University of Exeter
Department of Computer Science

Spatio-Temporal Heterogeneities in Human Mobility: A Gender Perspective

Mariana Gomes da Motta Macedo

Submitted by Mariana Gomes da Motta Macedo, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Computer Science, November, 2021.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signed: Mariana Macedo ....

Abstract

The movement of people (aka human mobility) impacts several sectors of society such as health, security, economics and politics. Studying specific patterns in mobility such as gender is a complex task because analysing a single dimension requires a rich understanding of the wide range of mechanisms and interactions that can shape the mobility of women and men. In this thesis, we focus on gender groups to broadly understand their mobility patterns, which might reflect (and influence) inequalities from cultural constructs and the labour market.

Cultural constructs impose different roles on women and men, translating into how each gender participates in the labour market. Women are usually pressured to participate in jobs related to care and household responsibilities, while men usually take jobs related to leadership and high financial return. In this way, job sectors have an imbalanced participation of each gender, which in turn, reflects in how people move in space and time.

This thesis presents the spatio-temporal patterns in the mobility of women and men for four countries: Brazil, Colombia, the United States of America and the United Kingdom. We contribute to the literature with a novel mobility pattern — mobility diversity — that estimates the spatial distribution of mobility across metropolitan areas and unveils differences in the mobility of women and men. Furthermore, we demonstrate the existence of two profiles of cities that share similar mobility patterns that can contribute to the understanding of systematic issues in mobility.

Cultural constructs, labour market, and mobility reinforce human behaviours and consequently affects inequalities feeding into the formation of new cultural constructs. To break this cycle, we study idiosyncrasies of the dimension with the most data availability: mobility.

Acknowledgements

Journeys are never easy, but they are always better when we have people by our side. Thank you all for being by my side. I am privileged to have amazing friends, colleagues and family by my side, and everything is a by-product of multiple social interactions.

First, I thank the people that were in direct contact with me in the development of this thesis. Dr. Hugo Barbosa was of fundamental importance for any word, table and figure of this document. PhD is more challenging than I could imagine, but Dr. Barbosa pushed me in all my limitations to expand my skills. Dr. Ronaldo Menezes was also part of the entire process of my PhD, even when it was indirect. His trust and desire to make me always better drove me in the entire process. Dr. Alessio Cardillo was also unexpectedly a major source of inspiration and knowledge to my research development. I really value his patience, caution and rigour that inspired me so much. I would like to especially thank Dr. Alessandra Sala, Dr. Alice Marascu and Dr. Laura Lotero for helping, guiding and supporting me in key moments of my PhD. In a field such as Computer Science that women are minority, it was an inspirational and incredible opportunity for having such amazing role models by my side.

Nothing would be possible without the support of people that helped me before the PhD, and continue to do so. I am always grateful for the support from Prof. Carmelo Bastos-Filho, Prof. Hugo Valadares, Prof. Fernando Buarque and Prof. Márcia Carréra.

I would like to thank the people who were part of my daily life in my Ph.D. In special, I am extremely grateful for all the support and friendship from Ana Maria Jaramillo, Paula Giestosa and Eloise de Carvalho. Also, I would like to thank the

BioComplexLab team, and especially Nandini Iyer and Arthur Vandervoort.

Last, I thank my family that in many ways blessed me with love, support and a lot of opportunities in life. In special, now and always, I am thankful for having such a special mother who inspired, invested, and helped me go after better professional journeys.

Contents

List of Figures

List of Tables

List of keywords and acronyms

`MDE` data from Medellín

`BGT` data from Bogotá

`SAO` data from São Paulo

`all` mobility related to any purpose

`work` mobility related to work purposes

`nonwork` mobility not related to work purposes

`women` mobility from women

`men` mobility from men

`single` mobility from travellers that are not married

`married` mobility from travellers that are married

`parent` mobility from travellers that have children

`without children` mobility from travellers that do not have children

`lower` mobility from lower income group

`middle` mobility from middle income group

`upper` mobility from upper income group

LODES Longitudinal Employer-Household Dynamics data

ACS American Community Survey data

Expansion factor a weight that projects an element (travel, person or family) to the real magnitude in a population

Hotspots set of areas that have individually similar fraction of travels

Hotspot level 1 set of areas that have individually high fraction of travels

Hotspot cutoff level 1 fraction of areas that represent the hotspot level 1

Levels in hotpots level 1 is the level with the highest mobility flow, and level n is the level with the smallest mobility flow

Mobility diversity spatial concentration of mobility over areas

Flow hierarchy mobility flow between hotspots

1. Introduction

Human mobility impacts several sectors of society such as health [1, 2], security [3, 4], economics [5] and politics [6]. Commuting to work, going to the gym, and grocery shopping are just a few of the ways we fulfil our demands using mobility. Mobility can be studied from different viewpoints and scales such as urban systems [7, 8], migration [9], and career mobility [10]. Regardless of the scale or type of mobility, cultural constructs impose different expectations around how women and men behave [7–10]. At the same time, cultural constructs are a set of norms, expectations and impositions directly and indirectly built from the individual decisions, perceptions, interactions and behaviours [11].

Historically, cultural constructs have imposed women and men to different roles in society [7–10, 12, 13]. Such roles affect how each gender participates in the labour market [14–19]. Occupations associated with caring, such as nursing and primary school teaching, are frequently performed by women [20–23]. On the other hand, positions associated with leadership and high wages are frequently occupied by men [20–24]. Thus, women and men might display different patterns in mobility because they take different decisions and roles. In this thesis, we focus on how cultural constructs that are entangled with gender affect human mobility in different ways, whilst disregarding any possible biological differences between gender.

In some countries, women tend to work closer to where they live mostly because of caring duties or lack of adequate public transportation [25, 26]. Such constraints can also be associated with lower job wages for women than men [27]. Moreover, women are more likely to take into account their safety when moving to and from locations. Consequently, women are less inclined to travel during the early mornings and nighttime [28]. Gendered cultural constructs, the labour market, and mobility are in a self-reinforcing relationship, in which each of the three concepts help solidify the others.

Besides, there are other sociodemographic characteristics that are key for the study of disaggregated patterns in mobility, given that gender as a sociodemographic category dynamically intersects with other categories such as socioeconomic status. In turn, these dynamics interact with the urban landscape (e.g., transportation infrastructure, available amenities and opportunities) leading to a complex system. For example, the choice of where to live, the type of transportation and where to work are not always an option for lower income groups, as affordability is key for them [29, 30]. Higher income levels in some countries indicate more accessibility to resources such as mode of transportation and opportunities, or well-located residence [30, 31]. In this way, affordability and characteristics of the urban landscape can influence the mobility of women and men.

In summary, cultural constructs, available modes of transportation, city development and spatial distribution of opportunities can impact mobility decisions. Therefore, our hypothesis is that there are gender and socioeconomic differences in the concentration of mobility, and that socioeconomic differences intensify the gender differences in mobility. We argue that the spatial concentration of mobility is a proxy to understand the spatial landscape of opportunities, and that inequalities in job sectors can be also seen in the mobility dimension.

