectives lay the groundwork for a multi-dimensional exploration of how environmental, geographical, and social factors interact with and influence mobility patterns. By developing new methods and metrics for assessing the complex relationship between individuals and their surroundings, this thesis aims to bridge existing gaps in human mobility research. Each objective, while distinct, contributes to the overarching aim of enriching our understanding of human mobility through a contextual lens. This integrated approach not only advances our methodological capabilities but also enhances our ability to interpret complex mobility data in a way that is meaningful for urban planning, public health, and environmental policy.

## 1.6 Structure of the dissertation

This dissertation is organized into six chapters, with the next four chapters structured as individual papers that are either published or under revision.

Chapter 2 explores the relationship between environmental factors and individual well-being, employing a novel method to capture the real-time emotional impact of urban environments during the daily travels on happiness and stress levels. This study highlights the significance of green and blue spaces and weather in shaping residents' emotional states during their daily travels.

Chapter 3 introduces the concept of a mobility deviation index (MDI), a measurement method for analyzing human mobility patterns across large areas, integrating geographical context to enable detailed comparison of human mobility patterns among individuals residing in different contexts. This method facilitates a deeper understanding of human mobility by providing different insights compared to traditional mobility measures, highlighting the importance of geographical context in interpreting mobility patterns and their association with socio-economic factors.

Chapter 4 introduces the local mobility index (LMI), a novel metric designed to quantify localness in human mobility patterns, integrating geographical context. This approach enables a refined analysis of urban movement trends, focusing on how individuals choose their destinations. By factoring in the spatial distribution of urban amenities, the LMI offers insights into the role of socio-economic and environmental factors in shaping local travel behaviors, contributing significantly to the field of urban mobility studies.

Chapter 5 proposes a network community structure similarity index (NCSSI), a method for evaluating community structures in complex networks, integrating both connectivity and the importance of connections. This approach addresses limitations of prior methods, enhancing network dynamics analysis and offering a clearer understanding of community structure evolution over time.

Chapter 6 concludes the dissertation, summarizing the key findings and discussions from the earlier chapters. It also suggests potential future research directions.

## References

- Antoniou V and Skopeliti A (2015) Measures and indicators of VGI quality: An overview. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences* 2. Copernicus GmbH: 345–351.
- Asgari F, Gauthier V and Becker M (2013) A survey on human mobility and its applications. *arXiv preprint arXiv:1307.0814*. Epub ahead of print 2013.
- Barbosa H, Barthelemy M, Ghoshal G, et al. (2018) Human mobility: Models and applications. *Physics Reports* 734. Elsevier: 1–74.
- Bergman NK and Fishman R (2020) Correlations of mobility and Covid-19 transmission in global data. *medRxiv*. Epub ahead of print 2020. DOI: 10.1101/2020.05.06.20093039.
- Blondel VD, Decuyper A and Krings G (2015) A survey of results on mobile phone datasets analysis. *EPJ data science* 4. Springer: 1–55.

- Borkowski P, Jażdżewska-Gutta M and Szmelter-Jarosz A (2021) Lockdowned: Everyday mobility changes in response to COVID-19. *Journal of Transport Geography* 90. Elsevier Ltd.
- Boyce DE and Williams HCWL (2015) *Forecasting Urban Travel: Past, Present and Future*. Edward Elgar Publishing.
- Brum-Bastos V, Łoś M, Long JA, et al. (2021) Context-aware movement analysis in ecology: a systematic review. *International Journal of Geographical Information Science*. Taylor & Francis: 1–23.
- Chen L, Grimstead I, Bell D, et al. (2021) Estimating vehicle and pedestrian activity from town and city traffic cameras. *Sensors* 21(13). MDPI: 4564.
- Chen Y, Xie N, Xu H, et al. (2023) A Multi-Context Aware Human Mobility Prediction Model Based on Motif-Preserving Travel Preference Learning. *IEEE Transactions on Intelligent Transportation Systems*. IEEE. Epub ahead of print 2023.
- Cuttone A, Lehmann S and González MC (2018) Understanding predictability and exploration in human mobility. *EPJ Data Science* 7. Springer: 1–17.
- Demissie MG, Phithakkitnukoon S, Kattan L, et al. (2019) Understanding human mobility patterns in a developing country using mobile phone data. *Data Science Journal* 18: 1.
- Demšar U, Long JA, Benitez-Paez F, et al. (2021) Establishing the integrated science of movement: bringing together concepts and methods from animal and human movement analysis. *International Journal of Geographical Information Science* 35(7). Taylor & Francis: 1273–1308.
- Dijst M and Vidakovic V (2000) Travel time ratio: the key factor of spatial reach. *Transportation* 27. Springer: 179–199.

- Elarde J, Kim J-S, Kavak H, et al. (2021) Change of human mobility during COVID-19: A United States case study. *PloS one* 16(11). Public Library of Science San Francisco, CA USA: e0259031.
- Fayyad U, Piatetsky-Shapiro G and Smyth P (1996) From data mining to knowledge discovery in databases. *AI magazine* 17(3): 37.
- Feng J, Li Y, Yang Z, et al. (2020) Predicting human mobility with semantic motivation via multi-task attentional recurrent networks. *IEEE Transactions on Knowledge and Data Engineering* 34(5). IEEE: 2360–2374.
- Ferreira DL, Nunes BAA and Obraczka K (2018) Scale-free properties of human mobility and applications to intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems* 19(11). IEEE: 3736–3748.
- Forghani M, Karimipour F and Claramunt C (2020) From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transportation Research Part C: Emerging Technologies* 117. Elsevier: 102666.
- Franch-Pardo I, Napoletano BM, Rosete-Verges F, et al. (2020) Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment* 739. Elsevier B.V.
- González MC, Hidalgo CA and Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196). Nature Publishing Group: 779–782.
- Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. *GeoJournal* 69. Springer: 211–221.
- Guzman LA, Arellana J, Oviedo D, et al. (2021) COVID-19, activity and mobility patterns in Bogotá. Are we ready for a ‘15-minute city’? *Travel Behaviour and Society* 24. Elsevier: 245–256.

- Hanson S and Johnston I (1985) Gender differences in work-trip length: explanations and implications. *Urban geography* 6(3). Taylor & Francis: 193–219.
- Havet N, Bayart C and Bonnel P (2021) Why do gender differences in daily mobility behaviours persist among workers? *Transportation research part A: policy and practice* 145. Elsevier: 34–48.
- Jia Y, Ge Y, Ling F, et al. (2018) Urban land use mapping by combining remote sensing imagery and mobile phone positioning data. *Remote Sensing* 10(3). MDPI: 446.
- König A and Axhausen KW (2002) The reliability of the transportation system and its influence on the choice behaviour. In: *Proceedings of the 2nd Swiss Transportation Research Conference, 2002*.
- Krekel C, Kolbe J and Wüstemann H (2016) The greener, the happier? The effect of urban land use on residential well-being. *Ecological economics* 121. Elsevier: 117–127.
- Laube P (2014) *Computational Movement Analysis*. 5th ed. Berlin: Berlin: Springer International Publishing.
- Leszczynski A (2018) Geoprivacy. In: *Understanding Spatial Media*. SAGE Publications Ltd, pp. 235–244.
- Long JA and Ren C (2022) Associations between mobility and socio-economic indicators vary across the timeline of the Covid-19 pandemic. *Computers, environment and urban systems* 91. Elsevier: 101710.
- Noi E, Rudolph A and Dodge S (2022) Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework. *International Journal of Geographical Information Science* 36(3). Taylor & Francis: 585–616.

- Noulas A, Scellato S, Lambiotte R, et al. (2012) A tale of many cities: universal patterns in human urban mobility. *PLoS one* 7(5). Public Library of Science San Francisco, USA: e37027.
- Oliver N, Matic A and Frias-Martinez E (2015) Mobile network data for public health: opportunities and challenges. *Frontiers in public health* 3. Frontiers Media SA: 189.
- Palmer JRB, Espenshade TJ, Bartumeus F, et al. (2013) New approaches to human mobility: Using mobile phones for demographic research. *Demography* 50(3). Duke University Press: 1105–1128.
- Purves RS, Laube P, Buchin M, et al. (2014) Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems*. Elsevier Ltd.
- Pykett J, Osborne T and Resch B (2020) From Urban Stress to Neurourbanism: How Should We Research City Well-Being? *Annals of the American Association of Geographers* 110(6). Taylor & Francis: 1936–1951.
- Rhee I, Shin M, Hong S, et al. (2011) On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking* 19(3). IEEE: 630–643.
- Roy A and Kar B (2020) Characterizing the spread of COVID-19 from human mobility patterns and SocioDemographic indicators. In: *Proceedings of the 3rd ACM SIGSPATIAL international workshop on advances in resilient and intelligent cities*, 2020, pp. 39–48.
- Sangiambut S and Sieber R (2016) The V in VGI: Citizens or civic data sources. *Urban planning* 1(2): 141–154.
- Sila-Nowicka K, Vandrol J, Oshan T, et al. (2016) Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science* 30(5). Taylor & Francis: 881–906.



- Simini F, González MC, Maritan A, et al. (2012) A universal model for mobility and migration patterns. *Nature* 484(7392). Nature Publishing Group UK London: 96–100.
- Song C, Qu Z, Blumm N, et al. (2010) Limits of predictability in human mobility. *Science* 327(5968). American Association for the Advancement of Science: 1018–1021.
- Tizzoni M, Bajardi P, Decuyper A, et al. (2014) On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Computational Biology* 10(7). Public Library of Science.
- Wang A, Zhang A, Chan EHW, et al. (2020) A review of human mobility research based on big data and its implication for smart city development. *ISPRS International Journal of Geo-Information* 10(1). MDPI: 13.
- Wang J, Kong X, Xia F, et al. (2019) Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD explorations newsletter* 21(1). ACM New York, NY, USA: 1–19.
- Wang P, Hunter T, Bayen AM, et al. (2012) Understanding road usage patterns in urban areas. *Scientific reports* 2(1). Nature Publishing Group UK London: 1001.
- Willberg E, Poom A, Helle J, et al. (2023) Cyclists' exposure to air pollution, noise, and greenery: a population-level spatial analysis approach. *International Journal of Health Geographics* 22(1). Springer: 5.
- Willberg E, Tenkanen H, Miller HJ, et al. (2023) Measuring just accessibility within planetary boundaries. *Transport Reviews*. Taylor & Francis: 1–27.
- Wu R, Luo G, Yang Q, et al. (2018) Learning individual moving preference and social interaction for location prediction. *Ieee Access* 6. IEEE: 10675–10687.
- Wu W, Wang J and Dai T (2016) The geography of cultural ties and human mobility: Big data in urban contexts. *Annals of the American Association of Geographers* 106(3). Taylor & Francis: 612–630.

- Xu F, Li Y, Jin D, et al. (2021) Emergence of urban growth patterns from human mobility behavior. *Nature Computational Science* 1(12). Nature Publishing Group US New York: 791–800.
- Xu H, Berres A, Tennille SA, et al. (2021) Continuous emulation and multiscale visualization of traffic flow using stationary roadside sensor data. *IEEE Transactions on Intelligent Transportation Systems* 23(8). IEEE: 10530–10541.
- Zheng Y, Li Q, Chen Y, et al. (2008) Understanding mobility based on GPS data. In: *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 312–321.
- Zhou F, Yue X, Trajcevski G, et al. (2019) Context-aware variational trajectory encoding and human mobility inference. In: *The world wide web conference*, 2019, pp. 3469–3475.

## Chapter 2

### 2 How does travel environment affect urban stress and happiness? A study using geographic ecological momentary assessment in England<sup>1</sup>

#### 2.1 Introduction

The health and wellbeing effects of urban living have become a major public health priority. Apart from being associated with urban lifestyle aspects such as sedentary behavior and psychological stress, urban living is often associated with higher exposure to environmental hazards such as air pollution, noise, and heat and less access to natural environments including greenspace. It is well documented that a variety of trip attributes such as mode of transportation (Duarte et al., 2010; Olsson et al., 2013), the duration of trips (Ettema et al., 2012; Olsson et al., 2013), and the type of activity (Archer et al., 2013; Ettema et al., 2012), can affect daily travel experiences (happiness, stress and satisfaction) during and following trips.

A large body of literature has investigated the individual-level determinants of subjective wellbeing, broadly defined as how people feel about how well or happy they are as opposed to measurable factors of health and welfare (Pain & Smith, 2010). More recent research has studied the impact of external factors on wellbeing such as environmental influences in addition to the more commonly studied internal factors and genetics (Bevilacqua & Goldman, 2011). As an environmental factor of wellbeing, access to green space at people's residential location has been studied showing the health and wellbeing benefits of 'greenness' (Dadvand et al., 2016).

Given that daily travel constitutes a significant part of people's routines, averaging 1 hour per day in 2019 (Dept. for Transport, 2021), the objective of this study is to understand

---

<sup>1</sup> A version of this chapter has been submitted for publication to Urban Studies.

how environmental context of daily travel influences urban stress and wellbeing. Studies on the attributes of daily travel in relation to wellbeing have not studied the environmental context of urban travel, on the one hand. Studies that have explored various environmental factors on wellbeing, on the other hand, have measured the wellbeing impact of environmental factors at the place where people live (Winters & Li, 2017), neglecting that most people are also exposed to environmental factors at other relevant places in their daily lives and while travelling. This applies specifically to workers who have to commute to a workplace (Clark et al., 2020). Connecting these two strands of literatures, will help better understand how urban wellbeing can be improved or negative effects mitigated.

Traditionally, studies on the environment and subjective wellbeing have used one-time questionnaire-based surveys which capture emotional snapshots within specific contexts (Guite et al., 2006; Singleton & Clifton, 2021). However, this approach is susceptible to recall bias, which emerges from the discrepancy between emotions recorded at the time of the survey and those actually experienced during the events in question (Kahneman et al., 2004; Robinson & Clore, 2002). Geographic Ecological Momentary Assessment (GEMA) has enabled the real-time tracking of everyday experiences (Pykett et al., 2020). Through the application of this approach, researchers can attain a deeper understanding of the complex interrelationships between urban elements like access to green spaces and noise levels, and the emotional states of individuals (Kou et al., 2020; Mennis et al., 2018; X. Zhang et al., 2020). However, even within this framework, environmental attributes that affect daily wellbeing experiences associated with travel remain poorly understood (see, for example, (Kirchner & Shiffman, 2016; Mennis et al., 2018; Parrish et al., 2020)). Since GEMA studies are resource-intensive, existing studies have used small and/or non-representative samples (e.g. (Pykett et al., 2020)).

In this study, we use GEMA to collect repeated information of the same individual on happiness (positive feeling/affect) and stress (negative affect) alongside other rich survey data in a large sample of individuals with workplaces in cities. The novel contributions of

this study are twofold. The first lies in linking daily travel with external (environmental) factors of wellbeing while the second contribution is methodological. We link the locational data and survey data to attributes characterizing the natural amenities while people travelled. Natural amenities are naturally-occurring locational attributes that enhance the desirability of a place and are expected to positively impact individual wellbeing (Winters & Li, 2017). We include as natural amenity factors that cover the proximity to the natural environment (green/blue spaces), exposure to weather (apparent temperature, visibility, daylight, and rain), and exposure to air pollution. By examining this range of factors within the context of daily travel in cities, we uncover relationships between urban stress and happiness ('wellbeing') and environmental context, and their relative effects beyond the stationary residential location.

## 2.2 Background

Daily travel, an integral aspect of daily urban life, is intimately tied to both the environment and wellbeing. Longer daily travel and commuting times have been associated with higher levels of stress, and lower mental wellbeing in several previous studies (Ettema et al., 2012; Stutzer & Frey, 2008). The stress of daily travel can also spill over into other areas of life, affecting work satisfaction, family relationships, and leisure activities (Clark et al., 2020). Moreover, accessibility to amenities, safety, and the overall pleasantness of the daily travel experience are key factors that influence commuters' wellbeing (Ettema & Schekkerman, 2016). Factors like temperature, precipitation, visibility (Böcker et al., 2016; Li et al., 2022), and air quality index (Ma et al., 2021) also play significant roles in influencing people's subjective wellbeing and overall health. Foggy weather and high humidity levels have been found to negatively affect individuals' moods (Čelić et al., 2019). Additionally, studies have indicated that transportation mode can significantly impact happiness (Eriksson et al., 2013; Mokhtarian & Pendyala, 2018).

The profound influence of green open spaces on subjective wellbeing has been extensively explored in previous studies. Previous evidence in research based on travel

surveys supports that trips occurring in greenspaces are associated with greater happiness levels (Wang et al., 2021). Access to green and open areas have also been linked with heightened physical activity and social interaction (Lee & Maheswaran, 2011; Maas et al., 2006; Matsuoka & Kaplan, 2008). Greater access to greenspace at residential locations has been related to increased life satisfaction and mental wellbeing, and reduced stress levels (Ambrey & Fleming, 2014; Fan et al., 2011; Smyth et al., 2008; Sugiyama et al., 2008; White et al., 2013). Such positive associations are further evident in studies connecting exposure to green space with reduced cortisol secretion, a hormone indicative of stress, suggesting a link to lower stress levels (Thompson et al., 2012).

## 2.3 Methods

### 2.3.1 Data

We collected data from 1017 participants across three cities in England: Brighton and Hove, Leeds, and Birmingham. Participants were recruited from October 2018 to May 2019 in Brighton & Hove and Leeds, and September 2019 to April 2020 in Birmingham. The sample included participants between 18 and 64 years old who were in employment (paid employment or self-employment) and who worked in the selected cities. While workers formed the sampling frame, the study aimed to capture all their daily travel experiences, not solely their commutes. We were conservative in application of data integrity measures which led to the exclusion of 411 participants due to insufficient data necessary for our statistical models. Therefore, our final analysis included data from  $n = 606$  individuals.

Participation in the study consisted of two parts: a questionnaire survey which captured sociodemographic characteristics and a mobile-phone app-based survey of approximately seven days which should be a 'normal' working week. Participants were asked to install a bespoke mobile application on their phones designed to facilitate data collection. This application was comprised of two core components: a location tracking feature and real-time survey delivered through push notifications. The former recorded participants' geographical positions using a dynamic motion-based approach, incorporating a two-state

rolling geofence system. The delineation between motion and stationary states was established by a geofence threshold of 50 meters. This approach allowed detailed GPS tracking of geographical movement during motion states while conserving resources (i.e., battery) during stationary phases. Participants' movement was continuously tracked, and the termination of any trip (i.e., transition to a stationary phase from a motion phase) prompted the delivery of a GEMA survey through the mobile app.

Trip is defined as the segment of data recorded between two consecutive stop points. Stop points are locations where participants remained stationary for  $\geq 5$  minutes, with movement between consecutive points  $< 75$  meters. We employed agglomerative clustering on the stop points to identify the home location of participants. This clustering method grouped nearby stop points within a defined distance threshold (75 meters), generating clusters that represented potential residential places. We assigned the cluster with the longest duration as the primary home location. To assess the precision of the GPS-derived locations, we cross-referenced them with the centroid locations of the postal codes that participants self-reported as their home postal codes. Our examination shows that the GPS-derived home locations matched the self-reported home postal code centroid within a 1 km radius in 81 % of instances. This method aligns with established practices in GPS tracking studies and allowed us to accurately determine participants' key locations based on their time spent at these sites (Kung et al., 2014; Siła-Nowicka et al., 2016). This enabled us to calculate the Euclidean distance between individuals' residential locations and the city center of the respective cities, allowing us to capture distinctions between those residing closer to the urban core and those located on the outskirts of the city.

Not all trips in the dataset are associated with a completed GEMA survey. We retained only those trips with a corresponding GEMA survey completed within one hour of travelling. The final dataset comprises 8,083 trips (25% completion rate) contributed by our 606 participants, which is an average of 13.3 trips per individual. In order to calculate

individual's 'baseline' happiness and stress levels, we also use GEMA surveys not associated with trips ( $n = 21,349$  GEMA survey responses).

## 2.3.2 Happiness and stress

We are focusing on the two GEMA response variables: happiness and stress. Both are measured using a 7-point Likert scale. Since individuals have differing 'baseline' happiness or stress levels, a comparison between individuals needs to control for the fact that some people are happier than others or cope with stressors differently (Nes & Røysamb, 2017). We therefore adjusted each participant's GEMA scores for happiness and stress by subtracting the mean response for each individual across all GEMA surveys (including those GEMA surveys not associated with a trip) from each response to get a relative measure (to their overall mean) of happiness and stress for each trip (Veenhoven, 1991).

## 2.3.3 Natural amenities

### 2.3.3.1 Green and blue spaces

We derive a measure of the exposure to green or blue space during individual trips as proportion of the trip through or close to green/blue space. We calculated the area of green and blue space present in a buffer of 50 meters around each trip's GPS data. The area of green and blue space related to each trip was divided by the area of the 50 m buffer to give a numerical proxy representing the proportion (between 0 and 1) of a trip (by area) within green and/or blue space (0 – no green or blue space, 1 – all green or blue space). We used the UK Centre for Ecology & Hydrology (UKCEH) land cover dataset to extract information on green and blue spaces (Morton et al., 2020). UKCEH uses Sentinel-2 Seasonal Composite Images reflecting the median reflectance for each season. The land cover dataset is comprised of 21 classes of land cover. Eleven green-related classes including deciduous woodland, coniferous woodland, arable, improve grassland, neutral grassland, calcareous grassland, acid grassland, fen, heather, heather grassland,



and bog, and two blue-related classes including saltwater and freshwater were merged to define the green-blue space.

The impact of green spaces on wellbeing can vary across seasons, as the presence or absence of leaves and other seasonal changes might affect our perception of these spaces (Mennis et al., 2018). However, quantifying the specific characteristics of green areas on a daily, monthly, or seasonal basis is a significant challenge. To address this, we have incorporated temperature and the season in which the trip was conducted as a proxy in our model to account for potentially seasonality effects of influence of green spaces on our happiness and stress levels.

### 2.3.3.2 Weather data

We linked weather condition data from the Meteorological Office Integrated Data Archive System (MIDAS) to each trip. We selected the timestamp at the midpoint of each trip to serve as a representative time of trips. The nearest weather station was identified that corresponded to the time of the trip.

MIDAS is a comprehensive weather database managed by the UK's national weather service (Met Office, 2006b, 2006a). With more than 200 stations, it covers the whole UK. Hourly data for rainfall (in millimeters) is used and converted into a binary variable to measure exposure to rain (yes/no) while travelling. Hourly horizontal visibility data, measured in meters, is also used directly from MIDAS without any modification. Previous research has demonstrated that apparent temperature is a useful variable for capturing how individuals experience weather, and therefore, we calculated the apparent temperature ( $AT$ ) (in Centigrade) (Brum-Bastos et al., 2018; Steadman, 1994)

$$AT = T + 0.33 * e - 0.70 * WS - 4.00 \quad 2.1$$

where  $AT$  is the apparent temperature in Centigrade,  $T$  is the air temperature in Centigrade,  $e$  is the water vapour pressure in hPa, and  $WS$  is the wind speed in m/s.

### 2.3.3.3 Air pollution

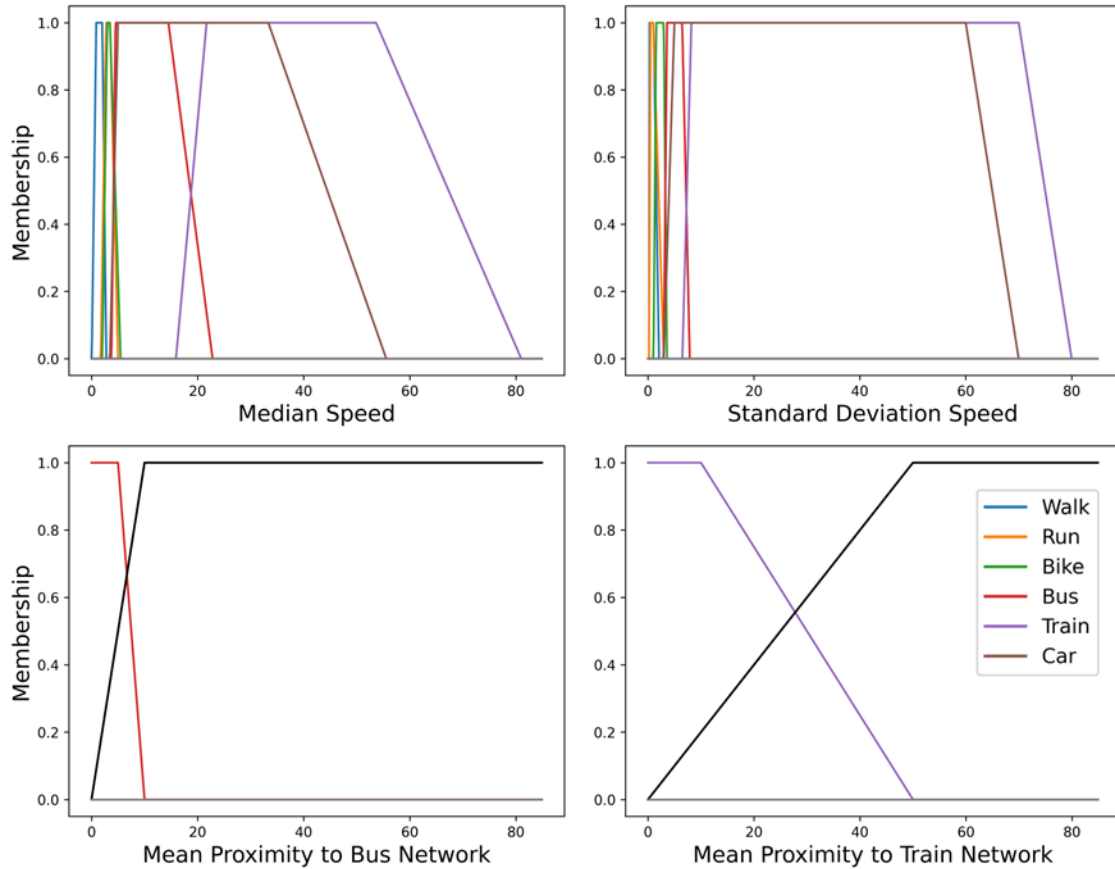
An air quality index score ranging from 1 to 10 (with 1 representing the least polluted air and 10 indicating extremely polluted air quality) is used to assess the impact of air quality on wellbeing in this study. We use daily data from the UK Air Information Resource (UK Air Information Resource, 2023) . Our analysis included five pollutants: sulfur dioxide (SO<sub>2</sub>), ozone (O<sub>3</sub>), nitrogen dioxide (NO<sub>2</sub>), particulate matter less than 10 µm in diameter (PM<sub>10</sub>), and particulate matter less than 2.5 µm in diameter (PM<sub>2.5</sub>). We employed a ten-point scale for each pollutant, in line with the standards set by the Committee on the Medical Effects of Air Pollutants (Ayres, 2011). The values of these pollutants were determined based on the closest monitoring site for each trip on the respective day. Following the recommended approach (Cheng et al., 2004; Tan et al., 2021), we then used the highest value among all pollutants to represent the air quality index for each trip.

### 2.3.4 Other travel and environmental variables

#### 2.3.4.1 Travel mode detection

We did not directly request participants to report their mode of travel. Thus, we infer the travel mode from participants' tracking data. We employed a Fuzzy Logic system based on Xu et al.'s approach to detect the travel modes (Xu et al., 2010). We used six transportation mode categories: walk, run, bike, bus, train, and car. In our fuzzy system we employed four variables: median speed, standard deviation speed, proximity to bus routes, and proximity to train routes. Incorporating four variables enabled us to distinguish between modes that are similar in one aspect but different in the other. For example, bus and car might have the same median speed, but their proximity to bus routes is different; consequently, our fuzzy system differentiates these two from each other. With six distinct modes to be identified using four variables, a total of 24 fuzzy membership functions were created (Figure 2-1). These membership functions were designed in a trapezoidal form. We employed min-max operation (minimum value in

each parameter and maximum value between all mode categories) to identify each mode of transportation.



**Figure 2-1 Trapezoidal fuzzy memberships for 6 different modes for 4 variables. The bottom plots feature black lines illustrating the overlapping membership functions of modes other than the bus on the left and the train on the right.**

To build the fuzzy system for mode detection, our initial step involved segmenting the trajectory into single-mode segments. This segmentation process relied on identifying trip start and end points and mode transition points, which were extracted from the pre-processed GPS data. Typically, walking served as the mode for transitioning between different modes (Xu et al., 2010). Our procedure began by classifying each data point as either walk or non-walk based on whether its speed was less than or equal to 2.5 meters per second or higher. Subsequently, we examined the data for segments where walk

points consistently persisted for a minimum duration of 120 seconds; these segments were labelled as walk segments. All consecutive points located between the start, walk segments, and end points of the trip were clustered as single-mode segments. These segments, including the walk segment, were then used as input for the fuzzy system. The inclusion of walk segments served as a reliability check to ensure the accurate identification of these segments. The outcome of this process was the detection of the travel mode for each segment and consequently, the whole trip, expressed as a percentage breakdown of the different modes utilized. For example, a trip could be entirely composed of walking, while another might consist of 25% walking, 50% car, and 25% bus usage.

To incorporate the proximity to bus routes and proximity to train routes in our model, we used the Open Street Map (OSM) dataset to extract train and bus routes of any kind. Though its coverage around the world varies (Mashhadi et al., 2013), it can be considered one of the most accurate and publicly available datasets (Barron et al., 2014; Girres & Touya, 2010; Haklay, 2010), and it has comprehensive coverage in England.

#### 2.3.4.2 Daylight

To determine whether a trip occurred during the day or night, we used the Astral package in Python (Kennedy, 2022). By considering the coordinates and timestamp of each trip's midpoint, we ascertained whether the trip took place during daylight or nighttime hours. This process generated a binary attribute for each trip, with a value of 1 signifying that the majority of the trip occurred during daylight hours (between sunrise and sunset), and a value of 0 indicating nighttime.

#### 2.3.5 Statistical analysis

Linear mixed-effect models were used to study self-reported happiness and stress levels in a multivariate modelling framework. Individual random effects were used to account for multiple trips of the same individual. Environmental factors were included as key independent variables all at once to measure their relative importance. Potential

confounding factors include trip attributes and destination attributes (hereafter control variables). The individual-level fixed effects incorporated into our analysis were gender, age, hometown, and the distance from the individual's home to their respective hometown city center. The trip attributes include the travel mode, the duration of the trip, and the season in which the trip occurred. The travel destination attributes include the type of destination, the activities in which the individual engaged at the destination, and whether the individual was alone or accompanied by others.

We ran three versions of the model: with the environmental variables only, with the control variables only, and with both the environmental and control variables. The model fit was assessed using the marginal  $R^2$  and conditional  $R^2$  values which determine, respectively, the proportional variance explained by the fixed effects-only, and the proportional variance explained by both the fixed and random effects (Nakagawa & Schielzeth, 2013).

### 2.3.6 Sample description

The sample contains a greater number of female participants and an age range primarily concentrated in the mid-career stages (Table 2-1). A larger proportion of participants was recruited in Leeds, followed by Brighton and Hove. Birmingham has the smallest sub-sample size due to a shorter recruitment period. Most participants lived relatively close to the city center.

**Table 2-1 Summary of individual-level variables of n = 606 individuals tracked by a mobile app.**

<b>Variable</b>		<b>Freq.</b>
<b>Gender</b>	Women	348
	Men	258
<b>Age</b>	18-24	75
	25-34	142
	35-44	168
	45-54	147
	55-64	74
<b>City</b>	Leeds	327
	Brighton and Hove	202

	Birmingham	77
<u>Distance from Home to City Center (km) – Mean (Sd)</u>		<u>6.04 (4.84)</u>

The reported happiness and stress levels after the trips averaged zero, which was expected as these values were adjusted based on individuals' average happiness and stress levels. The corresponding standard deviations were 1.02 and 1.11, respectively. Trip duration is on average 18.7 minutes (sd = 18.2), with 69% occurring during the day. A minority of trips (12%) took place in rainy conditions. On average, 13% of the area within a 50-meter buffer of each trip consists of green-blue spaces (Table 2-2). The most prevalent modes of transportation are walking and using the car. The number of trips was roughly the same in winter and autumn but trips in spring and summer are underrepresented in the data. The most common travel destination is the home, followed by work. The dominant activity reported at these destinations is work-related.

**Table 2-2 Summary statistics of environmental and control trip-level variables associated with n = 8083 trips by n = 606 individual tracked by a mobile app to study daily travel patterns in England.**

<b>Variable</b>	<b>Mean (Sd)</b>
<b>Dependent Variables</b>	
Happiness	0 (1.02)
Stress	0 (1.11)
<b>Independent Variables</b>	
<b>Environmental Factors</b>	
<b>Proximity to Nature</b>	
Exposure to Green-Blue Spaces (proportion of trip)	0.13 (0.19)
<b>Exposure to Weather</b>	
Apparent Temperature (c)	3.10 (4.58)
Visibility (m)	1188 (939)
Rain (binary)	0.12
<b>Exposure to Air pollution</b>	
Air Quality Index	3.39 (1.78)
<b>Control Variables</b>	
<b>Trip Attributes</b>	
<b>Travel mode</b>	
Walk	0.376 (0.39)
Run	0.016 (0.09)
Bike	0.045 (0.17)
Bus	0.043 (0.17)
Train	0.002 (0.04)
Car	0.516 (0.44)
Duration (min)	18.70 (18.2)

Daylight (binary)		0.69
		<b>Freq.</b>
<b>Season</b>	Winter	3200
	Spring	1785
	Summer	3
	Autumn	3095
<b>Destination Type</b>	Home	3475
	Work	1284
	Other	3324
<b>Destination Activity</b>	Work	5767
	Housework	451
	Leisure	1797
	Eating	23
	Other	45
<b>Presence of People at the destination</b>	Alone	4087
	Not Alone	3996

## 2.4 Results

When individual characteristics and trip and destination attributes are accounted for, the exposure to green and blue spaces during trips shows an association with both happiness and stress (**Table 2-3**). Increasing the area within a 50-meter buffer of each trip from zero green-blue space to completely composed of green-blue spaces, we see happiness levels increase by 0.17 points on the Likert scale while keeping all other variables constant, and stress levels decrease by 0.2 points on the Likert scale while keeping all other variables constant. These results support a relatively high effectiveness of green-blue spaces in improving happiness and reducing stress associated with urban travel. Creating a travel environment with higher levels of green-blue spaces is an achievable goal, particularly for short trips around residential or work areas where active transportation modes are typically used for daily activities.

Apparent temperature during trips was positively associated with happiness levels, where a 1-degree increase raised happiness by 0.01 points on the Likert scale while keeping all other variables constant, but showed no significant association with stress level (Table 2-3). The positive association we observed between apparent temperature and happiness may be explained by previous studies, which indicate that individuals engage more in

leisure and fun activities (Kim & Brown, 2022; Pivarnik et al., 2003), and have a better mood (Keller et al., 2005) during warmer days and seasons.

The positive association between visibility during trips and happiness levels, where a 1 km increase raises happiness by 0.05 points on the Likert scale while keeping all other variables constant, coupled with the negative association with stress levels, where a 1 km increase reduces stress by 0.04 points on the Likert scale while keeping all other variables constant, (Table 2-3). This suggests poor visibility may heighten feelings of uncertainty and vulnerability. Second, from a practical standpoint, better visibility is crucial for safe navigation and orientation during travel. Individuals likely experience increased comfort and reduced anxiety when they can clearly perceive their surroundings, anticipate potential hazards, and feel confident in their ability to respond to the environment.

Our findings revealed that travel in rainy weather can reduce stress levels, yet it showed no significant impact on happiness levels. This contrasts with Ettema et al.'s, (2017) research which indicated that rain or snow led to a decrease in participants' feelings of pleasure post-commute. The discrepancy in findings can be attributed to our inclusion of the horizontal visibility variable in the model. Rainy conditions inherently decrease visibility and increase travel hazards. By including the horizontal visibility variable in our model, we likely offset the impact of rain on visibility, explaining the absence of a relationship with happiness and the positive relationship with stress levels observed in our findings.

The air quality index during trips is neither related with happiness nor stress (Table 2-3). This contrasts previous studies that have shown that air pollution negatively affects overall subjective wellbeing (Lu, 2020). This is likely because our study primarily examined immediate happiness and stress levels rather than long-term health outcomes like respiratory issues or the prolonged psychological effects of pollution. By adjusting survey responses to the individuals' average, our analysis primarily captures the short-term influence of these environmental factors on happiness and stress, potentially sidelining their long-term wellbeing impacts.



**Table 2-3 Linear mixed-effect regressions of happiness and stress level with and without control variables**

Variables	Happiness		Stress		Happiness		Stress	
	$\beta$	p	$\beta$	p	$\beta$	p	$\beta$	p
(Intercept)	<b>-0.10</b>	<b>0.004</b>	<b>0.10</b>	<b>0.005</b>	0.14	0.102	-0.03	0.728
<b>Independent Variables</b>								
<b>Environmental Factors</b>								
<b>Proximity to Nature</b>								
Green-Blue Spaces	0.06	0.336	-0.08	0.248	<b>0.17</b>	<b>0.013</b>	<b>-0.20</b>	<b>0.010</b>
<b>Exposure to Weather</b>								
Apparent Temperature	0.00	0.062	-0.00	0.473	<b>0.01</b>	<b>0.001</b>	-0.00	0.174
Visibility	<b>0.06</b>	<b>&lt;0.001</b>	<b>-0.05</b>	<b>0.001</b>	<b>0.05</b>	<b>&lt;0.001</b>	<b>-0.04</b>	<b>0.005</b>
Rain	0.03	0.391	<b>-0.08</b>	<b>0.050</b>	0.04	0.251	<b>-0.08</b>	<b>0.039</b>
<b>Exposure to Air pollution</b>								
Air Quality Index	0.00	0.961	-0.01	0.238	-0.00	0.594	-0.01	0.420
<b>Control Variables</b>								
<b>Individual Factors</b>								
Gender [RC <sup>a</sup> : Female] Male					0.04	0.140	-0.00	0.880
<b>Age [RC: 18-24]</b>								
25-34					0.07	0.183	-0.11	0.061
35-44					0.06	0.236	<b>-0.13</b>	<b>0.029</b>
45-54					0.02	0.674	-0.09	0.130
55-64					0.01	0.823	-0.06	0.385
<b>City [RC: Leeds]</b>								
Brighton and Hove					-0.04	0.163	0.06	0.083
Birmingham					0.01	0.732	0.05	0.297
Distance from Home to City Center					-0.00	0.375	<b>0.01</b>	<b>0.002</b>
<b>Trip Attributes</b>								
<b>Travel Modes</b>								
Walk					0.02	0.761	-0.05	0.362
Run					0.04	0.746	-0.15	0.267
Bike					-0.04	0.629	-0.01	0.886
Bus					-0.14	0.070	0.08	0.352
Train					-0.30	0.286	0.21	0.491
Car					-0.08	0.091	0.01	0.840
Duration					-0.00	0.088	0.00	0.328
Daylight					<b>-0.28</b>	<b>&lt;0.001</b>	<b>0.23</b>	<b>&lt;0.001</b>
<b>Season [RC: Winter]</b>								
Spring					0.05	0.176	-0.06	0.113
Summer					0.26	0.657	-0.02	0.969
Autumn					<b>-0.08</b>	<b>0.012</b>	-0.03	0.428
<b>Destination Attributes</b>								
<b>Destination Type [RC: Home]</b>								
Work					0.05	0.179	-0.05	0.284
Other					0.01	0.821	0.02	0.634
<b>Destination Activity [RC: Work]</b>								
Housework					0.00	0.979	0.03	0.664
Leisure					0.00	0.994	0.07	0.059
Eating					-0.34	0.121	0.19	0.427
Other					0.04	0.832	-0.11	0.570
Presence of People at the Destination [RC: Alone] Not Alone					-0.02	0.556	-0.03	0.380
Number of observation	8083		8083		8083		8083	
Marginal R2 / Conditional R2	0.004 / 0.014		0.002 / 0.018		0.022 / 0.030		0.016 / 0.031	

Note: Bold numbers indicate significant association at a p-level of 0.05.

<sup>a</sup> RC: Reference Category

Overall, the models exhibited relatively low explanatory power with marginal  $R^2$  values  $\leq 0.022$  and conditional  $R^2$  values  $\leq 0.031$ . The control-only model had the lowest marginal  $R^2$  and conditional  $R^2$  for both happiness and stress levels (Appendix A). The model with both control and environmental factors (Table 2-3) yielded the highest marginal  $R^2$  and conditional  $R^2$  values for both happiness and stress during trips.

## 2.5 Discussion

Among the four environmental factors identified as significantly influencing stress and happiness, green-blue spaces are the most actionable in terms of policy changes in urban environments. The observed effects of green-blue spaces on happiness and stress during travel are consistent with previous findings, which demonstrate advantages for individuals who have access to, or are present in, such spaces. This points at a consistent impact of environmental exposure on wellbeing (De Vries et al., 2016; Kondo et al., 2020) and adds a critical role of green-blue spaces not only within residential areas but also along streets and in transit spaces throughout urban environments. Incorporating more green-blue areas along streets not only enhances perceptions of pleasantness, quietness, and safety (Nawrath et al., 2019; Zhu et al., 2022), but also promotes active travel behaviors (Wu et al., 2020).

Transportation mode was not significantly associated with self-reported happiness and stress levels after trips. Previous studies have suggested that certain modes of transportation, such as active transportation (walking, cycling) and private transportation (car use), can positively influence happiness levels and reduce stress (Chen et al., 2019; Duarte et al., 2010; Eriksson et al., 2013; Fan et al., 2019). Our results, however, indicate the mode of transportation itself does not significantly impact happiness or stress levels; a result that was consistent when we only considered the control variables (Supplementary Material II). This suggests that the surrounding environment during travel may play a more pivotal role in influencing subjective wellbeing than the mode of transportation. This finding is particularly significant for urban planners and policymakers, as it highlights the potential of environmental enhancements over transportation mode

changes to improve psychological wellbeing. This highlights the need for additional research into the direct effects of transportation mode on mental health. It is important for future studies to investigate the connections between transportation choices and mental health, taking into account the influence of environmental factors.

As happiness and stress are complex responses to capture in survey data and, therefore, often difficult to measure (Plutchik, 2001; Pykett et al., 2020), we tried to control this complexity by adjusting the happiness and stress levels by individuals' average responses. However, there are numerous other factors not controlled for in our analysis, which could potentially be additional determinants of stress and happiness levels that remain unaccounted for. The subjective nature of responses to surveys regarding happiness and stress are influenced not only by environmental factors but also by numerous other factors, including individual genetics and personal character (Bevilacqua & Goldman, 2011), interpersonal connections (Ogihara & Uchida, 2014), subjective characteristics, including perceived accessibility, safety, and social-spatial factors like inequality (Ettema & Schekkerman, 2016), as well as one's relative position within the neighborhood (Brodeur & Flèche, 2012). Therefore, it is likely that the complexity of individual happiness and stress levels may limit the explanatory power of our models (as observed here, overall model fit was low ( $R^2 < 3\%$ ). This finding aligns with previous studies that have used a variety of explanatory variables and techniques to explore the relationship between environmental factors such as greenspace and happiness and stress levels, which have also reported varying, but generally low, explanatory powers (Hazer et al., 2018; Roe et al., 2013).

We employed the UKCEH land cover dataset, derived from satellite images (Sentinel-2 Seasonal Composite Images) for extracting green areas, employing a top-down perspective. However, it is important to acknowledge that the perception of people on the street may differ because these satellite images disregard the vertical dimension of greenery (Gupta et al., 2012). To gain a better understanding of how individuals perceive greenery on the street, several efforts have been made to use street view images for

greenery extraction (Dong et al., 2018; Larkin & Hystad, 2019; Long & Liu, 2017). However, research by (Torkko et al., 2023) has shown that both street view-based methods and satellite imagery-based methods exhibit a high correlation with individuals' perceived greenery.

In future research directions, reducing subjectivity in analyses is paramount when assessing individuals' experiences within urban environments. Ambulatory assessment, with an emphasis on biosensing, emerges as an invaluable tool for this purpose. This methodology enables the simultaneous and real-time capture of both physiological indicators (such as heart rate and cortisol levels) using biosensors and psychological data (including emotional states and stress levels) using real-time surveys as individuals navigate urban surroundings (Pykett et al., 2020; Wilhelm & Grossman, 2010). However, it is crucial to acknowledge that the acquisition of such data can be challenging due to practical considerations such as the burden placed on individuals to carry specialized and often expensive biosensing devices dedicated to this purpose.

## 2.6 Conclusion

The primary aim of this study was to investigate how environmental factors influence individuals' self-reported happiness and stress levels in relation to daily travel. The emphasis has been on evaluating environmental factors related to trips, advancing previous studies that predominantly focused on understanding the impact of various environmental factors, particularly greenspaces, on happiness and/or stress levels based on individuals' residential location (Krekel et al., 2016; Wendelboe-Nelson et al., 2019; Y. Zhang et al., 2017), and not during daily travel. For instance, our measurement involves assessing the proportion of the trip covered by green-blue spaces based on area during a trip as a proxy to the exposure to green-blue spaces during the trip, rather than solely assessing access or presence at green-blue spaces. In our approach, we employed GEMA surveys and continuous individual tracking, offering a more comprehensive view of the travel experience. This entails not only detailed information about paths taken during trips but also the immediate post-trip happiness and stress levels. This nuanced

perspective provides a more holistic understanding of the interplay between travel experiences and happiness and stress.

The findings reveal that environmental factors, notably green and blue spaces, play a more substantial role in influencing urban happiness and stress levels than either the mode of transportation or individual trip characteristics. The consistent positive impact of green-blue spaces on wellbeing, regardless of the context—whether within residential areas, or along transit routes—underscores their importance in urban planning and public health strategies. This strategic placement of green-blue spaces not only within residential areas but also along routes and in transit areas throughout urban environments enhances the aesthetic and experiential quality of daily commutes and activities, contributing significantly to urban wellbeing. Hence, our research advocates for a holistic approach in urban planning and public health initiatives, emphasizing the need to distribute green and blue spaces across the urban landscape, including areas frequented by commuters and pedestrians. Such an approach ensures that the benefits of these natural spaces are accessible to a broader segment of the urban population, thereby maximizing their positive impact on public health and wellbeing.

## References

- Ambrey, C., & Fleming, C. (2014). Public greenspace and life satisfaction in urban Australia. *Urban Studies*, *51*(6), 1290–1321.
- Archer, M., Paleti, R., Konduri, K. C., Pendyala, R. M., & Bhat, C. R. (2013). Modeling the connection between activity-travel patterns and subjective well-being. *Transportation Research Record*, *2382*(1), 102–111.
- Ayres, J. G. (2011). *Review of the UK Air Quality Index: A Report by the Committee on the Medical Effects of Air Pollutants*.
- Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877–895.

Bevilacqua, L., & Goldman, D. (2011). Genetics of emotion. *Trends in Cognitive Sciences*, 15(9), 401–408.

Böcker, L., Dijst, M., & Faber, J. (2016). Weather, transport mode choices and emotional travel experiences. *Transportation Research Part A: Policy and Practice*, 94, 360–373.

Brodeur, A., & Flèche, S. (2012). Where the streets have a name: Income comparisons in the US. Available at SSRN 2149727.

Brum-Bastos, V. S., Long, J. A., & Demšar, U. (2018). Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*, 71, 131–152.

Čelić, I., Živanović, S., & Pavlović, N. (2019). The effects of weather conditions on the health of people living in urban and rural environments. *Економика Пољопривреде*, 66(1), 63–76.

Chen, S., Fan, Y., Cao, Y., & Khattak, A. (2019). Assessing the relative importance of factors influencing travel happiness. *Travel Behaviour and Society*, 16, 185–191.

Cheng, W.-L., Kuo, Y.-C., Lin, P.-L., Chang, K.-H., Chen, Y.-S., Lin, T.-M., & Huang, R. (2004). Revised air quality index derived from an entropy function. *Atmospheric Environment*, 38(3), 383–391.

<https://doi.org/https://doi.org/10.1016/j.atmosenv.2003.10.006>

Clark, B., Chatterjee, K., Martin, A., & Davis, A. (2020). How commuting affects subjective wellbeing. *Transportation*, 47, 2777–2805.

Dadvand, P., Bartoll, X., Basagaña, X., Dalmau-Bueno, A., Martinez, D., Ambros, A., Cirach, M., Triguero-Mas, M., Gascon, M., & Borrell, C. (2016). Green spaces and General Health: Roles of mental health status, social support, and physical activity. *Environment International*, 91, 161–167.

De Vries, S., Ten Have, M., van Dorsselaer, S., van Wezep, M., Hermans, T., & de Graaf, R. (2016). Local availability of green and blue space and prevalence of common mental disorders in the Netherlands. *BJPsych Open*, 2(6), 366–372.

Dept. for Transport, U. (2021). *National Travel Survey: 2021*. National Statistics. <https://www.gov.uk/government/statistics/national-travel-survey-2021>

Dong, R., Zhang, Y., & Zhao, J. (2018). How green are the streets within the sixth ring road of Beijing? An analysis based on tencent street view pictures and the green view index. *International Journal of Environmental Research and Public Health*, 15(7), 1367.

Duarte, A., Garcia, C., Giannarakis, G., Limão, S., Polydoropoulou, A., & Litinas, N. (2010). New approaches in transportation planning: happiness and transport economics. *NETNOMICS: Economic Research and Electronic Networking*, 11, 5–32.

Eriksson, L., Friman, M., & Gärling, T. (2013). Perceived attributes of bus and car mediating satisfaction with the work commute. *Transportation Research Part A: Policy and Practice*, 47, 87–96. <https://doi.org/https://doi.org/10.1016/j.tra.2012.10.028>

Ettema, D., Friman, M., Gärling, T., Olsson, L. E., & Fujii, S. (2012). How in-vehicle activities affect work commuters' satisfaction with public transport. *Journal of Transport Geography*, 24, 215–222.

Ettema, D., Friman, M., Olsson, L. E., & Gärling, T. (2017). Season and weather effects on travel-related mood and travel satisfaction. *Frontiers in Psychology*, 8, 140.

Ettema, D., & Schekkerman, M. (2016). How do spatial characteristics influence well-being and mental health? Comparing the effect of objective and subjective characteristics at different spatial scales. *Travel Behaviour and Society*, 5, 56–67.

Fan, Y., Brown, R., Das, K., & Wolfson, J. (2019). Understanding trip happiness using smartphone-based data: the effects of trip-and person-level characteristics. *Findings*.

- Fan, Y., Das, K. V., & Chen, Q. (2011). Neighborhood green, social support, physical activity, and stress: Assessing the cumulative impact. *Health & Place, 17*(6), 1202–1211.
- Girres, J., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS, 14*(4), 435–459.
- Guite, H. F., Clark, C., & Ackrill, G. (2006). The impact of the physical and urban environment on mental well-being. *Public Health, 120*(12), 1117–1126.  
<https://doi.org/https://doi.org/10.1016/j.puhe.2006.10.005>
- Gupta, K., Kumar, P., Pathan, S. K., & Sharma, K. P. (2012). Urban Neighborhood Green Index—A measure of green spaces in urban areas. *Landscape and Urban Planning, 105*(3), 325–335.
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design, 37*(4), 682–703.
- Hazer, M., Formica, M. K., Dieterlen, S., & Morley, C. P. (2018). The relationship between self-reported exposure to greenspace and human stress in Baltimore, MD. *Landscape and Urban Planning, 169*, 47–56.  
<https://doi.org/https://doi.org/10.1016/j.landurbplan.2017.08.006>
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*(5702), 1776–1780.
- Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., Conway, A., & Wager, T. (2005). A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological Science, 16*(9), 724–731.
- Kennedy, S. (2022). *Astral Package* (3.0). <https://astral.readthedocs.io/en/latest/>



Kim, Y., & Brown, R. (2022). Effect of meteorological conditions on leisure walking: a time series analysis and the application of outdoor thermal comfort indexes. *International Journal of Biometeorology*, *66*(6), 1109–1123.

Kirchner, T. R., & Shiffman, S. (2016). Spatio-temporal determinants of mental health and well-being: advances in geographically-explicit ecological momentary assessment (GEMA). *Social Psychiatry and Psychiatric Epidemiology*, *51*, 1211–1223.

Kondo, M. C., Triguero-Mas, M., Donaire-Gonzalez, D., Seto, E., Valentín, A., Hurst, G., Carrasco-Turigas, G., Masterson, D., Ambròs, A., & Ellis, N. (2020). Momentary mood response to natural outdoor environments in four European cities. *Environment International*, *134*, 105237.

Kou, L., Tao, Y., Kwan, M.-P., & Chai, Y. (2020). Understanding the relationships among individual-based momentary measured noise, perceived noise, and psychological stress: A geographic ecological momentary assessment (GEMA) approach. *Health & Place*, *64*, 102285.

Krekel, C., Kolbe, J., & Wüstemann, H. (2016). The greener, the happier? The effect of urban land use on residential well-being. *Ecological Economics*, *121*, 117–127.

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PloS One*, *9*(6), e96180.

Larkin, A., & Hystad, P. (2019). Evaluating street view exposure measures of visible green space for health research. *Journal of Exposure Science & Environmental Epidemiology*, *29*(4), 447–456.

Lee, A. C. K., & Maheswaran, R. (2011). The health benefits of urban green spaces: a review of the evidence. *Journal of Public Health*, *33*(2), 212–222.

- Li, X., Wang, Y., Tang, J., Shi, L., Zhao, T., & Chen, J. (2022). Emotional wellbeing in intercity travel: Factors affecting passengers' long-distance travel moods. *Frontiers in Public Health, 10*, 1046922.
- Long, Y., & Liu, L. (2017). How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View. *PloS One, 12*(2), e0171110.
- Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology, 32*, 52–65.
- Ma, J., Liu, G., Kwan, M.-P., & Chai, Y. (2021). Does real-time and perceived environmental exposure to air pollution and noise affect travel satisfaction? evidence from Beijing, China. *Travel Behaviour and Society, 24*, 313–324.
- Maas, J., Verheij, R. A., Groenewegen, P. P., De Vries, S., & Spreeuwenberg, P. (2006). Green space, urbanity, and health: how strong is the relation? *Journal of Epidemiology & Community Health, 60*(7), 587–592.
- Mashhadi, A., Quattrone, G., & Capra, L. (2013). Putting ubiquitous crowd-sourcing into context. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 611–622.
- Matsuoka, R. H., & Kaplan, R. (2008). People needs in the urban landscape: analysis of landscape and urban planning contributions. *Landscape and Urban Planning, 84*(1), 7–19.
- Mennis, J., Mason, M., & Ambrus, A. (2018). Urban greenspace is associated with reduced psychological stress among adolescents: A Geographic Ecological Momentary Assessment (GEMA) analysis of activity space. *Landscape and Urban Planning, 174*, 1–9.
- Met Office. (2006a). *MIDAS UK Hourly Rainfall Data*.  
<https://catalogue.ceda.ac.uk/uuid/bbd6916225e7475514e17fdbf11141c1>

- Met Office. (2006b). *MIDAS: UK Hourly Weather Observation Data*.  
<https://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c>
- Mokhtarian, P. L., & Pendyala, R. M. (2018). Travel satisfaction and well-being. *Quality of Life and Daily Travel*, 17–39.
- Morton, R. D., Marston, C. G., O'Neil, A. W., & Rowland, C. S. (2020). *Land Cover Map 2019 (20m classified pixels, GB)*. NERC Environmental Information Data Centre.  
<https://doi.org/10.5285/643eb5a9-9707-4fbb-ae76-e8e53271d1a0>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
- Nawrath, M., Kowarik, I., & Fischer, L. K. (2019). The influence of green streets on cycling behavior in European cities. *Landscape and Urban Planning*, 190, 103598.
- Nes, R. B., & Røysamb, E. (2017). Happiness in behaviour genetics: An update on heritability and changeability. *Journal of Happiness Studies*, 18, 1533–1552.
- Ogihara, Y., & Uchida, Y. (2014). Does individualism bring happiness? Negative effects of individualism on interpersonal relationships and happiness. *Frontiers in Psychology*, 5, 135.
- Olsson, L. E., Gärling, T., Ettema, D., Friman, M., & Fujii, S. (2013). Happiness and satisfaction with work commute. *Social Indicators Research*, 111, 255–263.
- Pain, R., & Smith, S. J. (2010). Introduction: Geographies of wellbeing. *The SAGE Handbook of Social Geographies*, 299–308.
- Parrish, E. M., Depp, C. A., Moore, R. C., Harvey, P. D., Mikhael, T., Holden, J., Swendsen, J., & Granholm, E. (2020). Emotional determinants of life-space through GPS

and ecological momentary assessment in schizophrenia: what gets people out of the house? *Schizophrenia Research*, 224, 67–73.

Pivarnik, J. M., Reeves, M. J., & Rafferty, A. P. (2003). Seasonal variation in adult leisure-time physical activity. *Medicine & Science in Sports & Exercise*, 35(6), 1004–1008.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344–350.

Pykett, J., Osborne, T., & Resch, B. (2020). From Urban Stress to Neurourbanism: How Should We Research City Well-Being? *Annals of the American Association of Geographers*, 110(6), 1936–1951. <https://doi.org/10.1080/24694452.2020.1736982>

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128(6), 934.

Roe, J. J., Thompson, C. W., Aspinall, P. A., Brewer, M. J., Duff, E. I., Miller, D., Mitchell, R., & Clow, A. (2013). Green space and stress: Evidence from cortisol measures in deprived urban communities. *International Journal of Environmental Research and Public Health*, 10(9), 4086–4103.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5), 881–906. <https://doi.org/10.1080/13658816.2015.1100731>

Singleton, P. A., & Clifton, K. J. (2021). Towards measures of affective and eudaimonic subjective well-being in the travel domain. *Transportation*, 48(1), 303–336. <https://doi.org/10.1007/s11116-019-10055-1>

- Smyth, R., Mishra, V., & Qian, X. (2008). The environment and well-being in urban China. *Ecological Economics*, 68(1–2), 547–555.
- Steadman, R. G. (1994). Norms of apparent temperature in Australia. *Aust. Met. Mag*, 43, 1–16.
- Stutzer, A., & Frey, B. S. (2008). Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics*, 110(2), 339–366.
- Sugiyama, T., Leslie, E., Giles-Corti, B., & Owen, N. (2008). Associations of neighbourhood greenness with physical and mental health: do walking, social coherence and local social interaction explain the relationships? *Journal of Epidemiology & Community Health*, 62(5), e9–e9.
- Tan, X., Han, L., Zhang, X., Zhou, W., Li, W., & Qian, Y. (2021). A review of current air quality indexes and improvements under the multi-contaminant air pollution exposure. *Journal of Environmental Management*, 279, 111681.
- Thompson, C. W., Roe, J., Aspinall, P., Mitchell, R., Clow, A., & Miller, D. (2012). More green space is linked to less stress in deprived communities: Evidence from salivary cortisol patterns. *Landscape and Urban Planning*, 105(3), 221–229.
- Torkko, J., Poom, A., Willberg, E., & Toivonen, T. (2023). How to best map greenery from a human perspective? Comparing computational measurements with human perception. *Frontiers in Sustainable Cities*, 5, 1160995.
- UK Air Information Resource. (2023). *UK-AIR database*. [https://uk-air.defra.gov.uk/data/data\\_selector\\_service?](https://uk-air.defra.gov.uk/data/data_selector_service?)
- Veenhoven, R. (1991). Is happiness relative? *Social Indicators Research*, 24, 1–34.
- Wang, R., Feng, Z., Pearce, J., Zhou, S., Zhang, L., & Liu, Y. (2021). Dynamic greenspace exposure and residents' mental health in Guangzhou, China: From over-head

to eye-level perspective, from quantity to quality. *Landscape and Urban Planning*, 215, 104230.

Wendelboe-Nelson, C., Kelly, S., Kennedy, M., & Cherrie, J. W. (2019). A scoping review mapping research on green space and associated mental health benefits. *International Journal of Environmental Research and Public Health*, 16(12), 2081.

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological Science*, 24(6), 920–928.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, 84(3), 552–569.

Winters, J. V., & Li, Y. (2017). Urbanisation, natural amenities and subjective well-being: Evidence from US counties. *Urban Studies*, 54(8), 1956–1973.

Wu, J., Wang, B., Ta, N., Zhou, K., & Chai, Y. (2020). Does street greenery always promote active travel? Evidence from Beijing. *Urban Forestry & Urban Greening*, 56, 126886.

Xu, C., Ji, M., Chen, W., & Zhang, Z. (2010). Identifying travel mode from GPS trajectories through fuzzy pattern recognition. *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, 2, 889–893.

Zhang, X., Zhou, S., Kwan, M.-P., Su, L., & Lu, J. (2020). Geographic Ecological Momentary Assessment (GEMA) of environmental noise annoyance: The influence of activity context and the daily acoustic environment. *International Journal of Health Geographics*, 19(1), 1–13.

Zhang, Y., Van den Berg, A. E., Van Dijk, T., & Weitkamp, G. (2017). Quality over quantity: Contribution of urban green space to neighborhood satisfaction. *International Journal of Environmental Research and Public Health*, *14*(5), 535.

Zhu, M., Sze, N. N., & Newnam, S. (2022). Effect of urban street trees on pedestrian safety: A micro-level pedestrian casualty model using multivariate Bayesian spatial approach. *Accident Analysis & Prevention*, *176*, 106818.

## Chapter 3

### 3 Mobility deviation index: Incorporating geographical context into analysis of human mobility<sup>1</sup>

#### 3.1 Introduction

Spatial and temporal patterns of the movement of individuals are shaped by the configuration of the environment in which they live, and this configuration is, in turn, shaped by people's movements (Haggett, 1966). The structure of the places within which an individual moves, including the affordances they provide (Jordan et al., 1998), can constrain or limit movement opportunities, or provision easy movement and access to activities (Sharif & Alesheikh, 2017). Therefore, when studying human mobility patterns, it is important to consider the relationship between observed movement patterns and the geographical context within which they occur (Brum-Bastos et al., 2021).

Failure to capture the context within which human movement occurs may alter our understanding and interpretation of human mobility. However, it is very rare in movement analysis to specifically adjust for the effects of the surrounding local environment when directly measuring human mobility levels. Therefore, measured mobility levels, and the outcomes of subsequent analyses (e.g., investigating the associations of human mobility with socio-economic variables), may be more reflective of the structural nature of the places in which the individual lives and moves. To elaborate, consider person A living in a suburban area who has to travel long distances to reach daily tasks due to the lack of amenities in their proximal environment. In contrast, person B lives in an urban area, and all the amenities they need are relatively proximal. If both persons A and B exhibit similar mobility patterns and levels, existing measurement

---

<sup>1</sup> A version of this chapter has been published (Malekzadeh, M., & Long, J. A. Mobility deviation index: incorporating geographical context into analysis of human mobility. *J Geogr Syst* (2024). <https://doi.org/10.1007/s10109-024-00444-1>). This chapter is included in this dissertation with permission under the Springer Nature License Number 5832511011688, issued on July 19, 2024, to Milad Malekzadeh.



methods quantify their mobility levels equally. However, it is clear that person A had fewer proximal options for activities and thus required higher mobility levels to participate in those activities, while person B had proximal choices but chose activities at greater distances resulting in similar mobility levels to person A. Quantifying and, consequently, comparing these two individuals' mobility patterns without appraising geographical context may lead to spurious associations in the resulting inferences and erroneous interpretations about their mobility patterns relative to other individuals and areas.

The geographical context within which a person moves will significantly influence the analytical results of geographical studies (e.g., human mobility studies) when attempting to associate an individual's movement with a single place (e.g., through aggregation). Purves et al., (2014) put forward different alternatives for defining geographical context in movement studies, such as (1) information derived from additional data which are collected concurrently with movement data, (2) information on the spaces within which movement can occur, and (3) information gained by comprehensive understanding of the processes being studied used to guide initial hypotheses. Brum-Bastos et al., (2021) also present a general definition of context as one or more variables that characterize the external or internal conditions that led to a movement decision or caused specific movement behavior. In this study, we focus only on the external conditions of individual movement – which we call geographical context, as we do not consider the internal conditions of the individual. To be specific, here, our current focus is assessing the availability of nearby amenities as a means to integrate contextual factors into the measurement of human mobility.

Currently, the prevailing practice often involves classifying an individual's residential region to quantify the extent of urbanization or suburbanization (Noi et al., 2022; Turner & Niemeier, 1997), or incorporating the geographical context indirectly into statistical models through proxies such as proximity to downtown (Cheng et al., 2022; Ma et al., 2018) or density of amenities (Barthelmes et al., 2022; Yue et al., 2017). Despite their

utility, these approaches possess certain limitations. The categorization of areas is typically executed through arbitrary definitions, frequently relying on factors like population density (e.g., the official definition of population centers in Canada (Statistics Canada, 2017)). This introduces variability that can impact outcomes in different cases. Employing distance to downtown may lose effectiveness when the study area encompasses a mix of rural, suburban, and urban zones, as many of these regions lack a city center or downtown core for reference, or there are multiple centers.

In this paper, we address this research limitation in human mobility studies by integrating geographical contextual data directly into the measurement of human mobility. In doing so, we aim to rescale observed human mobility levels by their respective expected mobility level, computed by a model into which geographical contexts are integrated. We introduce the concept of *mobility deviation index* (MDI) which is both a general and relative measure of human mobility that aims to rescale observed human mobility measurements to account for geographical contextual factors.

The rest of the paper is organized as follows. We first introduce the concept of *mobility deviation index*. Then we define how we compute expected mobility, which serves as a measure to compare the observed mobility level against. We use a case study in Ontario, Canada, to demonstrate how mobility deviation index differs from traditional methods, when investigating human mobility's relationship with socio-economic variables. In our discussion, we elaborate on the added value of using mobility deviation index compared to existing mobility measurements and discuss some limitations and possible future directions for the use of mobility deviation index in human mobility studies.

## 3.2 Materials and methods

### 3.2.1 Human mobility data

Mobility data were obtained from a large, aggregated and de-identified network mobility dataset provided by TELUS Communications Inc. (see (Long & Ren, 2022)). In Ontario, the TELUS mobile network includes around 90,000 distinct tower receivers. As mobile

devices move, they transition their connection from one tower receiver to another, typically connecting to the closest one. The data comprises the beginning and end times of each tower connection, as well as the geographical location of the corresponding tower receiver. By analyzing the sequence of tower connections for each device, we can infer a basic trajectory of movement, identifying the series of tower receivers to which the device was connected. The data are then used to compute estimates of the mobility levels aggregated to Canadian census units (aggregated dissemination areas – ADAs; (Statistics Canada, 2016)). Aggregated dissemination areas usually consist of populations of ~5000-10000 individuals. The dataset comprises average estimates of mobility for the week of 02 February 2020 for each ADA. We limited our analysis to the  $n=1542$  ADA regions (out of 1685 in Ontario) which were determined as having sufficient data ( $> 100$  mobile devices) to both protect individual privacy and provide a reliable area-level estimate of mobility (Long & Ren, 2022).

### 3.2.2 Mobility Deviation Index

We take a practical definition for *mobility deviation index* simply as the ratio of observed mobility relative to the expected level of mobility given the context of a specific location. We adopt this formulation of the classical OE ratio:

$$MDI = \frac{\textit{Observed mobility}}{\textit{Expected mobility}} \quad 3.1$$

The value of MDI is technically bounded on  $[0, +\infty)$ , but has the simple interpretation that mobility deviation index is 1 if the observed value is equal to the expected value, greater than 1 if it is greater than expected, and less than 1 if it is lower than expected. The magnitude of MDI can be interpreted as the proportional difference in observed mobility relative to expected mobility. For example, a MDI value of 0.5 is interpreted as having half the level of mobility as expected, while a MDI value of 2 as double the expected level of mobility. In this sense, the distribution of MDI is expected to be asymmetric around 1. The  $\log(MDI)$  is anticipated to follow a normal distribution with a mean of 0. Therefore, interpretations should align with those of traditional OE ratios.

While in this study we use a measure of human mobility that captures the range of movement in our case study, it is worth mentioning that MDI is not limited to this type of measurement and the concept of MDI can be applied broadly on any type of human mobility measurement. The challenge in using MDI in analysis then involves deriving a suitable method for determining expected mobility for a given individual or location based on the chosen measure of human mobility. In the following, we first discuss the method with which we measured human mobility, and then we introduce four models to compute the expected mobility.

### 3.2.3 Observed mobility

We use a popular measure of human mobility – the radius of gyration (ROG) (González et al., 2008). ROG is defined as:

$$ROG = \sqrt{\frac{\sum_i^n (d_{i,c})^2}{n}} \quad 3.2$$

Where  $i$  is each activity location selected by an individual,  $c$  is the mean center of all the activity points in the trajectory,  $n$  is the number of activity points in the trajectory,  $d_{i,c}$  is the distance between the mean center to point  $i$ . ROG is computed regardless of transportation mode as its computation uses only activity points and their distance to the central point. ROG has been widely used in mobility studies due to its operability and low computational cost (Shi et al., 2015). Specifically, we employ a modified version of the classical ROG measure, as implemented by Long et al., (2021). We replaced the centroid with an estimate of home location in our modified version of ROG because of the importance of the home as central activity place in mobility patterns (Song et al., 2010). Extracting home locations is an essential component of many studies that wish to characterize human mobility (Chen et al., 2016; Hoteit et al., 2016; Xu et al., 2015) since much of human mobility behavior is tied to the home location, and several human movement models are developed by assuming humans' preferential return to their home or workplace (Song et al., 2010). These models make the assumption that people tend to

work in jobs closer to their home, at least in the distance they can easily commute to and from their workplace as well as purchasing goods for ease/convenience. Home-Work trips often influence other trips, too (Shearmur, 2006). This assumption is a cornerstone of the space-time prism concept (Hägerstrand, 1970) since shorter commutes augment individuals' available time budget, consequently increasing individuals' accessible area and enabling them to have more freedom in their activity selection.

### 3.2.4 Expected mobility

Previous efforts on simulating mobility patterns are either based on historical or statistical mobility information to predict the next location (Monreale et al., 2009; Song et al., 2017), or they are developed to support routing protocols in mobile networks (Ghosh et al., 2007; Mei & Stefa, 2009). Neither of these approaches are ideal in our case because we do not have information on individual movement trajectories (because of the nature of the network mobility data we are working with). Also, we aim to integrate geographical context into our measurement rather than find the next possible locations for individuals. Therefore, the process we employed is relatively simple in that we simulate an activity pattern by sampling from a set of possible activities, and then computing the ROG mobility measure for each simulation, which we call expected ROG (EROG). Generally, individuals choose locations for desired activities based on distance and attractiveness factors. We draw on previous models studying human mobility to determine two main criteria that influence the selection set of activities to be included in the calculation: (1) types of activities, and (2) geographical distance (Chen & Jia, 2019; Piovani et al., 2018; Tao et al., 2018; Vickerman, 1974). In our model, to simulate an activity pattern for each point, two primary inquiries need sequential addressing: "Which type of amenity should be chosen?" followed by "Which specific point of that type should be selected?" It is upon these two questions that we formulate a straightforward method for simulating mobility patterns.

Our approach is fundamentally a two-step approach: in the first step, we determine the number of activities of each type that a simulated mobility pattern should encompass. We

used the 2016 Transportation Tomorrow historical survey data conducted in the Greater Golden Horseshoe area of Southern Ontario, which is home to 7.826 million people, accounting for more than half of Ontario's total population of 14.57 million, (Government of Ontario, 2016) to guide how many activities of each type each simulated mobility pattern should include. Specifically, to assign relative proportions of trip types based on the survey data (Government of Ontario, 2016), we used the total number of trips to and from different activities such as work ( $\Phi_{work} = 42.2\%$ ), school ( $\Phi_{school} = 14.5\%$ ), shopping-related ( $\Phi_{shopping-related} = 17.3\%$ ), and other trips ( $\Phi_{other} = 25.9\%$ ). However, because of the nature of the data we used, we did not have information on work locations, therefore, we assumed that work locations were associated with available POIs and were evenly distributed between them. This assumption led to the determination of the final weights for each category, which are school ( $\Phi_{school} = 25\%$ ), shopping-related ( $\Phi_{shopping-related} = 29.9\%$ ), and other trips ( $\Phi_{other} = 44.8\%$ ). As an example, if a simulated mobility pattern is to include  $N=10$  activities, there would be  $n_{school} = 2$  chosen from the school category,  $n_{shops} = 3$  chosen from the shops category, and  $n_{other} = 5$  chosen from the other category (Figure 3-1).

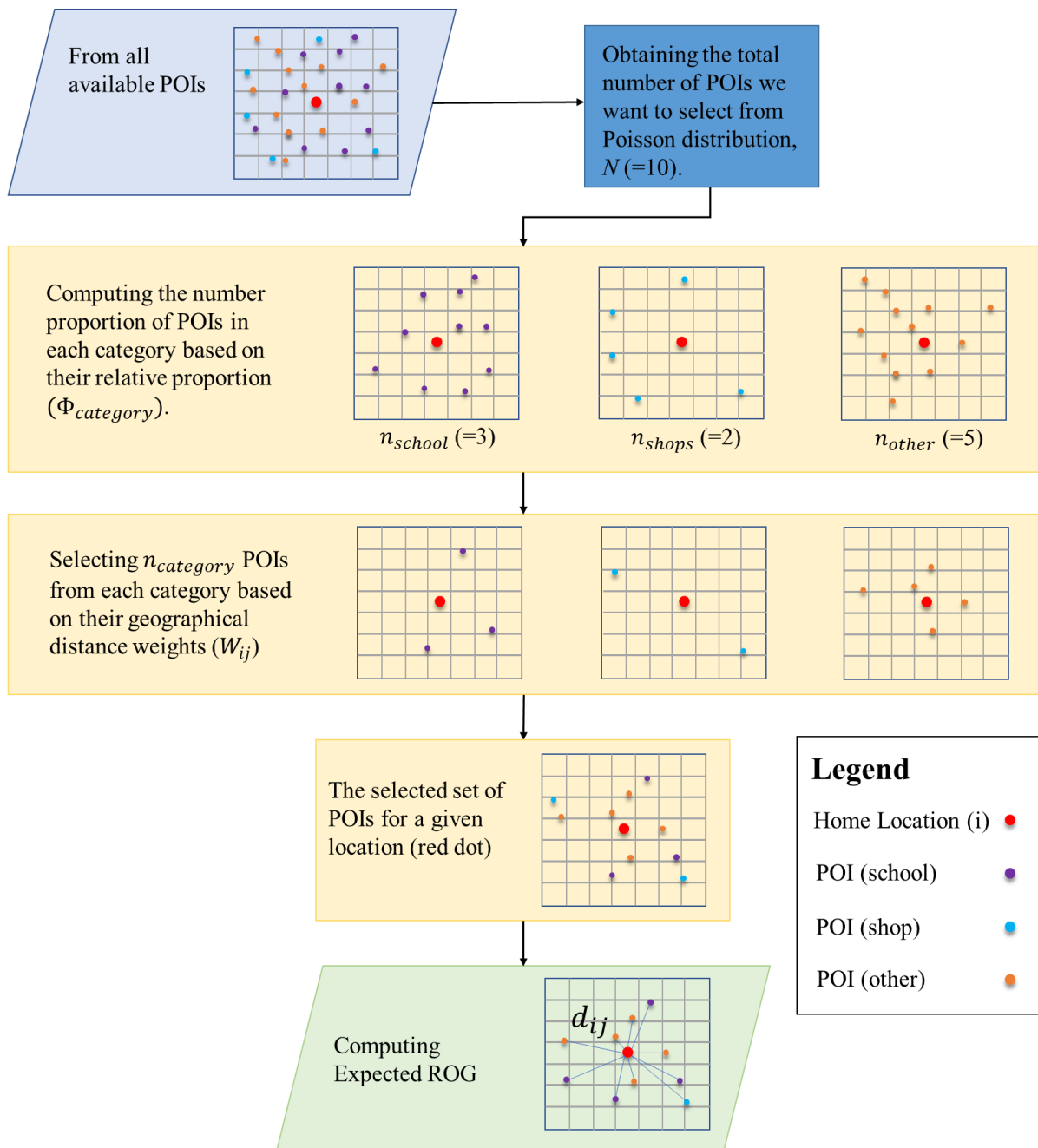
Having established the relative proportions of each type, we need to calculate the total number of activities ( $N$ ) within an activity pattern before proceeding to compute the number of activities for each type. We employed the Poisson distribution as the discrete probability distribution of the number of activities to simulate the total number of activities. The gamma parameter for the Poisson distribution was estimated by the average number of activities observed from the sample data (Zhao et al., 2020). Following this, we utilize the relative proportions attributed to each type to calculate the number of activities within each type.