We argue that studying mobility might help to have better interventions and improving access to food, education, job and any other opportunities. In most cities, population growth is closely followed by an increasing carbon footprint from transportation systems, traffic congestion, and traffic stress to name a few [32–35]. These problems are not separate from, but intertwined with, the social dynamics and cultural expectations that we study in this thesis. By studying the differences in mobility that potentially impact inequalities, stress, economic constraints and environmental pollution, we contribute to the broader understanding necessary to tackle these issues. The United Nations, together with the support of several countries, designated 17 sustainable development goals in favour of general well-being [36]. This thesis relates to the academic development in 5 of 17 goals: (3) Good health and well-being (5) Gender Equality (8) Decent work and economic growth (10) Reduced Inequalities and (11) Sustainable Cities and Communities.

The existing literature mainly focuses on the universalities of human mobility [37–42]. The number of works that focus on analysing disaggregated patterns in mobility for different cities and countries has been increasing over the years [13, 43–48]. However, all these works are limited in spatial and temporal scope. Therefore, we focus on expanding the spatial and temporal scope by including a larger number of cities across a larger time period in order to identify possible systematic and robust mobility patterns of sociodemographic groups across cities that can contribute to the broader understanding of mechanisms that can affect gender and social inequalities. Our key contribution to this effort is a new pattern — mobility diversity — that captures the uneven distribution of mobility across genders, and that might be a proxy of gender differences and inequalities in the labour market. Finally, we demonstrate the existence of two main profiles on the cities of the United States that reflect a major separation in urban development, sociodemographic characteristics and, consequently, mobility characteristics.

In summary, this thesis tackles the following open questions:

RQ1: What are the differences in the urban mobility of gender and socioeconomic groups across cities? (Chapter 3)

RQ2: How do travellers from each gender and socioeconomic group temporally organise their travels? (Chapter 3)

RQ3: Do women concentrate their travels in urban areas similarly to men? (Chapter 4)

RQ4: Do people across socioeconomic groups explore urban areas in different manners? (Chapter 4)

RQ5: Do women have shorter commuting travel time than men, regardless of the city and household arrangements? (Chapter 5)

RQ6: Do women and men have similar exposure to opportunities? (Chapter 5)

RQ7: How do cities with similar mobility characteristics reflect urban development and gender gaps? (Chapter 6)

This thesis found that differences in mobility patterns across sociodemographic groups are systematic across cities, and that some cities profiles emerge from the urban mobility concentration. We study the mobility of more than 50 cities based on datasets of 4 countries: United States, United Kingdom, Brazil and Colombia. However, we also found that there is a shortage of sociodemographically disaggregated data, which is crucial for designing better interventions.

The thesis is divided into the following chapters:

Chapter 2: Human Mobility We present an overview of the data types that allow us to study and understand mobility from the three aspects: spatial, temporal and social. Then, we summarise the main findings in the literature regarding overall human mobility behaviour, and the patterns in the mobility of gender and socioeconomic groups. This chapter focuses on the main concepts of human mobility, including the advantages and disadvantages of using certain data types, methodologies, models and frameworks to study patterns in mobility.

Chapter 3: Characterising urban mobility We study urban mobility in four countries: Brazil, Colombia, the United States of America and the United Kingdom, analysing the distinct mobility patterns of travellers from different gender and socioeconomic groups. This chapter describes the spatio-temporal characteristics of mobility for women and men, and people belonging to the lower, middle and upper income groups.

Chapter 4: Estimating mobility diversity We propose a methodology using mobility diversity metrics to capture a new mobility pattern that unveils gender and socioeconomic differences. This methodology can be used to study the uneven distribution of job opportunities across gender and socioeconomic groups. This chapter shows that men and middle income travellers tend to distribute their mobility more evenly through all urban areas compared to other gender and socioeconomic groups, likely affording them greater access to job opportunities across a given metropolitan area.

Chapter 5: Estimating mobility cost and reward We present how different sociodemographic groups can benefit via mobility from urban systems across different countries. This chapter demonstrates that different gender and socioeconomic groups show unique mobility patterns: cost (commuting travel time) and reward (quantity and diversity of amenities/opportunities).

Chapter 6: Profiling cities based on mobility characteristics We define a unique set of mobility characteristics (feature vector) that represent the concentration of mobility in several cities. Based on this feature vector, we create profiles of cities to study how their sociodemographic groups are affected by the urban system. This chapter demonstrates that cities with specific urban mobility characteristics are associated with particular development, infrastructures and sociodemographic indicators.

Chapter 7: Discussion, conclusion & limitation In this final chapter, we summarise our contributions, the impact of our work and the limitations of our analyses. In this thesis, we highlight the importance of studying mobility patterns across gender and socioeconomic groups in order to inform the development of sustainable cities and communities.

2. Human mobility

The study of human mobility is fundamental for the understanding of multiple aspects of our life such as migration [9], epidemics [49] and labour market [16]. In this chapter, we summarise the main findings within human mobility, specifically in the context of data, identifying patterns that emerged from the general population and detecting specific patterns belonging to sociodemographic groups. We point out the major methods and frameworks related to understanding mobility patterns, particularly considering the temporal, spatial and social aspects.

2.1 Data

Human mobility behaviour can be extracted from several data sources such as global positioning systems (GPS), government censuses, social media activity, third-parties surveys, internet service providers (ISP), mobile service providers (MSP) and credit card providers (CCP). Each source reveals its specific data quality [3], and we classify them by their common type into four categories:

Spatio-Temporal data: contain information about *where* the person was in what *time*. There are five common data sources: Call Detailed Records (CDRs) [1, 2, 50–54], Credit Card Records (CCRs) [43, 55, 56], Bills (banknotes) [37], Wi-Fi [57–60] and GPS trackers [53, 60–64]. The CDRs are mobile phone activities that specify the timestamps for any call and text messages. CDRs are accurate on who is calling and when the call was started and completed, but dense areas provide high spatial granularity in contrast to rural areas (constrained by the antenna capabilities and quantities). Moreover, inactive periods follow active burst periods of calls and messages presenting temporal sparsity. CCRs are information regarding when a person purchases products or services in a particular location using a debit or credit card. CCRs are biased towards age and income characteristics; for example, in general, rich adult people would dispose of more money to spend than young or poor people. CCRs also experiences inactive periods because the majority of people do not spend money continuously. The tracking of bills like dollars or euros provides information about the trajectory of people paying for services and products, but long gaps are perceived because of the difficulty of tracking all the transactions made by a particular code bill. Moreover, the movement (trajectory) of such banknotes represent an aggregation of the trajectories of several individuals given that the notes change hands often. Wi-Fi data can track people inside predefined areas such as universities, and airports. This source of data is commonly private, spatially limited to small areas and contains a few visitors. GPS trackers are a powerful device for temporal and regular precision of displacements, but they rely on a battery, may generate position errors and not be effective for indoor information. Some works show that social or contextual information can be inferred from many of these types of datasets. For instance, the call and text message distribution and information between people may infer their relationship: friends, family or colleagues.

Socio-Temporal data: contain information about the *relationship* between people and their activities in *time*. The data sources are generally from Online Social Networks (OSNs) [65] but could also come from traditional longitudinal studies performed by social scientists (collection of data via surveys). OSNs like Facebook are highly accurate about friendships between people and provide timely information about a person or group activity. However, people on OSNs do not necessarily provide information about their location. This type of data also experiences inactive periods because it depends on the user's active posting. However, integrating socio-temporal data with other types of data contributes to establishing broad social information.

Socio-Spatio-Temporal data: contain information about the *relationship* between people and their activities in *time* and *place*. Examples of this category are: Location-based Social Networks (LBSNs) [62, 66–68], Census [1, 50, 54, 69, 70], Surveys [4, 55, 56, 71] and Mobile Flow Records (MFRs) [51]. LBSNs are social networks in which the information is geotagged (e.g., text, image or videos). LBSNs are limited by the users' activity but are frequently public and rich in contextual information. Although Facebook is a private source, Facebook can also be considered as LBSN because they support geotagged information [60]. The literature more widely analyses twitter data as they are freely accessible to anyone, but the social information is not as rich as Facebook. On Facebook, a person points out the relationship with another individual (e.g., mother, father, husband, wife), making the platform one of the most reliable data sources for social ties. Census and surveys are reports about a selected group of people that represents the entire population. The government census, for example, includes public information on socioeconomic aspects, which are rarely available by other sources. MFRs are mobile phone information (e.g., system logs, use of applications, calls, messages and offline activities) that capture high-quality temporal information. MFRs can extract non-uniform sampling information. MFRs passively record mobile activities, and depending on the mobile phone settings, MFRs can also passively record spatial locations.

Contextual data: contain information about the *characteristics* of a person such as age, gender, friendship, income and others. Credit Card Records (CCRs), Online Social Networks (OSNs), Location-based Social Networks (LSBNs), Census, Surveys and Mobile Flow Records (MFRs) are examples of accessing extra information related to a person. Some specific peculiarities may reshape the classification of human mobility patterns. For instance, people with higher income might purchase more services and products, increasing the number of locations collected by the CCRs, so people who do not spend as much money might be underestimated on the number of visited locations.

The no free lunch theorem applies to human mobility data types. Each data type contains different challenges. However, one aspect that all these types have in common is data sparsity and level of granularity. In this thesis, we will use the surveys and censuses of Brazil [72], Colombia [73], the United States of America [74–76] and the United Kingdom [77]. Our data selection is based on the availability of contextual information of individuals associated with their mobility characteristics. Moreover, we will also integrate data from OpenStreetMap [78] that contain information about amenities, highways and buildings located in multiple cities. The OpenStreetMap data allow us to understand the urban landscape, as the surveys and censuses allow us to understand the population and how the urban landscape is used. All these data together allow us to study mobility in multiple cities.

The temporal, spatial and social aspects regarding human mobility are generally studied separately in the literature. In the following sections, we explore each of these aspects in the disaggregation of human mobility patterns.

2.2 Temporal aspect

Temporal analyses of mobility face the limitations of data sparsity and waiting time measurement between locations [50, 64, 67, 79]. The data sparsity restricts the predictability of human mobility patterns because, between two consecutive records, displacements are not being captured [54]. For example, individuals do not call from every location they go, so Call Detailed Records (CDR) data only capture some visited locations [2, 51, 80, 81].

In this way, small observation periods are usually aggregated to larger periods to fill in the temporal gaps and normalise the different lengths between all the users' trajectories. To predict the next locations, trajectories are usually represented by a constant length [4]. Similarly, Alhasoun et al. [52] incorporate the idea that people who do not necessarily know each other share similar mobility characteristics, and these similarities help to fill the gaps with incomplete trajectories. The similarity of users can be calculated using the temporal closeness metrics allowing a clearer temporal picture about the heterogeneity and sample limitations.

Secondly, the waiting time between two locations draws a distinction between passing-by and stayed-in locations. The establishment of waiting time is complex because each person stays at a place in a different range of time and days [51, 54]. For instance, people probably stay at a coffee shop on Monday for a shorter period (at most 15 minutes) than on Sunday (more than 15 minutes) because people usually have different daily responsibilities on weekdays and weekends. Therefore, the temporal aggregation of waiting time affects the data analysis because of heterogeneities in individuals trajectories [4, 50, 67, 79]. Liu et al. [51] show that if the waiting time is equal to 2 hours, the observed behaviour would be completely different from the behaviour extracted from the aggregated data from less than 2 hours. Interestingly, aggregating data for a waiting time higher than 2 hours do not affect the observed behaviour. This highlights the importance of analysing mobility using a waiting time of less than 2 hours.

Temporal aggregations and analyses also take into account issues related to regularities. Mobility is repeated in many ways, such as daily, weekly, biweekly, monthly, and seasonally. Zufiria et al. [4] investigate these various periods, and point that there is a seasonal and cultural effect playing a role in human mobility patterns.

Wang et al. [67] discovered that severe winter storms affect human mobility behaviour, which highlights the importance of taking into account seasonality while studying mobility. They showed that a log-normal distribution better characterises mobility under either regular and irregular conditions, but the largest deviations were unveiled on storm weeks. Similarly, Kondor et al. [50] detected weekly long-term trends but different ones seasonally. Holidays tend to change human behaviour. In the winter holidays, there is a larger decrease in call activity, and in the summer holidays, there is a reduction in call activity but on a smaller scale. Indeed, the analysis of the time series of activity should account for this seasonality; interestingly, monthly periods did not show a significant effect.

Various works contribute to the fact that weekdays and weekends present different patterns [2, 50, 54, 64, 67, 82–84]. The predictability of weekdays, for example, is higher because of daily responsibilities such as work or study (return based-activities) [50, 84]. In Li et al. [68], busy hours provide poor predictability because there are more alternatives in people's free time like 19:00hs (food, events, nightclubs, art and entertainment). The analysed data were from New York City, Chicago and Los Angeles that contain various commercial locations and are also places highly visited for tourism. Statistical analyses can be performed to take into account these unbalance in data. Ponte et al. [63] used the Gini coefficient to calculate the statistical time dispersion to establish the heterogeneity of travelling times from the bus trips dataset. Otherwise, larger routes would have larger time delays, which interferes with the prediction.

2.3 Spatial aspect

When dealing with the spatial aspect of mobility, there are three main concerns: data sparsity, presence of hotspots, and differences in modes of transportation. Spatial data sparsity occurs because some regions naturally will have less information. For instance, areas covering mountains, rivers and forests might be rarely visited and occupied by people. Secondly, hotspots refer to some areas that will tend to have most of the information about mobility because hotspots attract a high volume of people. Examples of hotspots are airports and city centres. Last, the modes of transportation change the velocity in which the mobility is performed between two timestamps. In this way, the mode of transport affects mobility measurements such as travel distances and times. If these three concerns are not well dealt with, the captured human mobility patterns may be misleading or, worse, lead to incorrect patterns.

The spatial scope is seen as locations, trajectories, areas, grids and clustered regions [50, 51, 55, 62, 64]. Trajectories can be inferred using locations by their category (home, work) or by specific points (latitude, longitude) [52, 59, 67, 85]. Some works assign and categorise locations considering demographic information from census or third-party information to incorporate relationships and *a priori* division of areas [1, 56, 69, 86].

When information about locations are not available in the data, researchers apply methodologies and algorithms to infer locations, areas and trajectories. For example, Jiang et al. [54] grouped nearby points to define locations. This technique aims to prevent such points from being interpreted as different locations. One place (e.g., restaurant, airport) includes a larger space when we look at specific latitudes and longitudes. Among several techniques, clustering is widely used to group points; one example of a clustering technique is the Density-Based Spatial Clustering of

Applications with Noise (DBSCAN) [67, 79, 84]. However, differences in elevation (from floors in a building) are poorly investigated in the literature. In fact, Wi-Fi data allow researchers to have a better understanding of indoor mobility as the presence of routers in different floors allows distinguishing latitude, longitude, and elevation [57–60].