We use point-of-interest (POI) data (from Open Street Map (OSM); (OpenStreetMap contributors, 2017)) as the activity points. We categorized the OSM POI data into the same three categories: school, shopping-related, and other (see Appendix B).

The second step is to use geographical distance to assign weights to POIs (Figure 3-1); the farther a POI, the less attractive it becomes. To incorporate distance decay, we modeled the weighting of POIs as an inverse relationship with distance (Haggett, 1966). We used a negative exponential function as the distance decay function, which is the most commonly used approach (Skov-Petersen, 2001) and is efficient in distinguishing local interactions (Y. Chen, 2008) 3.3.

$$W_{ij} = e^{-\theta d_{ij}} \quad 3.3$$

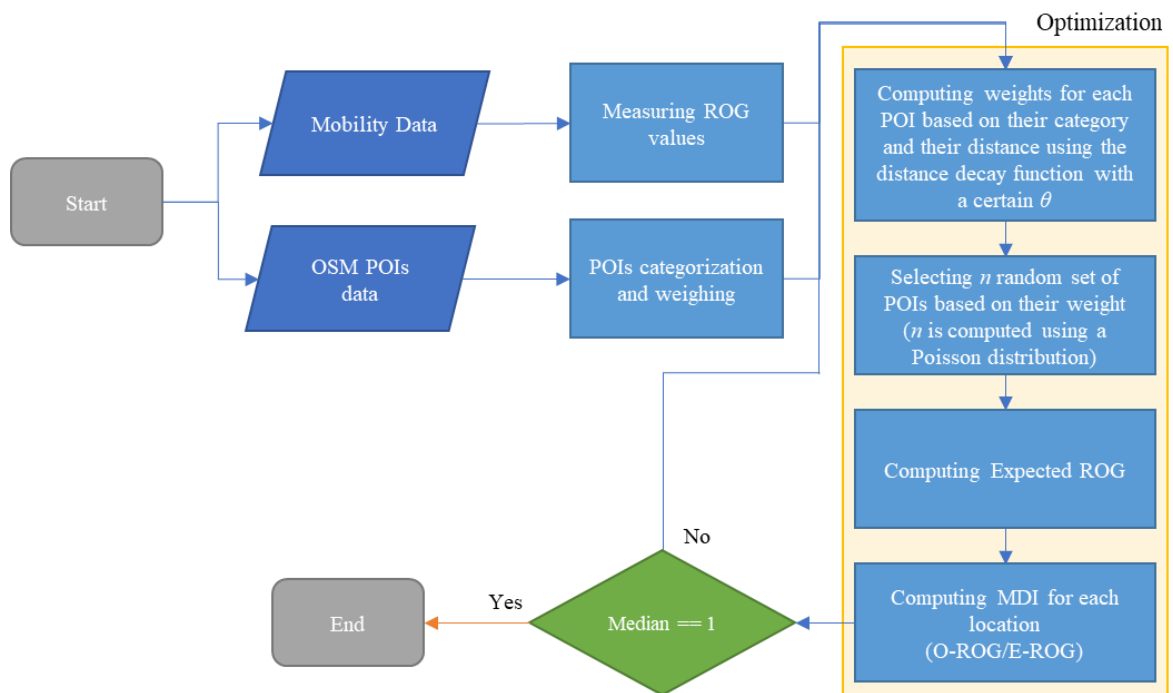
Where  $W_{ij}$  is the weight of point  $j$  for the chosen location  $i$ ,  $\theta$  is the parameter of the distance decay function, and  $d_{ij}$  is the distance between the POI and the chosen location  $i$ . We used the geographical distance over network or time distance functions due to the nature of the mobility data we were working with (large mobile phone data analyzed only in an aggregate form) and the fact that deriving network or time distances would be computationally difficult in such a simulation study. The use of geographical distance (over network distance) is also related to our chosen mobility measure (ROG) which is a simple measure of movement range (González et al., 2008) based on the geographical distances from a known center point.



**Figure 3-1 Example showing a single simulation of a mobility pattern (EROG). The flowchart demonstrates the process of selecting a set of POIs for the computation of expected mobility. The numbers in the parentheses are simply an example and vary based on the local data. We ran the simulation  $s=100$  times for each ADA to simulate expected mobility.**



Selecting an appropriate distance decay formulation (specifically the selection of  $\theta$  in equation 3) is necessary to model this geographical relationship. Finding and using an optimal distance decay function has a long history in geography and social science studies (Haggett, 1966; Ravenstein, 1885; Taylor, 1971). A high value of  $\theta$  would result in a narrower selection of POIs, i.e., the POIs will be selected more locally, and a low value of  $\theta$  would result in a selection of more distant POIs. We employed a bisection algorithm (Burden et al., 2015) to iteratively optimize the distance decay parameter ( $\theta$  in equation 3) for the distance decay function weights for expected mobility. We determined the optimal  $\theta$  value when the median MDI value was found to be 1 (+/- 0.001). Optimizing the  $\theta$  using all observed mobility values incorporates people's behavior in choosing the activities in the simulation process. The simulation process is presented as a flowchart to diagram the manner in which mobility patterns were simulated (Figure 3-2).



**Figure 3-2 Implementation flowchart for how activity patterns were simulated.**

Simulations of observed mobility patterns exhibit a degree of randomness, and therefore we simulate  $s$  activity patterns for each chosen location. We then used the mean ROG of

the  $s$  simulations as the Expected ROG for that chosen location. We conducted a sensitivity analysis to help choose the number of simulations  $s$ , as lower values will reduce computational time. We found that the mean expected mobility value stabilized after approximately  $s = 50$  simulations (Appendix C). Therefore, we chose a conservative value of  $s = 100$  as our number of simulated activity patterns for each location. Employing multiple activity patterns for each ADA aligns with the inherent nature of our observed data, wherein the observed mobility values of all mobile devices within each ADA were aggregated into a singular value.

### 3.2.5 Statistical analysis of mobility deviation index and Socio-Economic Factors

To highlight how the calculation of mobility deviation index might influence inferences made in human mobility analyses we tested whether using the ROG or MDI models resulted in different associations when compared with socio-economic factors commonly shown to be associated with observed mobility levels at the aggregated level. Given that prior researchers employed POI density as a proxy to control geographical context in their models, we aim to assess the extent to which its inclusion could impact outcomes. As a result, we employ two models for ROG: one without POI density and another incorporating it as an independent variable. POI density is characterized by the ratio of the total count of POIs over the area of the region. In this analysis, we also use six chosen socio-economic variables (Table 3-1 **Error! Reference source not found.**) that are widely used in previous studies (Chakrabarti et al., 2021; Luo et al., 2016; Pappalardo et al., 2015; Ruktanonchai et al., 2021).

**Table 3-1 Six socio-economic variables chosen to compare with observed mobility (ROG) and mobility deviation index based on different models of expected mobility. All descriptions are based on Statistics Canada, (2017a).**

Variables	Mean	Min	Max	SD	Description
Average Age	41.1	25	57.8	3.93	Average age of the population (years)

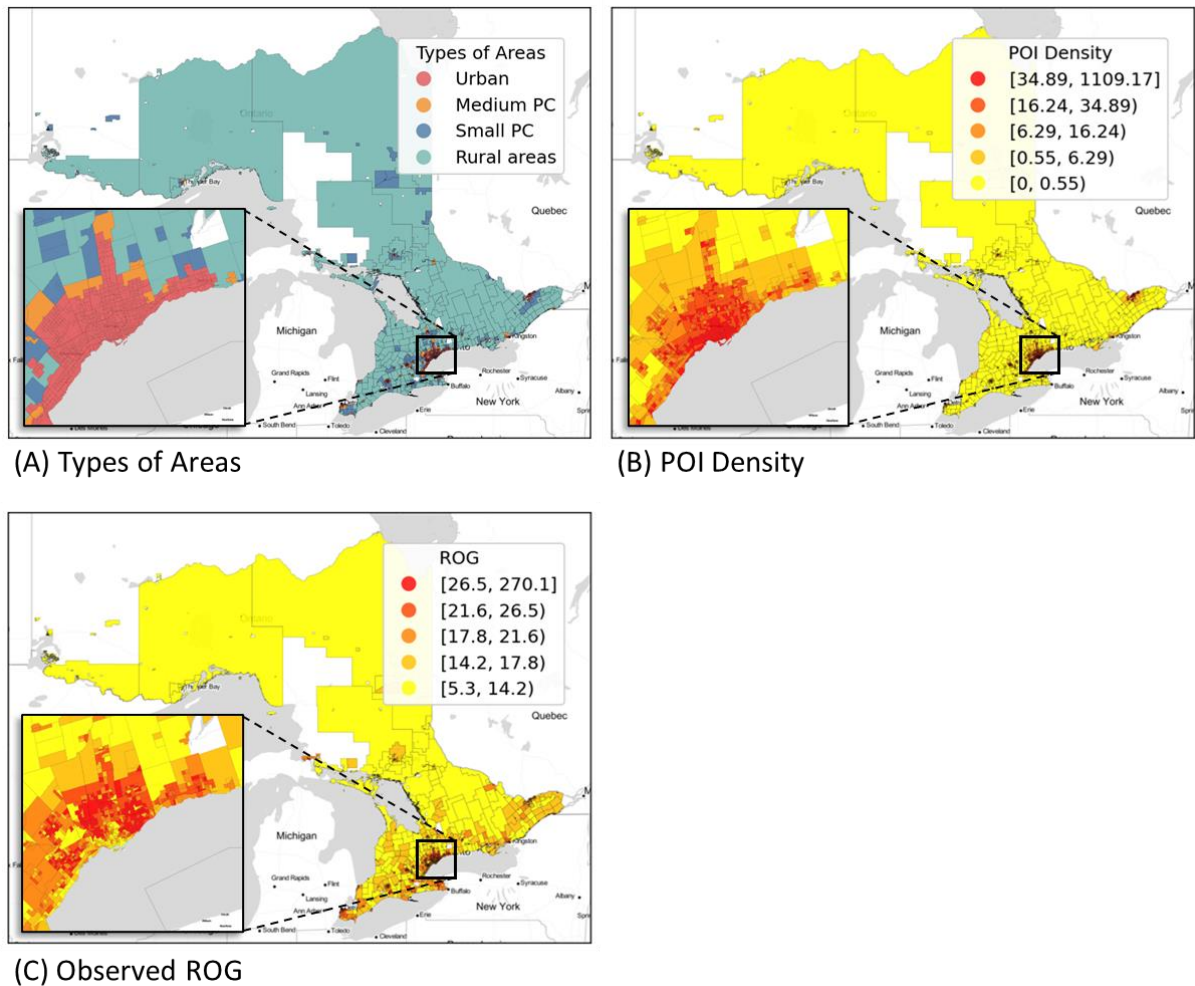
Median Income		34577	8232	68510	8321	In Canadian dollars (CAD)
Proportion of Visible Minority		28%	0%	97.6%	27.2%	Refers to whether a person belongs to a visible minority group as defined by the <i>Employment Equity Act</i> (Government of Canada, 1995)
Proportion of Postsecondary Graduates		64.3%	6.5%	90.6%	11.1%	Refers to a person who holds 'Postsecondary certificate, diploma or degree'
Proportion of Detached Households		57.5%	0%	99.7%	26.8%	Refers to households that live in a single-detached house in terms of its structural type.
Proportion of Private Transport Commuters		25.4%	0%	46.2%	10.5%	Represents the proportion of the labor force aged 15 and over who have a usual workplace or no fixed workplace address and primarily use a private car, truck, or van as their mode of commuting.

We considered observed ROG and the MDI as dependent variables separately in regression models and included all six socio-economic factors and POI density. We first ran the Variance Inflation Factor (VIF) test to assess multicollinearity of our six socio-economic variables and no value greater than 3 was found, which suggests multicollinearity is not present (Rogerson, 2019). Given that we are studying mobility aggregated into geographical units, the assumption of independent errors should be further examined in light of spatial autocorrelation which can lead to biased regression estimates (Anselin, 1988). We ran a general regression model and fit the model by ordinary least squares. We then ran a Moran's I test to examine the spatial autocorrelation of the residual errors using a classical queen contiguity definition of the spatial neighborhood as a common practice based on previous research (Long & Ren, 2022; Tokey, 2021). If the residual errors were found to exhibit significant positive spatial autocorrelation, we then chose to employ a spatial regression model; to adjust for the observed spatial autocorrelation in the regression model (Anselin, 1988). We then ran Lagrangian Multiplier (LM), and robust Lagrangian Multiplier (RLM) tests to choose between a spatial lag model and spatial error model (Wang & Mu, 2018). According to Anselin's, (1988) framework, the model with the larger test statistics between RLM tests

will be chosen as the suitable model. We assessed model goodness of fit using pseudo R<sup>2</sup> values, which are defined as the square correlation between the dependent variable and its prediction (Anselin, 1988). In addition, no skewness in the residuals of the models was found. All analysis was conducted in python, using the PySAL package (Rey & Anselin, 2010) for fitting spatial regression models.

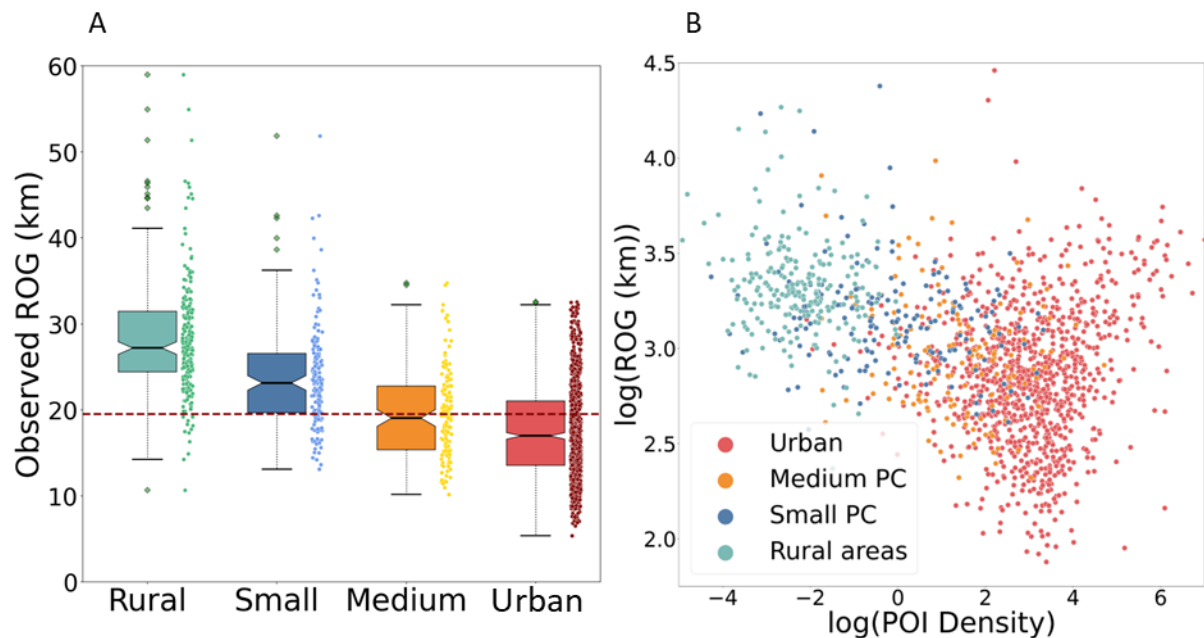
### 3.3 Results

In order to enhance result clarity and facilitate seamless comparisons, we use Statistics Canada's definition of population centers to delineate the four types of areas along the urban-rural gradient (Statistics Canada, 2017): large urban population center, medium population center, small population center, and rural areas (Figure 3-3-A). As anticipated, the density of POIs is greater in urban areas compared to other area types (Figure 3-3-B). Observed mobility levels (ROG) varied considerably throughout Ontario, Canada (Figure 3-3-C). This variability underscores the multifaceted nature of human mobility, shaped by geographical, infrastructural, and socio-economic influences. This emphasizes the necessity for a comprehensive province-wide comparison, rather than focusing solely on individual cities or urban regions.



**Figure 3-3 A) Map of types of areas based on Statistics Canada’s definition of population centers: urban areas, medium population centers (PC), small population centers (PC), and rural areas Statistics Canada, (2017a). These population centers are defined based on a specific range of population and population density per square kilometer. B) Map of POI Density defined as the ratio of the number of POIs over the area of ADA in Ontario, Canada using on OSM POI data (OpenStreetMap contributors, 2017). C) ROG values for the week of February 02, 2020, to February 08, 2020 in Ontario, Canada. Mobility data sourced from a large, de-identified network mobility dataset provided by TELUS Communications Inc.**

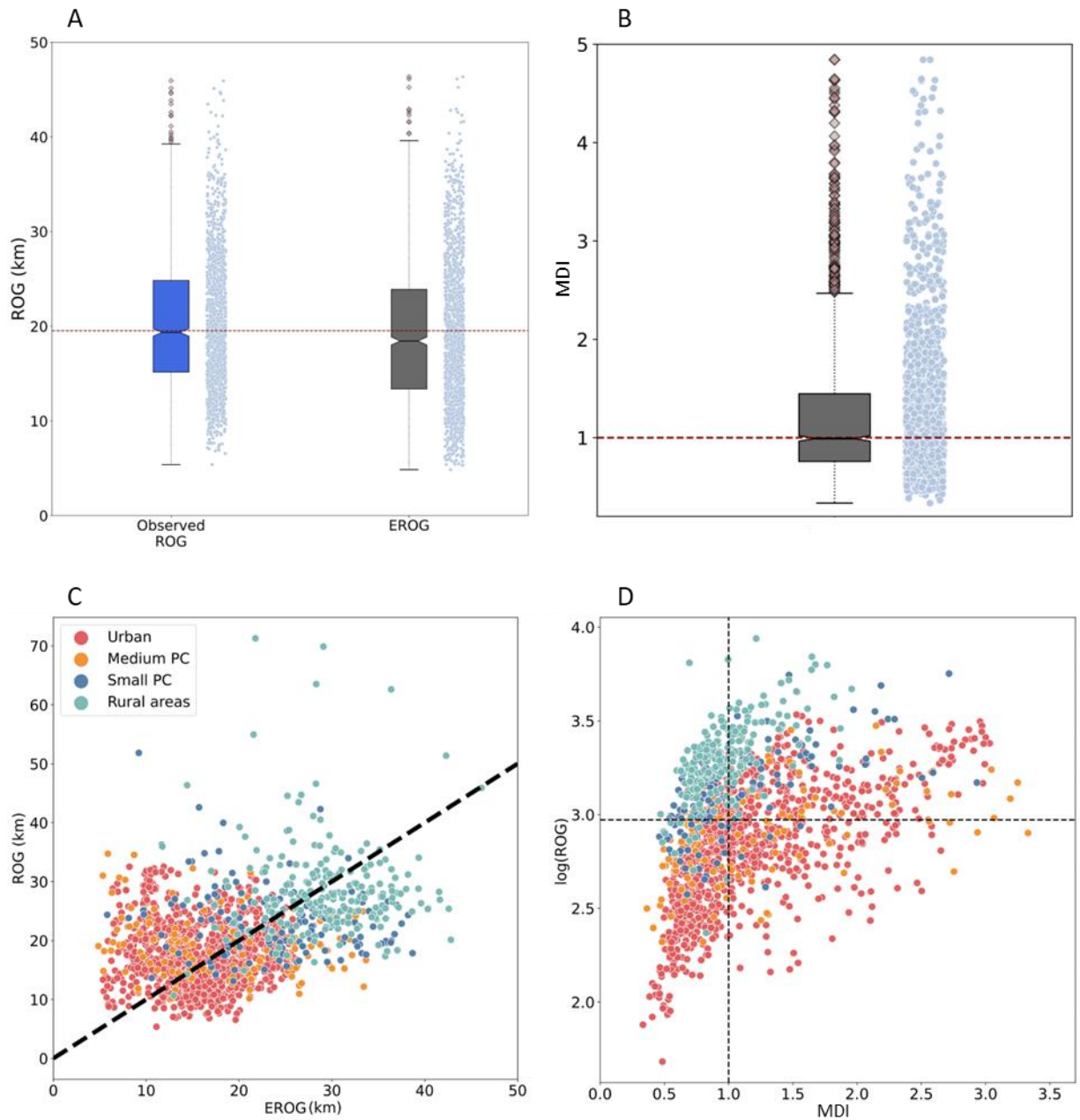
Variance in observed ROG was also greatest in rural areas ( $\mu = 29.74$ ,  $\sigma^2 = 154.25$ ) compared to small population centers ( $\mu = 23.73$ ,  $\sigma^2 = 34.63$ ), medium population centers ( $\mu = 19.46$ ,  $\sigma^2 = 26.61$ ), and large urban centers ( $\mu = 17.52$ ,  $\sigma^2 = 29.84$ ). We observe almost all rural areas (95%; 217 out of 229) demonstrate a mobility level above the provincial median, while a large proportion of large urban population centers (64%; 643 out of 996) show mobility levels below the provincial median (Figure 3-4-A). We found a strong negative correlation between POI density and observed ROG (using a log-log scale –  $r = -0.41$ ; Figure 3-4-B).



**Figure 3-4 A) Boxplots of observed ROG in different types of population centers. The red dotted line is the provincial median ROG. B) Scatter plot of the log of ROG values against the log of POI density for each ADA. The different colors show different types of population centers. The definition of population centers is based on Statistics Canada.**

In the expected ROG values derived from the simulation, we observe a range of values similar to the observed values (Figure 3-5-A), indicating no systematic positive or negative skewness compared to the observed values. The MDI values predominantly

cluster around 1, yet they exhibit a substantial span of variability ranging from close to 0 to above 9. (Figure 3-5-B, Table 3-2). For a more comprehensive understanding of the model's outcomes, we plotted ROG values against EROG values (Figure 3-5-C), as well as the natural logarithm of ROG values against MDI values (Figure 3-5-D). It is evident that a significant proportion of ADAs situated within urban areas and medium population centers exhibit ROG values below the median ROG level, while many of these ADAs demonstrate MDI values exceeding the median MDI level. In contrast, ADAs located in rural areas tend to display ROG values surpassing the median ROG level, yet concurrently exhibit MDI values lower than the median MDI level. This observation underscores the influence of incorporating geographical context on shaping our comprehension of mobility levels. Furthermore, we extended our examination to geographical representations, with maps depicting ROG, EROG, and MDI distributions across the Greater Toronto Area (see Appendix D).



**Figure 3-5 A) box plots of ROG values: observed and expected. The red dashed line is the provincial median of observed ROG values. B) box plots of MDI values. The red dashed line is associated with the value of MDI=1 (indicating an OE ratio of 1). C) Scatter plot of ROG values against EROG values. We colored each node based on their area type. D) Scatter plot of the natural logarithm of ROG values against**



**MDI values. The black horizontal dashed line is the provincial median of observed ROG values, and the black vertical dashed line is where MDI=1.**

**Table 3-2 Summary statistics of observed and computed values for ROG and MDI resulted from different models.**

Model	Metric	Mean	Median	SD
Observed	ROG	21.33	19.53	12.40
Simulation	ROG	19.38	18.44	8.28
	MDI	1.28	0.99	0.91

We performed a multivariate regression of ROG values and MDI against six socio economic factors (Table 1). We found positive and significant values for Moran's I tests on the model residuals when fit using ordinary least squares (Table 3-3). The Lagrangian Multiplier (LM) tests were significant for the ROG and MDI models (Table 3-3); therefore, we employed the Robust Lagrangian Multiplier (RLM) tests to choose the best model. The results of RLMs for all models were significant also (Table 3-3), thus we chose the spatial error model over the spatial lag model as the RLM test of the spatial error model had larger test statistics in ROG and MDI models (Table 3).

**Table 3-3 Summary of the model selection procedure using Moran's I test of spatial autocorrelation on ordinary least squares regression model residuals, and the Lagrangian multiplier (LM) and robust Lagrangian multiplier (RLM) tests for the spatial error and spatial lag models.**

Dependent variable	Moran's I	LMerr	LMlag	RLMerr	RLMlag
ROG	0.37**	572.94**	403.36**	288.52**	118.94**
MDI - Simulation	0.56**	1285.87**	1233.51**	68.60**	16.24*

\*\* p value < 0.01, \* 0.01 < p value ≤ 0.5

From the results of the spatial regression models (spatial error model) we identify some notable differences between the observed ROG models (with and without POI density as a covariate) and the MDI model (Table 3-4). While models' fit levels varied, the pseudo R-squared of the MDI model was the highest ( $R^2 = 0.145$ ). The spatial autocorrelation parameter ( $\lambda$ ) was significant and positive in all models, suggesting a strong spatial autocorrelation effect was present.

Although numerous prior studies have employed POI density as a proxy to incorporate geographical context into their models, our observation indicates that the inclusion of POI density in our model did not yield discernible distinctions. All variables with the exception of median income exhibited similar coefficients, standard errors, and p-values in both ROG model I and ROG model II.

However, comparing the ROG models and the MDI model, we observe a substantial change in both the significance levels and directions of associations (Table 3-4). To be more specific, the association between average age and mobility in the observed ROG model was negative and significant, meaning mobility levels tend to be higher in ADAs containing younger people. Median income was a negative and not-significant predictor of mobility in the observed ROG model. The association between the ratio of visible minority and mobility in the observed ROG model was negative and significant. This indicates that ADAs with a higher minority ratio move less than those with a lower minority ratio. The ratio of postsecondary graduates was a positive and not-significant predictor of mobility in the observed ROG model, meaning that the level of education is not associated with mobility level. The association between the ratio of detached households and mobility was found to be positive and significant in the observed ROG model. This association indicates that people in these ADAs tend to move more than those who are in urban settings living in apartments or other types of households. The association between the ratio of private transport commuters and mobility in the observed ROG model was negative and significant. This highlights that in areas where the

predominant mode of transport for commuting is by private car, truck, or van, individuals tend to exhibit lower mobility level.

In contrast, we observed different statistical associations in five out of six of the model variables when employing the MDI model. The age variable was not a significant predictor of mobility in the MDI model. Median income was a positive and significant predictor of mobility in the MDI model, suggesting that ADAs containing people with higher income exhibit higher than expected mobility. We observed the same result as the observed ROG model for the visible minority variable in the MDI model. The proportion of visible minority showed a significant negative association with mobility deviation index. While the level of education was not significant in the observed ROG model, in the MDI model, we observed a positive and significant association between the proportion of postsecondary graduates and mobility. The proportion of detached households was a negative and significant predictor of mobility in the MDI model. Although the proportion of commuters using private transport was significant in the observed ROG model, it was not observed to be significant in the MDI model.

**Table 3-4 Spatial error models summaries using radius of gyration (ROG) and mobility deviation index (MDI) values computed by different models as dependent variables.**

Dependent Variable	ROG			ROG			MDI		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
Intercept	<b>0.03</b>	<b>0.004</b>	<b>0.000</b>	<b>0.03</b>	<b>0.004</b>	<b>0.000</b>	<b>0.00</b>	<b>0.004</b>	<b>0.026</b>
Average Age	<b>-0.40</b>	<b>0.141</b>	<b>0.004</b>	<b>-0.39</b>	<b>0.141</b>	<b>0.005</b>	0.01	0.127	0.905
Median Income	-0.00	0.097	0.99	0.01	0.097	0.891	<b>0.23</b>	<b>0.091</b>	<b>0.011</b>
% Visible Minority	<b>-0.23</b>	<b>0.035</b>	<b>0.000</b>	<b>-0.23</b>	<b>0.035</b>	<b>0.000</b>	<b>-0.15</b>	<b>0.040</b>	<b>0.000</b>
% Postsecondary Graduates	0.08	0.133	0.545	0.02	0.135	0.828	<b>0.57</b>	<b>0.129</b>	<b>0.000</b>
% Detached Households	<b>0.17</b>	<b>0.039</b>	<b>0.000</b>	<b>0.19</b>	<b>0.041</b>	<b>0.000</b>	<b>-0.09</b>	<b>0.037</b>	<b>0.009</b>
% Private Transport Commuters	<b>-0.11</b>	<b>0.03</b>	<b>0.000</b>	<b>-0.10</b>	<b>0.032</b>	<b>0.001</b>	-0.02	0.028	0.44

POI Density				<b>0.04</b>	<b>0.018</b>	<b>0.034</b>			
Lambda	0.56	0.027	0.000	0.55	0.027	0.000	0.78	0.018	0.000
Number of observation	1543			1543			1543		
Pseudo R-squared	0.129			0.140			0.145		
Log Likelihood	4817.96			4820.14			4966.04		

To gain insight into how MDI and ROG differ when compared to socio-economic variables, we plotted all ADAs' observed ROG values and MDI values against these socio-economic variables (see Appendix E).

### 3.4 Discussion

Mobility deviation index is a broad concept which can be employed with different measures of human mobility – not just using ROG as done here. A rigorous method for defining and calculating expected mobility is fundamental to estimating mobility deviation index. Importantly then the derivation of the expected mobility when calculated mobility deviation index should be based on the measure of mobility being used. For example, if the observed mobility is being measured as the average travel time for an individual, then the expected mobility level should represent a relevant estimate of the expected average travel time for that same individual based on geographical context. There are a number of ways this can be done in practice, for example, using model-based methods (Jiang et al., 2016; Yin et al., 2017), model-free methods (Kulkarni et al., 2018; Ouyang et al., 2018), simulations (Camp et al., 2002; Rhee et al., 2011), or by some combination of these. Here, we chose to use ROG as a spatial measure of mobility and derived expected mobility by using a simulation approach that was specifically relevant to ROG.

ROG is one of the most commonly employed measures of human mobility, especially with modern big mobility datasets (Yan et al., 2011). ROG is an effective measure of the range or spatial extent of mobility, and is able to distinguish how far an individual's activity locations are from their center of mass (E.g., home location) (Asgari et al., 2013). However, ROG (as a standalone measure) is unable to capture the effect of geographical

context on human mobility. Therefore, by adjusting and rescaling existing metrics for geographical context, through the use of MDI, we are able to highlight and capture a different dimension of human mobility. Specifically, in urban areas and medium-sized population centers, a considerable number of ADAs display ROG values that fall below the median. Conversely, these same ADAs often have MDI values that are above the median MDI level. This indicates that although these areas show low mobility levels when measured by ROG, they exhibit a higher level of mobility when context is integrated into the measurement.

We conducted spatial regression models to illustrate the potential influence of various approaches to measuring human mobility on the interpretations of the results. Our models showed that four out of six socio-economic factors (average age, median income, proportion of postsecondary graduates, proportion of private transport commuters) changed from significant to not-significant or not-significant to significant (at the 95% level) between the ROG models, and the MDI. The only consistent variable in all three models was the proportion of visible minorities which was significant and negative. This clear contrast provides strong evidence that interpreting human mobility levels without considering geographical context will directly influence inferences being made. While many studies have studied the relationship between people's mobility behavior and socio-economic factors (Chakrabarti et al., 2021; Luo et al., 2016; Pappalardo et al., 2015; Ruktanonchai et al., 2021), these studies do not adjust the measured mobility indicators to account for geographical context. In doing so, the inferences made may be incorrectly associated with geographical contextual factors and not representative of overall mobility levels. For instance, in our models using observed ROG values as the dependent variable, the ratio of detached households has a positive association with human mobility at the aggregate level, while in our MDI model, it is found to be a negative indicator of mobility deviation index. This makes sense as higher detached household levels might be associated with more suburban regions, but after accounting for the geographical distribution of amenities in these areas (i.e., expected mobility) these mobility levels may be relatively lower than expected. Therefore, the contrasting interpretation here is relative

to the geographical context of the location. When mobility data are used by decision-makers from different disciplines to improve and solve problems such as inequity in transportation accessibility, urban design and planning, and long-term governmental policies (Hidayati et al., 2021) it is important that analytical models of human mobility properly account for geographical contextual factors. It is important to note that the six socio-economic factors selected in our study serve as examples demonstrating how MDI can enhance the understanding of the relationship between human mobility and socio-economic variables. While these factors are illustrative, there are other commonly used variables in similar studies that we did not include in our model.

The MDI offers valuable insights for urban policy and transportation planning. For instance, changes in local amenities, such as the opening or closing of schools, restaurants, or post offices, can alter human mobility patterns (Gupta et al., 2021). To examine this effect, we can employ MDI as a valuable tool. If people do not change their mobility pattern, ROG remains the same, but the MDI will show a difference in value due to the altered context, and vice versa. This property of MDI is helpful for evaluating hypotheses regarding the effects of modifications to local amenities or broader urban transformations on mobility patterns. As MDI provides a tool to quantify mobility behavior based on the local geographical context, another application of MDI is optimizing services to achieve a targeted MDI value. For example, urban planners can aim for an MDI value of 1 or less by optimizing infrastructure or promoting more local mobility behavior.

In our model to compute the expected mobility values, we aimed to simulate EROG values by choosing a set of activity points and their distance to individuals' home locations (here, the centroid of each ADA as an aggregated estimation of home locations of all people living in each ADA). Since work locations account for a large portion of human activities, they should be incorporated into the calculation of the simulated expected mobility values. One of the constraints of working with big data is being unable to access/derive individuals' home and work location information because of how data

are stored and aggregated to maintain individual privacy. Due to this constraint, we assumed that work locations were associated with the POIs and we did not explicitly incorporate a work location in our simulation of expected mobility.

In our case study, we decided to use the OSM POI dataset, since it is one of the most accurate and extensive publicly available datasets on POIs (Barron et al., 2014; Girres & Touya, 2010; Haklay, 2010). However, its coverage depends on different socio-economic factors of contributors and is not uniformly distributed (Mashhadi et al., 2013). Other potential options could be explored whether at a global scale, such as Foursquare POI datasets (McKenzie et al., 2015), or local scale, such as Baidu Map Services POIs dataset in China (Zhai et al., 2019), and Ordnance Survey POIs dataset in the UK (Ordnance Survey, 2022). The dataset of POIs must contain needed information with regard to the measurement method of the observed value. For instance, if one uses a time-based measurement method, the dataset must include information on authority constraints of these POIs (i.e., closing/opening time), to have a more realistic estimation of expected values.

In our simulation method, we used the Transportation Tomorrow historical survey data to reference people's daily activity destinations (Government of Ontario, 2016). This survey classified activity destinations into 7 classes: work, attend school, daycare pickup/dropoff, shopping/retail service, facilitate passenger, return home, and other. We used all classes except return home and facilitate passenger since the locations of these two are not considered an activity location in ROG computation. In Transportation Tomorrow survey, return home refers to the proportion of trips that ended at home, and facilitate passenger to the proportion of trips in which people travel to a destination only to facilitate someone else's trip (e.g., driving someone to the train station). We also merged daycare and attend school since OSM POI dataset's definition of daycare and school overlaps. Within the OSM POI data there is no standard classification for defining activity locations and POIs (McKenzie et al., 2015), we had to merge similar classes of the OSM POI dataset to align them with the Transportation Tomorrow historical survey

analogy. Undoubtedly, this process of categorizing POIs and simulating trips/activities may vary depending on the dataset used and the study purpose.

During the simulation process, essentially, we employed distance as a proxy for attraction level. However, POIs in the same category might have different attraction levels based on other attributes. The POI dataset we used did not have attraction-level information. Therefore, the probability of selecting a small boutique cafe is the same as a large chain restaurant. Alternatively, where possible, the attraction level of POIs can be derived from different sources or datasets directly or by using methods to extract this information from available datasets, such as using the number of check-ins in the Foursquare platform (McKenzie et al., 2015). Nevertheless, access to such datasets was not readily available in this case.

We chose ROG as the measure of mobility and an essential part of its calculation is how to compute the distance between activity points. The distance values differ significantly with regard to the embedded space and the distance type. We used Euclidean distance as an estimate of the distance between two points because our observed mobility data was derived from millions of mobile-phone locations, network-based ROG was computationally limiting. To avoid inconsistencies in our comparison, we also used Euclidian distance in our expected mobility simulations as well. However, we are aware that, especially in urban regions, the distance between two locations can rarely be approximated by their Euclidean distance (Buczowska et al., 2019). Using travel time as the distance between two points might change the degree of mobility, especially in areas with high traffic congestion, but is highly challenging to compute when using large mobility datasets such as was done in this paper.

Aggregated mobility data is an average of a local population's behavior, and therefore we cannot infer individual patterns. The aggregation level of data can influence the final results and smooth over the variation in mobility patterns within local regions (Yan et al., 2013). While we observed varying levels of mobility throughout the study area (among all ADAs), the data for each ADA will not show outliers or anomalies that would be



observed within individual-level data. When studying individual-level data, the data may contain highly-varying ranges of mobility levels within a single areal unit which may influence the simulation procedure of expected mobility values, and subsequently the interpretation of the results. This highlights a key challenge in spatial analysis: the selection of an appropriate scale, often referred as the modifiable areal unit problem (Fotheringham & Wong, 1991; Manley, 2021). In the context of mobility deviation index, analyzing the effects of data aggregation level and the scale of geographical unit on MDI values is crucial. Such analysis is fundamental to understanding the broader applicability and interpretive validity of our findings across varying spatial contexts.

### 3.5 Conclusion

In this study, we aimed to quantify the role of geographical context in human mobility measurement methods. The primary question we address is how geographical context affects the measurement of human mobility levels and subsequently associations with socio-economic variables. Specifically, we introduce the concept of mobility deviation index, defined as the ratio of observed mobility to expected mobility. We proposed a model to compute the expected mobility using OSM POI data. A case study, using de-identified and aggregated network mobility data to measure mobility (using radius of gyration as a mobility indicator) was used to compare mobility deviation index to observed mobility levels. Our results demonstrated that when we rescale observed mobility measures (in this case, the radius of gyration – ROG) by model estimates of expected mobility, the associations with socio-economic factors believed to influence mobility can change. We showed that when using MDI in comparison to the observed ROG, different associations were found with 5/6 socio-economic variables we employed in our models. This is even after we control for a similar contextual variable (in this case POI density). Our results highlight these differences across urban, suburban, and rural configurations which are likely to influence human mobility patterns. Thus, here we have demonstrated that geographical context is not a peripheral factor in human mobility, and rather that when properly controlled for in human mobility analysis, may lead to new

inferences and interpretations about how socio-economic factors and the built environment are associated with levels of human mobility.

Future research could further refine the MDI measure by incorporating network distance, which reflects actual travel routes along the road network, offering a more realistic view of urban mobility patterns (Buczowska et al., 2019). Additionally, integrating potential work locations into the MDI analysis could provide a deeper understanding of individual mobility, considering significant daily travel destinations like places of employment (Sila-Nowicka et al., 2023). Exploring alternatives to the ROG could also enhance mobility pattern analysis. Metrics that capture visit frequency or the diversity of visited places (Song et al., 2010), along with incorporating a temporal dimension into the MDI measure, would allow for the examination of mobility changes across different times and conditions, contributing to a dynamic portrayal of human mobility.

## References

- Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.
- Asgari, F., Gauthier, V., & Becker, M. (2013). A survey on human mobility and its applications. *ArXiv Preprint ArXiv:1307.0814*.
- Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS, 18*(6), 877–895.
- Barthelmes, L., Heilig, M., Klinkhardt, C., Kagerbauer, M., & Vortisch, P. (2022). The effects of spatial characteristics on car ownership and its impacts on agent-based travel demand models. *Procedia Computer Science, 201*, 296–304.
- Brum-Bastos, V., Łoś, M., Long, J. A., Nelson, T., & Demšar, U. (2021). Context-aware movement analysis in ecology: a systematic review. *International Journal of Geographical Information Science, 1–23*.

Buczowska, S., Coulombel, N., & de Lapparent, M. (2019). A comparison of euclidean distance, travel times, and network distances in location choice mixture models. *Networks and Spatial Economics*, 19(4), 1215–1248.

Burden, R. L., Faires, J. D., & Burden, A. M. (2015). *Numerical analysis*. Cengage learning.

Camp, T., Boleng, J., & Davies, V. (2002). A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5), 483–502.

Chakrabarti, S., Hamlet, L. C., Kaminsky, J., & Subramanian, S. v. (2021). Association of human mobility restrictions and race/ethnicity-based, sex-based, and income-based factors with inequities in well-being during the COVID-19 pandemic in the United States. *JAMA Network Open*, 4(4), e217373–e217373.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. In *Transportation Research Part C: Emerging Technologies* (Vol. 68, pp. 285–299). Elsevier Ltd. <https://doi.org/10.1016/j.trc.2016.04.005>

Chen, X., & Jia, P. (2019). A comparative analysis of accessibility measures by the two-step floating catchment area (2SFCA) method. *International Journal of Geographical Information Science*, 33(9), 1739–1758.

Chen, Y. (2008). A wave-spectrum analysis of urban population density: entropy, fractal, and spatial localization. *Discrete Dynamics in Nature and Society*, 2008.

Cheng, L., Wang, K., De Vos, J., Huang, J., & Witlox, F. (2022). Exploring non-linear built environment effects on the integration of free-floating bike-share and urban rail transport: A quantile regression approach. *Transportation Research Part A: Policy and Practice*, 162, 175–187.

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7), 1025–1044.

Ghosh, J., Philip, S. J., & Qiao, C. (2007). Sociological orbit aware location approximation and routing (SOLAR) in MANET. *Ad Hoc Networks*, 5(2), 189–209.

Girres, J., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782.  
<https://doi.org/10.1038/nature06958>

Government of Canada. (1995). *Employment Equity Act (S.C. 1995, c. 44)*. Government of Canada. <https://laws-lois.justice.gc.ca/eng/acts/e-5.401/>

Government of Ontario. (2016). *Transportation Tomorrow historical survey data*. <https://data.ontario.ca/en/dataset/transportation-tomorrow-historical-survey-data>

Gupta, S., Nguyen, T., Raman, S., Lee, B., Lozano-Rojas, F., Bento, A., Simon, K., & Wing, C. (2021). Tracking public and private responses to the COVID-19 epidemic: evidence from state and local government actions. *American Journal of Health Economics*, 7(4), 361–404.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24, 7–21.

Haggett, P. (1966). *Locational analysis in human geography*.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4), 682–703.

Hidayati, I., Tan, W., & Yamu, C. (2021). Conceptualizing mobility inequality: Mobility and accessibility for the marginalized. *Journal of Planning Literature*, 36(4), 492–507.

Hoteit, S., Chen, G., Viana, A., & Fiore, M. (2016). Filling the gaps: On the completion of sparse call detail records for mobility analysis. *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, 45–50. <https://doi.org/10.1145/2979683.2979685>

Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., & González, M. C. (2016). The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37), E5370–E5378.

Jordan, T., Raubal, M., Gartrell, B., & Egenhofer, M. (1998). An affordance-based model of place in GIS. *8th Int. Symposium on Spatial Data Handling, SDH*, 98, 98–109.

Kulkarni, V., Tagasovska, N., Vatter, T., & Garbinato, B. (2018). Generative models for simulating mobility trajectories. *ArXiv Preprint ArXiv:1811.12801*.

Long, J. A., Malekzadeh, M., Klar, B., & Martin, G. (2021). Do regionally targeted lockdowns alter movement to non-lockdown regions? Evidence from Ontario, Canada. *Health & Place*, 102668.

Long, J. A., & Ren, C. (2022). Associations between mobility and socio-economic indicators vary across the timeline of the Covid-19 pandemic. *Computers, Environment and Urban Systems*, 91, 101710.

Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, 70, 11–25.

Ma, X., Ji, Y., Jin, Y., Wang, J., & He, M. (2018). Modeling the factors influencing the activity spaces of bikeshare around metro stations: A spatial regression model. *Sustainability*, 10(11), 3949.

Manley, D. (2021). Scale, aggregation, and the modifiable areal unit problem. In *Handbook of regional science* (pp. 1711–1725). Springer.

Mashhadi, A., Quattrone, G., & Capra, L. (2013). Putting ubiquitous crowd-sourcing into context. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 611–622.

McKenzie, G., Janowicz, K., Gao, S., Yang, J.-A., & Hu, Y. (2015). POI pulse: A multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 50(2), 71–85.

Mei, A., & Stefa, J. (2009). SWIM: A simple model to generate small mobile worlds. *IEEE INFOCOM 2009*, 2106–2113.

Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). Wherenext: a location predictor on trajectory pattern mining. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 637–646.

Noi, E., Rudolph, A., & Dodge, S. (2022). Assessing COVID-induced changes in spatiotemporal structure of mobility in the United States in 2020: a multi-source analytical framework. *International Journal of Geographical Information Science*, 36(3), 585–616.

OpenStreetMap contributors. (2017). *Planet dump* retrieved from <https://planet.osm.org> .

Ordnance Survey. (2022). *Points of Interest*. <https://www.ordnancesurvey.co.uk/business-government/products/points-of-interest>

Ouyang, K., Shokri, R., Rosenblum, D. S., & Yang, W. (2018). A non-parametric generative model for human trajectories. *IJCAI*, 3812–3817.

Pappalardo, L., Pedreschi, D., Smoreda, Z., & Giannotti, F. (2015). Using big data to study the link between human mobility and socio-economic development. *2015 IEEE International Conference on Big Data (Big Data)*, 871–878.

Piovani, D., Arcaute, E., Uchoa, G., Wilson, A., & Batty, M. (2018). Measuring accessibility using gravity and radiation models. *Royal Society Open Science*, 5(9), 171668.

Purves, R. S., Laube, P., Buchin, M., & Speckmann, B. (2014). Moving beyond the point: An agenda for research in movement analysis with real data. In *Computers, Environment and Urban Systems* (Vol. 47, pp. 1–4). Elsevier Ltd. <https://doi.org/10.1016/j.compenvurbsys.2014.06.003>

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 48(2), 167–235.

Rey, S. J., & Anselin, L. (2010). PySAL: A Python library of spatial analytical methods. In *Handbook of applied spatial analysis* (pp. 175–193). Springer.

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., & Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 19(3), 630–643.

Rogerson, P. A. (2019). *Statistical methods for geography: a student's guide*. Sage.

Ruktanonchai, C. W., Lai, S., Utazi, C. E., Cunningham, A. D., Koper, P., Rogers, G. E., Ruktanonchai, N. W., Sadilek, A., Woods, D., & Tatem, A. J. (2021). Practical geospatial and sociodemographic predictors of human mobility. *Scientific Reports*, 11(1), 1–14.

Sharif, M., & Alesheikh, A. A. (2017). Context-awareness in similarity measures and pattern discoveries of trajectories: a context-based dynamic time warping method. *GIScience & Remote Sensing*, 54(3), 426–452.

Shearmur, R. (2006). Travel from home: an economic geography of commuting distances in Montreal. *Urban Geography*, 27(4), 330–359.

Shi, L., Chi, G., Liu, X., & Liu, Y. (2015). Human mobility patterns in different communities: a mobile phone data-based social network approach. *Annals of GIS*, 21(1), 15–26. <https://doi.org/10.1080/19475683.2014.992372>

Sila-Nowicka, K., Fotheringham, A. S., & Demšar, U. (2023). Activity triangles: a new approach to measure activity spaces. *Journal of Geographical Systems*, 1–29.

Skov-Petersen, H. (2001). Estimation of distance-decay parameters: GIS-based indicators of recreational accessibility. *ScanGIS*, 237–258.

Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10), 818–823. <https://doi.org/10.1038/nphys1760>

Song, X., Shibasaki, R., Yuan, N. J., Xie, X., Li, T., & Adachi, R. (2017). DeepMob: learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data. *ACM Transactions on Information Systems (TOIS)*, 35(4), 1–19.

Statistics Canada. (2016). *The aggregate dissemination area (ADA): a new census dissemination geographic area*. Statistics Canada. <https://www12.statcan.gc.ca/census-recensement/2016/geo/ADA/adainfo-eng.cfm>

Statistics Canada. (2017). *Ontario [Province] and Canada [Country] (table). Census Profile 2016 Census*. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/Page.cfm?Lang=E&Geo1=PR&Code1=35&Geo2=&Code2=&SearchText=Ontario&SearchType=Begin&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=35&type=0>

Tao, Z., Yao, Z., Kong, H., Duan, F., & Li, G. (2018). Spatial accessibility to healthcare services in Shenzhen, China: improving the multi-modal two-step floating catchment area



method by estimating travel time via online map APIs. *BMC Health Services Research*, 18(1), 1–10.

Taylor, P. J. (1971). Distance transformation and distance decay functions. *Geographical Analysis*, 3(3), 221–238.

Tokey, A. I. (2021). Spatial association of mobility and COVID-19 infection rate in the USA: A county-level study using mobile phone location data. *Journal of Transport & Health*, 22, 101135.

Turner, T., & Niemeier, D. (1997). Travel to work and household responsibility: new evidence. *Transportation*, 24, 397–419.

Vickerman, R. W. (1974). Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning A*, 6(6), 675–691.

Wang, M., & Mu, L. (2018). Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems*, 67, 169–175.

Xu, Y., Shaw, S. L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, 42(4), 625–646. <https://doi.org/10.1007/s11116-015-9597-y>

Yan, X.-Y., Han, X.-P., Wang, B.-H., & Zhou, T. (2013). Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, 3(1), 1–5.

Yan, X.-Y., Han, X.-P., Zhou, T., & Wang, B.-H. (2011). Exact solution of the gyration radius of an individual's trajectory for a simplified human regular mobility model. *Chinese Physics Letters*, 28(12), 120506.

Yin, M., Sheehan, M., Feygin, S., Paiement, J.-F., & Pozdnoukhov, A. (2017). A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems*, *19*(6), 1682–1696.

Yue, Y., Zhuang, Y., Yeh, A. G. O., Xie, J.-Y., Ma, C.-L., & Li, Q.-Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, *31*(4), 658–675.

Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z.-R., & Gu, C. (2019). Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Computers, Environment and Urban Systems*, *74*, 1–12.

Zhao, J., Zhang, F., Zhao, C., Wu, G., Wang, H., & Cao, X. (2020). The Properties and Application of Poisson Distribution. *Journal of Physics: Conference Series*, *1550*(3), 032109.

## Chapter 4

### 4 Quantifying local mobility patterns in human mobility data<sup>1</sup>

#### 4.1 Introduction

Current urban planning strategies aim to encourage localized mobility patterns by adopting smart growth interventions that shift from monocentric city structures to polycentric configurations (Brezzi & Veneri, 2015). These interventions, including the concept of a 15-minute city, emphasize the provision of various local amenities within a relatively small radius (Graells-Garrido et al., 2021; Pozoukidou & Chatziyiannaki, 2021; Veneri, 2018).

Individuals tend to prefer travelling within relatively close distances to their primary locations with a power law distribution (Brockmann et al., 2006; Schneider et al., 2013; Song, Koren, et al., 2010). However, various factors contribute to longer distance mobility behaviors. For example, although people often prefer to choose residences close to their workplaces or vice versa in an effort to reduce commuting time and optimize daily routines (Shuttleworth & Gould, 2010), there are trade-offs. In high-priced urban areas, households tend to accept longer commutes for larger houses or lower housing costs further from central urban locations (Alonso, 1964; Muth, 1961).

Despite the importance of understanding local mobility patterns, there is a notable gap in directly measuring and quantifying the localness of individuals' mobility patterns, which takes into account their choices. When choosing where to shop, eat out or partake in other types of activity, individuals may exhibit distinct behaviors influenced by their own set of preferences. Extensive research has examined various factors that impact consumers'

---

<sup>1</sup> A version of this chapter has been submitted for publication to International Journal of Geographical Information Science.

choice preferences when selecting retail locations, including pricing, quality or service of products, sales personnel, location convenience, sales promotions, advertising, store ambiance, and reputation (Bearden, 1977; Belk, 1975; Degeratu et al., 2000; Doyle & Fenwick, 1974; Erdem et al., 1999; Hackett et al., 1993; Lindquist, 1974; Yilmaz, 2004; Yilmaz et al., 2007). However, these aspects of preference and choice have not been considered in human mobility research. Traditional time-based methods, such as measuring time spent outside the home or workplace (Demissie et al., 2019), or travel time ratio (TTR) (Dijst & Vidakovic, 2000), do not capture the degree of localness, as individuals may spend varying amounts of time locally or further afar. Similarly, diversity-based measures, like the Shannon diversity index (Song, Qu, et al., 2010), fail to accurately assess localness since proximity or distance is not explicitly considered. Additionally, classical mobility measures, such as the number of trips taken (Zhao et al., 2016), do not provide insights into the localness aspect of mobility.

Among the commonly used measures, range-based approaches, such as the farthest distance from home or the maximum distance between any two points visited by individuals (Isaacman et al., 2011; Mir et al., 2013), the size of the activity space (Hirsch et al., 2014), the radius of gyration (ROG) (González et al., 2008), or the local travel index (LTI) (Managh & El-Geneidy, 2012), offer the potential to quantify localness in mobility patterns. However, these measures are limited by two common characteristics of individual mobility data. The first is that these measures do not directly account for the influence of the relative proximity/accessibility to amenities on individuals' mobility patterns (Malekzadeh & Long, 2023). This limitation relates to well-known geographical disparities in equitable access to different services and activities within cities, which manifests in different mobility patterns owing to differential spatial access to necessary activities (Frey, 2017; Logan et al., 2021; Pandey et al., 2022; Purifoy, 2021; Rigolon, 2016). As a result, synthesizing an outcome solely based on the raw geometry of observed movement patterns without accounting for the underlying geographical distribution of activities will bias measurement towards individuals with greater access to activities and amenities (Malekzadeh & Long, 2023). Crucially, this overlooks the equity

dimension in urban mobility, as it fails to capture how spatial inequalities influence individual mobility patterns. Addressing this gap is vital for a comprehensive understanding of human mobility behavior, advocating for measures that recognize and rectify these disparities. The second limitation is that very long trips (which are often rare) can significantly impact the outcome of these measures, potentially skewing the representation of an individual's actual mobility behavior (Sahasrabudde et al., 2021). To address these two limitations, it is necessary to quantify local mobility patterns using methods that are both less sensitive to rare, long-distance trips, while at the same time able to capture the geographical disparity in locally-available amenities on individuals' mobility patterns. Hence, in this paper, we propose a new method termed the Local Mobility Index (LMI) for measuring the relative *localness* of individual mobility patterns that overcomes limitations of existing methods.

The structure of this paper is as follows: Section 4.2 presents our methodology where we outline the theoretical foundations and calculation of the Local Mobility Index (LMI), alongside detailing our validation process and the statistical analyses employed. In Section 4.3, we present a case study in England using a GPS tracking dataset. Section 4.4 discusses the findings of the presented application, interpreting their significance in the broader context of urban mobility and planning, while also acknowledging the limitations of our study. Finally, the paper concludes with Section 5, summarizing our key insights and contributions.

## 4.2 Methodology

### 4.2.1 Local Mobility Index (LMI)

LMI is conceptualized as a metric reflecting local mobility behavior, designed to incorporate geographical disparities in access to amenities into its measurement and minimize the impact of outliers (i.e., singular long trips). We define local mobility behavior as an individual consistently opting for the nearest available option across various activities relative to their significant locations (i.e., home and second-place), drawing on the core geographical concept of nearest neighbor distances. Increasing

deviations from the choice of the nearest amenity signify an increasing degree of non-local mobility behavior. In the LMI framework, we assess the localness for each activity choice, factoring in dwell time as a proxy for activity importance to mitigate outlier effects. Summarizing localness values for all activity choices yields a single LMI value for each individuals' mobility pattern.

Implementing the LMI requires two key datasets. First, we require a dataset which captures individual mobility, and specifically represents spatially an individual's activity (stop) points. Practically, these data may be derived from GPS (or other form of) tracking data which is represented as a series of stops and moves (Spaccapietra et al., 2008). Alternatively, many modern mobility datasets exclusively capture individual activity points only (Chen et al., 2011). Second, we need a dataset that characterizes the geographical distribution of amenities within an individual's range. Typically, we will use available point-of-interest (POI) GIS data, which can be derived from any of a variety of sources. To simplify the analysis POI data are here used to map potential activities (or amenities) which should be categorized into different types. The categorization of POIs into activity types may be situationally dependent, and may reflect local behaviors or specific study questions.

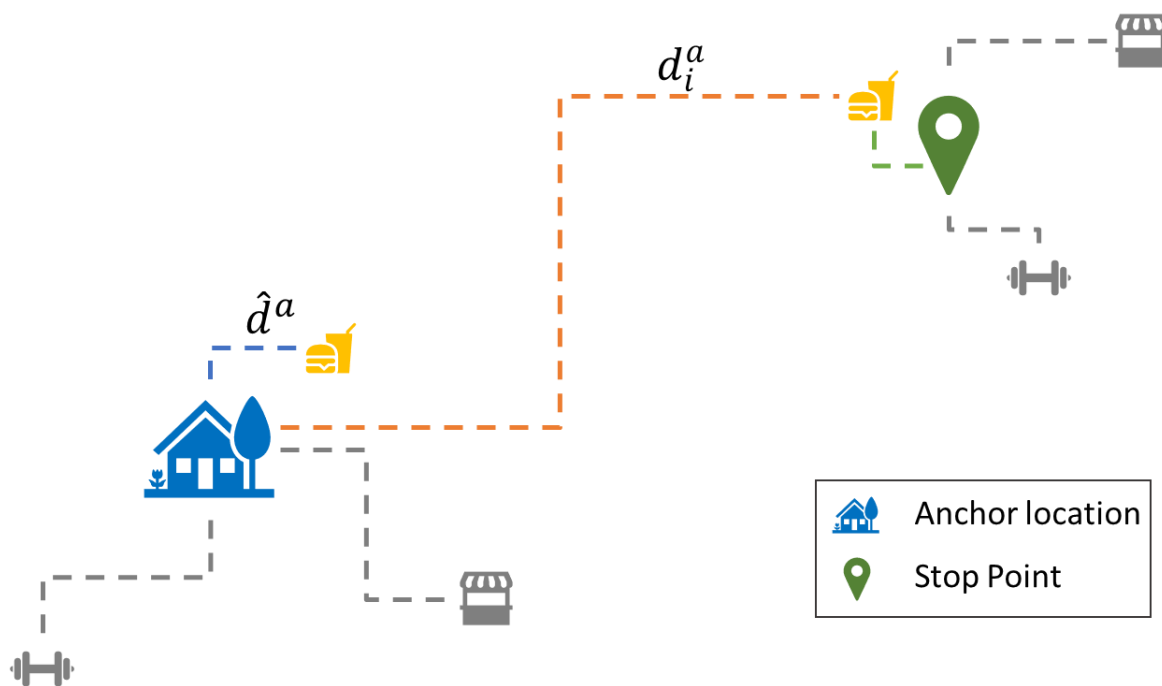
To measure an individual's LMI, we calculate a localness score for each activity (stop) point. Localness is defined relative to an anchor; therefore, we choose the home location of the individual as the anchor point from which to assess localness. Other choices for the anchor point are possible (e.g., work location, geometric centroid of activities), but for most applications it will make sense to use the home location. Similarly, multiple anchor points can be chosen (see section 4.2.2) to capture the multi-centric nature of many individuals' mobility patterns associated with commutes to a workplace or more anchors (Ahas et al., 2010; Peng et al., 2012).

To calculate the localness score of stop points, we compare the distance of a chosen activity location (from the anchor) to the nearest available activity of the same type. In doing so, first, we find the closest activity of the same type to the anchor (e.g., home).

Then we calculate the ratio of the distance between the anchor location (e.g., home) and the nearest available activity to the distance between the anchor location (e.g., home) and the chosen activity (Equation 4.1),

$$L_i^a = \frac{\hat{d}^a}{d_i^a} \quad 4.1$$

where  $L_i^a$  is the localness score for activity stop  $i$  with POI activity type  $a$  (the nearest POI within search radius  $r$ ),  $\hat{d}^a$  is the distance from the anchor location (e.g., home) to the nearest activity POI of type  $a$ , and  $d_i^a$  is the distance from the anchor location (e.g., home) to the chosen activity stop  $i$ . This ratio varies from 0 to 1, with a value of 1 indicating a very local choice (i.e., choosing the nearest POI of that type). Values near 0 signify non-local behavior, (i.e., choosing a POI much further than the nearest POI of that type). It is important to note that by utilizing the ratio, we consider variations in the availability of POIs in individuals' activity space. For instance, comparing someone residing in an area with limited POI availability to another person in a densely populated POI environment, the former might cover longer distances geometrically. However, they could still exhibit a high level of localness by opting for closer POIs. In contrast, the latter might travel shorter distances but opt for non-local choices, resulting in a lower level of localness. Figure 4-1 shows how this process works in practice.



**Figure 4-1** An example scenario of the components of LMI. The stop point is surrounded by several nearby POIs. Among these, the restaurant (a), the nearest POI to the stop point (i), is identified as the destination at the stop point. Accordingly, the analysis identifies the closest POI of the same type as i (a) from the anchor location.  $\hat{d}^a$  is the distance from the anchor location (e.g., home) to the nearest activity POI of type a, and  $d_i^a$  is the distance from the anchor location (e.g., home) to the chosen activity stop i. All other POI types depicted in gray, are not considered for this particular stop point in the calculation.

In practice, the definition used in Equation 4.1 (Figure 4-1) is imperfect because not all stops are found to be close to a POI. This can be due to a variety of reasons, but is likely owing to either activities not being associated with a POI (such as visiting a friend or relative) or due to errors/missing data in the chosen POI dataset. Identifying such instances requires that we choose a POI search radius ( $r$ ) at which to define when an activity stop is associated with any POI. Typically, the prevalence of non-POI stops may be associated with the nature of mobility behavior and spatial accuracy of the mobility



data used to identify activity locations. We hereafter refer to activity stop points with a POI within the specified search radius as POI stop points, and those without a POI as non-POI stop points.

To incorporate non-POI stop points into the analysis of local mobility behavior, we need a method for assigning a localness score (between 0 and 1) to non-POI stop points in a similar fashion to Equation 4.1. To do so, we use a simple heuristic based on a critical distance threshold ( $\hat{d}_c$ ) that is used to generate a similar localness score (Equation 4.2),

$$\begin{cases} L_i^b = 1 - \frac{d_i}{\hat{d}_c} & d_i \leq \hat{d}_c \\ L_i^b = 0 & d_i > \hat{d}_c \end{cases} \quad 4.2$$

where  $L_i^b$  is the localness score for stop  $i$  that has no POI within the search radius  $r$  (termed a non-POI stop),  $d_i$  is the distance between the anchor location (e.g., home) and non-POI stop  $i$ , and  $\hat{d}_c$  is the critical distance threshold. In practice,  $\hat{d}_c$  can be chosen to be relevant based on local mobility behaviour, representing some upper bound on travel distance, or can be estimated based on the data.

To compute an individual's overall LMI score we take the weighted sum of the localness scores for each activity ( $i$ ). We choose weights as the proportion of time spent at each activity point. The temporal weighting is used to quantify the significance of each stop, representing that the time spent on an activity reflects its importance. The temporal weighting helps further mitigate the impact of rare outlier trips. The sum of these weighted ratios yields an individual's LMI measurement (Equation 4.3).

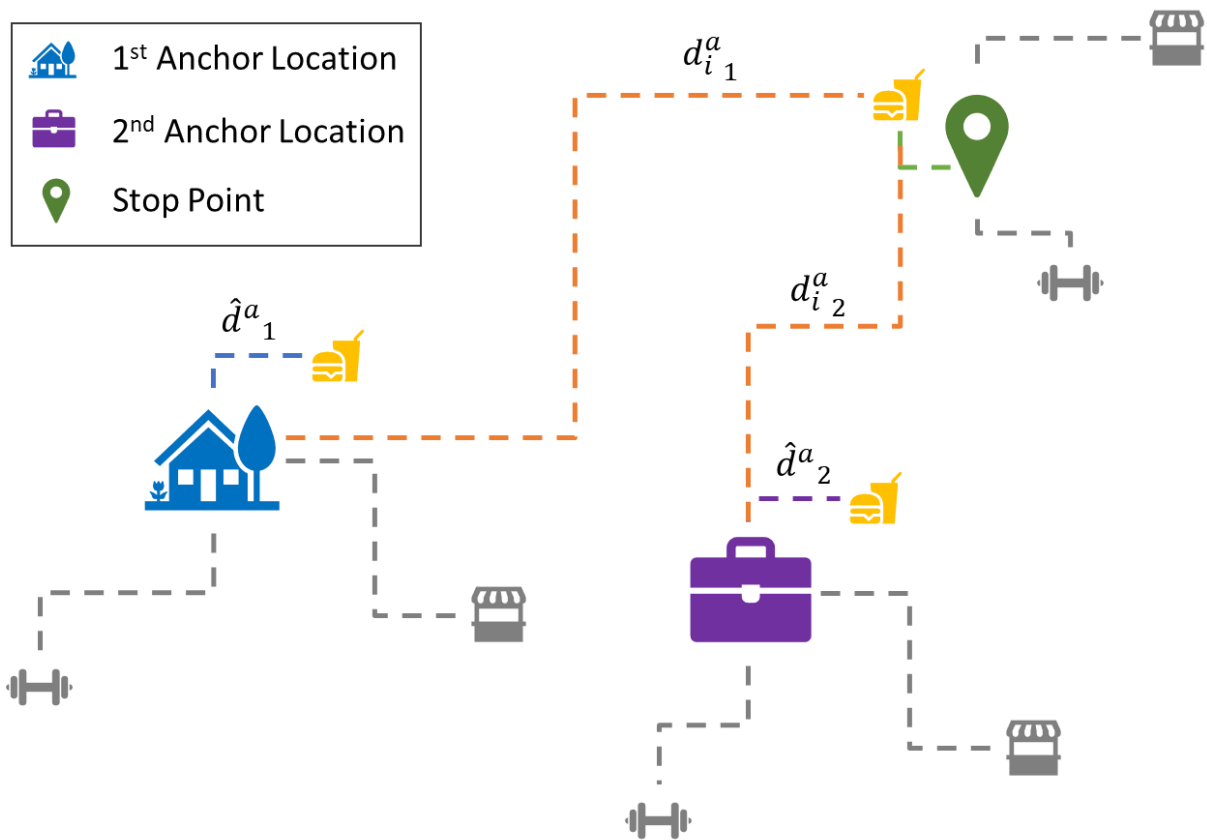
$$LMI = \sum L_i^{\{a,b\}} * \frac{t_i}{T} \quad 4.3$$

where  $L_i^{\{a,b\}}$  is the localness score of activity  $i$  associated with activity type  $a$  (defined in equation 1) or non-POI stops  $b$  (defined in equation 2),  $t_i$  is the dwell time of activity  $i$ , and  $T$  is the cumulative sum of all dwell times for the individual ( $T = \sum t_i$ ).

### 4.2.2 Incorporating multi-centric activity patterns into measures of local mobility

In this section, we refine our methodology to better capture multi-centric activity patterns. We recognize that individuals' activities are frequently distributed across several significant locations (Ahas et al., 2010; Sila-Nowicka et al., 2023), often extending beyond just the residential home location, such as workplaces and schools. This understanding is crucial for analyzing patterns of local mobility, considering how these diverse centers influence individuals' activity choices and contribute to reducing daily travel times including, for example, commutes. Our revised approach evaluates the proximity of chosen destinations to these multiple significant locations. For example, an individual might choose an activity destination that is far from their home but close to their workplace or school. This decision reflects a preference for activities located near these key centers, and despite the distance from home, it is still considered a manifestation of local behavior.

To compute localness scores considering more than one anchor location, we compute the distances between the chosen activity point  $I$  to each anchor locations, selecting the one with the nearest distance as  $d_i^a$  in equation 4.1. We then determine the distance between the closest available POI of the same type to the chosen anchor location  $\hat{d}^a$  in Equation 4.1. Calculating  $L_i^b$  for non-POI stops follows an identical format as in Equation 4.2 where again we compute the distances between the chosen destination to each anchor locations, selecting the one with the shortest distance as  $d_i$  in Equation 4.2. The calculation of the LMI follows identically thereafter, again using a temporal weighting scheme. Figure 4-2 presents a schematic of a multi-centric activity pattern with two anchor locations used in the calculation of  $L_i^a$ .



**Figure 4-2** An example scenario of the components of multi-centric LMI with two anchor locations. The stop point is surrounded by several nearby POIs. Among these, the restaurant ( $a$ ), the nearest POI to the stop point, is identified as the destination at the stop point. Accordingly, the analysis identifies the closest POI of the same type as  $a$  (a restaurant) from the two anchor locations.  $d^a_{i_1}$  and  $d^a_{i_2}$  are the distances between the first and second anchor locations and the nearest POI to the stop point.  $\hat{d}^a_1$  and  $\hat{d}^a_2$  are the distances between the first and second anchor locations and the closest POI of the type  $a$ . All other POIs, depicted in gray, are not considered for this particular stop point in the calculation.

To promote the reproducibility of the LMI method, we have developed a Python script for computing the LMI statistic for any set of individual mobility data globally (<https://figshare.com/s/5b26511ccb33d2de7c4e>). The function requires as input a set of

activity locations, with dwell times attached. A further set of parameters such as home location and/or second-place is also required. The script automatically downloads and processes the necessary network and POI data from OpenStreetMap (OpenStreetMap Wiki, 2022) using a buffer around the submitted mobility data as the bounding box.

## 4.3 Case study

### 4.3.1 GPS tracking data and study area

We collected GPS data from 1017 participants using a bespoke mobile phone app across three cities in the United Kingdom: Brighton and Hove, Leeds, and Birmingham. The dataset was collected during two distinct sampling periods: September 2018 to May 2019 for Leeds and Brighton & Hove, and September 2019 to March 2020 for Birmingham. The dataset provides a comprehensive view of participants' daily mobility behaviors, allowing us to explore the relative localness in individual movement patterns. The sampling frame were 18-64-year-olds who were in employment in one of the target cities. Participants were recruited from the Dun & Bradstreet business directory which contains addresses of firms, including very small firms with no employee other than the owner. This sampling frame was used to study the daily mobility of city workers. The study sample comprised individuals from a range of age groups and occupational backgrounds, reflecting the diversity of the workforce in these cities.

Participants were requested to install a customized mobile app designed to track their movement over seven days of a normal week, following their explicit consent. The mobile app consisted of two main components: a location tracking feature and real-time survey questions delivered through push notifications, which also recorded the participants' location. Our developed mobile application employed a motion-based approach, utilizing a two-state rolling geofence system to differentiate between the moving and stationary states. The geofence had a threshold set at 50 meters, determining whether participants were actively moving or stationary based on their location updates. Within the moving state, location data were recorded at a high temporal frequency to capture detailed movement patterns. Conversely, in the non-moving state, location

positions were only recorded when participants received real-time survey questions or during state transitions from stationary to moving. In addition to GPS data, an onboarding survey captured individual-level socio-economic data including attributes, such as age, sex, and occupation, characteristics of their daily routines, and the postal code of their homes (the postcode of the workplace was included in the Dun and Bradstreet data).

We excluded participants who contributed less than 5 days of data. For participants with data spanning more than 10 days, we selected a 10-day window frame yielding the highest number of data points. This resulted in a dataset containing data spanning from 5 to 10 days for each individual. We conducted a typical post-processing procedure that involved the removal of outlier and low-accuracy GPS location points. During the analysis and processing of the data, we removed a total of 258 participants due to insufficient data or missing sociodemographic attributes necessary for our models. These exclusions were necessary to ensure the integrity and reliability of our findings by focusing only on participants with complete data ( $n=759$  individuals). We followed a typical GPS tracking data workflow that focuses on identifying stops (e.g., activities) and trips connecting stops (Gong et al., 2014). Stop points were defined as locations where participants remained stationary for at least 5 minutes, with the movement between consecutive GPS points measuring less than 75 meters (Long et al., 2023).

To accurately determine the participants' home and work locations, we employed agglomerative clustering on the stop points recorded during the study. This clustering approach allowed us to group together spatially proximal stop points within a distance threshold of 75 meters, providing clusters that represent potential home and second-place locations. We assigned the mean location of the cluster with the longest duration as the participant's home location, representing their primary residential area, and the second longest duration as their second place (which would typically be considered their work location). This methodology aligns with previous studies that have successfully identified home locations from GPS tracking data using similar measures, such as the longest

duration of time spent (Sifa-Nowicka et al., 2016) or the longest duration of time spent at night (Kung et al., 2014).

To validate the accuracy of the GPS-derived home locations, we compared them with the self-reported home postal code (from the survey questionnaire) as well as the centroid locations of the postal codes provided by the United Kingdom Office for National Statistics. Our analysis revealed that the GPS-derived home locations were within 1 km of the self-reported home postal code centroid in approximately 80.7% of cases. Similarly, we compared the GPS-derived second-place locations to the provided work postal code centroid and found they were within 1 km of the provided work postal code centroid in about 50% of cases. This proximity suggests that these second-place locations often correspond to an individual's workplace; however, they are not exclusively representative of work locations. This is likely especially true in the context of people working from home for whom the second-place might not be their workplace. Therefore, these locations more broadly reflect a significant place in an individual's daily routine beyond their home.

We employed the OpenStreetMap (OSM) road network dataset for the calculation of all distances between location points (i.e., as defined in equations 1-3). The OSM road network is globally more than 80% complete, providing a comprehensive and reliable source of road network data (Barrington-Leigh & Millard-Ball, 2017). This extensive coverage is even more pronounced in Western countries like the United Kingdom, where the road network in OSM is considered a reliable representation of the true road network (Barrington-Leigh & Millard-Ball, 2017). We also used the OSM POI dataset to represent the geographical distribution of available amenities and activities. The OSM POI dataset is known for its global coverage, although the level of detail may vary in different regions (Mashhadi et al., 2013). However, it is regarded as one of the most accurate and publicly accessible datasets for points of interest (Barron et al., 2014; Girres & Touya, 2010; Haklay, 2010). This dataset provided us with comprehensive coverage of various types of establishments and amenities (number of types=188) in the United Kingdom.