The aggregation of points into areas can be done in many ways. Using mobile phone data, Kondor et al. [50] broadly analyse the impact on five levels of spatial aggregation: whole city (all antenna-level), boroughs (based on official divisions $39km^2$), wards (based on official divisions $1.87km^2$), pixels (divide the area equally in $500m^2$) and clusters (based on similarity on the activity time series). The difference between spatial granularity shows that irregularities disappear while larger areas are aggregated as heterogeneities are hidden. The clustering predictability provided better results than the pixels, wards and boroughs because the local and smooth similarities between areas become homogeneous. Even though city-level aggregation is widely used in the literature, clustering techniques usually better separate the irregularities coming from hotspots and empty areas. The selection of a spatial aggregation should depend on the problem. For instance, some data on sociodemographic indicators of profit, tourism and crime might be only available in a city level aggregation, establishing the aggregation a priori.

Moreover, spatial aggregation can also use contextual information from the environment and population. To improve the predictability of next locations where people may be found, Alhausoun et al. [52] aggregated users with similar patterns based on spatio-temporal closeness (city scale) like daily home-work trips. With a similar perspective, Yu et al. [84] proposed a dynamic method to predict the next location people would visit by extracting the activity of contextual features. The method dynamically updates the transition probability of locations by adjusting the location weight based on the user's historical trajectories volume.

Other works used the idea of attractiveness to identify locations. Yan et al. [85] differentiated locations arguing that people return more to previous or highly-visited locations (also established by other works [50, 59, 60, 67, 68, 83, 84]). Similarly, Liu et al. [62] demonstrated that specific locations attract numerous people (aka hotspots), which uncover a burst phenomenon. When considering differences between residents and visitors patterns, short-distance movements are independently detected, which hinders the identification of visitors' patterns. Thus, attractiveness, hotspots and bursty phenomena are relevant concepts for understanding mobility.

In addition, Barbosa et al. [40] demonstrated that the last visited locations play a role in mobility, together with the role of the most frequent locations [38, 39, 87]. Then, Alessandretti et al. [60] pointed out that individuals' sets of visited locations grow temporally, but the number of new locations saturates. Routines evolve gradually in time, but they hold a typical size of 25 locations. The social life is related to the set of preferred locations which corroborates the returners and explorers dichotomy [87] and highlights a correlation between spatial exploration and the number of social ties. Thus, people are attracted to popular places and return to previously frequent and recent locations. This attraction to popular places might be a collective mechanism that influences mobility indicating that individual patterns in mobility can also be influenced by society and sociodemographic characteristics [26, 43, 88].

2.4 Social aspect

People are influenced to visit new places as their friends, family, and colleagues recommend places. In this way, social ties can play a role in introducing unexpected places to the routine of individuals. As a consequence, people are also driven by other people's mobility.

In the literature, the establishment of a relationship between people based on the visited locations is usually superficial because it only divides into two categories: friends or not friends. In reality, we distinguish people between several categories such as family, parents, close friends, childhood friends and colleagues. However, this kind of data is rarely available to research. Frameworks such as Facebook and Instagram are the ones that have a better perception of social ties. For the case of censuses and surveys, there is available data about household composition that allows researchers to study family dynamics. Thus, the literature available on the influence of social ties is growing as data are available for researchers.

In general, people are usually influenced by others in different amounts. For example, parents typically decide the locations that children visit until their children become independent. As people age, their friends and partners might play a major role in mobility than their family. Rubrichi et al. [2] claimed that friends impact differently depending on the context; as the amount of shared time increases, the chance of sharing similarities also increases. However, several works point out that people without any relationship to the others also display similar patterns regardless of who is influencing who [52, 89].

Researchers assumed relationships between people based only on the co-location of users [90], and interestingly, Alhasoun et al. [52] show that similar strangers (people that potentially do not know each other) information help to improve the predictability of human mobility patterns. These strangers do not necessarily determine real friendship, but grouping these users can exhibit virtual (online platform behaviour) and real similarities. Chen et al. [89] present in fact that co-located people provide as much predictability in mobility as social ties.

Using a different approach, Alessandretti et al. [60, 91] explored the relationship between social and spatial domain, demonstrating that people tend to be explorers socially and spatially in a similar way. Besides, they show that women and extroverted individuals are more likely to be explorers. Likewise, Jiang et al. [53] point out that social influence is driven by semantic information, not only by locations but what the locations mean to people.

In summary, social ties tend to influence individuals mobility, people that share similar traits in the social domain can display similar traits in the mobility domain, and people who share the same locations can also share similar behaviours in mobility. All these components are part of complex interactions between people and how cultural constructs play a role in these interactions. Any human behaviour can influence not only the behaviour of one person but also can be translated to the norms and expectations in society [11].

2.5 Modelling general patterns of mobility

Even though mobility can be affected by spatial, temporal and social aspects, several works model mobility using one or a combination of two aspects. In fact, the focus on one aspect of mobility benefits the accuracy and robustness of models. In this section, we list the most popular models in human mobility.

Human Mobility can be modelled in different levels [3]: Individual-level, Population-level and Intermodality. These models show the importance of major characteristics in mobility. For instance, distance is a recurrent dimension that plays a role in mobility for most models.

In individual-level, the models reproduce the movement of each person based on their pattern. Examples of this level are: (i) *Brownian motion*, replicates series of irregular and random movements; (ii) *Lévy Flight*, replicates short movements followed by a few large movements; (iii) *Continuous Time Random Walks* adjust the models (e.g., Brownian motion, Lévy Flight, Ordinary Diffusion and Ambivalent Processes) to consider a random time between movements [92, 93]; (iv) *Preferential Return*, replicates the tendency of people coming back to well-known locations [38]; (v) *Recency Model*, replicates the return to places recently visited [40]; and (vi) *Container Model*, replicates individual movements of different spatial scales [41].

All the population-level, the most well-known models are *Gravity model*, *Radiation model* and *Intervening Opportunities model.* The *Gravity model* reproduces the migration and urban flows based on the population quantity in different regions [94–96]. The *Radiation model* captures the average flow of travellers over regions based on the number of opportunities and the distance between locations [42, 97, 98]. The *Intervening Opportunities model* highlights the direct relationship of the number of opportunities and not directly related to the distance between places [99–104].

The intermodality level replicates the transport types as different layers of human mobility [105–107]. Cars, trains and people, for example, move and are surrounded by different environments and temporal limitations. Here, multilayer networks provide the cumulative relationship between locations at different levels.

Lévy flights and continuous time random walks diver accurate predictions but fail to reproduce the individuals' tendency to revisit locations [88]. The majority of models in human mobility accounts for the stationary long-term mobility, and they do not take into account the purpose of travel and its importance to mobility [88].

All these models summarise mobility in their major components very well, but they do not explain the peculiarities of some groups of people, such as women and men. For instance, Alessandretti et al. [41] shows that women and men are better characterised by different parameters for the container model. Thus, inequalities that come from mobility might be hidden in these models, and further and different studies are necessary to understand well mobility in a disaggregated way. The analyses of disaggregated patterns in mobility will also take into account distance, mode of transportation and opportunities probably considering these dimensions differently.