### 4.3.2 Data analysis

For each individual with complete data ( $n = 759$ ), we calculated the local mobility index using, first, only their home location as an anchor point (LMI1), and second, using both the home location and the second-place as two anchor locations reflecting a bi-centric activity pattern (LMI2). To associate stop points with nearby POIs from the OSM POI dataset, we used a search radius of  $r = 75\text{m}$ . Our analysis showed a low sensitivity to the POI search radius, with only minor variations observed across different distances (Appendix F). In all cases, we used the mean distance of all trips (for the entire dataset;  $\hat{d}_c = 10050\text{ m}$ ) as the critical distance for quantifying the localness score of non-POI stops in equation 4.2. In our analysis, 70.4 % of all activities (stops) were associated with a nearby POI (using  $r = 75\text{m}$ ), and therefore the influence of non-POI stop points on LMI in our study was relatively low.

To determine the extent to which the LMI is capturing information already present in existing measures, we compare the newly derived LMI with six existing and widely used mobility and activity space measures (Table 4-1). We calculated Pearson's correlation coefficient between LMI and each of the six commonly employed mobility metrics from Table 4-1 and visually explored their relationship using scatter plots.

**Table 4-1 Summary of currently available mobility metrics often used to capture relative local mobility patterns and a selected reference.**

Measure	Description	Selected Reference
Average number of daily stops (NDS)	Reflects the average number of stops each participant makes per day, serving as an indicator of individual travel frequency. The calculation involves counting the number of stops made by each individual per day, defined by a time threshold of 5 minutes and a distance	(Yang et al., 2024)

---

	threshold of 75 meters. The average of these daily stop counts is then calculated for each participant to determine their NDS.	
Average total daily travel distance (TDD)	Reflects the average total daily distance traveled by participants, which provides insights into their mobility. The distance is computed from the GPS data of each individual for each day.	(Yang et al., 2024)
Farthest distance from home (FDH)	Measures the maximum travel distance of participants. It is determined by calculating the Euclidean distance from the home location to each stop point and then selecting the maximum distance among these.	(Yang et al., 2024)
Radius of gyration (ROG)	Quantifies the spatial dispersion of participants' movements, offering insight into the extent of their mobility range. The classical Radius of Gyration is calculated using the centroid of all stop points, with the Euclidean distance between the centroid and each stop point determining the radius.	(González et al., 2008)
Convex hull area (CHA)	Represents the individuals' activity space, calculated by creating a convex hull using all stop points for each individual and then measuring the area of this convex hull.	(Williams et al., 2013)
Local Travel Index (LTI)	A previously established metric for evaluating local mobility behavior. It is calculated using the areas of the convex hulls, assessing the spatial dispersion and standardization of these areas. For detailed calculation methods, refer to the provided reference.	(Manaugh & El-Geneidy, 2012)

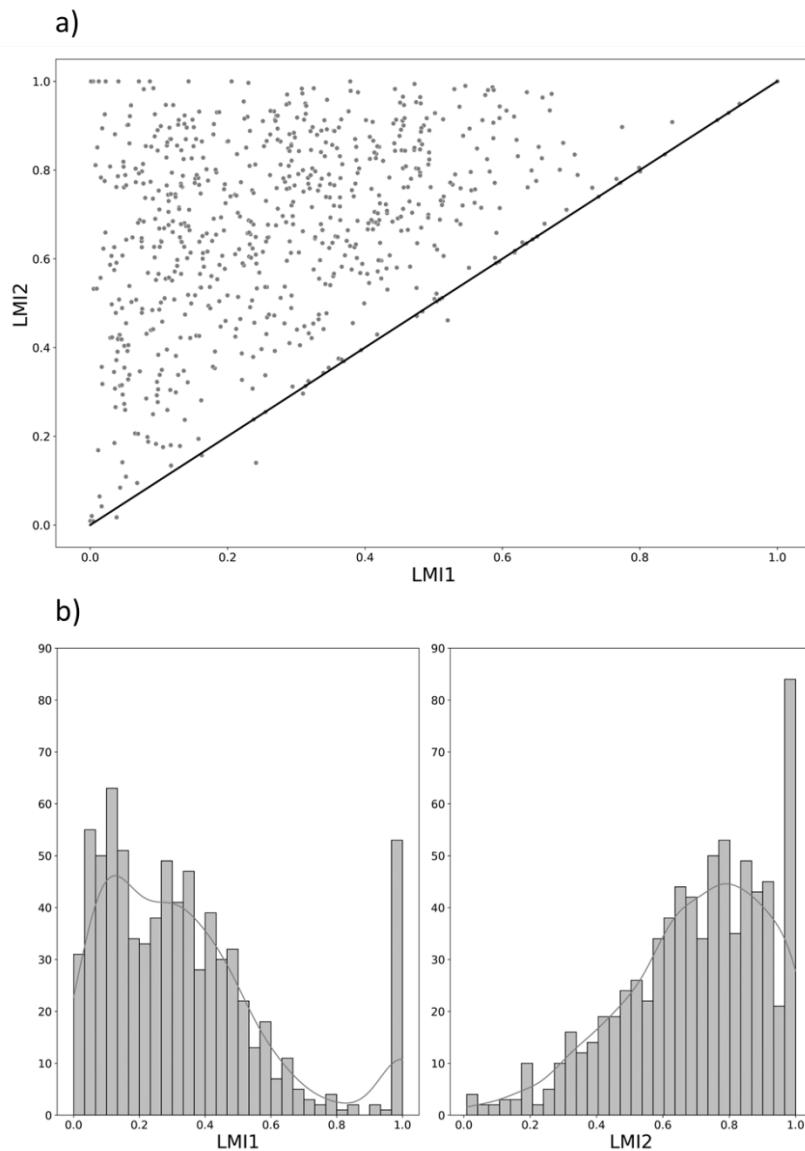
---



To explore the relationship between the LMI (and other commonly used mobility measures) and individual-level sociodemographic attributes, we use a generalized linear model. We included 14 independent variables included in the analysis: self-reported gender, age (in range categories), education, occupation, having a health issue that limits mobility, presence of children in the household, household type, car access and employment characteristics (part-time/full-time employed, employment status). Additionally, we incorporated a variable for distance from home to city-center to capture differences between urban core and city periphery residents. We also included the metric percentage of time homeworking, which represents the proportion of each individual's total working hours spent at home. The number of days of survey effort (ranging from 5 to 10) was also controlled for in our analysis. To assess multicollinearity among the selected covariates, we used the variance inflation factor (VIF). The results indicated no evidence of multicollinearity, as all chosen independent variables had VIF values below 3. We ran 8 versions of the same model using LMI1 and LMI2 as the dependent variable, along with the six commonly reported spatial measures of mobility (Table 4-1). We assessed goodness-of-fit of each model using the adjusted  $R^2$ . All statistical analysis was performed in R.

### 4.3.3 Results

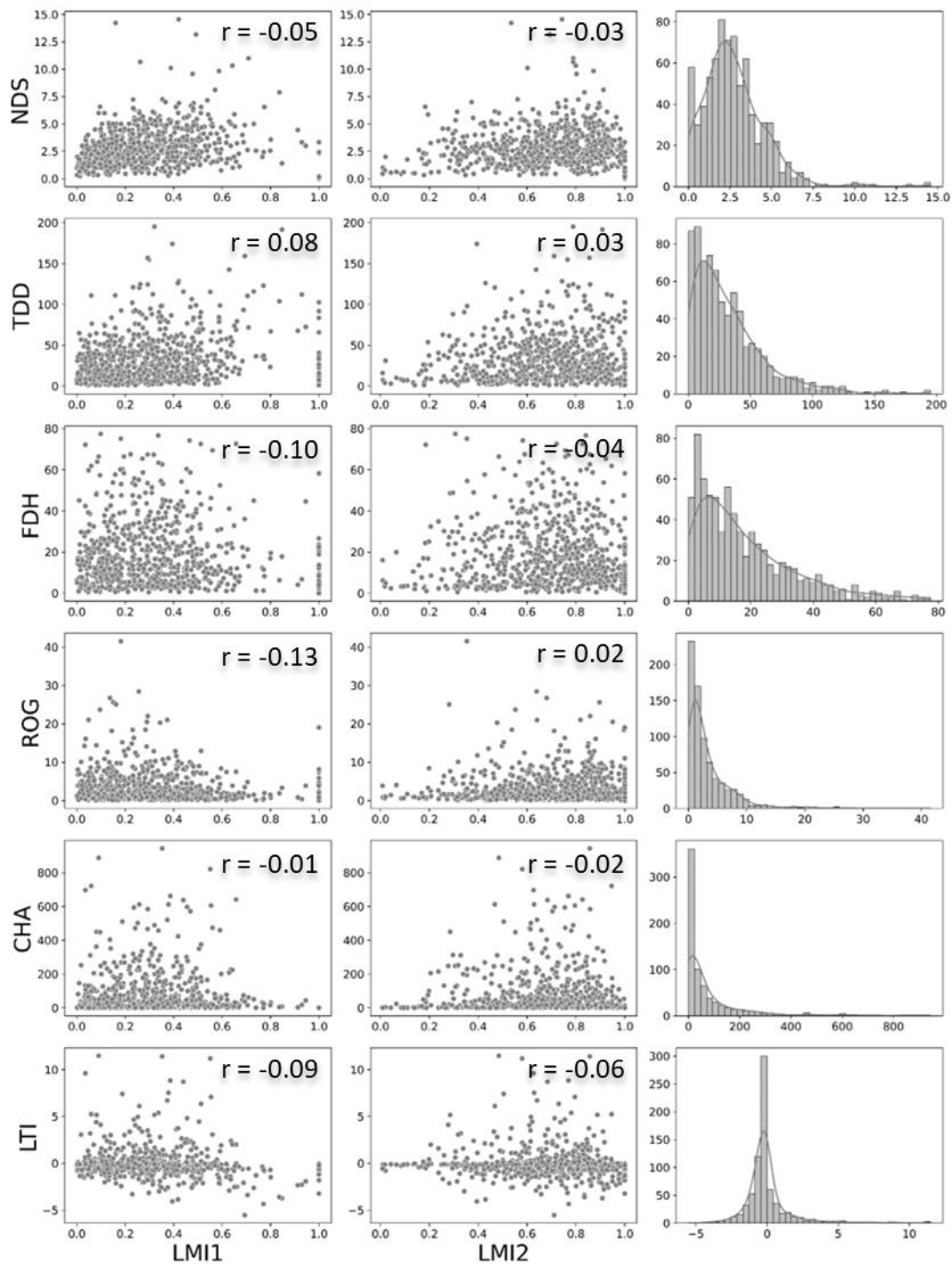
In our dataset we found LMI values spread across the full range of LMI values from 0 to 1 (Figure 4-3). We found a large difference in the mean LMI values between the two LMIs with LMI2 on average twice as large as LMI1 (mean LMI1 =0.33, mean LMI2 = 0.69; Figure 4-3-a). There was a relatively strong, positive correlation between LMI1 and LMI2 measures ( $r=0.50$ ). However, it was observed that the distribution of LMI1 was positively skewed (skewness=1.16), while the distribution of LMI2 exhibited a negative skew (skewness=-0.67) (Figure 4-3-b). There was a large peak at LMI = 1 in both frequency distributions (Figure 4-3-b) suggesting that these individuals have hyper-local movement patterns. Further inspection of the data revealed that these users typically had only a few activities in their data, associated with the home-only or home and second-place.



**Figure 4-3 LMI1: local mobility index to home; and LMI2: local mobility index to home and second-place. a) Scatter plot of LMI1 and LMI2, showing that almost all LMI2 values (with the exception of 26 data points) are equal to or greater than LMI1 values. b) Histograms of LMI1 and LMI2, where a positive skewness is observed in LMI1 and a negative skewness in LMI2.**

Correlation analysis between LMI1 and LMI2 and the six commonly employed mobility measures revealed that in all cases the LMI is uncorrelated with these existing measures

( $r < 0.15$  in all cases; Figure 4-4). The strongest relationship observed was a weak negative correlation ( $r = -0.13$ ) between the LMI and the ROG metric (Figure 4-4). The negative direction of the weak correlations in most cases in these plots are expected based on the definition of the LMI vs the other metrics, where increasing values represent increasingly local behavior in the LMI. The magnitude of these comparatively weak correlations supports that LMI is providing altogether different information of local mobility behavior relative to the selected six established measures.



**Figure 4-4 Correlations between two LMI metrics and six commonly used mobility. LMI1: local mobility index to home; LMI2: local mobility index to home and second-place; NDS: average number of daily steps; TDD: average total daily travel**

**distance (km); FDH: farthest distance from home (km); ROG: radius of gyration (km); CHA: convex hull area (km<sup>2</sup>); LTI: local travel index.**

In our analysis, we investigated how various mobility metrics, including two different LMI measurement methods, correlate with various social and demographic attributes (Table 4-2). We also created boxplots for employment characteristics in the LMI models to provide a clearer visualization of the variability within each category of these variables. Those with part-time jobs have a more local behavior centered around their home than full-time employees in the LMI1, but there is no significant difference between part-time and full-time employment when home and second-location (LMI2) are used (Table 4-2; Figure 4-5-a). Employees show a more local behavior compared to self-employed individuals or business owners on both measures (Table 4-2; Figure 4-5-b). Moreover, when considering only the home location as the anchor, the greater the distance of the home to the city center, the less local are individual's mobility patterns, but when both the home and the second-location are measured, those reside closer to the city centre show more local patterns. (Table 4-2; Figure 4-6). Age, gender, education level, parenthood status, employment characteristic, having a health issue that limits mobility, household status, car access, city of residence, percentage of time spent homeworking, and survey effort showed no significant association with LMI1 and LMI2.

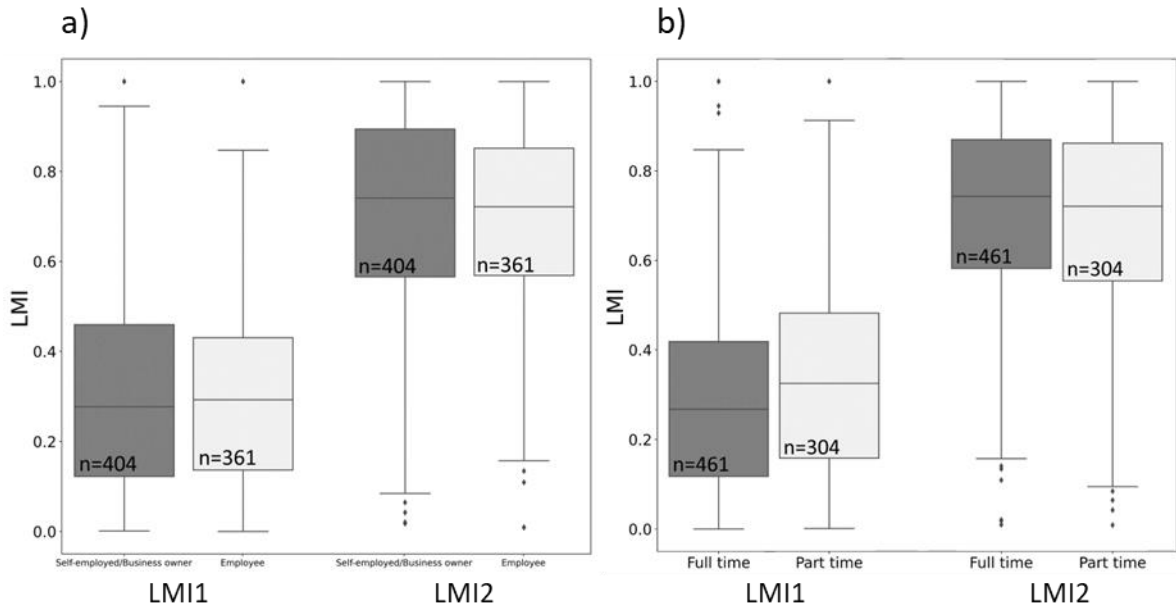
**Table 4-2 Results ( $\beta$  coefficients) from eight linear regression models with the eight mobility metrics as the dependent variable. Bold relationships indicate  $p < 0.05$**

	LMI1	LMI2	Num. of Daily Stops	Total Daily Distance	Farthest Distance	ROG	Convex Hull	LTI
<b>Variable</b>								
(Intercept)	<b>0.40***</b>	<b>0.40***</b>	1.33	<b>-2.03***</b>	<b>-2.77***</b>	<b>-3.11***</b>	<b>-3.31***</b>	<b>-3.46***</b>
Male [RC <sup>a</sup> : Female]	0.02	0.01	<b>0.53***</b>	<b>0.28***</b>	<b>0.22**</b>	<b>0.17*</b>	<b>0.29***</b>	<b>0.35*</b>
<b>Age</b> [RC: 18-24]								
25-34	0.05	0.02	-0.30	-0.09	-0.22	-0.02	-0.14	-0.11
35-44	0.02	0.04	-0.12	-0.12	-0.24	-0.17	-0.10	0.07
45-54	0.04	0.03	-0.50	-0.13	<b>-0.28*</b>	-0.12	-0.20	-0.13
55-64	0.05	0.06	-0.05	-0.08	<b>-0.52**</b>	<b>-0.35*</b>	<b>-0.34*</b>	-0.44
No higher education [RC: Higher education]	0.02	0.02	0.20	0.02	-0.02	-0.13	-0.03	-0.13
Has child/ren [RC: No child]	0.04	0.02	<b>0.36*</b>	0.02	-0.08	-0.10	-0.09	-0.12
Part-time job [RC: Full-time job]	<b>0.06**</b>	0.00	0.09	-0.10	0.01	<b>-0.18*</b>	0.09	<b>0.31*</b>
Has a mobility issue [RC: No mobility issue]	-0.02	-0.03	-0.04	-0.10	0.01	0.03	-0.17	-0.09
<b>Occupation</b> [RC: Managers, Directors and Senior Officials]								
Administrative and Secretarial	-0.02	-0.00	0.29	0.00	-0.16	-0.08	-0.08	-0.05
Associate Professional and Technical	-0.01	<b>-0.06*</b>	0.06	0.06	-0.19	-0.07	0.00	0.03
Caring, Leisure and Other Service	0.04	-0.01	<b>0.62*</b>	-0.06	-0.27	-0.16	-0.01	-0.06
Elementary	-0.04	<b>-0.12**</b>	-0.47	-0.09	<b>-0.43*</b>	-0.04	-0.10	-0.05
Process, Plant and Machine Operatives	-0.10	-0.04	<b>2.58***</b>	<b>0.97**</b>	-0.13	-0.23	0.18	-0.49
Professional	0.01	-0.01	0.21	0.21	-0.01	-0.08	0.06	-0.09
Sales and Customer Service	-0.07	-0.00	-0.31	-0.24	<b>-0.42*</b>	0.01	-0.21	-0.08
Skilled Trades	-0.02	0.03	0.25	0.13	-0.17	-0.01	-0.01	-0.16
Other/Insufficient detail to code	0.06	0.01	-0.22	0.09	-0.39	-0.03	-0.00	0.05
Single household [RC: Multiple occupants]	-0.03	-0.05	0.38	-0.13	0.06	-0.15	0.04	0.27
Has access to car [RC: No access to car]	0.03	0.02	0.12	<b>0.44***</b>	<b>0.45***</b>	<b>0.26**</b>	<b>0.33***</b>	0.12
Employee [RC: Self-employed/Business owner]	<b>0.06*</b>	<b>0.04*</b>	-0.21	-0.15	-0.15	0.07	<b>-0.21*</b>	-0.29
<b>City</b> [RC: Brighton and Hove]								
Birmingham	0.03	0.05	<b>0.68**</b>	0.17	-0.15	-0.19	-0.10	-0.17

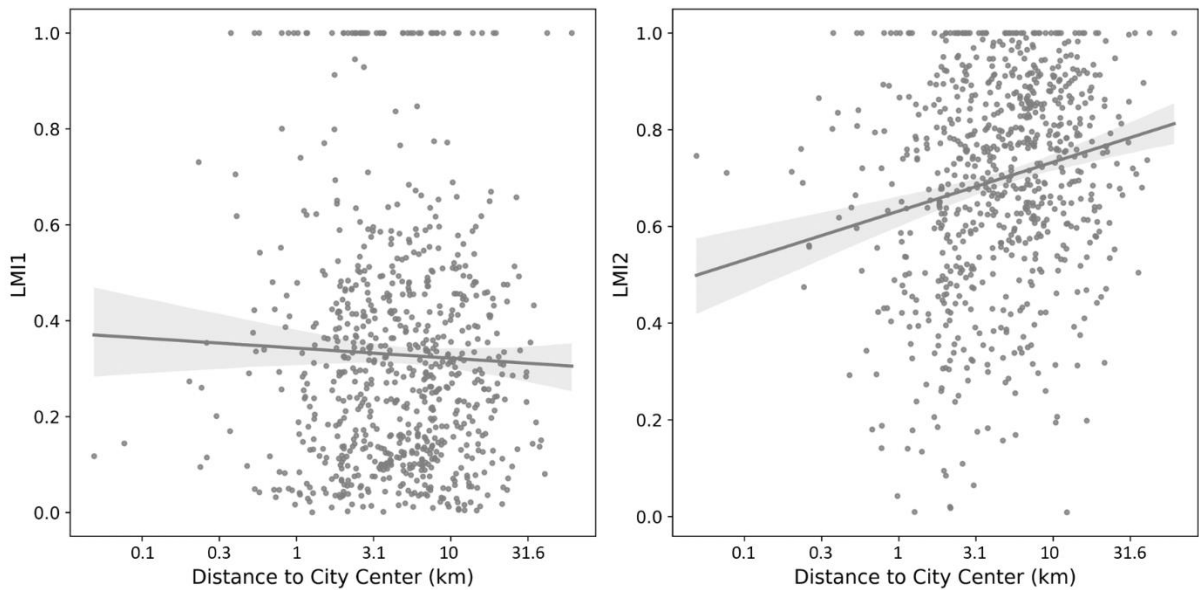
Leeds	0.01	0.03	0.11	-0.07	-0.07	-0.12	-0.01	0.19
Distance to city center	<b>-0.02*</b>	<b>0.02*</b>	0.05	<b>0.16***</b>	<b>0.26***</b>	<b>0.33***</b>	<b>0.29***</b>	<b>0.38***</b>
% of time homeworking	0.01	-0.01	-0.07	-0.00	-0.04	-0.02	-0.02	-0.03
Survey effort	-0.00	0.00	0.05	<b>0.06*</b>	<b>0.09***</b>	<b>0.06**</b>	<b>0.11***</b>	0.02
Number of observation <sup>b</sup>	759	759	759	759	759	759	737	737
R <sup>2</sup>	0.053	0.098	0.109	0.196	0.184	0.155	0.183	0.087
R <sup>2</sup> adjusted	0.020	0.066	0.078	0.167	0.155	0.125	0.153	0.053

<sup>a</sup> Reference Category. \*\*\* for  $p < 0.001$ , \*\* for  $p < 0.01$ , and \* for  $p < 0.05$ .

<sup>b</sup> LTI and Convex Hull methods necessitates a minimum of three data points. Participants with fewer than three points yield NA values and were consequently excluded from the regression analysis.



**Figure 4-5 a) Boxplots of employment characteristics against LMI models. b) Boxplots of worker type categories against LMI models. SE/BO: self-employed/business owner.**





**Figure 4-6 Scatter plots and best fit lines of distance to city center against LMI1 and LMI2 (n=765). To enhance clarity in visualization, we have applied a log<sub>10</sub> transformation to the distance values before plotting.**

Individuals who identified as men were associated with a greater number of daily stops, longer total travel distances, farther distances from home, higher ROG values, larger convex hull areas, and higher LTI values compared to women. Having access to a car significantly increases total daily distance, farthest distance, ROG, and convex hull measurements. Furthermore, the distance to the city center was positively associated with all other range-based mobility metrics except with the average number of daily stops. Survey effort (number of days in the survey) was positively associated with total daily distance, farthest distance from home, ROG, and convex hull.

The  $R^2$  values for the LMI to home and LMI to home and second place models are notably low, at 0.020 and 0.066 respectively. This indicates a limited variance in the dependent variable being explained by the independent variables in these models. In contrast, the total daily distance, farthest distance from home, and convex hull models exhibited a higher R-squared value among all the models, suggesting a stronger explanatory power of the independent variables for the variation in these metrics.

## 4.4 Discussion

In this paper, we developed a new local mobility index (LMI) which can be used to quantify the ‘relative localness’ of an individual’s activity pattern through linking individuals’ choices for localized destinations with their travel patterns. While previous research focused on comparing mobility levels among individuals (Manaugh & El-Geneidy, 2012), addressing ‘local relative to whom,’ our approach shifts the emphasis to ‘local relative to where,’ evaluating activity choices against the nearest available options. We demonstrate how the anchor location used in LMI (typically the home) can be altered to capture multi-centric activity patterns (e.g., home and second place; (Ahas et al., 2010; Sila-Nowicka et al., 2023)). In our data we focused on individuals who are workers, and

therefore over 50% of the second most significant places (anchor points) in our dataset are associated with individual's work location. Other applications of LMI could creatively choose the anchor location to explore other local-mobility hypotheses, for example associated with key urban features (Kloosterman & Musterd, 2001), second-homes (Hiltunen & Rehunen, 2014), or the home-work-corridor (Adler & Ben-Akiva, 1979; Hanson, 1980; X. Wang et al., 2013). Increasing the number of anchor points is expected to raise the LMI values, tending closer to 1.

When quantifying mobility patterns, it is essential to recognize that people residing and working in different environments often have different levels of access to amenities (Newman & Kenworthy, 2006; Vine et al., 2012), which in turn impacts mobility patterns. While accessibility methods assess the accessibility to amenities by concentrating on the static interplay of supply and demand (Geurs & Van Wee, 2004; Guagliardo, 2004; Hansen, 1959; Luo & Wang, 2003; F. Wang, 2012), they do not consider individual movement and cannot be regarded as mobility metrics. The proposed LMI has addressed this limitation in existing methods by incorporating both individual movement patterns and the availability of amenities within their respective environments.

We found no strong correlation between LMI metrics and six established mobility metrics. This outcome underscores the distinctiveness of LMI as a novel and complementary measure that stands apart from these conventional mobility indicators and provides a potentially new dimension (i.e., localness) of individual mobility patterns (Fillekes et al., 2019; Long et al., 2023). This divergence highlights the necessity of considering LMI alongside commonly used metrics to comprehensively capture and interpret individual mobility patterns. This suggests that assumptions underlying traditional metrics—such as considering longer distances as indicative of greater mobility (Barbosa et al., 2018)—may not fully account for the complexity of local travel behaviors, which include not just the extent but the nature of engagement with the local environment.

It is noteworthy that LMI metrics offer new insights, especially when considering the relationship with sociodemographic factors. Variables, such as sex and age, which are frequently associated with variations in travel behavior and choices (Benito & Oswald, 2000; McQuaid & Chen, 2012; Rouwendal & Rietveld, 1994; Schmöcker et al., 2005), did not emerge as significant predictors of LMI. We only included individuals who are in employment which could be one reason for the observed differences. However, it could be, and this should be investigated further using different data, whether LMI produces different findings such as in relation to sex and age differences in travel behavior. The relatively low  $R^2$  values, particularly in the context of LMI metrics, underscore the complexity and multifaceted nature of mobility behaviors, suggesting that they cannot be fully explained by a limited set of sociodemographic variables.

Part-time employment was observed to increase localness according to LMI1, but this effect was not observed with LMI2. This difference suggests that full-time workers, who often have their workplace as the second significant location measured in our study, display less localized movement patterns compared to part-time workers when only the home location is considered. This discrepancy arises because when the second-place location is not included in the LMI calculation, movements related to and around the workplace—significant for full-time employees due to their longer hours and higher engagement—are likely to be categorized as non-local compared to their home location. Consequently, this leads to an apparent decrease in localness for full-time workers in the first model. However, when second places are considered in LMI2, movements related to work for full-time workers are recognized as local, diminishing the observed difference in localness between part-time and full-time workers. This highlights the importance of inclusion or exclusion of significant places in shaping the perceived localness values.

Our findings highlight that self-employed individuals and business owners exhibit a lower degree of localness in their mobility behavior compared to employees. This is consistent with findings from studies such as those by Long and Reuschke (Long & Reuschke, 2021), which showed that the daily mobility patterns of small business owners

and self-employed individuals demonstrate higher daily mobility compared to premise-based employees. This could be attributed to a variety of reasons, including the structured nature of premise-based employees' work, which typically necessitates travel to a fixed defined location. Exhibiting a higher degree of localness in premise-based employees' mobility potentially results in shorter travel times and lower travel costs compared to self-employed individuals and small business owners. This implies a different set of transportation and urban planning needs for self-employed individuals and small business owners group compared to employees who commute to a fixed workplace. Understanding these distinctions is crucial for developing tailored policies and infrastructure that support the diverse mobility needs of urban populations, thereby enhancing overall urban mobility efficiency and sustainability.

We observed that not accounting for second-places in LMI1 leads to a decrease in measured localness for individuals living farther from the city center. In contrast, when second-places are included in LMI2, localness increases for those residing at a greater distance from the city center. This change occurs because, in the first model, second places are counted as stops and included in the calculation, often leading to decreased localness due to their distance from home. Conversely, in the second model, second-place stops are excluded from the calculation, and any other stops near the second-place are measured in relation to second place location and its nearby POIs. This method reveals that incorporating second places when measuring relative local mobility behavior that may be especially important to consider when individuals have regular daily mobility patterns around their second-places.

LMI1 and LMI2 exhibit no correlation with traditional range-based indices, underscoring the LMIs' capacity to reveal distinct aspects of mobility patterns that range-based metrics fail to capture (Figure 4-4). This observation suggests that relying solely on range-based indices to evaluate localness could lead to misinterpretations, as these indices might either overestimate or underestimate localness levels. This is further substantiated by the varied range values associated with different levels of localness, demonstrating that high

or low ranges are not reliable indicators of localness (Figure 4-4). This variability underscores the importance of incorporating contextual information, such as the availability of nearby activities, into mobility assessments. Localness is inherently relative (Manaugh & El-Geneidy, 2012), emphasizing the necessity for measures like LMI that consider the geographical context to accurately reflect individuals' mobility patterns.

While our methodology involved selecting the closest available POI to participants' stop points representing the intended destination, similar to (Bohte & Maat, 2009), we acknowledge that there may be cases where individuals chose a different yet nearby POI. This variation in POI selection can potentially impact the calculation of LMI, as we compare the distance to the nearest POI of the same type. If the closest POI differs from the one the individual actually visited, it could alter the resulting LMI value. However, it is important to note that with the data at hand, we are unable to always accurately determine the exact destination chosen by individuals – especially in urban areas with a high density of POIs. Additionally, it is possible that individuals visited multiple destinations during a single stop, exploring such multi-destination scenarios represents an area of opportunity for future developments in studying relative localness and urban travel behavior.

Although our study focused on individual patterns, future research could broaden the scope to include the impact of geographical contexts, such as differences in home neighborhoods and workplaces, on LMI. This expanded analysis could inform infrastructure optimization and urban accessibility improvements (El-Geneidy & Levinson, 2006; Grengs et al., 2010; Handy, 1993). Specifically, by aggregating LMI scores for different areas, urban planners can tailor their strategies to local needs. For example, high aggregate LMI scores, reflecting strong local mobility, suggest a focus on enhancing resources for active transportation like walking and biking. Conversely, low aggregate scores highlight areas where local mobility is limited, prompting an investigation into the barriers to local travel. Addressing these barriers might involve

improving pedestrian infrastructure and expanding public transit options. Additionally, examining aggregated LMI scores can reveal how mobility patterns affect socioeconomic factors, including local business growth and community engagement. Analyzing the relationship between neighborhood LMI scores and access to essential services or levels of community participation could guide community development efforts and urban policy decisions (Guzman et al., 2021; Zhang et al., 2021).

## 4.5 Conclusion

We developed a novel approach to measuring local mobility, termed the local mobility index (LMI) which offers a more comprehensive understanding of individual movement behaviors through considering the unequal distribution (and accessibility) of urban amenities. We were able to uncover distinct patterns of localized mobility preferences among individuals. By bridging the gap between spatiotemporal movement indicators and accessibility measurement methods, the LMI contributes to various fields, including transportation planning and urban studies. The lack of substantial correlations between LMI and commonly used mobility metrics emphasizes its uniqueness as a distinct measure that captures a different facet of mobility behavior.

Our analysis highlights the importance of various demographic and social factors in influencing local mobility behavior. Although some variables were significant in both LMI models and other metrics, we observed a diverse range of variables with differing levels of significance. These findings underscore the distinct nature of LMI compared to other metrics. It is crucial to clarify that this does not imply LMI is a replacement for existing measures, but rather, we aim to emphasize the unique information and insights LMI provides in analyzing human mobility behavior.

In light of these insights, our study contributes to a more nuanced understanding of mobility behaviors, underscoring the intricate interplay between individual preferences, social dynamics, and urban characteristics. The methodology we introduced opens new

avenues for research into understanding the factors that shape local mobility behaviors, which, in turn, can inform urban planning, policy-making, and resource allocation.

## References

- Adler, T., & Ben-Akiva, M. (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, 13(3), 243–257.
- Ahas, R., Silm, S., Järv, O., Saluveer, E., & Tiru, M. (2010). Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, 17(1), 3–27. <https://doi.org/10.1080/10630731003597306>
- Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Harvard university press.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1–74.
- Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PloS One*, 12(8), e0180698.
- Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895.
- Bearden, W. O. (1977). Determinant attributes of store patronage-downtown versus outlying shopping centers. *Journal of Retailing*, 53(2), 15.
- Belk, R. W. (1975). Situational variables and consumer behavior. *Journal of Consumer Research*, 2(3), 157–164.
- Benito, A., & Oswald, A. J. (2000). *Commuting in Great Britain in the 1990s*.

- Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, 17(3), 285–297.
- Brezzi, M., & Veneri, P. (2015). Assessing polycentric urban systems in the OECD: Country, regional and metropolitan perspectives. *European Planning Studies*, 23(6), 1128–1145.
- Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075), 462–465.
- Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., & Jia, Q. (2011). Exploratory data analysis of activity diary data: a space–time GIS approach. *Journal of Transport Geography*, 19(3), 394–404.
- Degeratu, A. M., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, 17(1), 55–78.
- Demissie, M. G., Phithakkitnukoon, S., Kattan, L., & Farhan, A. (2019). Understanding human mobility patterns in a developing country using mobile phone data. *Data Science Journal*, 18, 1.
- Dijst, M., & Vidakovic, V. (2000). Travel time ratio: the key factor of spatial reach. *Transportation*, 27, 179–199.
- Doyle, P., & Fenwick, I. (1974). How store image affects shopping habits in grocery chains. *Journal of Retailing*, 50(4), 39–52.
- El-Geneidy, A. M., & Levinson, D. M. (2006). *Access to destinations: Development of accessibility measures*.



- Erdem, O., Ben Oumlil, A., & Tuncalp, S. (1999). Consumer values and the importance of store attributes. *International Journal of Retail & Distribution Management*, 27(4), 137–144.
- Fillekes, M. P., Giannouli, E., Kim, E.-K., Zijlstra, W., & Weibel, R. (2019). Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research. *International Journal of Health Geographics*, 18(1), 1–20.
- Frey, N. (2017). Equity in the distribution of urban environmental amenities: the case of Washington, DC. *Urban Geography*, 38(10), 1534–1549.
- Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2), 127–140.
- Girres, J., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.
- Gong, L., Morikawa, T., Yamamoto, T., & Sato, H. (2014). Deriving personal trip data from GPS data: A literature review on the existing methodologies. *Procedia-Social and Behavioral Sciences*, 138, 557–565.
- González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779–782. <https://doi.org/10.1038/nature06958>
- Graells-Garrido, E., Serra-Burriel, F., Rowe, F., Cucchiatti, F. M., & Reyes, P. (2021). A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in Barcelona. *PloS One*, 16(5), e0250080.
- Grengs, J., Levine, J., Shen, Q., & Shen, Q. (2010). Intermetropolitan comparison of transportation accessibility: sorting out mobility and proximity in San Francisco and Washington, DC. *Journal of Planning Education and Research*, 29(4), 427–443.

- Guagliardo, M. F. (2004). Spatial accessibility of primary care: concepts, methods and challenges. *International Journal of Health Geographics*, 3(1), 1–13.
- Guzman, L. A., Arellana, J., Oviedo, D., & Aristizábal, C. A. M. (2021). COVID-19, activity and mobility patterns in Bogotá. Are we ready for a ‘15-minute city’? *Travel Behaviour and Society*, 24, 245–256.
- Hackett, P. M. W., Foxall, G. R., & Van Raaij, W. F. (1993). Consumers in retail environments. In *Advances in psychology* (Vol. 96, pp. 378–399). Elsevier.
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4), 682–703.
- Handy, S. (1993). *Regional versus local accessibility: Implications for nonwork travel*.
- Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73–76.
- Hanson, S. (1980). The importance of the multi-purpose journey to work in urban travel behavior. *Transportation*, 9(3), 229–248.
- Hiltunen, M. J., & Rehunen, A. (2014). Second home mobility in Finland: Patterns, practices and relations of leisure oriented mobile lifestyle. *Fennia-International Journal of Geography*, 192(1), 1–22.
- Hirsch, J. A., Winters, M., Clarke, P., & McKay, H. (2014). Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis. *International Journal of Health Geographics*, 13(1), 1–14.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Ranges of human mobility in Los Angeles and New York. *2011*

*IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 88–93.

Kloosterman, R. C., & Musterd, S. (2001). The polycentric urban region: towards a research agenda. *Urban Studies*, 38(4), 623–633.

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PloS One*, 9(6), e96180.

Lindquist, J. D. (1974). Meaning of image-survey of empirical and hypothetical evidence. *Journal of Retailing*, 50(4), 29-+.

Logan, T. M., Anderson, M. J., Williams, T. G., & Conrow, L. (2021). Measuring inequalities in urban systems: An approach for evaluating the distribution of amenities and burdens. *Computers, Environment and Urban Systems*, 86, 101590.

Long, J. A., Lee, J., & Reuschke, D. (2023). Activity graphs: Spatial graphs as a framework for quantifying individual mobility. *Journal of Geographical Systems*, 1–26.

Long, J., & Reuschke, D. (2021). Daily mobility patterns of small business owners and homeworkers in post-industrial cities. *Computers, Environment and Urban Systems*, 85. <https://doi.org/10.1016/j.compenvurbsys.2020.101564>

Luo, W., & Wang, F. (2003). Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design*, 30(6), 865–884.

Malekzadeh, M., & Long, J. A. (2023). *Mobility Deviation Index: Incorporating geographical context into analysis of human mobility*, [Manuscript submitted for publication]. Department of Geography, Western University.

Manaugh, K., & El-Geneidy, A. M. (2012). What makes travel 'local'? Defining and understanding local travel behavior. *Journal of Transport and Land Use*, 5(3), 15–27.

- Mashhadi, A., Quattrone, G., & Capra, L. (2013). Putting ubiquitous crowd-sourcing into context. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 611–622.
- McQuaid, R. W., & Chen, T. (2012). Commuting times—The role of gender, children and part-time work. *Research in Transportation Economics*, 34(1), 66–73.
- Mir, D. J., Isaacman, S., Cáceres, R., Martonosi, M., & Wright, R. N. (2013). Dp-where: Differentially private modeling of human mobility. *2013 IEEE International Conference on Big Data*, 580–588.
- Muth, R. F. (1961). The spatial structure of the housing market. *Papers of the Regional Science Association*, 7(1), 207–220.
- Newman, P., & Kenworthy, J. (2006). Urban design to reduce automobile dependence. *Opolis*, 2(1).
- OpenStreetMap Wiki. (2022, October 11). *Points of Interest - OpenStreetMap*. OpenStreetMap. [https://wiki.openstreetmap.org/wiki/Points\\_of\\_interest](https://wiki.openstreetmap.org/wiki/Points_of_interest)
- Pandey, B., Brelsford, C., & Seto, K. C. (2022). Infrastructure inequality is a characteristic of urbanization. *Proceedings of the National Academy of Sciences*, 119(15), e2119890119.
- Peng, C., Jin, X., Wong, K.-C., Shi, M., & Liò, P. (2012). Collective human mobility pattern from taxi trips in urban area. *PloS One*, 7(4), e34487.
- Pozoukidou, G., & Chatziyiannaki, Z. (2021). 15-Minute City: Decomposing the new urban planning eutopia. *Sustainability*, 13(2), 928.
- Purifoy, D. M. (2021). North Carolina [Un] incorporated: Place, race, and local environmental inequity. *American Behavioral Scientist*, 65(8), 1072–1103.
- Rigolon, A. (2016). A complex landscape of inequity in access to urban parks: A literature review. *Landscape and Urban Planning*, 153, 160–169.

- Rouwendal, J., & Rietveld, P. (1994). Changes in commuting distances of Dutch households. *Urban Studies*, 31(9), 1545–1557.
- Sahasrabuddhe, R., Lambiotte, R., & Alessandretti, L. (2021). From centre to centres: polycentric structures in individual mobility. *ArXiv Preprint ArXiv:2108.08113*.
- Schmöcker, J.-D., Quddus, M. A., Noland, R. B., & Bell, M. G. H. (2005). Estimating trip generation of elderly and disabled people: Analysis of London data. *Transportation Research Record*, 1924(1), 9–18.
- Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84), 20130246.
- Shuttleworth, I., & Gould, M. (2010). Distance between home and work: a multilevel analysis of individual workers, neighbourhoods, and employment sites in Northern Ireland. *Environment and Planning A*, 42(5), 1221–1238.
- Sila-Nowicka, K., Fotheringham, A. S., & Demšar, U. (2023). Activity triangles: a new approach to measure activity spaces. *Journal of Geographical Systems*, 1–29.
- Sila-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5), 881–906. <https://doi.org/10.1080/13658816.2015.1100731>
- Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10), 818–823. <https://doi.org/10.1038/nphys1760>
- Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, 327(5968), 1018–1021.

- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1), 126–146. <https://doi.org/https://doi.org/10.1016/j.datak.2007.10.008>
- Veneri, P. (2018). Urban spatial structure in OECD cities: Is urban population decentralising or clustering? *Papers in Regional Science*, 97(4), 1355–1374.
- Vine, D., Buys, L., & Aird, R. (2012). The use of amenities in high density neighbourhoods by older urban Australian residents. *Landscape and Urban Planning*, 107(2), 159–171.
- Wang, F. (2012). Measurement, optimization, and impact of health care accessibility: a methodological review. *Annals of the Association of American Geographers*, 102(5), 1104–1112.
- Wang, X., Grengs, J., & Kostyniuk, L. (2013). Visualizing travel patterns with a GPS dataset: How commuting routes influence non-work travel behavior. *Journal of Urban Technology*, 20(3), 105–125.
- Williams, N. E., Thomas, T., Dunbar, M., Eagle, N., & Dobra, A. (2013). Measurement of human mobility using cell phone data: developing big data for demographic science. *Population Association of America Annual Meeting*.
- Yang, X., Li, J., Fang, Z., Chen, H., Li, J., & Zhao, Z. (2024). Influence of residential built environment on human mobility in Xining: A mobile phone data perspective. *Travel Behaviour and Society*, 34, 100665.
- Yilmaz, V. (2004). Consumer behavior in shopping center choice. *Social Behavior and Personality: An International Journal*, 32(8), 783–790.
- Yılmaz, V., Aktaş, C., & Celik, H. E. (2007). *Development of a scale for measuring consumer behavior in store choice*.

Zhang, S., Yang, Y., Zhen, F., Lobsang, T., & Li, Z. (2021). Understanding the travel behaviors and activity patterns of the vulnerable population using smart card data: An activity space-based approach. *Journal of Transport Geography*, *90*, 102938.

Zhao, K., Tarkoma, S., Liu, S., & Vo, H. (2016). Urban human mobility data mining: An overview. *2016 IEEE International Conference on Big Data (Big Data)*, 1911–1920.

## Chapter 5

# 5 A Network Community Structure Similarity Index for Weighted Networks<sup>1</sup>

## 5.1 Introduction

Due to their complex structure, the identification of communities continues to be an essential part of the analysis of networks (Fortunato, 2010a). Accordingly, measuring similarities between communities is a fundamental part of analyzing community structure in different (usually related) networks (Dongen, 2000; Fan et al., 2007; Gusfield, 2002; Gustafsson et al., 2006; Meilă, 2007; Mirkin, 1996; Rand, 1971; Traud et al., 2008; Wallace, 1983; Zhang et al., 2006). Methods for assessing network community similarity can be employed to assess network models (e.g., comparing to known communities) (Hu et al., 2008; Lancichinetti & Fortunato, 2009), comparing different states of the system (e.g., before and after an intervention), or time series analysis in networks (Duan et al., 2009).

Although several methods exist for quantifying network community similarity, such as the widely-used Jaccard index (Meilă, 2007) and mutual information (Danon et al., 2005), they predominantly focus on community labels and neglect the intricate relationships encoded within edge weights. A critical drawback arises when a uniform metric treats dissimilarities in edges and their corresponding weights equivalently. For example, in a subway network, the cost of incorrectly clustering the central station (with a high number of edges and traffic flows due to its functionality) is the same as the cost of incorrectly clustering an intermediate station (with a low number of edges and traffic flows). This gap in the existing literature underscores the need for a novel approach that

---

<sup>1</sup> A version of this chapter has been published (Malekzadeh, M., & Long, J. A. (2023). A network community structure similarity index for weighted networks. *Plos one*, 18(11), e0292018. <https://doi.org/10.1371/journal.pone.0292018>).



effectively incorporates edge weights into the community similarity calculation. Additionally, the application of current methods is not possible when comparing networks with different number of nodes.

We propose a novel network community structure similarity index (NCSSI) as a similarity index for comparing network communities. The formulation of the NCSSI is based on the concept of edit distance, whereby the cost of each node changing its community is computed based on their edges and corresponding weights. By taking this approach, the NCSSI provides an advantage over existing methods in that it is able to incorporate community labels, edges and their weights into a single measurement of community similarity. Employing simulated data and real-life New York Yellow Taxi flows data, we highlight how NCSSI compares to existing approaches and provides an alternative view of how similarity structure can be measured between communities.

The structure of the article is as follows: in section 2 we provide a general background on network structure, community detection, and similarity of communities. Then we introduce NCSSI and its implementation. In section 4 we experiment with NCSSI and simulated data in two different scenarios: on benchmarks to compare with existing similarity methods, and on benchmarks where we change the weights of random nodes to highlight the limitation of existing similarity methods. Then in section 5 we implement NCSSI on a real-life mobility dataset, New York Yellow Taxi flows data. In the discussion and conclusion, we discuss the advantages of the proposed method and the results, and we conclude the study.

## 5.2 Background/Literature Review

Many complex systems can be modeled by networks (also commonly termed graphs) (Fortunato, 2010b). A network consists of a set of nodes (or vertices) and a set of edges (Saoub, 2021). Edges can include varying weights or direction and nodes can have multiple attributes. Different systems can be modeled by different structures of graphs

such as directed (Barabasi & Oltvai, 2004), weighted (Bast et al., 2016), overlapping and clustered (Palla et al., 2005), or hierarchical graphs (Kemp & Tenenbaum, 2008).

Unlike random graphs, in which the probability of having an edge between two nodes among all pairs of nodes is equal and the distribution of edges is homogenous (Erdős & Rényi, 1960), real-life networks show a high level of heterogeneity in edge distribution (Fortunato, 2010a). These heterogeneities are often represented as clusters within which intra-cluster edge density is much higher than inter-cluster edges. These clusters are typically referred to as communities in networks (Girvan & Newman, 2002) and these two terms are usually used interchangeably. In real-life networks, nodes of communities usually share similar characteristics and/or similar interests (e.g., social networks; (Girvan & Newman, 2002)) and/or have higher levels of interactions (e.g., transportation networks; (Zheng et al., 2014)). Finding communities in networks can provide new insights into the structure of networks and enable us to better understand the underlying complex system.

Different algorithms have been developed to detect communities from networks. Graph partitioning methods divide the graph into a specific number of communities with a specified number of nodes to minimize the number of inter-community edges (Barnes, 1982; Kernighan & Lin, 1970). Hierarchical algorithms comprise two types (Hastie et al., 2009): bottom-up (agglomerative) in which all similar communities are merged until reaching a specified threshold and top-down (divisive) which adhere to the opposite direction—dividing dissimilar communities until reaching a specified threshold. A very well-known divisive community detection method was proposed by Girvan and Newman (Girvan & Newman, 2002; Newman & Girvan, 2004). In Girvan and Newman's method, edges are iteratively removed based on their value of betweenness which expresses the frequency of shortest paths between all pairs of nodes that pass along the edge. An alternative set of methods are modularity-based algorithms (Clauset et al., 2004; Newman, 2004b). Modularity, as a quality measurement of community detection, is computed based on the comparison between the existing structure of edges in the

subgraph and the density of edges in the subgraph if the nodes were attached irrespective of the community structure (Newman, 2004a; Newman & Girvan, 2004). Spectral algorithms (Donetti & Munoz, 2004), methods based on statistical inferences (Newman & Leicht, 2007), and algorithms that use deep learning methods (Kipf & Welling, 2016) are the other alternatives. For a more comprehensive review of clustering and community detection algorithms, refer to (Fortunato, 2010b; Jin et al., 2021).

Methods for measuring the similarity (or differences) between two sets of communities have been employed in methodological work to test the performance of a community detection method on a benchmark (Huang et al., 2010), or to cross-check the results of community detection using different methods (Lancichinetti & Fortunato, 2009). In applied settings, network community similarity is used across various domains such as studying similarity of sets of communities over time (Yang et al., 2011), and comparing different types of networks (Taya et al., 2016). While there is extensive literature on community detection methods, there is a lack of attention to similarity measurements for network communities (Fortunato, 2010a).

There are four primary types of measures for assessing network community similarity: pair counting, community matching, information theory, and distance. Pair counting methods identify corresponding communities in two different sets. If a node is identified in the same community (two communities with the highest overlapping) in two different sets, it is considered correctly detected. The Rand index (Rand, 1971) and its adjusted version (Meilă, 2007), the Jaccard index and its adjusted version (Meilă, 2007), and the Mirkin metric (Mirkin, 1996) are all pair counting methods. Community matching measures are based on finding the largest overlap between two sets of communities. However, these methods may only consider the largest similar portion of communities, potentially disregarding certain parts of the community set as a whole. Classification error, defined by Meilă & Heckerman (Meilă & Heckerman, 2001), and the normalized Van Dongen metric (Dongen, 2000) are examples of community matching measures. A third type of similarity measure is based on information theory (MacKay, 2003). Using

information theory, if two sets are alike, there is little information needed to derive one set from the other. Less similarity between the two sets indicates that more information is needed to infer one from the other. A commonly used measure of this type is the normalized mutual information (Danon et al., 2005). Lastly, distance-based measures in which the number of movements and divisions are considered to calculate the distance between two sets. Movement is defined as the minimum number of nodes that are needed to be moved from one community to the meet set of two communities to match, and divisions are defined as the number of divisions needed for one community to match with the meet set of two communities (Gusfield, 2002; Gustafsson et al., 2006).

### 5.3 Network Community Structure Similarity Index (NCSSI)

We first define notation that will be used throughout the paper. Let:

$V$  be a set of  $n$  nodes in network  $N$

$E$  be a set of edges that connect network nodes

$W$  be a set of weights associated with each edge

A community set is defined as a set of  $>1$  communities derived from a given network. The goal of NCSSI is to quantify the similarity between two community sets. NCSSI is derived from the edit distance concept in which the distance is computed based on the total cost to transform one community set into another. The critical feature of NCSSI is the cost function which is associated with each of the necessary edits (inserts and removes) to perform this transformation. We should emphasize that in NCSSI, we are not trying to transform one network to another but rather to transform one community set into another. This is an important distinction as transforming the whole structure of the graphs is unnecessary when the principal focus is only on the community structures. Hence, NCSSI focuses on nodes whose labels differ between the two sets. Then, for each of these nodes, we calculate the cost associated with the minimum edits needed to change the community labels. Calculation of NCSSI follows a three-step method outlined below.

### 5.3.1 Step 1 – Pairing communities

To find the similarity between two community sets, we must find the most similar community pairs in each of these two sets. Since both community sets have different community labels, to find the most similar community pairs, we must calculate the most overlapping communities in each set in terms of both nodes and weighted edges. To calculate this, assume we have two communities  $x$  and  $y$ , from the community sets  $X$  and  $Y$ , respectively, where the intersection set of their nodes is  $\cap_{xy}$  and the union set is  $\cup_{xy}$ . We focus on incorporating edge weight information into the measurement of NCSLI. Therefore, we sum all edge weights ( $W$ ) between the nodes in  $\cap_{xy}$  and  $\cup_{xy}$  and adjust them by dividing it by the total weights of edges in the network. Since we have two community sets with different configurations regarding their edges and associated weights, we need to calculate the sum of all edge weights for both Networks. Finally, the overlap score (OS) of communities  $x$  and  $y$  is defined as the ratio of the sum of all adjusted edge weights in  $\cap_{xy}$  to  $\cup_{xy}$ :

$$OS_{xy} = \frac{\frac{\sum_{i,j \in \cap_{xy}} W_{ij}^X}{\sum_{i,j} W_{ij}^X} + \frac{\sum_{i,j \in \cap_{xy}} W_{ij}^Y}{\sum_{i,j} W_{ij}^Y}}{\frac{\sum_{i,j \in \cup_{xy}} W_{ij}^X}{\sum_{i,j} W_{ij}^X} + \frac{\sum_{i,j \in \cup_{xy}} W_{ij}^Y}{\sum_{i,j} W_{ij}^Y}} \quad 5.1$$

Where  $OS_{xy}$  is the overlap score between communities  $x$  and  $y$  from the community sets of  $X$  and  $Y$ ;  $\cap_{xy}$  and  $\cup_{xy}$  are the intersection and union sets of nodes from communities  $x$  and  $y$ ; and  $W_{ij}^X$  and  $W_{ij}^Y$  are the edge weights between  $i$  and  $j$  in the community sets  $X$  and  $Y$ , respectively.

To find the paired communities, we need to calculate all the possible OS values between communities of two community sets, resulting in an OS matrix. The unordered bijective pairs are chosen by the maximum OS, meaning each community in a set is paired only with one community from the other set with the highest overlap. It is possible to have different numbers of communities between the two community sets. If, due to the

difference in the number of communities in sets, a community (let us assume  $x$ ) is not paired with a community in the other set, the pair will be denoted as  $(x, \phi)$ .

### 5.3.2 Step 2 – Calculating edit costs of nodes

Based on the definition of communities, the stronger the intra-community links, and the weaker the inter-community links, the stronger the structure of communities. Hence, in calculating NCSSI, we concentrate on the nodes' labels and the edge weights associated with the nodes.

In this step, we need to calculate the minimum cost associated with the edits needed to transform the node  $i$ 's membership (Equation 5.2 and 5.3). In equation 5.2 and 5.3, we assume that we have community set  $X$ , which contains communities  $x_1$  and  $x_2$ , and  $Y$ , which contains communities  $y_1$  and  $y_2$ . Additionally,  $x_1$  and  $y_1$ , and,  $x_2$  and  $y_2$ , are paired with each other.

$$EC_i^X = \left| \sum_{j \in x_1} W_{ij} - \sum_{j \in x_2} W_{ij} \right| \quad 5.2$$

$$EC_i^Y = \left| \sum_{j \in y_1} W_{ij} - \sum_{j \in y_2} W_{ij} \right| \quad 5.3$$

Where  $EC_i^X$  and  $EC_i^Y$  is the adjusted edit cost of node  $i$  in the community sets  $X$  and  $Y$ , respectively; Within Equation 5.2,  $x_1$  is the community in which  $i$  is a member; and  $x_2$  is the community which is the paired community of the changed community  $y_2$ ; similarly within Equation 5.3,  $y_2$  is the community in which  $i$  is a member; and  $y_1$  is the community which is the paired community of the changed community  $x_1$ ;  $j$  is a node member of the communities  $x_1$ ,  $x_2$ ,  $y_1$ , and  $y_2$ ; and  $W_{ij}$  is the weight associated with the edge between nodes  $i$  and  $j$ .

These costs reflect the necessary modifications required to transform the membership of a node from one community to another while considering the associated edge weights.

However, it is important to note that these costs may not be on the same scale for both community sets  $X$  and  $Y$ , as they depend on the specific characteristics of the community sets and the corresponding edge weights. To calculate edit costs on a common scale, it is necessary to adjust them by the sum of intra-community edge weights associated with node  $i$  and the sum of weights between  $i$  and paired community nodes. To adjust edit costs, we need to compute the adjustment factors for both community sets  $X$  and  $Y$  (Equation 5.4 and 5.5). It should be noted that the adjustment factors include weightings from all nodes irrespective of whether the node's community membership remains unchanged or changed.

$$AF^X = \sum_i \left( \sum_{j \in x_1} W_{ij} + \sum_{j \in x_2} W_{ij} \right) \quad 5.4$$

$$AF^Y = \sum_i \left( \sum_{j \in y_1} W_{ij} + \sum_{j \in y_2} W_{ij} \right) \quad 5.5$$

Where  $AF^X$  and  $AF^Y$  are the adjustment factor associated with community  $X$  and  $Y$ , respectively. The notation for  $i, j, x_1, x_2, y_1, y_2$ , and  $W_{ij}$  remains the same as previously defined.

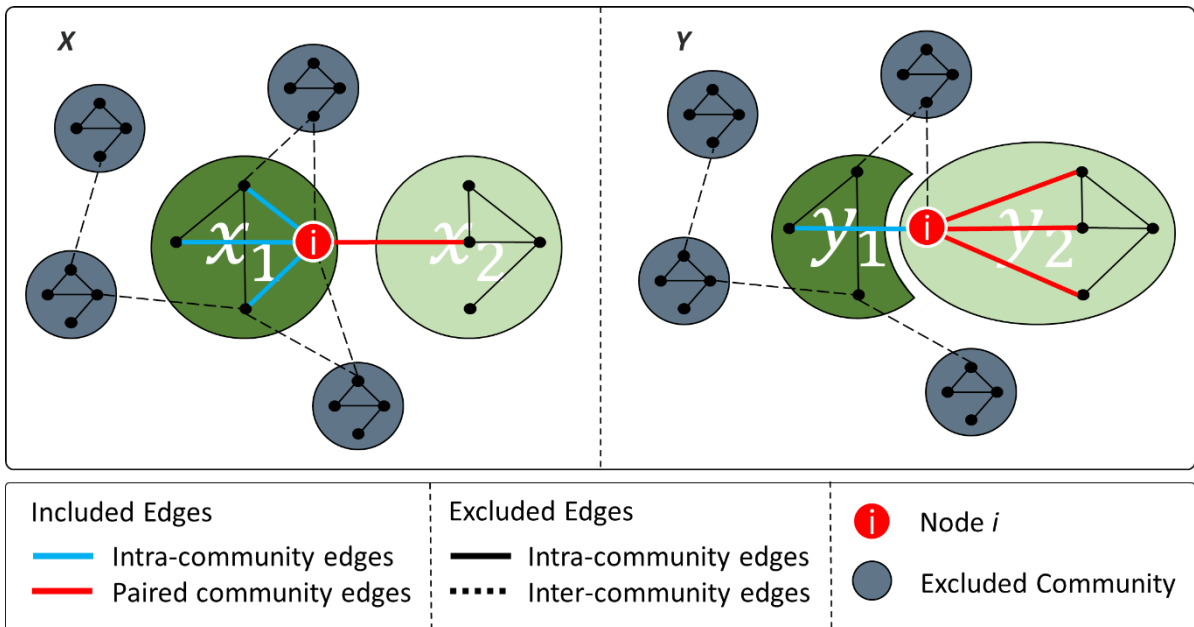
By dividing the edit costs by the adjustment factors, we obtain the adjusted edit costs on a common scale for both community sets (Equation 5.6 and 5.7). These adjusted edit costs are bounded between 0 and 1 for each community set, ensuring that both community sets are on the same scale and comparable. It is important to note that the sum of all adjusted edit costs for each community set cannot exceed 1, as they represent the relative transformation costs within the community sets.

$$AEC_i^X = \frac{EC_i^X}{AF^X} \quad 5.6$$

$$AEC_i^Y = \frac{EC_i^Y}{AF^Y} \quad 6.7$$

Where  $AEC_i^X$  and  $AEC_i^Y$  are the adjusted edit cost of node  $i$  in the community sets  $X$  and  $Y$ , respectively. The notation for  $EC_i^X$ ,  $EC_i^Y$ ,  $AF^X$  and  $AF^Y$  remains the same as previously defined.

As an example, with the same assumption as before, we have community set  $X$ , which contains communities  $x_1$  and  $x_2$ , and  $Y$ , which contains communities  $y_1$  and  $y_2$ . Additionally,  $x_1$  and  $y_1$ , and,  $x_2$  and  $y_2$ , are paired with each other. If node  $i$  is a member of  $x_1$  in community set  $X$  and a member of  $y_1$  in community set  $Y$ , as the two communities are paired with each other, the node  $i$  has not changed its community membership— $NEC_i$  is zero. However, if node  $i$  is a member of  $x_1$  in community set  $X$  and a member of  $y_2$  in community set  $Y$  (Figure 5-1), as these two communities are not paired with each other, it means that the node has changed its community membership and we need to consider it in our calculation. In this case, in each community set (let us first consider community set  $X$  and then we must repeat this procedure for community set  $Y$ ) we have two communities that we need to concentrate on;  $x_1$  in which  $i$  is a member, and  $x_2$  which is the paired community of the changed community  $y_2$  (Figure 5-1).





**Figure 5-1 An example of edit cost calculation when node  $i$  has changed its community membership. Blue lines and red lines represent the edges that are considered in the calculation of the edit cost related to node  $i$ , as intra-community edges and edges to the paired community. Dashed black lines represent the inter-community edges and black lines represent the intra-community edges that were not considered in the calculation of the edit cost related to node  $i$ . The paired communities in different community sets are colored the same (dark green and light green). All the other communities that are not considered in the edit cost calculation for node  $i$ , are colored gray.**

### 5.3.3 Step 3 – NCSSI calculation

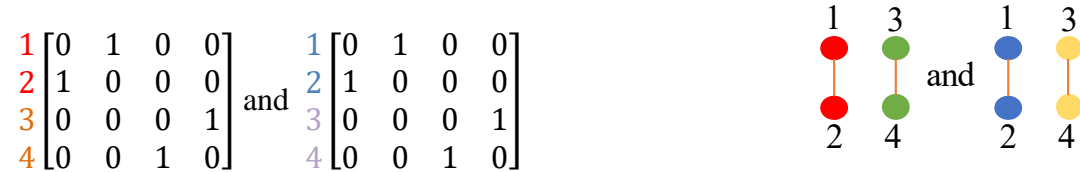
NCSSI is the total cost associated with transforming a node's membership and is equal to the sum of the adjusted edit costs of the node in each of the two sets. Consequently, to get the total cost of transforming one community set into another, we need to sum up all nodes' *NECs* from each community set. Then the total cost is obtained by averaging the total adjusted edit costs of both community sets (Equation 5.8). The selection of the average as the method for combining the adjusted edit costs is justified by several factors. Firstly, taking the average ensures a balanced consideration of both community sets, allowing for equal evaluation of their respective edit costs. Secondly, it preserves the information from both sets by incorporating the collective impact of changes, rather than favoring one set over the other (unlike selecting the minimum or maximum). Thirdly, the average maintains sensitivity to changes in both community sets, capturing the nuanced transformation between the structures. Finally, since the adjustment factors and edit costs are already on the common scale, the average does not introduce bias towards a specific range.

$$NCSSI_{XY} = 1 - \frac{\sum_i AEC_i^X + \sum_i AEC_i^Y}{2} \quad 5.8$$

Where  $NCSSI_{XY}$  is the Network Community Structure Similarity Index of community sets  $X$  and  $Y$ ;  $i$  is a node from the set of nodes that have changed their community

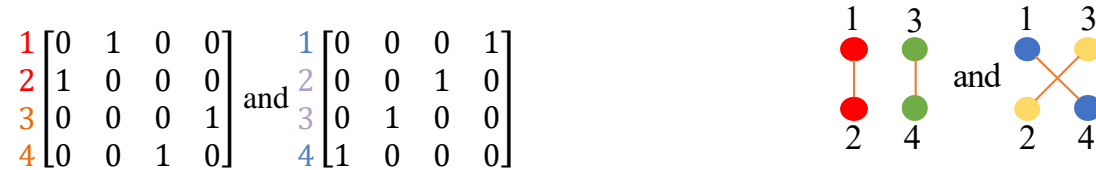
memberships. This results in a value between 0 to 1, which highlights the distance between two community sets. To convert distance to similarity, it is then subtracted from 1. The final value is then defined as the NCSSI of the two sets of communities ranging from 0 to 1 where 0 represents low similarity and 1 identical sets of communities

An example of identical sets of communities:



In which community pairs are (red, blue;  $OS=1$ ) and (green, yellow;  $OS=1$ ) and  $NCSSI=1$

An example of the most dissimilar sets of communities:



In which community pairs are (red, blue;  $OS=0$ ) and (green, yellow;  $OS=0$ ) and  $NCSSI=0$

In these examples, all weights assigned to the edges were 1 for simplicity.

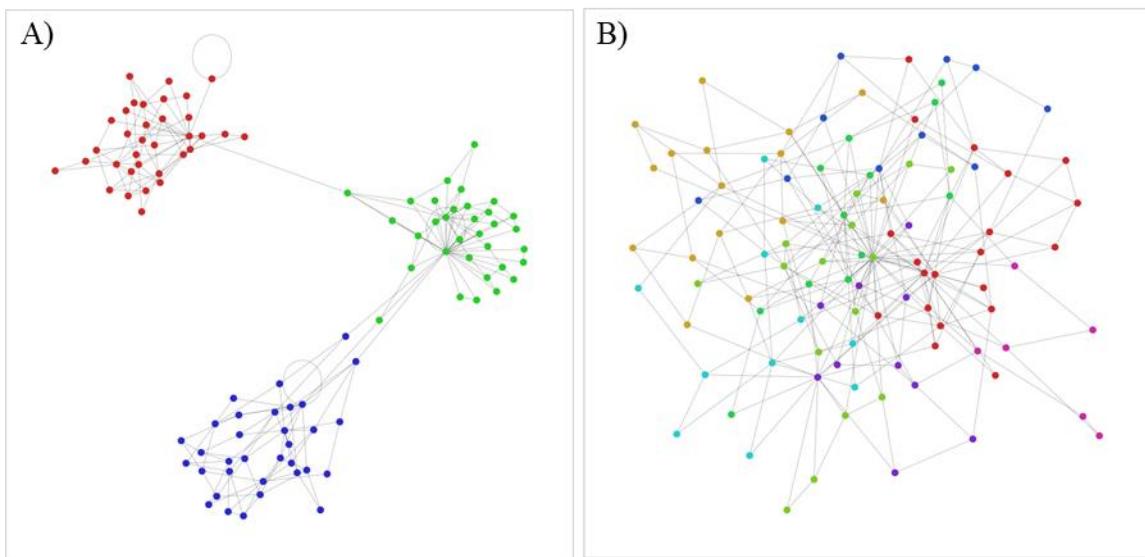
## 5.4 Simulation Tests

### 5.4.1 Benchmark

A benchmark network is a network in which all nodes' community labels are known. These benchmarks were originally introduced to enable researchers to test the results of community detection methods. Here, we are not interested in testing the clustering and community detection methods, rather, we want to test the community similarity measures using these benchmarks. We use computer-generated benchmarks as they enable us to adjust the network and its attributes with specifically chosen parameters. The most popular class of computer-generated benchmarks works based on  $p_{in}$  as the probability of

having intra-community links for nodes and  $p_{out}$  as the probability of having inter-community links while these two probabilities are independent (Condon & Karp, 2001). A specific case of this type was proposed by Girvan and Newman (Girvan & Newman, 2002) in which a graph with 4 communities and 32 nodes in each community is considered. This benchmark is not an ideal representative of real-life networks as the number of communities and the number of nodes in each community are constant and nodes' degrees are almost similar in each community.

The Lancichinetti-Fortunato-Radicchi (LFR) benchmark introduced by Lancichinetti et al., (Lancichinetti et al., 2008) is a modified version of this model in which the heterogeneity of nodes' degrees, the community sizes, and the number of communities is taken into account (Figure 5-2). In the LFR the distribution of nodes' degrees and the size of communities are determined by power law functions with two parameters of  $\tau_1$  and  $\tau_2$ , respectively. The intra-community degrees are determined by the fraction of  $1 - \mu$  of nodes' degree. Thus, the inter-community degrees are computed based on the fraction of  $\mu$  of nodes' degrees. When  $\mu$  is 0 we have highly separated communities but when  $\mu$  is 1 we have highly interconnected communities. Here, we use the LFR benchmark in our experiments.



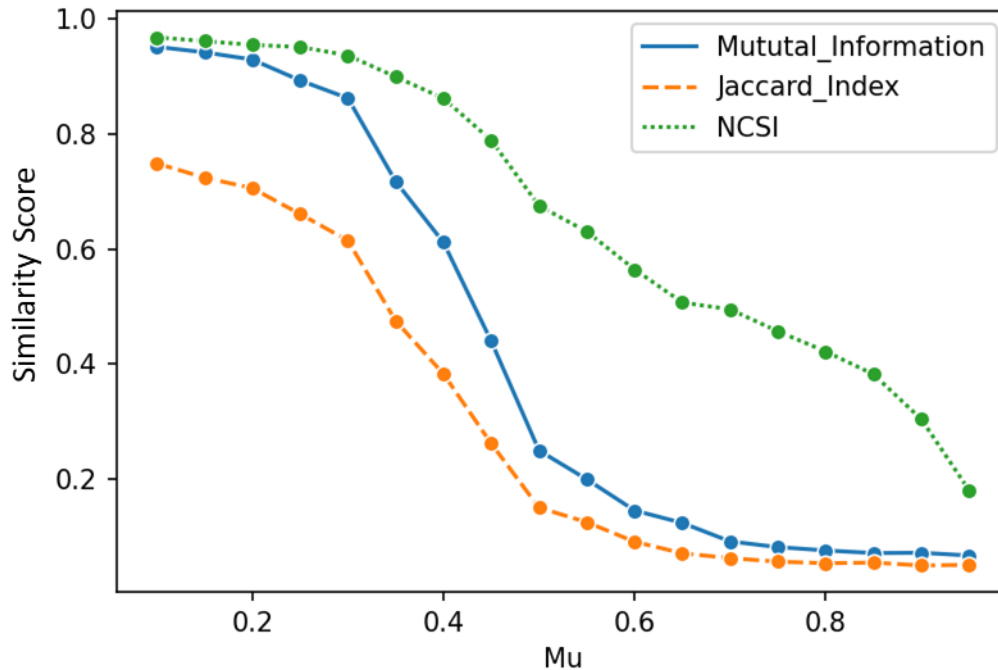
**Figure 5-2 A) LFR benchmark (Lancichinetti et al., 2008) when  $\mu$  equals 0.05, and B) LFR benchmark when  $\mu$  equals 0.95. Each color is representative of the nodes' communities.**

## 5.5 Different community detection methods

We generated simulated community graphs for comparison using the LFR benchmark. We varied  $\mu$  from 0.1 to 0.95 and for each  $\mu$  we ran the process 100 times as suggested by (Lancichinetti et al., 2008) to avoid random bias. We used  $n = 1000$  nodes in the graph and  $\tau_1 = 1.5$  and  $\tau_2 = 3$ , as the parameters of the LFR benchmarks. We deployed the Louvain community detection algorithm (Blondel et al., 2008) to derive communities from each benchmark. We computed the similarity index using the Jaccard index, mutual information, and NCSSI by comparing the communities detected by the algorithm to the known benchmark communities. We repeated this process deploying Clauset's algorithm (Clauset et al., 2004), asynchronous label propagation algorithm (Raghavan et al., 2007), and semi-synchronous label propagation algorithm (Cordasco & Gargano, 2010) (Appendix G).

The Jaccard index and mutual information resulted in low similarity between two sets of communities when  $\mu > 0.5$  (Figure 5-3). However, considering the network's structure in the similarity measurement, we observe that increasing  $\mu$  decreases the similarity between known and computed community sets. The relationship between  $\mu$  and NCSSI is more linear and monotonic than with the Jaccard index and mutual information measures. The reason that we still observe a low-level of similarity when  $\mu > 0.5$ , is that even in networks with low community structure, there is an inevitable random degree of similarity across the community sets. Based on LFR benchmark algorithm, when  $\mu$  equals 0.5, the number of inter-links and cumulative intra-links (links with all the other communities) will be equal. Although community structure is commonly perceived to weaken after  $\mu = 0.5$ , it is important to note that in networks with a substantial number of nodes and communities, community structure can still be discernible even when  $\mu > 0.5$ . This is because the presence of a large number of nodes and communities ensures that the

intra-links within each community outnumber the inter-links to individual communities. Thus, despite the higher fraction of inter-community edges, the overall network can still exhibit noticeable community structure.



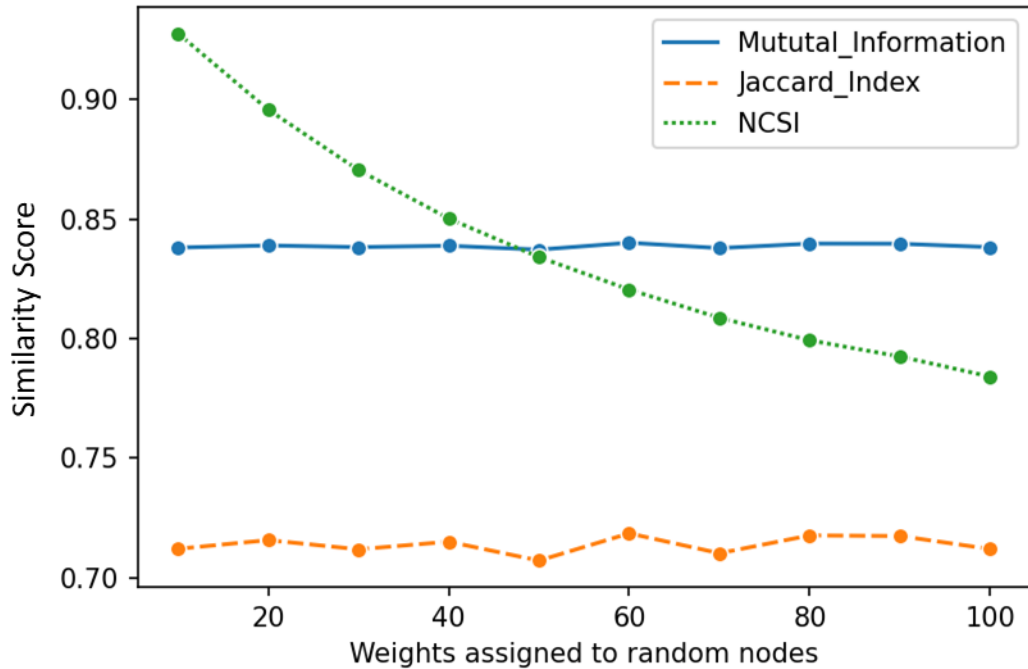
**Figure 5-3 Similarity scores deploying Louvain similarity measure on LFR benchmarks using two existing similarity measurement methods (the Jaccard index and mutual information) and the proposed method NCSI.**

### 5.5.1 Manually changing nodes' community memberships and their weights

In this example, we used an LFR benchmark ( $n = 1000$ ,  $\tau_1 = 1.5$ ,  $\tau_2 = 3$ , and  $\mu = 0.2$ ), in which all edges had the same weight of 1. We randomly selected a subset of 10 nodes, which constituted 1% of all nodes. To assess the impact of weights, we increased the weights of all the edges connected to these nodes incrementally from 10 to 100 (with an interval of 10). Then we changed these nodes' communities as an intervention (intentional misclustering). We aimed to examine how different similarity measures can capture the difference between the communities before and after the intervention. We repeated this

process for each weight 100 times and took the mean similarity score of the 100 repetitions.