2.6 Disaggregated patterns in mobility

Cultural constructs in work and education responsibilities, childcare, accessibility, security, and social relationships shape human mobility. Because of the responsibilities of childcare, women tend to avoid long commuting travel time, whereas men are more likely to take job opportunities farther from their home [27]. When it comes to mobility, women account for factors like safety and accessibility more than what men do [25, 27, 28, 108–110]. For instance, 65% of women in Mexico City have suffered some offence while using the public transportation system [25]. Accounting for safety and accessibility translates into higher travel costs not only for travellers but also for authorities themselves since they have to provide surveillance services to avoid women being subject to offences [28]. However, everyone should have, in theory, the same access to every part of the city. This means that public transportation must have a fair cost regardless of gender, age, disability, and social status [25].

The literature on the subject has shown that women are more likely to perform short travels [111], are more likely to work at home [112], are more likely to return home more frequently [48], have paths that are more sinuous (deviates more from the main trajectory) [47], make more trips with distinct purpose [28], travel longer than men for household-serving purposes [111], tend to perform trip-chaining (multipurposes) [25], and are more likely to choose cheaper transportation modes [46]. Also, women are more likely to have flexible or irregular work schedules [108], opt to part-time jobs because of childcare responsibilities [45, 113], visit fewer unique locations [26], are more prone to take career breaks [108], accept more easily less-qualified jobs [27], and spend their time more homogeneously across different locations [48]. Women are more likely to be imposed to stereotypes than men, causing some difficulties for women to reach high-paid jobs [114, 115]. Low participation of women in some job sectors is a reality even in high-income countries [115]. Women are less often trained in elite research groups, are promoted more slowly, and are more likely to leave STEMM careers [115, 116].

A family shares responsibilities and changes in their life, so their mobility patterns influence each other [111, 117]. In fact, having children impact more women's mobility than men's mobility [43]. Changes in the travel behaviour of women and men are affected by marriage or divorce, parenthood, entering or leaving the labour market, changing jobs or moving houses [111, 117]. Cities where children have more ability to move without their parents generate less burden in parents' lives. All the aforementioned factors shape women's mobility aside from their socioeconomic status, and even higher gender gaps in mobility patterns are found when socioeconomic status is taken into account [25, 26].

Socioeconomic status impacts the mobility patterns as well as household distribution because they are strongly linked to each other [118]. People choose to live in affordable locations and in areas with (or close to) higher concentrations of facilities related to work, study, and retail. Travellers belonging to different socioeconomic groups have different travelling capabilities [31, 44, 119]. For instance, people/travellers with low income can not travel as far and as much as those with high income [120]. The low rate of accessibility impact mobility's demand [121], and only large-scale centralised actions may overcome these issues.

Mobility is also seen as an indicator of power as it relates to how much accessibility people have to opportunities [112]. Milan et al. [122] demonstrate that new transit development is more environmentally friendly and offer equal access to public transportation for different groups of people.

Taken together, all the works in our related topic are converging to a set of different mobility patterns that are shaped when considering gender, socioeconomic status and other dimensions. Gender is a dynamic cultural construct that takes into account institutional, intersectional and structural components of society [123] that should be studied and analysed. In this thesis, we contribute with a new mobility pattern disaggregated by gender, a new revisitation of overall mobility characteristics seen from a network and data science perspective, and a new methodology to understand similarities across cities.

3. Characterising urban mobility

Via mobility, people are able to fulfil some of their responsibilities such as working, studying and going to the doctor. As cultural constructs impose different roles on people, women and men may have different expectations and responsibilities, which is reflected in their mobility behaviour [12, 124–127]. In the literature, women and men present differences in how they move, with respect to travel time [111], trip-chaining [128], and mode of transportation [26]. However, as cultural constructs evolve over time, in different manners, across various cities, mobility patterns may shift as a result [11]. Therefore, we revisit and expand the study of urban mobility from travellers of different genders and socioeconomic groups in Brazil, Colombia, the United Kingdom and the United States of America. We unveil in this chapter that: (i) urban landscape plays an important role in mobility, (ii) women and men distribute their travels differently over space and time, (iii) travellers from different socioeconomic groups display major differences in how travels are distributed over space and time, and (iv) gender differences in the structure of urban mobility are amplified when accounting for the temporal dimension or household arrangements.

3.1 Data

This thesis explores data from household and national travel surveys of different cities from four countries: Brazil (BR) [72], Colombia (CO) [73, 129], the United Kingdom (UK) [77], and the United States of America (USA) [74, 76]. We summarise the available information for the data of each country in Table 3.1, and we describe their details in this section.

These datasets detail the routine of the population in each country or city, considering household and economic arrangements, employment status and mobility. We have an *expansion factor* available that allows us to project the sample onto the real number of residents within each area/zone (spatial partitioning established in each dataset) — the expansion factor projects the number of individuals, households or travels for a given area. One state, city or metropolitan region is divided into several areas/zones, and one area/zone is the smallest spatial partitioning available in the mobility data. Thus, the *expansion factor* is the weight associated to a variable (e.g. travel or person) that allows us to estimate better characteristics of the sample in relation to the entire population [130]. This is necessary as census and surveys are not able to collect data with the entire population as it is highly expensive and time costly.

Table 3.1: Summary of available information on the datasets of each country.

Type	Information	Countries			
		BR	CO	USA	UK
Temporal	Number of years	3	2	5	5
	Travel time	x	x	x	x
	Departure time	x	x	x	x
	Arrival time	x	x	x	x
Spatial	Regions resolution	x	x	x	x
	Cities resolution	x	x	x	-
	Zones resolution	x	x	x	-
	Place of work	x	x	x	-
	Place of residence	x	x	x	-
Purpose	work	x	x	x	x
	not related to work	x	x	-	-
Contextual	Individual Gender	x	x	x	x
	Individual Income	-	-	x	x
	Household Income	x	x	x	x
	Marital Status	x	x	x	x
	Family composition	x	x	x	x

In the case of the USA and the UK, we have information about the mobility performed only for the purpose of work, such as commuting travel times. The available data for both countries already have the commuting characteristics computed from the individual mobility and contain limited information about mobility trajectories. Moreover, for the United Kingdom, the spatial partitioning of the entire country only allows us to see at most 11 zones, which is insufficient for computing several analyses of this thesis. However, when it is appropriate, we use the data from the United Kingdom to compare with the results from the other countries. As one of the key contributions of this thesis is indicating systematic, robust and general patterns in mobility across countries, we found the United Kingdom data important to be included in this thesis.

For the USA, we have data for the entire country via the American Community Survey (ACS) [76], IPUMS USA [75], Longitudinal Employer-Household Dynamics data [74], and National Household Travel Survey [131]. We have individual, household, and commuting travel information that we group by city or state. Some data from the American Community Survey (ACS), such as the place of work, was easily available on the IPUMS USA framework [75]. We analyse the data of 50 cities in the United States from 2015 until 2019. Table A.1 in the appendix displays a summary of general statistics in the data from the United States for the top 50 cities considering the number of zones in the available spatial partitioning of the Longitudinal Employer-Household Dynamics data. These 50 cities are the most populated ones in the USA as the spatial partitioning considers population volume, sociodemographic characteristics and household distribution. To be consistent across cities, we divide the data into three socioeconomic groups based on the percentiles of the individual income in each city and year.

The data from Brazil and Colombia have information about travels made for different purposes, such as study and leisure, and includes the travel origin and destination on a spatial partitioning defined by the sociodemographic characteristics of each region. These data have a detailed spatio-temporal granularity, which enables us to estimate a wider range of characteristics in mobility. The Latin American data are from three cities: Medellín (`MDE`) [73], Bogotá (`BGT`) [129] and São Paulo (`SAO`) [72]. For each region in Brazil and Colombia, we analyse the data collected in different years: {2005, 2017} for `MDE`, {2012, 2019} for `BGT`, and {1997, 2007, 2017} for `SAO`, respectively (see Table 3.2 for data summary). Throughout this thesis, these data will be widely used to compute in-depth analyses of urban mobility across sociodemographic groups.

For the Colombian datasets, the surveys have a consistent number of socioeconomic groups across years and cities. Households are grouped into six strata, with

Table 3.2: Data description of the sample datasets analysed in our study. For each region, we have the number of zones into which it is divided, N_Z. Then, for each year we have the number of travellers N_P, the fraction of men (women) travellers F^M (F^W), the number of travels N_t, and the fraction of travels made by men (women) F_t^M (F_t^W).

Country	**Region**	Year	N_Z	N_P	F^M	F^W	N_t	F_t^M	F_t^W
CO	MDE	2005	403	55,681	0.48	0.52	126,164	0.48	0.52
		2017	541	38,048	0.49	0.51	87,614	0.49	0.51
	BGT	2012	911	58,313	0.46	0.54	122,361	0.45	0.55
		2019	1084	66,820	0.48	0.52	152,310	0.48	0.52
BR	SAO	1997	374	98,780	0.48	0.52	199,647	0.48	0.52
		2007	452	91,405	0.47	0.53	196,698	0.48	0.51
		2017	511	86,318	0.47	0.53	183,092	0.50	0.50

stratum 1 corresponding to people with the lowest income and stratum 6 corresponding to people with the highest income. We then map these strata to: `lower` (strata 1 and 2), `middle` (strata 3 and 4), and `upper` (strata 5 and 6), respectively.

In São Paulo, the socioeconomic classification of the population changed over time to better capture the current picture of the population characteristics. In 1997, for example, respondents were classified into five socioeconomic groups labelled as *A* (upper), *B*, (mid-upper), *C* (middle), *D* (mid-lower), and *E* (lower). Table 3.3 presents our map of socioeconomic groups defined by the Brazilian institutes such as IBGE (Brazilian geography and statistics institute) [132] and ABEP (Brazilian association for population studies) [133]. Even though there are available more socioeconomic subdivisions in one year than the other, our mapping considers the meaning of subdivisions (lower, middle and upper). Further data details from Colombia and Brazil are displayed in Tables A.2 and A.3 in the appendix.

Table 3.3: Mapping of the Brazilian classification scheme into the `lower`, `middle`, and `upper` socioeconomic groups (SES) for the three years of the survey.

	Year		
SES	1997	2007	2017
`lower`	*D, E*	*C2, D, E*	*C2, D, E*
`middle`	*B, C*	*B1, B2, C1*	*B1, B2, C1*
`upper`	*A*	*A1, A2*	*A*

We also have collected data from OpenStreetMap [78], which provides spatial data for amenities (e.g., workplaces, transportation systems and buildings). The OpenStreetMap data can be used, for example, to indicate possible attractors of mobility for certain zones that are rich in amenities. It is important to highlight that this framework has evolved over the years, and the most recent versions are the most developed and accurate. Therefore, temporal analyses using OpenStreetMap are not always possible and available.

3.2 Mobility

Mobility, in this thesis, is defined as the displacement between locations within metropolitan regions. On a map, locations are seen as points, areas, neighbourhoods, businesses, buildings, counties, cities, states, or countries. In this manner, when a person goes from home to their workplace, their movement can be analysed from all of these spatial scales.

The spatial partitioning embedded in our data is defined by the organisations and governments of each country. Therefore, we are limited by their spatial partitioning. We argue that this spatial partitioning is robust and reliable because it considers the sociodemographic characteristics of each metropolitan region, which is the key component of our analyses.

Nevertheless, we point out that the data from each region have a particular limitation. For example, the urban mobility of the cities in the USA and UK is not on the same scale as those in Colombia and Brazil. Mobility characteristics will be analysed, when possible, for each data. For the USA and UK, we will have a macro-scale analysis of the human mobility patterns. As for Colombia and Brazil, we extract more detailed mobility patterns from the data because we have the routine movements, including origins and destinations, available coupled with rich metadata related to individuals, households and travels (see Table 3.1).

We then, in this chapter, map the mobility onto a weighted spatial network where the nodes are zones/areas, and the travels/journeys between the nodes represent the edges [134]. Each edge/journey is associated with the number of travels between zones. Figure 3.1 displays the nodes, Zone A and Zone B, and the edge between the two zones that is the fraction of travels made by a group X, f_t^X.

For each city and year, we build a network from the mobility of each gender and

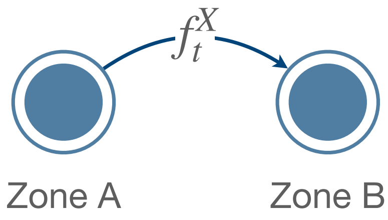

Figure 3.1: The mobility network is weighted by the fraction of travels, f_t^X, performed by a group, X (e.g. all, women and men), between two nodes, zones (e.g. Zone A and Zone B). The sum of all the fraction of travels, f_t^X, is equal to 1.

socioeconomic group in order to understand the main characteristics of their urban organisation. The network analyses will focus mainly on the data from Colombia and Brazil because of the high resolution of mobility data that is available for those countries. Moreover, we will analyse a weighted network for each city, year and sociodemographic group. In our analyses, we represent the gender groups as men (`men`) and women (`women`), and the socioeconomic groups as the lower-income group (`lower`), middle-income group (`middle`) and upper-income group (`upper`). We point out that the socioeconomic grouping reflects income level and housing conditions, such as access to water and energy. Therefore, higher-income groups may afford better living and mobility conditions.

Then, we will analyse how the commuting patterns of all four countries are spread over space and time. We unveil how, despite variations in spatial partitioning, the differences in mobility are consistent across several cities in the four countries. Our analyses are limited to the data that we have available, but we argue that the systematic results ensure that our analyses and conclusions are valid for multiple regions.

3.3 Spatial patterns

By mapping the spatial distribution of mobility in urban areas to a weighted network, we analyse network properties (e.g., density and clustering) from a spatial perspective. These network analyses describe the urban organisation of mobility for different sociodemographic groups. We compute well-established metrics in the literature [135–137] for our weighted networks, displayed in Table 3.4.

First, we look at the average inflow — the average number of travels made in the direction of an area — for the network of each sociodemographic group. We see in Table 3.4 that the mobility network of men travellers has a higher average inflow for `MDE` and `SAO`, indicating that men tend to have a higher number of travels to the majority of the areas. This is true regardless of women having a higher fraction of travels than men (see Table 3.2). We observe that `BGT` shows a higher average inflow for women than for men, which is different from the other cities. We observe that the upper income group has the smallest average inflow for the socioeconomic groups, indicating that the upper income group does not frequently visit most areas. In fact, we see areas having 0% of travels from upper income group.

Using indegree centrality, we take into account not only the inflow but also the degree of areas. The degree of an area, i, is the number of areas connected to it (mobility established between areas). The indegree of an area i is the number of areas with mobility going to the area i. In this way, areas that receive more visitors from a wider range of areas will have a higher value of indegree. High values of indegree centrality represent that areas receive a higher fraction of travels from several origin zones. Thus, we observe in Table 3.4 that men and middle income group are more likely to show high indegree centrality in their mobility network, which suggests that both groups visit areas that tend to be highly visited locations.