The Jaccard index and mutual information measures were not sensitive to the changes in the weights of nodes as increasing the weights associated with these nodes did not affect similarity scores (Figure 5-4). In contrast, NCSSI captured the effect of increasing weights associated with the misclassified nodes, leading to lower similarity between community sets. These results clearly demonstrate that NCSSI is sensitive to changes in edge weights and the subsequent effect on network community similarity, while Jaccard index and mutual information measures are not.



**Figure 5-4 Similarity values for an LFR benchmark while changing the weights associated with 10 randomly chosen nodes using two existing similarity measurement methods (the Jaccard index and mutual information) and the proposed method NCSSI.**

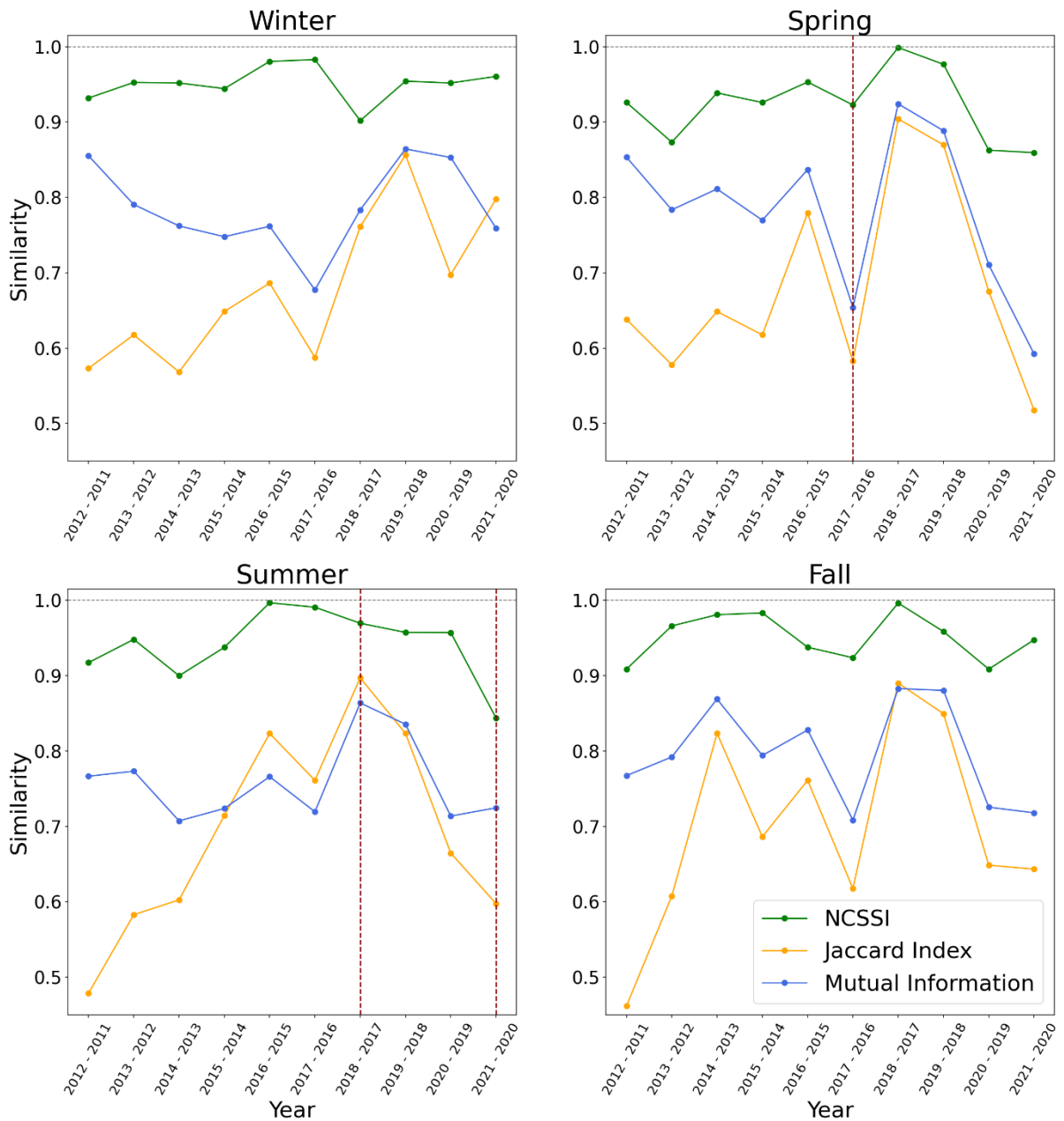
## 5.6 Case study: NYC Taxi Flow dataset

As a real-life example of how NCSSI can be implemented, we compared how community structures computed based on human mobility flows differ from 2011 to 2021 using Yellow Taxi Pick-up/Drop-off data provided by the New York City Taxi and Limousine Commission, USA (New York Taxi & Limousine Commission, 2023). This dataset contains more than 1.3 billion trips for 13,587 taxis in New York City, USA. The locational information in the dataset is aggregated into 263 taxi zones. Each data entry includes a pick-up zone and time, and drop-off zone and time, along with other attributes. We aggregated the data seasonally and created a flow (Origin-Destination) matrix between taxi zones for each season which is a form of a spatial network where zones represent nodes and taxi trip frequencies between zones represent edge weights. We then detected the communities employing Louvain's algorithm for each season. To avoid seasonal effects, we compared each season's community structure to the previous year's. There is no true community structure to compare to, however we can make some hypotheses. First, we do not expect to see community structures change drastically over time, and year to year communities should be relatively stable. However, we do expect to see strong differences associated with the COVID-19 pandemic and associated changes in human mobility patterns in 2019-2020. Further, the advent of ridesharing services (e.g., Uber, Lyft) which originated in 2011 may have an impact over time as they increase in popularity, especially in the less dense areas of NYC. Seven taxi zones appeared to have no trips for at least one season throughout the study period. This causes an inconsistency in the number of nodes in the networks. As mutual information and Jaccard index require the same number of nodes in both networks, this limitation forced us to eliminate inconsistent zones in the networks to compare the results of different methods.

We observe a relatively low similarity level between all pairs of consecutive years throughout 2011 to 2021 using the Jaccard index (Figure 5-5), but this is perhaps not surprising as Jaccard index always presented the lowest similarity between benchmarks. However, the similarity scores of the Jaccard index showed the greatest variation (i.e.,

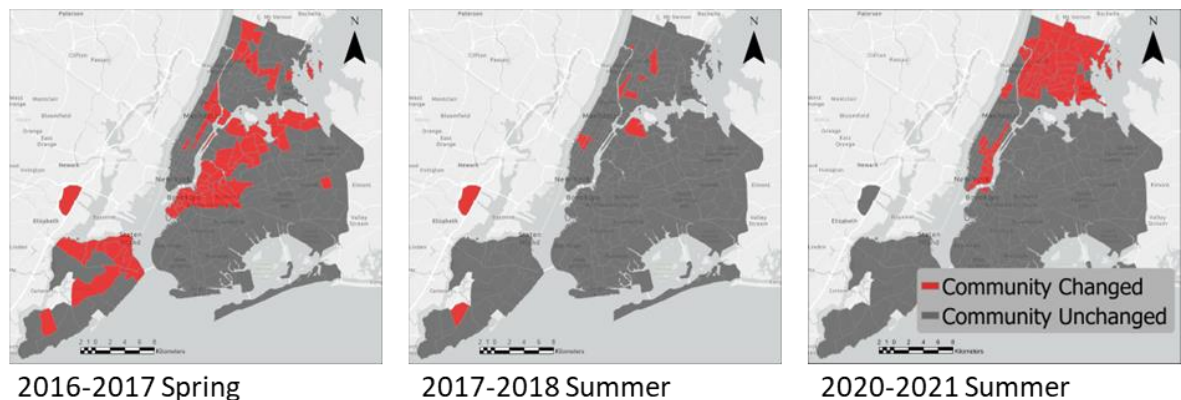
lack of consistency in similarity) between consecutive pairs of years (Figure 5-5). In general, mutual information found higher similarity values relative to Jaccard index (Figure 5-5), but again showed highly variable results from consecutive pairs of years. NCSSI exhibited the highest similarity values in each pair of years, and more interestingly, NCSSI demonstrated more stable similarity values in all pairs of years leading up to 2019. We see the effect of the COVID-19 pandemic on mobility result in a drop in similarity values in 2019-2020 spring for all measures. In the winter, fall, and spring we see that Jaccard index and mutual information measures show that this similarity continues to be low in the 2020-21 comparisons, but this pattern is not observed using NCSSI.





**Figure 5-5** The seasonal level of similarity in community structures for pairs of consecutive years using different indices from 2011 to 2021. The vertical axis demonstrates the similarity score. The dotted red line highlights the three periods of time we discuss in this section (spring 2016-2017; summer 2016-2017; and summer 2020-2021).

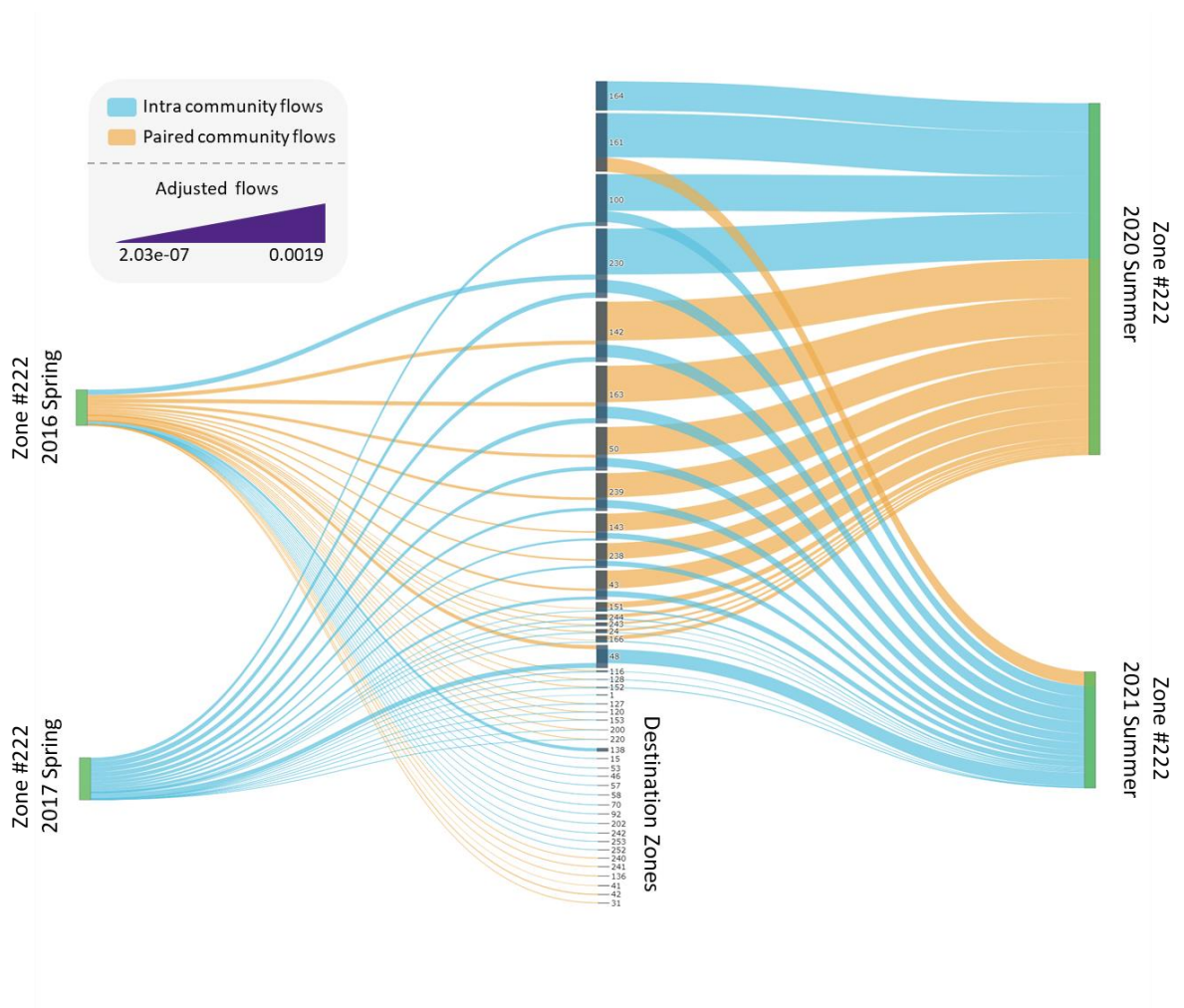
We chose three periods of time to compare different scenarios and mapped them to visually evaluate the results (Figure 5-6). The first period is the spring of 2017 compared to the spring of 2018, where the Jaccard index and mutual information resulted in a low level of similarity, but NCSSI showed a high level of similarity. Next is the summer of 2017 compared to the summer of 2018, where all measures demonstrated a high level of similarity. Finally, as an interesting period related to the Covid-19 pandemic, we compared the summer of 2020 to the summer of 2021 where all measures showed a relatively low level of similarity. We observe when the number of zones that changed their communities are higher, as in 2017-2018 Spring and 2020-2021 Summer, the Jaccard index and mutual information resulted in a low level of similarity. However, while the number of communities that changed their community was high, NCSSI did not result in a low level of similarity as the weights of the intra-community flows and flows to the paired community were not substantial in this case, resulting in low edit costs (Figure 5-7). On the other hand, when the flows substantially changed (as in the summer of 2020-2021), this results in high edit costs, and NCSSI captures a relatively low similarity level (Figure 5-7).



**Figure 5-6** Maps of New York communities computed based on NY Yellow Taxis flow data. Taxi zones colored red represent the areas that changed their communities, and gray zones, represent the areas that did not. Base map and its related data are from OpenStreetMap and OpenStreetMap Foundation (OpenStreetMap Wiki, 2022).

We plotted intra-community adjusted flows and adjusted flows to paired communities for the spring of 2016 and 2017, and the summer of 2020 and 2021, choosing taxi zone #222, which had a community change in all three periods of interest (Figure 5-7). It is important to note that the calculated adjusted flows presented in the plots are obtained by dividing the flows by the corresponding adjustment factors associated with each season and year. It is crucial to emphasize that these adjusted flows should be distinguished from actual flows. Despite the fact that the actual flows during the 2020-2021 period experienced a significant decrease, the higher values of adjusted flows compared to the 2016-2017 period is attributed to the flow adjustment process. This process is exclusively employed for the purpose of facilitating visualization, enabling us to visually compare the adjusted flows and providing clarification on how NCSSI incorporates edge weights in its calculations.

We observe that the adjusted flows in the summer of 2020 and 2021 have a higher values relative to the spring of 2016 and 2017. While the number of taxi zones that changed their communities is almost the same in both cases (2016-2017 Spring = 190, and 2020-2021 Summer = 193), the adjusted flow levels differed greatly. For instance, the intra-community and paired community adjusted flows of the zone taxi #222 for the spring of 2016 are 0.003, and 0.141, for the spring of 2017 are 0.017, and 0, for the summer of 2020 are 0.066, and 0.26, and for the summer of 2021 are 0.043, and 0.18 respectively. Since the Jaccard index and mutual information only consider the difference in community memberships and not the flows, they resulted in low values of similarity in both scenarios.



**Figure 5-7 Intra community adjusted flows and adjusted flows to the paired community for taxi zone number 222 for the spring of 2016 and 2017, and the summer of 2020 and 2021. This zone is chosen since it changed its community in all the chosen periods. Intra-community adjusted flows are presented with blue and adjusted flows to the paired community are presented with orange. The width of the flows is based on their intensity; the thicker, the higher rate of adjusted flows.**

## 5.7 Discussion and conclusion

In the simulated examples, we changed the fraction of inter-community edges ( $\mu$ ) of the benchmarks from 0.1 to 0.95. The Jaccard index and mutual information methods

exhibited low similarity after a certain point ( $\mu > 0.5$ ) (Lancichinetti & Fortunato, 2009), while NCSSI had a more consistent decrease across the range of the fraction of inter-community edges (from 0.1 to 0.95). Therefore, NCSSI seems to be more closely associated with the fraction of inter-community edges than both the Jaccard index and mutual information methods. This observation underscores the intrinsic sensitivity of these methods to changes in network configurations. While the Jaccard index and mutual information methods exhibited diminishing sensitivity when  $\mu > 0.5$ , NCSSI consistently responded across the full  $\mu$  range, capturing variations in edge structures. This behavior highlights NCSSI's relationship with  $\mu$ , distinguishing it from the Jaccard index and mutual information methods.

In another experiment, we intentionally changed the labels of several nodes and their associated weights to create an intentional difference in the community sets. We observed that Jaccard index and mutual information could capture the differences created by mislabeled nodes, but changing weights did not affect their similarity values as they are primarily sensitive only to changes in the node labels rather than changes in the weights. In contrast, NCSSI demonstrated a distinct advantage in capturing the difference when the weights of intervened nodes change. NCSSI effectively considered both the label and weight information in the communities, enabling it to capture the impact of changes in weights in measuring the similarity between community sets. While there has been considerable focus on integrating weights into community detection methods (Blondel et al., 2008; Meo et al., 2011; Steinhäuser & Chawla, 2008), little attention has been given to incorporating edge weights into similarity measurement methods for community sets. This experiment highlights the distinct advantage of NCSSI over mutual information and the Jaccard index methods, where both label and weight variations are related factors in measuring similarity between community sets.

We employed NCSSI on the New York Yellow Taxi flows dataset as a real-life case study. We computed similarity scores for each season between each pair of consecutive years. Different factors such as technological advancements (Surry & Baker III, 2016),

economic development (Summers & Branch, 1984), political and environmental factors (Duerden, 2004; Fischer et al., 2013; Schoennagel et al., 2017) can influence human communities. The pace of these changes is insufficient (Marsden et al., 2020) to expect major variations in human community structures within two consecutive years. However, we observed a relatively high rate of change in the Jaccard index and mutual information results. On the other hand, NCSSI demonstrated a more stable similarity degree between the communities, except after the outbreak of COVID-19 in early spring 2020.

It is common that entities in complex systems change (established or annihilated) (Negre et al., 2011). For example, people might die or be born in social networks, or through policy changes, the number of census tracts might differ over time. While in most cases, community detection is conducted for the same set of nodes (Fortunato, 2010a), there are cases where the number of nodes changes (e.g., social networks, friendship, or professional networks). The NCSSI can handle this without requiring any modifications. To elaborate, when a node is added or removed from one community set to another, the corresponding edge weights connected to that node will also change (added or removed). As NCSSI effectively captures the transformation of community sets by pairing similar communities and calculating the edit costs associated with node membership and their edge weight changes, node additions or removals and consequently edge weights additions and removals will directly impact the NCSSI result. Hence, NCSSI is well-equipped to handle cases where the networks being compared have different numbers of nodes.

In cases when the NCSSI yields high similarity scores for two sets of communities, it is important to recognize that the underlying community set topologies could still differ. It is crucial to emphasize that the NCSSI does not assess the overall similarity of network (graph) structures. Instead, its focus is specifically on quantifying the similarity between the structures of the community sets themselves.

In conclusion, NCSSI offers a new approach to better measure and assess community similarities. By considering both node labels and edge weight variations in communities,

NCSSI enhances the accuracy of the similarity measurement between community sets. It also allows measuring similarity between communities with different number of nodes. This capability is particularly valuable where the community sets are of varying sizes and compositions. It is worth noting that NCSSI incorporates the principles of edge importance, weight awareness, and edge submodularity, introduced in (Koutra et al., 2016) primarily for graph similarity measurements, since it effectively captures the impact of edge changes only on the edges associated with nodes that change their communities, acknowledges the significance of edge weights in weighted graphs, and accounts for the varying importance of edge changes in dense and sparse networks.

## References

- Barabasi, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101–113.
- Barnes, E. R. (1982). An algorithm for partitioning the nodes of a graph. *SIAM Journal on Algebraic Discrete Methods*, 3(4), 541–550.
- Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., & Werneck, R. F. (2016). Route planning in transportation networks. *Algorithm Engineering: Selected Results and Surveys*, 19–80.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.
- Condon, A., & Karp, R. M. (2001). Algorithms for graph partitioning on the planted partition model. *Random Structures & Algorithms*, 18(2), 116–140.

- Cordasco, G., & Gargano, L. (2010). Community detection via semi-synchronous label propagation algorithms. *2010 IEEE International Workshop on: Business Applications of Social Network Analysis (BASNA)*, 1–8.
- Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008.
- Donetti, L., & Munoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(10), P10012.
- Dongen, S. (2000). *Performance criteria for graph clustering and Markov cluster experiments*. CWI (Centre for Mathematics and Computer Science).
- Duan, D., Li, Y., Jin, Y., & Lu, Z. (2009). Community mining on dynamic weighted directed graphs. *Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management*, 11–18.
- Duerden, F. (2004). Translating climate change impacts at the community level. *Arctic*, 204–212.
- Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1), 17–60.
- Fan, Y., Li, M., Zhang, P., Wu, J., & Di, Z. (2007). Accuracy and precision of methods for community identification in weighted networks. *Physica A: Statistical Mechanics and Its Applications*, 377(1), 363–372.
- Fischer, A. P., Paveglio, T., Carroll, M., Murphy, D., & Brenkert-Smith, H. (2013). Assessing social vulnerability to climate change in human communities near public forests and grasslands: A framework for resource managers and planners. *Journal of Forestry*, 111(5), 357–365.



- Fortunato, S. (2010a). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174.
- Fortunato, S. (2010b). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821–7826.
- Gusfield, D. (2002). Partition-distance: A problem and class of perfect graphs arising in clustering. *Information Processing Letters*, 82(3), 159–164.
- Gustafsson, M., Hörnquist, M., & Lombardi, A. (2006). Comparison and validation of community structures in complex networks. *Physica A: Statistical Mechanics and Its Applications*, 367, 559–576.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
- Hu, Y., Chen, H., Zhang, P., Li, M., Di, Z., & Fan, Y. (2008). Comparative definition of community and corresponding identifying algorithm. *Physical Review E*, 78(2), 026121.
- Huang, J., Sun, H., Han, J., Deng, H., Sun, Y., & Liu, Y. (2010). Shrink: a structural clustering algorithm for detecting hierarchical communities in networks. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 219–228.
- Jin, D., Yu, Z., Jiao, P., Pan, S., He, D., Wu, J., Yu, P., & Zhang, W. (2021). A survey of community detection approaches: From statistical modeling to deep learning. *IEEE Transactions on Knowledge and Data Engineering*.
- Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105(31), 10687–10692.
- Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 49(2), 291–307.

- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *ArXiv Preprint ArXiv:1609.02907*.
- Koutra, D., Shah, N., Vogelstein, J. T., Gallagher, B., & Faloutsos, C. (2016). Deltacon: Principled massive-graph similarity function with attribution. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(3), 1–43.
- Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical Review E*, 80(5), 056117.
- Lancichinetti, A., Fortunato, S., & Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4), 046110.
- MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
- Marsden, P. V, Smith, T. W., & Hout, M. (2020). Tracking US social change over a half-century: The general social survey at fifty. *Annual Review of Sociology*, 46, 109–134.
- Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5), 873–895.
- Meilă, M., & Heckerman, D. (2001). An experimental comparison of model-based clustering methods. *Machine Learning*, 42(1), 9–29.
- Meo, P. De, Ferrara, E., Fiumara, G., & Proveti, A. (2011). Generalized Louvain method for community detection in large networks. *2011 11th International Conference on Intelligent Systems Design and Applications*, 88–93. <https://doi.org/10.1109/ISDA.2011.6121636>
- Mirkin, B. (1996). *Mathematical classification and clustering* (Vol. 11). Springer Science & Business Media.

- Negre, E., Missaoui, R., & Vaillancourt, J. (2011). Predicting a social network structure once a node is deleted. *2011 International Conference on Advances in Social Networks Analysis and Mining*, 297–304.
- New York Taxi & Limousine Commission. (2023). *TLC Trip Record Data*. New York Taxi & Limousine Commission. <https://www.nyc.gov/site/tlc/about/data.page>
- Newman, M. E. J. (2004a). Analysis of weighted networks. *Physical Review E*, 70(5), 056131.
- Newman, M. E. J. (2004b). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 066133.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.
- Newman, M. E. J., & Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23), 9564–9569.
- OpenStreetMap Wiki. (2022, August 13). *Basemap*. <https://wiki.openstreetmap.org/w/index.php?title=Basemap&oldid=2371479>
- Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814–818.
- Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), 036106.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846–850.
- Saoub, K. R. (2021). *Graph theory: an introduction to proofs, algorithms, and applications*. CRC Press.

- Schoennagel, T., Balch, J. K., Brenkert-Smith, H., Dennison, P. E., Harvey, B. J., Krawchuk, M. A., Mietkiewicz, N., Morgan, P., Moritz, M. A., & Rasker, R. (2017). Adapt to more wildfire in western North American forests as climate changes. *Proceedings of the National Academy of Sciences*, *114*(18), 4582–4590.
- Steinhaeuser, K., & Chawla, N. V. (2008). Community detection in a large real-world social network. In *Social computing, behavioral modeling, and prediction* (pp. 168–175). Springer.
- Summers, G. F., & Branch, K. (1984). Economic development and community social change. *Annual Review of Sociology*, *10*(1), 141–166.
- Surry, D. W., & Baker III, F. W. (2016). The co-dependent relationship of technology and communities. *British Journal of Educational Technology*, *47*(1), 13–28.
- Taya, F., de Souza, J., Thakor, N. V., & Bezerianos, A. (2016). Comparison method for community detection on brain networks from neuroimaging data. *Applied Network Science*, *1*(1), 1–20.
- Traud, A. L., Kelsic, E. D., Mucha, P. J., & Porter, M. A. (2008). Community structure in online collegiate social networks. *ArXiv Preprint ArXiv:0809.0960*.
- Wallace, D. L. (1983). A method for comparing two hierarchical clusterings: comment. *Journal of the American Statistical Association*, *78*(383), 569–576.
- Yang, T., Chi, Y., Zhu, S., Gong, Y., & Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Machine Learning*, *82*, 157–189.
- Zhang, P., Li, M., Wu, J., Di, Z., & Fan, Y. (2006). The analysis and dissimilarity comparison of community structure. *Physica A: Statistical Mechanics and Its Applications*, *367*, 577–585.

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3), 1–55.

## Chapter 6

### 6 Conclusion

#### 6.1 Summary

This dissertation presents four studies on human mobility, addressing critical aspects of how environmental factors, geographical context, and community structures influence urban travel behaviors. Presented as standalone peer-reviewed manuscripts, this body of work sheds light on the complex relationships between human mobility patterns and their broader contexts. The research is grounded in a multidisciplinary approach, merging insights from urban planning and public health with advanced spatial and statistical analyses. Collectively, these studies contribute to our understanding of human mobility, offering perspectives beyond traditional mobility analyses.

In chapter 2, I examine the impact of environmental factors on urban stress and happiness during daily travel. I employed Geographic Ecological Momentary Assessment (GEMA) to collect data on from a large sample of workers from cities in the United Kingdom: Birmingham, Leeds, and Brighton and Hove. This approach allowed us to track individuals' experiences of stress and happiness in direct response to their environment. The study focused on key environmental variables such as the presence of green and blue spaces, daylight, visibility, and varying weather conditions, exploring how these factors influence individuals' experiences of stress and happiness during urban travel.

Results from this study indicate a significant relationship between environmental factors and variations in urban stress and happiness. Exposure to green and blue spaces during daily travels was identified as a key variable, showing a positive association with heightened happiness and diminished stress levels. This finding underscores the importance of natural spaces in urban environments for enhancing emotional well-being. These insights are particularly relevant for urban planning and public health policy, suggesting that thoughtful design and structuring of urban environments can significantly

influence daily experiences of stress and happiness among city residents. The application of GEMA in this context provided a detailed understanding of the dynamic relationship between environmental factors and urban well-being.

In chapter 3, I investigate the complex relationship between socio-economic factors and human mobility, emphasizing the significant role of geographical context. I introduce the concept of *Mobility Deviation Index* (MDI) to quantify the relative level of observed mobility in comparison to expected mobility for a given location, with expected mobility accounting for nearby amenities. This method represents a significant advancement in the study of human mobility, as it enables a more detailed understanding of mobility behaviors by incorporating context directly into measurement. I use a large network mobility dataset, aggregating mobile device connections to cell tower receivers across Ontario, Canada. This data provides a comprehensive view of mobility patterns, offering insights into how different geographical contexts influence human movement.

The findings from this research demonstrate that considering geographical context leads to significant shifts in our interpretation of human mobility and its association with socio-economic factors. In our statistical analysis, I found that when geographical context is accounted for, the associations between mobility levels and neighborhood-level socio-economic variables, such as income, age, and minority status, differed from when using traditional methods. This highlights the importance of geographical context in interpreting human mobility data, challenging existing methodologies that often overlook this critical factor. The concept of Mobility Deviation Index provides a more accurate and context-sensitive understanding of human movement, especially in large landscapes, making it a valuable tool for planners and policymakers.

In chapter 4, I introduce the *Local Mobility Index* (LMI), a novel metric designed to quantify the extent of localized mobility behavior. This index is innovative in its approach, as it not only considers the geometry of individual travel patterns but also integrates the availability and accessibility of amenities in the surrounding area. I use GPS data from a large sample of workers across three UK cities, along with

OpenStreetMap POI data, to calculate the LMI. This metric is calculated by assessing the proximity of individuals' chosen destinations to the nearest available amenities, with a higher LMI score indicating a greater tendency towards localized travel. The study's analytical approach highlights the variability in local mobility behavior, shedding light on how individuals' choices of destinations are influenced by the distribution of amenities in their environment. I demonstrate that the LMI provides insights into local mobility patterns, capturing aspects of travel behavior that are not addressed by conventional mobility metrics. The study's findings have important implications for urban planning and policy, as they highlight the need to consider the geographical distribution of amenities when assessing and facilitating local mobility patterns.

In chapter 5, I introduce a novel approach to assess similarities between community structures called the *Network Community Structure Similarity Index* (NCSSI). I specifically address the limitations of current methods that overlook the effects of edge weights and are unable to evaluate the similarities between networks with varying numbers of nodes. The NCSSI is formulated based on the edit distance concept, where the similarity between two community sets is quantified based on the total cost to transform one community set into another. This cost function incorporates both community labels and edge weights, allowing for a better comparison of network community structures. Analyzing community structure over time is crucial for understanding the dynamic evolution of urban communities derived from human mobility patterns. This knowledge enhances our quantitative understanding of urban development, facilitating improvements in commercial activities, public safety, social interactions, and the overall creation of sustainable, vibrant living spaces (Wang et al., 2018).

The study demonstrates the application of NCSSI through simulated data and a case study analysis of New York Yellow Taxi flows, comparing its performance with other commonly used methods like mutual information and the Jaccard index. The results from applying NCSSI show its effectiveness in capturing the impact of label and edge weight changes on community structures, which are aspects not addressed by existing methods.



In the case study with New York Yellow Taxi flows, I used NCSSI to compare community structures derived from taxi trip data over different years. This application highlighted how NCSSI could discern changes in community structures over time, including the significant shifts caused by external factors such as the COVID-19 pandemic. Unlike other methods, NCSSI's sensitivity to both label and weight changes in networks allows for a more comprehensive understanding of community dynamics. These findings position NCSSI as a valuable tool for analyzing complex networks, offering new perspectives in understanding how community structures evolve.

## 6.2 Contributions

### 6.2.1 Empirical findings

In chapter 2, I conducted an empirical study examining the impact of environmental factors on happiness and stress levels during daily travel. This study filled a gap in the literature by leveraging granular GPS tracking data alongside surveys of happiness and stress at the end of each trip, using a large dataset. This approach was motivated by the significant portion of life spent traveling and the potential for enhancing mental well-being through improved travel environments. The findings revealed that when environmental factors are considered, other factors previously thought to significantly influence stress and happiness become less important. This underscores the necessity for detailed, longitudinal studies to minimize recall bias and accurately assess travel routes and their environments, highlighting a shift away from traditional cross-sectional studies.

In the case studies for both MDI and LMI (chapters 3 and 4), integrating context into mobility measurements revealed contrasting results to those obtained using previous metrics, especially for individuals in suburban and rural areas. Contrary to earlier methods which suggested higher mobility levels in less urbanized areas, incorporating the availability of local amenities revealed a reverse trend. Specifically, rural area residents displayed lower MDI levels compared to their urban counterparts. This finding underscores the significance of including contextual factors in mobility analysis, as discarding them can lead to fundamentally different interpretations and outcomes in

social science research where we are interested to learn the underlying derives of mobility patterns.

In the NCSSI (chapter 5) case study using New York taxi data, we noted not just changes in mobility flows and volumes, as observed in other studies (Dueñas et al., 2021), but also alterations in community structures due to COVID-19. While previous methodologies reported consistent dissimilarities in community structures in the years leading up to the COVID-19 pandemic, the NCSSI specifically highlighted significant dissimilarities only in the year the pandemic struck. This indicates NCSSI's exceptional sensitivity to significant events capable of altering community dynamics. The dissimilarity in community structures found by the NCSSI underscores the profound impact of a pandemic on both mobility patterns and the structural composition of communities, highlighting a deeper level of change than previously understood.

### 6.2.2 Conceptual and methodological contributions

In chapters 3 and 4, I introduced two concepts – MDI and LMI – addressing the gap in integrating context into the direct measurement of human mobility patterns. The MDI merges contextual and mobility data, establishing a baseline for expected mobility against which observed mobility can be compared. Previously, there was no established concept for measuring deviation from expected mobility. Researchers often compared the observed mobility of individuals or areas in aggregated forms against a baseline, typically defined as the mean or median over a certain period (Long & Ren, 2022). The MDI represents an advancement by offering a refined analysis that captures deviations from expected mobility more effectively than these earlier approaches. Additionally, I formulated a methodological framework to compute the expected mobility, simulating mobility behaviors using the radius of gyration concept. This framework represents a methodological advancement within a specific case study context. Nonetheless, it is crucial to distinguish between the methodological and conceptual contributions of this study. The MDI offers a broad conceptual insight, enhancing our understanding of human

mobility through contextual analysis; however, the methodological contribution is specific, applying this concept in a detailed case study.

Furthermore, I introduced the concept of Local Mobility Behavior, enhancing the conceptual framework by integrating context into the understanding of local behavior. "Local" is inherently relative, prompting the question: local relative to what? Previous research focused on comparing mobility levels among individuals (Manaugh & El-Geneidy, 2012), addressing "local relative to whom." My approach shifts the emphasis to "local relative to where," evaluating activity choices against the nearest available options. This shift offers insights into the reasons behind an individual's non-local behavior, whether due to infrastructure, environmental design, or personal characteristics. Additionally, I enhanced the methodological framework by extracting activity choices from tracking data and comparing these to the nearest alternatives.

In chapter 5, my contribution focused on advancing methodology to address the longstanding issue of comparing sets of communities. By considering all features of network structure, rather than limiting focus to member labels as done in previous methods (Danon et al., 2005; Meilă, 2007), the NCSSI proved highly valuable. This approach, using the concept of edit distance tailored for network analysis, allowed for more precise measurement of community similarities than previous methods. This methodological innovation enhances our ability to compare community networks effectively.

### 6.2.3 Open-source code and reproducibility

Throughout my dissertation, I have prioritized open-source code and reproducibility. Although most datasets used were private and sensitive, I made the codes for chapters 2, 3, and 4 publicly accessible. The code for the MDI is publicly available (<https://doi.org/10.6084/m9.figshare.20405808.v2>). It requires a dataset as input to optimize its parameters, thereby calculating the MDI level for each area. This functionality allows for a detailed analysis of mobility behavior by area, leveraging

dataset information to enhance the precision of MDI calculations. The LMI code is also available (<https://github.com/miladmzdh/LMI>) as a standalone function that computes LMI levels from individual trajectories, automatically incorporating contextual data, downloading OSM POI dataset. Additionally, I intend to contribute the NCSSI function (<https://doi.org/10.6084/m9.figshare.24021324.v1>) to the NetworkX library (Hagberg et al., 2008), widely used for network analysis. This commitment ensures that the research advancements made in my dissertation are accessible to all.

### 6.3 Future research

Building on my empirical, conceptual, and methodological work, I aim to further explore the impact of urban environments on well-being, altering my analytical perspective and the types of data used for environmental factors. Unlike the top-down approach common in greenery studies (Torkko et al., 2023), I plan to assess how individuals experience environments on the ground, considering a combination of high-resolution greenery data from not only satellite imagery (Li, Zhang, Li, Kuzovkina, et al., 2015; Yang et al., 2009), but also Google Street View (Li, Zhang, Li, Ricard, et al., 2015), and LiDAR (Klingberg et al., 2017). Additionally, while traditionally studies have relied on subjective measures to understand the environmental impact on well-being, reflecting our direct perception, they often overlooks unconscious effects on our health. By incorporating longitudinal objective measures (Pykett et al., 2020; Wilhelm & Grossman, 2010), I can bridge this gap, providing a clearer, more comprehensive understanding of environmental influences on well-being beyond our immediate awareness.