In terms of average clustering, we quantify how locally well-connected areas are through the mobility of each group of travellers. This metric computes the tendency of having mobility between three areas. We see that women and the upper income group tend to have higher values of average clustering, indicating that travellers from these groups are more likely to travel between three areas (local organisation).

Next, we compute the networks' density to study the extent to which globally well-connected areas emerge for each group of travellers. Contrary to locally well-connected areas, we see that high-density values of mobility are more common for men and travellers in the middle income group. Combining average clustering and density, we see that some groups (men, middle income group) are more likely to show global organisation in their mobility, and other groups (women and upper income group) tend to show a local organisation.

We then look at the average neighbour indegree that computes how many areas on average the neighbours of area i receives travels. This metric takes into account the degree and the fraction of travels to the neighbours of the area i. We see that the areas that men and middle income groups usually visit are well-connected, but the neighbouring areas that both groups visit are well-connected. This finding indicates that both groups are more likely to be attracted to or reside in highly-visited areas. Indeed, in Chapter 5, we present that the men and middle income group are more likely to have a higher number of travels in the majority of the areas, and they are more likely to visit all the areas in `SAO`, `BGT` and `MDE` evenly.

In summary, looking at the main network properties in Table 3.4, we see that the values extracted from the overall network are closer to the ones extracted from the networks of each gender than from the networks of each socioeconomic group. This might be a consequence of socioeconomic segregation within metropolitan areas [138–141].

The upper income group and women tend to have smaller values of average inflow, in-degree centrality and average neighbour indegree than the other groups. This implies that both groups might have higher likelihoods of concentrating their travels in a few areas (further analyses in Chapter 5).

The small differences across gender groups might be the case because women and men form a similar global urban structure that might emerge from the infrastructure and the concentration of opportunities and amenities in the urban area. In fact, we see, for example, that in our data, women and men have similar probabilities of living and working in the same area (shown in Table 3.7). However, this is not the case for socioeconomic groups because of the uneven distribution of where they live and where they travel to. For example, we may see differences in the residential distribution for different income groups because the upper income group may tend to live in areas that are not affordable for the lower income group. [120]. The same happens to the travels related to work. The job opportunities of the upper income group might be concentrated in a few areas that are not necessarily highly visited by lower income group (shown in Chapter 5). For the case of the middle income group, their travels are distributed more evenly throughout the metropolitan region (shown in Chapter 5).

Table 3.4: Network properties from the mobility of `all`, `men`, `women`, `lower` income group, `middle` income group and `upper` income group travellers.

Metrics	Region	Year	`all`	`women`	`men`	`lower`	`middle`	`upper`
Average inflow	`MDE`	2005	11259	4945	5009	4668	5411	1330
		2017	120	64	74	89	73	55
	`BGT`	2012	16063	9099	6964	8469	6848	833
		2019	16466	9957	9383	8085	6862	822
	`SAO`	1997	71302	4690	7958	17860	44235	2328
		2007	47032	2647	3418	11192	30805	2678
		2017	51916	2548	3332	8010	33085	3091
Indegree centrality	`MDE`	2005	0.7662	0.6234	0.6475	0.5050	0.5079	0.2799
		2017	0.4732	0.3239	0.3407	0.1088	0.1621	0.1875
	`BGT`	2012	0.2634	0.1746	0.1493	0.2004	0.1603	0.2375
		2019	0.3210	0.2147	0.1976	0.1818	0.2165	0.2011
	`SAO`	1997	0.6091	0.3551	0.4905	0.2656	0.5366	0.2857
		2007	0.5497	0.3898	0.4055	0.2163	0.5180	0.2270
		2017	0.3315	0.2318	0.2338	0.0921	0.2930	0.1865
Average clustering	`MDE`	2005	0.4620	0.3890	0.3640	0.3340	0.3880	0.4180
		2017	0.3080	0.2850	0.2640	0.2650	0.1720	0.2480
	`BGT`	2012	0.2380	0.2170	0.1810	0.1810	0.2280	0.3160
		2019	0.1880	0.1580	0.1350	0.2000	0.2520	0.3320
	`SAO`	1997	0.3800	0.4300	0.4130	0.1940	0.1220	0.3510
		2007	0.4130	0.4150	0.3680	0.2770	0.1870	0.3390
		2017	0.3770	0.4160	0.3700	0.3090	0.2350	0.4090
Density	`MDE`	2005	0.1727	0.1222	0.1396	0.0797	0.1097	0.0528
		2017	0.0126	0.0532	0.0655	0.0532	0.0531	0.0342
	`BGT`	2012	0.0440	0.0423	0.0434	0.0205	0.0317	0.0193
		2019	0.0525	0.0239	0.0236	0.0238	0.0329	0.0101
	`SAO`	1997	0.2126	0.1327	0.1744	0.0839	0.1671	0.0532
		2007	0.0869	0.0293	0.0342	0.0307	0.0728	0.0390
		2017	0.0619	0.0203	0.0245	0.0171	0.0510	0.0282
Average neighbour indegree	`MDE`	2005	122	122	123	109	121	99
		2017	84	82	85	76	80	45
	`BGT`	2012	62	62	63	56	62	39
		2019	100	61	60	55	65	41
	`SAO`	1997	117	118	123	149	210	189
		2007	89	69	76	110	125	178
		2017	74	68	69	78	87	89

We then use the Portrait Divergence metrics to compare the network structures of gender and socioeconomic groups [142]. Portrait Divergence quantifies the similarity between two network structures, varying its values between 0 and 1. Values closer to 0 indicate similarities between network structures, while values closer to 1 indicate different network structures.

Mathematically, the Portrait Divergence compares networks represented as portraits. A portrait comprises the information about the network structure into a matrix, B. Each element, $B_{l,k}$, of this B-matrix is the number of nodes k at a distance l. For the case of weighted networks, the distance takes into account the weight using a binning strategy. Then, the portrait divergence computes the distance between the portraits using the Jensen-Shannon (JS) divergence [143]. JS computes the similarity of two probability distributions that in the network context it turns into:

$$PD_{JS}(G_1, G_2) = \frac{1}{2}KL(P(G_1)||M) + \frac{1}{2}KL(P(G_2)||M), \tag{3.1}$$

where,

$$M = \frac{1}{2}P(G_1) + P(G_2), \tag{3.2}$$

and,

$$P(G_1) = \frac{kB_{l,k}}{N^2}. \tag{3.3}$$

where G_1, G_2 are the two networks being compared, and KL is the Kullback-Leibler (KL) divergence [144] which computes the distance between distributions. Using PD, we can compute how much the mobility network for each sociodemographic group can be different.

In Table 3.5, we display the values of Portrait Divergence (PD) when comparing the networks from the mobility of each gender and socioeconomic group. We also generate a null model to compare the magnitude of PD expected to emerge from the network of the population if we randomly assign the gender and socioeconomic status to the individuals. We repeat the null model 1000 times to extract the average PD value from randomly assigned groups in our data. The values are shown in Table 3.5 inside parenthesis.