Incorporating feature detection and AI models (Khosla et al., 2014) into my research opens a new avenue for understanding the effect of urban environment's features on individuals' well-being. Manually evaluating large areas and their features is nearly impossible, necessitating the automatic extraction of features through feature detection models. I can then assess the relationship between people's subjective perceptions, alongside the objective responses to the environment with these automatic extracted features, using statistical analysis.

Furthermore, exploring the urban environment's impact on individuals with mental and physical disorders offers a significant research avenue. For instance, individuals with autism spectrum disorder (ASD) may perceive urban environments uniquely, necessitating further investigation to foster inclusive and accessible city designs. This aspect is particularly crucial in public transit, which is a predominant mode of transportation for people with ASD (Falkmer et al., 2015). Understanding their specific needs and perceptions can guide the development of more inclusive urban planning and public transportation systems.

Expanding on the Local Mobility Index concept, I plan to apply this framework in environmental studies, specifically to investigate the carbon footprint associated with non-local movement behavior. While urban design often receives attention for its role in carbon emissions (Bartholomew, 2009; Kinigadner et al., 2020; Sun & Leng, 2021), the contribution of city residents to their carbon footprint through non-local mobility behavior remains understudied. This avenue of research will explore the extent of individuals' responsibility for their carbon footprint versus the influence of city design in prompting higher-than-expected emissions, bridging a gap in understanding the interplay between local mobility behavior and environmental impact.

I plan to advance the MDI methodology by incorporating metrics based on other mobility measures, including time-based (Demissie et al., 2019) and entropy-based (Song et al., 2010) metrics. This expansion will enable the MDI concept to be applicable across a broader range of studies with diverse measurement methods. Additionally, I aim to enhance the NCSSI methodology by integrating the hierarchical structure of networks into its measurement, offering a better understanding of community structures similarities and differences, particularly in urban mobility networks where hierarchical structures are common (Louail et al., 2015; Murali et al., 2016; Yildirimoglu & Kim, 2018).

## References

- Bartholomew, K. (2009). Cities and Accessibility: The Potential for Carbon Reduction and the Need for National Leadership. *Fordham Urb. LJ*, 36, 159.
- Danon, L., Diaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09), P09008.
- Demissie, M. G., Phithakkitnukoon, S., Kattan, L., & Farhan, A. (2019). Understanding human mobility patterns in a developing country using mobile phone data. *Data Science Journal*, 18, 1.
- Dueñas, M., Campi, M., & Olmos, L. E. (2021). Changes in mobility and socioeconomic conditions during the COVID-19 outbreak. *Humanities and Social Sciences Communications*, 8(1).
- Falkmer, M., Barnett, T., Horlin, C., Falkmer, O., Siljehav, J., Fristedt, S., Lee, H. C., Chee, D. Y., Wretstrand, A., & Falkmer, T. (2015). Viewpoints of adults with and without Autism Spectrum Disorders on public transport. *Transportation Research Part A: Policy and Practice*, 80, 163–183.
- Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- Khosla, A., An An, B., Lim, J. J., & Torralba, A. (2014). Looking beyond the visible scene. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3710–3717.
- Kinigadner, J., Büttner, B., Wulfhorst, G., & Vale, D. (2020). Planning for low carbon mobility: Impacts of transport interventions and location on carbon-based accessibility. *Journal of Transport Geography*, 87, 102797.

- Klingberg, J., Konarska, J., Lindberg, F., Johansson, L., & Thorsson, S. (2017). Mapping leaf area of urban greenery using aerial LiDAR and ground-based measurements in Gothenburg, Sweden. *Urban Forestry & Urban Greening*, *26*, 31–40.
- Li, X., Zhang, C., Li, W., Kuzovkina, Y. A., & Weiner, D. (2015). Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. *Urban Forestry & Urban Greening*, *14*(4), 751–759.
- Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, *14*(3), 675–685.
- Long, J. A., & Ren, C. (2022). Associations between mobility and socio-economic indicators vary across the timeline of the Covid-19 pandemic. *Computers, Environment and Urban Systems*, *91*, 101710.
- Louail, T., Lenormand, M., Picornell, M., Garcia Cantu, O., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature Communications*, *6*(1), 6007.
- Manaugh, K., & El-Geneidy, A. M. (2012). What makes travel 'local'? Defining and understanding local travel behavior. *Journal of Transport and Land Use*, *5*(3), 15–27.
- Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, *98*(5), 873–895.
- Murali, P., Ordóñez, F., & Dessouky, M. M. (2016). Modeling strategies for effectively routing freight trains through complex networks. *Transportation Research Part C: Emerging Technologies*, *70*, 197–213.

- Pykett, J., Osborne, T., & Resch, B. (2020). From Urban Stress to Neurourbanism: How Should We Research City Well-Being? *Annals of the American Association of Geographers*, *110*(6), 1936–1951. <https://doi.org/10.1080/24694452.2020.1736982>
- Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021.
- Sun, L., & Leng, J. (2021). Research on influencing factors of travel in underground space based on multi-source data: Spatial optimization design for low-carbon travel. *Energy and Buildings*, *253*, 111524.
- Torkko, J., Poom, A., Willberg, E., & Toivonen, T. (2023). How to best map greenery from a human perspective? Comparing computational measurements with human perception. *Frontiers in Sustainable Cities*, *5*, 1160995.
- Wang, P., Fu, Y., Zhang, J., Li, X., & Lin, D. (2018). Learning urban community structures: A collective embedding perspective with periodic spatial-temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *9*(6), 1–28.
- Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, *84*(3), 552–569.
- Yang, J., Zhao, L., McBride, J., & Gong, P. (2009). Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, *91*(2), 97–104.
- Yildirimoglu, M., & Kim, J. (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic. *Transportation Research Part C: Emerging Technologies*, *89*, 254–267.



## Appendices

### Appendix A: Results of linear mixed-effect regression models of happiness and stress level

**Table A.1 Results of linear mixed-effect regression models of happiness and stress level with only control variables**

Variables	Happiness		Stress	
	B	p	B	p
(Intercept)	0.21	<b>0.008</b>	-0.12	0.169
<b>Control Variables</b>				
<b>Individual Factors</b>				
Gender [RC: Female] Male	0.04	0.161	-0.00	0.921
<b>Age [RC: 18-24]</b>				
25-34	0.07	0.191	-0.11	0.055
35-44	0.06	0.269	-0.13	<b>0.031</b>
45-54	0.02	0.679	-0.09	0.128
55-64	0.01	0.835	-0.06	0.385
<b>Hometown [RC: Leeds]</b>				
Brighton and Hove	-0.02	0.526	0.05	0.163
Birmingham	0.01	0.849	0.07	0.164
Distance from Home to City Center	0.00	0.893	0.01	<b>0.016</b>
<b>Trip Attributes</b>				
<b>Travel Modes</b>				
Walk	0.01	0.801	-0.04	0.425
Run	0.03	0.819	-0.14	0.295
Bike	-0.05	0.499	0.00	0.967
Bus	-0.14	0.067	0.08	0.343
Train	-0.26	0.361	0.16	0.609
Car	-0.08	0.104	0.01	0.915
Duration	-0.00	0.435	0.00	0.995
Daylight	-0.27	<b>&lt;0.001</b>	0.23	<b>&lt;0.001</b>
<b>Season [RC: Winter]</b>				
Spring	0.06	0.100	-0.07	0.098
Summer	0.26	0.656	-0.00	0.994
Autumn	-0.06	<b>0.031</b>	-0.03	0.378
<b>Destination Attributes</b>				
<b>Destination Type [RC: Home]</b>				
Work	0.05	0.198	-0.05	0.288
Other	-0.01	0.640	0.03	0.417
<b>Destination Activity [RC: Work]</b>				
Housework	-0.01	0.917	0.03	0.646
Leisure	-0.00	0.897	0.07	<b>0.045</b>
Eating	-0.33	0.132	0.18	0.465
Other	0.03	0.880	-0.11	0.570
Presence of People at the Destination [RC: Alone] Not Alone	-0.02	0.524	-0.03	0.374
Marginal R2 / Conditional R2		0.018 / 0.025		0.014 / 0.028

## Appendix B: MDI POI categorization

In our analysis, we faced the challenge of aligning the diverse POI types from the OSM dataset with the limited destination types specified in the 2016 Transportation Tomorrow historical survey data conducted in the Greater Golden Horseshoe area of Southern Ontario. The OSM dataset offers a wide range of POI categories. To make our analysis compatible with the survey data, we grouped the extensive OSM categories into broader classes that correspond to the survey's three destination types: school, shop, and other. This merging was necessary due to the differing classification schemes in the two datasets and was performed with the aim of capturing the essence of each destination type while maintaining the integrity of our analysis. Table B.1 outlines how specific OSM categories were amalgamated into each survey category, providing a structured and logical framework for our approach.

**Table B. 1 Categories of OSM POI datasets and their according category in the survey**

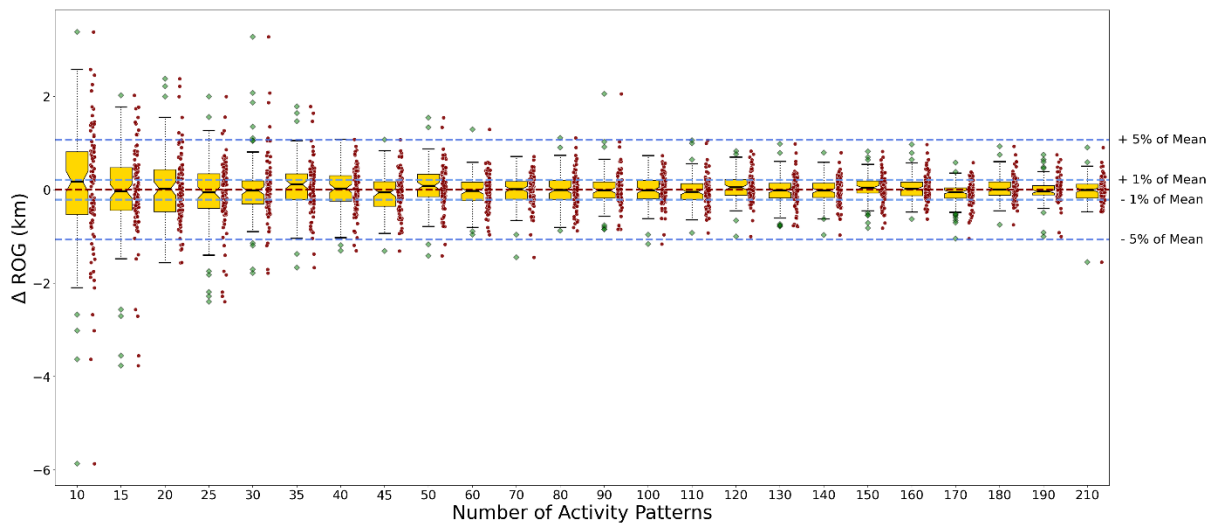
Category in the survey	Categories in OSM
school	college, library, school, university, kindergarten
shop	atm, bank, beauty_shop, bicycle_rental, bicycle_shop, bookshop, butcher, car_dealership, car_rental, car_sharing, car_wash , chemist, clothes, computer_shop, convenience, convenience, department_store, florist, furniture_shop, garden_centre, gift_shop, greengrocer, hairdresser, jeweller, laundry, mall , mobile_phone_shop, newsagent, outdoor_shop, pharmacy, shoe_shop, sports_shop, stationery, supermarket, toy_shop, vending_any, vending_machine, video_shop
other	archaeological, arts_centre, artwork, attraction, bakery, battlefield, bench, camp_site, caravan_site, chalet, cinema, community_centre, dog_park,

fountain, golf\_course, graveyard, guesthouse, hostel, hotel, hunting\_stand,  
ice\_rink, lighthouse, memorial, monument, motel, museum, nightclub,  
observation\_tower, park, picnic\_site, pitch, playground, ruins, shelter,  
swimming\_pool, theatre, theme\_park, toilet, tourist\_info, tower,  
town\_hall, travel\_agent, vending\_parking, viewpoint, water\_mill,  
water\_tower, water\_well, water\_works, wayside\_shrine, windmill, zoo,  
track, sports\_centre, biergarten, bar, beverages, cafe, drinking\_water,  
fast\_food, food\_court, pub, restaurant, hospital, post\_office, optician,  
public\_building, recycling, recycling\_clothes, recycling\_glass,  
recycling\_metal, recycling\_pape, veterinary, courthouse fire\_station,  
police, prison, wastewater\_plant, doctors, dentist airfield, airpor,  
bus\_station, ferry\_terminal , helipad, railway\_halt, railway\_station, taxi

---

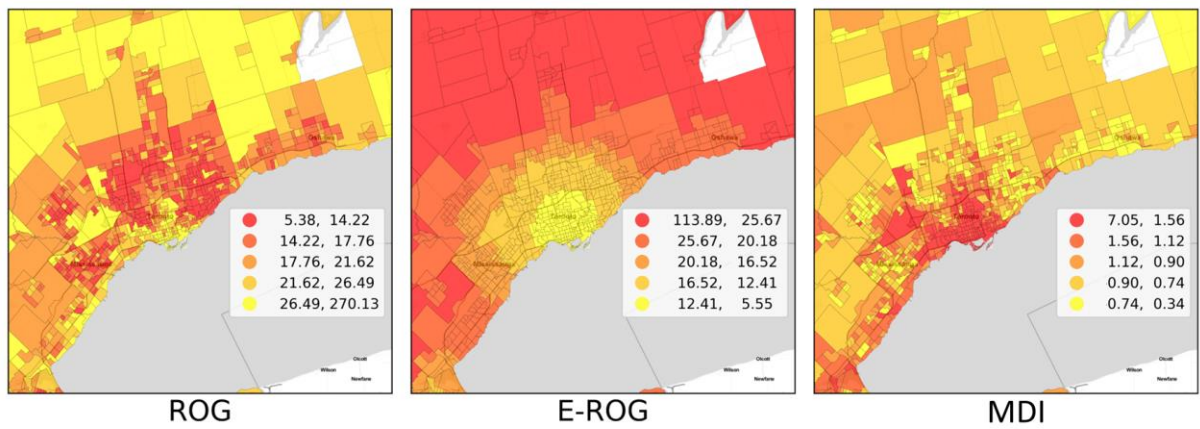
## Appendix C: MDI sensitivity analysis

We conducted a sensitivity analysis to test how the number of potential activity patterns can influence our results (Figure C-1). We subtracted each set of expected ROG from the previous one (starting from 5 potential activity patterns to 200). We found that the mean expected mobility value stabilized after approximately  $s = 50$  simulations. After 100 activity patterns, we observed that the lower and upper quartiles stayed in the range of (-1, +1) percent of the mean value of all observed ROG. Also, lower and upper whiskers are in the range of (-5, +5) percent of the mean value of all observed ROG. Therefore, we chose a conservative value of  $s = 100$  as our number of simulated activity patterns for each location



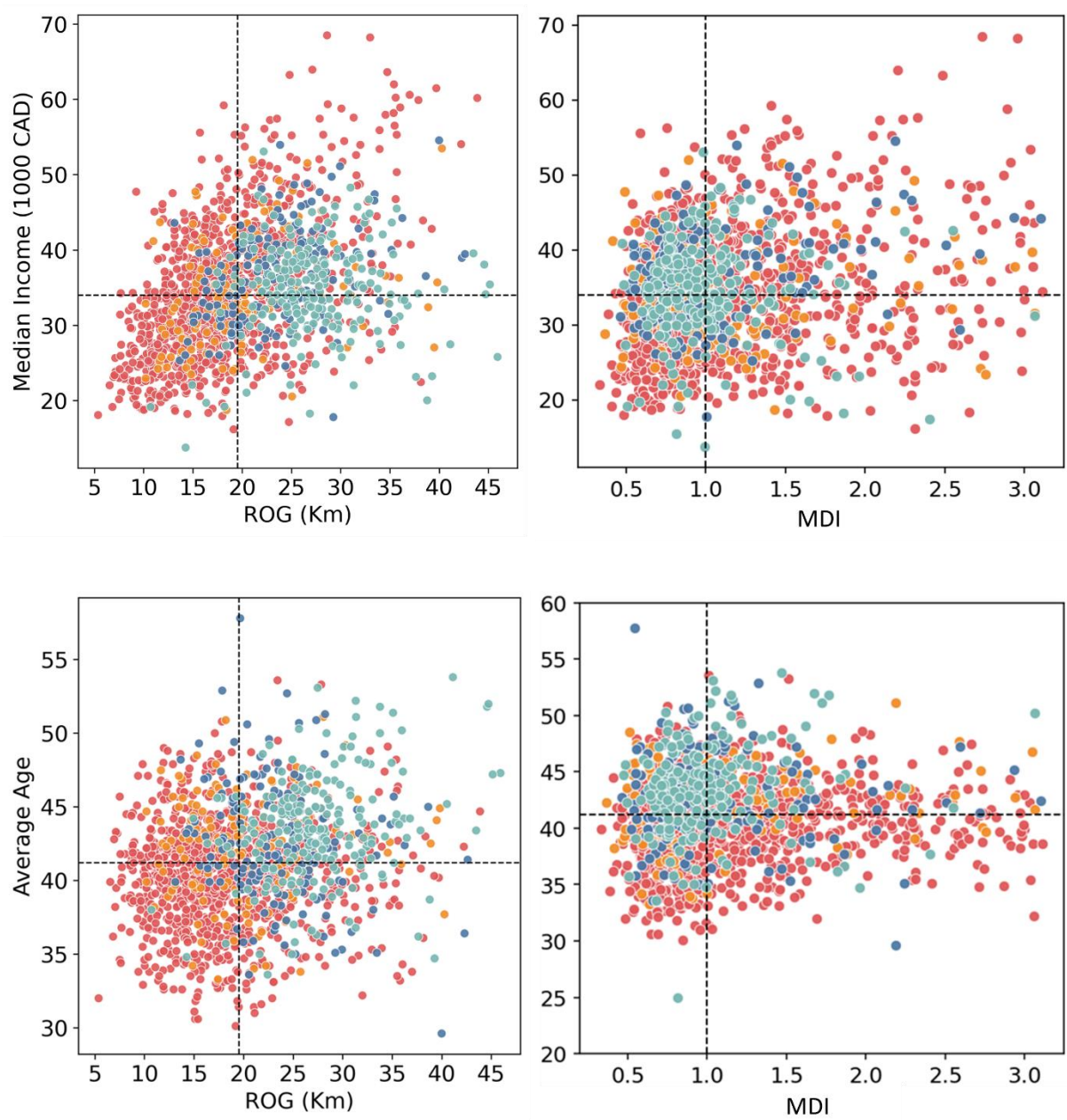
**Figure C-1 Sensitivity analysis for finding the optimum number of activity patterns. Each box plot shows the results of the subtraction of each set from the previous one. To better understand how these differences can be interpreted, four vertical lines (blue lines) are plotted to show the thresholds of -5, -1, +1, and +5% of the mean of all observed ROG values.**

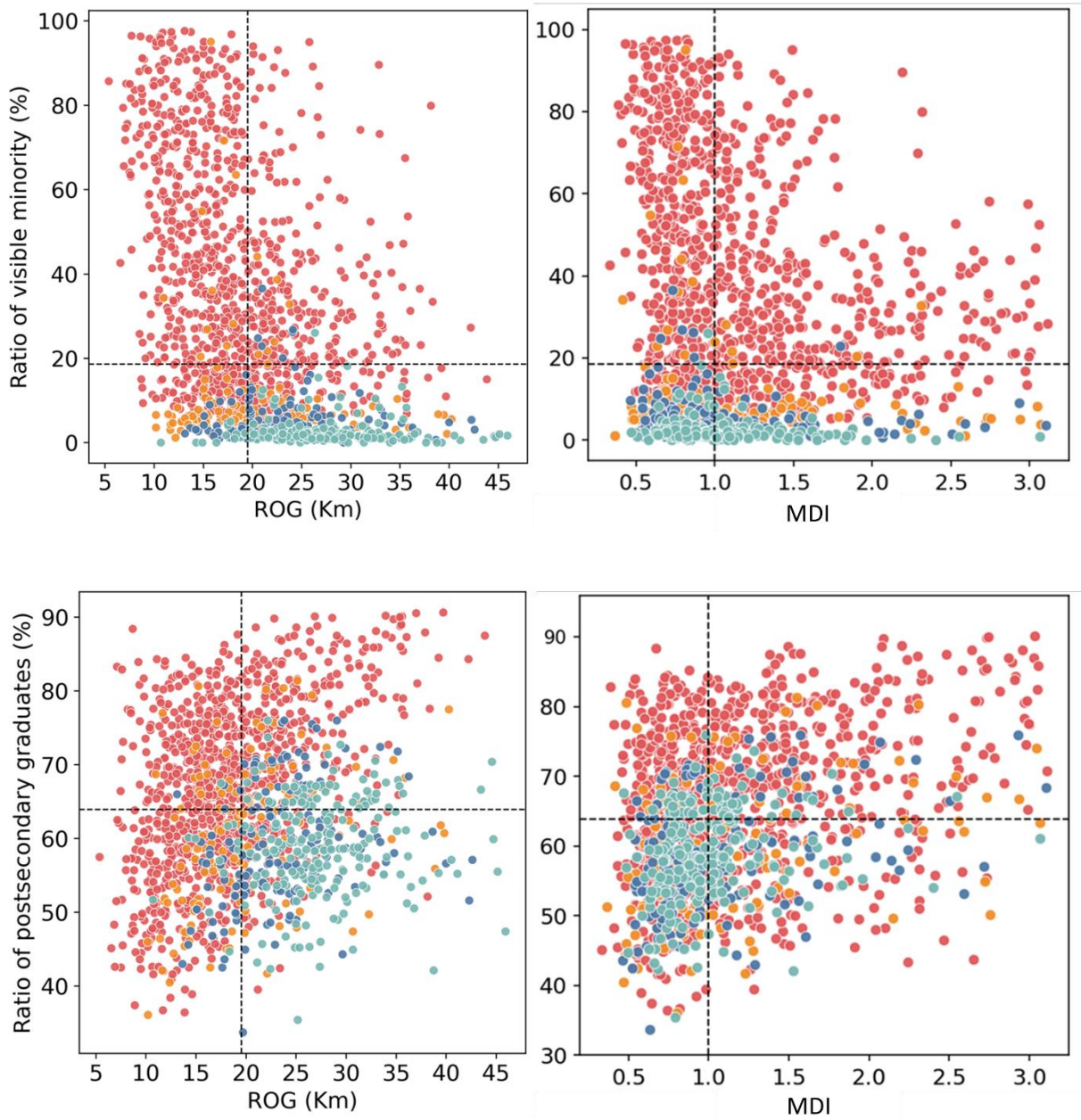
## Appendix D: ROG, EROG, and MDI maps

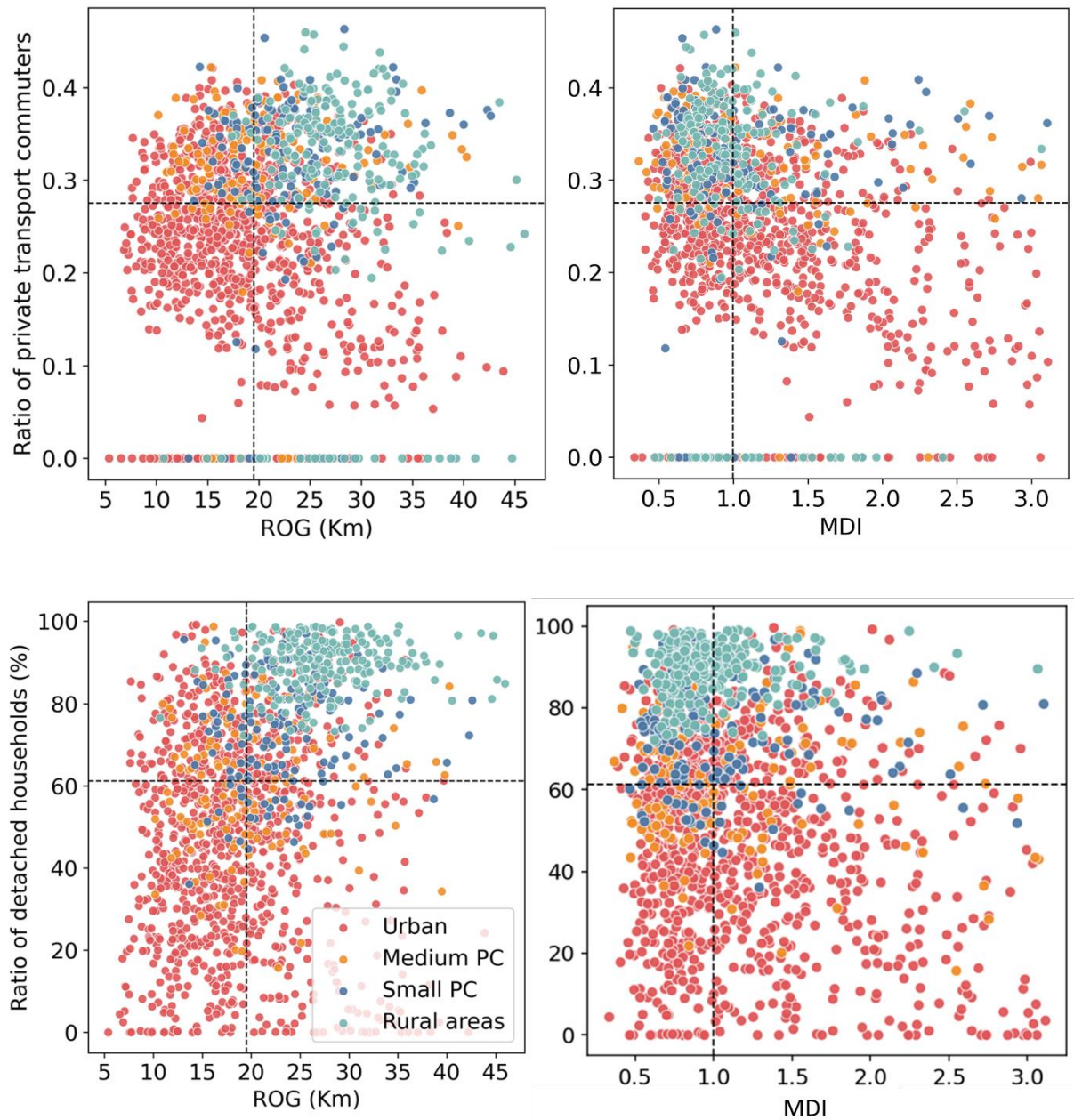


**Figure D-1 Maps illustrating the spatial distribution of ROG, EROG, MDI values across the Greater Toronto Area for each model.**

## Appendix E: Socio-economic factors-MDI scatter plot







**Figure E-1 Scatter plots of four socioeconomic factors against observed ROG values and MDI values. Different colors demonstrate the area type of each point.**



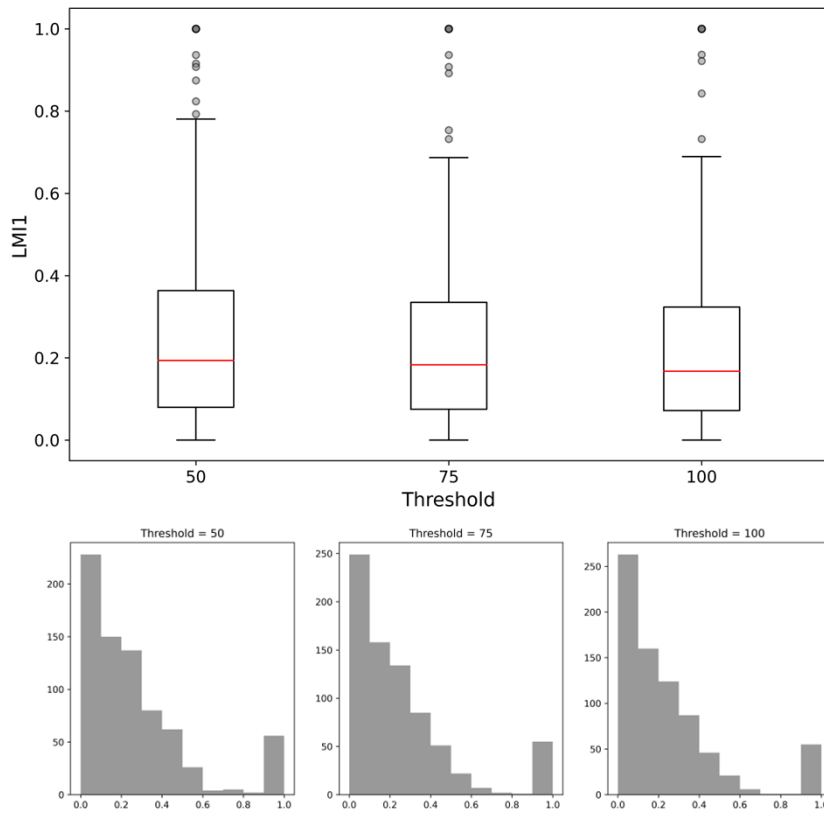
## Appendix F: LMI search radius sensitivity analysis

To determine an appropriate threshold for calculating LMI, we computed the LMI values for all individuals using varying search radius. Since calculating LMI is computationally intensive, we considered search radius of 50, 75, and 100 meters to evaluate how the results would change and identify the best option (Table F.1). Choosing a larger search radius provides more POI stop points, while smaller search radius may result in more non-POI stop points. However, larger search radius could potentially associate POIs with stop points that were not actually visited by the individual. For both LMI1 and LMI2, the mean and median values showed a slight decrease, and their standard deviation remained relatively stable (Table F.1).

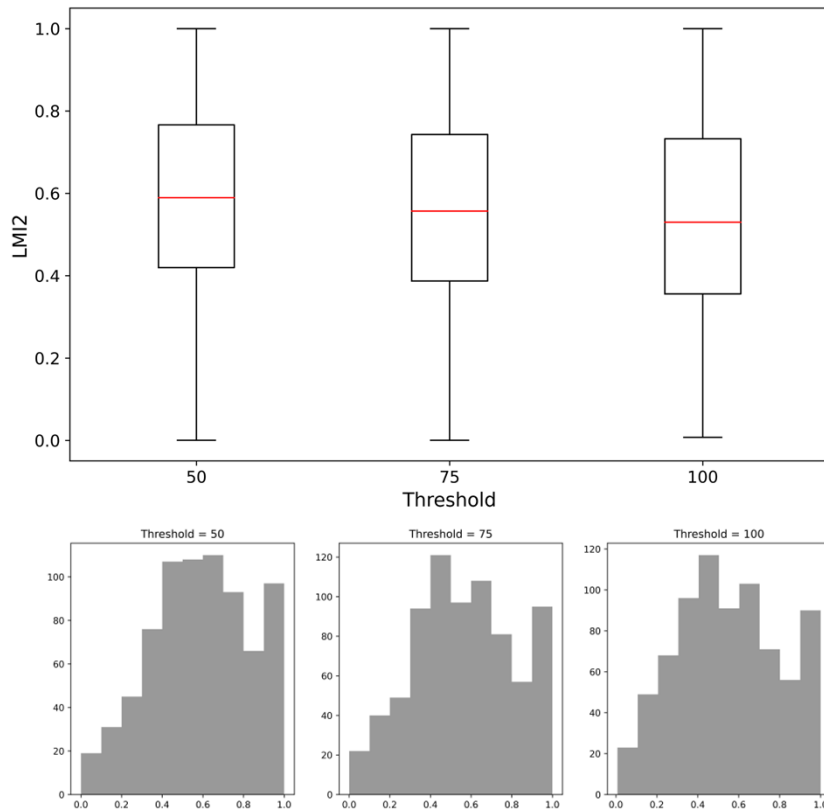
**Table F. 1 Summary Statistics of LMI Models using varying search radius**

Threshold	LMI1			LMI2		
	Mean	Median	sd	Mean	Median	sd
50	0.26	0.19	0.25	0.59	0.58	0.24
75	0.25	0.18	0.25	0.56	0.55	0.25
100	0.24	0.16	0.25	0.55	0.53	0.25

We visualized the distribution of each LMI metric at different search radius using boxplots and histograms (Figure F-1 and F-2). For both LMI1 and LMI2, we observe a modest decrease in LMI values with increasing the search radius.



**Figure F-1 Distribution of LMI1 with varying search radius via boxplots and histograms**

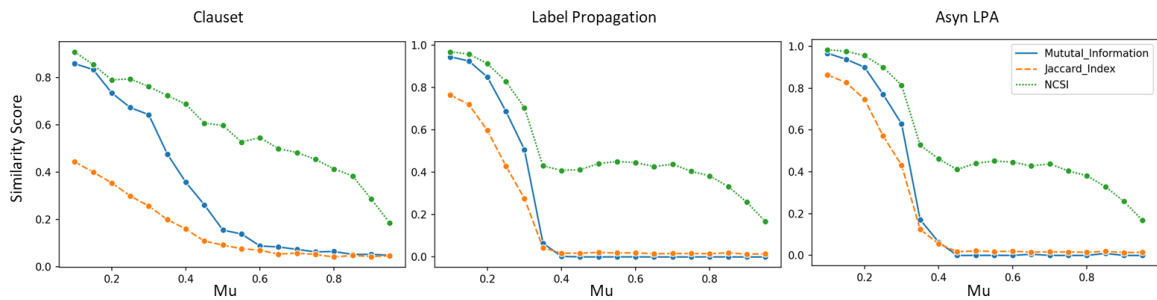


**Figure F-2 Distribution of LMI2 with varying search radius via boxplots and histograms**

Given that these two models exhibit a modest decrease with increase in search radius, we established a criterion for selecting the appropriate search radius. Considering 75-meter threshold used in the stop detection algorithm and the 75 meters threshold for cellphone GPS error used in data cleaning, we opted for a more conservative choice of 75 meters as the suitable search radius.

## Appendix G: Community detection algorithms' Similarity scores

We computed different similarity scores using different similarity methods on LFR benchmarks. Here are the results of three different community detection algorithms.



**Figure G- 1 Similarity scores of three different community detection algorithms (Clusset's algorithm, asynchronous label propagation algorithm, and semi-synchronous label propagation algorithm) on LFR benchmarks using three different similarity measurement methods.**

## Appendix H: Ethics Approval Letters

For Chapters 2 and 4, where the WorkAndHome dataset was used, the data usage adhered to the ethical guidelines and regulations set forth by the University of Southampton's Ethics and Research Governance Online (ERGO) system (Ethics number: 30058).

In Chapter 3, where the network mobility dataset provided by TELUS Communications Inc. was used, the data usage complied with the ethical guidelines and regulations established by the Western University Non-Medical Research Ethics Board (Ethics number: 115997).

Chapter 5 did not require ethics approval for data usage.

## Curriculum Vitae

**Name:** Milad Malekzadeh

**Post-secondary Education and Degrees:** University of Tehran  
Tehran, Iran  
2012-2016 B.Sc.

University of Tehran  
Tehran, Iran  
2017-2020 M.Sc.

The University of Western Ontario  
London, Ontario, Canada  
2020-2024 Ph.D.

**Honours and Awards:** Utrecht Excellence Scholarship  
2016

SustainMV scholarship  
2022

Best TA of the year, Western University  
2023

Student Achievement Award, ESRI Canada Centres of Excellence  
2023

Western University Inspiring Mind Award  
2023

**Related Work Experience** Field surveyor  
Tosee Paydar Sahar  
2014-2016

CEO and Founder  
Apply Plus  
2017-2023

Sessional Lecturer  
The University of Western Ontario  
2023

Teaching/Research Assistant  
The University of Western Ontario  
2020-2024

GIS Analyst  
The University of Western Ontario  
2023-2024

Researcher  
University of Helsinki  
2024-Present

## **Publications:**

### **Peer-reviewed journal articles**

- Malekzadeh, M., & Long, J. A. Mobility deviation index: incorporating geographical context into analysis of human mobility. *J Geogr Syst* (2024). <https://doi.org/10.1007/s10109-024-00444-1>
- **Malekzadeh, M., & Long, J. A.** (2023). A network community structure similarity index for weighted networks. *Plos one*, 18(11), e0292018. <https://doi.org/10.1371/journal.pone.0292018>
- Long, J. A., **Malekzadeh, M.**, Klar, B., & Martin, G. (2023). Do regionally targeted lockdowns alter movement to non-lockdown regions? Evidence from Ontario, Canada. *Health & Place*, 79, 102668. <https://doi.org/10.1016/j.healthplace.2021.102668>

### **Peer-reviewed conference papers**

- **Malekzadeh, M.**, Reuschke, D., & Long, J. A. (2023). How Does Travel Environment Affect Mood? A Study Using Geographic Ecological Momentary Assessment in the UK. In 12th International Conference on Geographic Information Science (GIScience 2023). Schloss-Dagstuhl-Leibniz Zentrum für Informatik.
- **Malekzadeh, M.**, & Long, J. A. (2023). Non-linear associations between human mobility and sociodemographic and contextual factors. *Spatial Knowledge & Information*.
- **Malekzadeh, M.**, Javanmard, R., & Karimipour, F. (2019). Extracting point of interests from movement data using kernel density and weighted k-means. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 717-720.

- Javanmard, R., Esmaili, R., **Malekzadeh, M.**, & Karimipour, F. (2019). Investigating the Rhythms of Human Movements in Geneva Lake Region Using Mdc Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 583-588.