For the gender groups, we would say, in general, that the network has similar network structures because both genders have similar residential distributions over the areas (Spearman correlation higher than 0.50 when comparing the probability of women and men living in the zones). However, we see that the values of PD are higher than those of the null model, indicating that gender may affect the macro-level organisation of mobility. We argue that if we had a higher resolution of data, the

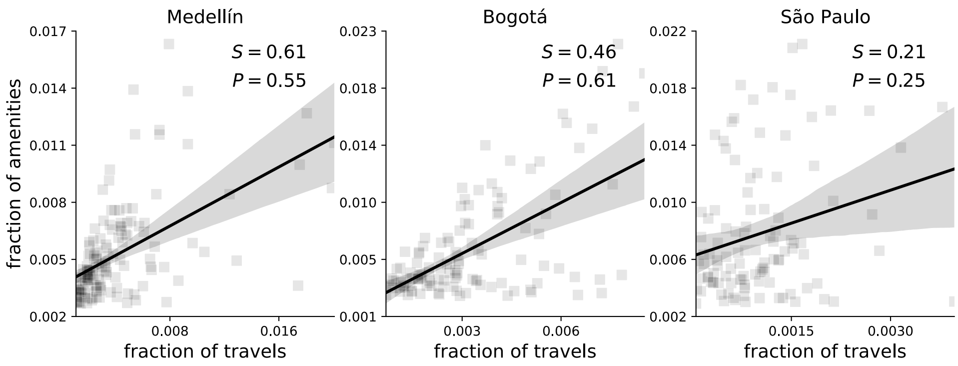

Figure 3.2: Relationship between fraction of travels and fraction of amenities for the zones in Medellín, Bogotá and São Paulo. We show the Spearman S and Pearson P correlations between the fraction of travels and amenities for each city.

difference could be more evident in the network because we could have a more detailed trajectory of individuals allowing us to distinguish locations inside zones such as home and work. In fact, Alessandretti et al. [41] showed that women's mobility is better characterised within smaller areas than for men, and women also move more between areas. Therefore, flexible spatial partitioning might uncover a deeper understanding of mobility across gender groups.

Moreover, using our data, we show that some gender differences are hidden in the network (Chapter 4). Nevertheless, one interesting point is that the urban landscape for different gender groups plays a major role in mobility. We observe that a high fraction of travels is concentrated in areas that have a high fraction of amenities related to transport, work, education and leisure (estimated using OpenStreetMap data 2021 as shown in Figure 3.2). In Chapter 4, we further investigate how women and men are exposed to amenities.

We observe that the values of PD are close to 1 for comparing the networks across socioeconomic groups, indicating that travellers organise their mobility uniquely across groups. We also notice that the values of the null models are much smaller than the average PD values extracted from our original data. This indicates that the differences between the network structures are not a byproduct of random processes, and that gender and socioeconomic dimensions may play a role in how mobility is organised.

For the case of gender, we also extract more information about mobility. Table 3.6 displays how men are more likely to have a higher fraction of travels in the

Table 3.5: Portrait Divergence between the mobility network from travellers of each gender (`women`,`men`) and socioeconomic (`lower`,`middle`,`upper`) group, $PD(X_{i=1}, X_{i=2})$, where $X_i \epsilon \{$`women`, `men`, `lower`, `middle`, `upper` $\}$. The values extracted from the null models are in parentheses. The null models shuffle the gender and socioeconomic categories.

Region	Year	PD(`women`,`men`)	PD(`lower`,`middle`)	PD(`lower`,`upper`)	PD(`middle`,`upper`)
`MDE`	2005	0.3500 (0.1764)	0.7082 (0.3628)	0.9635 (0.6103)	0.8943 (0.6119)
	2017	0.4126 (0.2352)	0.7779 (0.4199)	0.9646 (0.5768)	0.9071 (0.5876)
`BGT`	2012	0.4521 (0.2084)	0.8269 (0.4899)	0.9667 (0.6679)	0.9121 (0.6720)
	2019	0.5526 (0.2314)	0.8856 (0.5683)	0.9755 (0.7514)	0.9363 (0.7526)
`SAO`	1997	0.3676 (0.2128)	0.4627 (0.2274)	0.8494 (0.3868)	0.7774 (0.3738)
	2007	0.2111 (0.2468)	0.2764 (0.2909)	0.6678 (0.3658)	0.5165 (0.3256)
	2017	0.1721 (0.2637)	0.2751 (0.3295)	0.6565 (0.4019)	0.5165 (0.3416)

majority of the zones for overall and work travels, but for `nonwork` related travels, we observe the opposite. Moreover, the differences seen in the network are not solely coming from mobility differences within zones or residential areas, as gender groups have similar likelihoods (Table 3.7).

Table 3.6: Percentage of areas for which the fraction of travels performed by men is higher than the same quantity computed for women for `all`, `work` and `nonwork` travels ($P^M_{all,area} > P^W_{all,area}$, $P^M_{work,area} > P^W_{work,area}$, $P^M_{nonwork,area} > P^W_{nonwork,area}$).

City	Year	$P^M_{all,area} > P^W_{all,area}$	$P^M_{work,area} > P^W_{work,area}$	$P^M_{nonwork,area} > P^W_{nonwork,area}$
`MDE`	2005	61.46%	79.67%	43.20%
	2017	62.67%	91.12%	36.40%
`BGT`	2012	28.21%	63.79%	21.55%
	2019	26.84%	77.09%	15.02%
`SAO`	1997	63.88%	87.20%	39.84%
	2007	46.85%	86.12%	26.77%
	2017	50.00%	86.69%	34.00%

Table 3.7: Percentages of the travels for which the origin and destination zones, P^X, are the same: the percentages of travels performed by `all` (A), `men` (M) and `women` (W) in all travels (P^X_{all}), the percentages of work travels performed by `all` (A), `men` (M) and `women` (W) (P^X_{work}), and the percentages of work travels performed by `all` (A), `men` (M) and `women` (W) at the same zone that travellers live ($P^X_{live=work}$).

City	Year	P^A_{all}(%)	P^M_{all}(%)	P^W_{all}(%)	P^A_{work}(%)	P^M_{work}(%)	P^W_{work}(%)	$P^A_{live=work}$(%)	$P^M_{live=work}$(%)	$P^W_{live=work}$(%)
`MDE`	2005	18.34	17.52	19.20	1.71	1.90	1.26	7.76	8.02	7.38
	2017	18.99	16.79	21.62	14.39	24.31	15.73	22.04	26.41	17.32
`BGT`	2012	28.43	25.34	30.94	1.92	2.60	1.51	13.72	14.37	12.92
	2019	1.16	1.19	1.14	1.71	2.06	1.58	10.81	10.54	11.15
`SAO`	1997	43.83	41.05	46.90	12.45	13.99	10.75	23.31	21.83	25.73
	2007	37.77	35.45	40.15	11.53	12.03	11.01	20.23	18.66	22.32
	2017	38.69	37.75	39.63	12.49	13.91	11.07	20.63	20.23	21.11

Lastly, we show in Table A.4 the main network properties for the mobility of each socioeconomic group for women and men travellers. We can not see any systematic pattern combining gender and socioeconomic dimensions from the mobility patterns. Even though we see that women and men from the upper-income group tend to show similar properties in their mobility network, the network structures are different using Portrait Divergence (Table A.5). Thus, we argue that women and men from each socioeconomic group have different network structures from their mobility, but these differences in the structure vary across cities and years.

3.4 Temporal patterns

We now focus our attention on another aspect of mobility: temporal. In fact, people typically have a daily routine, following a circadian rhythm, to fulfil their responsibilities such as commuting and returning home [3, 145]. Therefore, movements are repeated at different hours of the day. Outcomes can emerge from the temporal concentration of mobility, such as traffic jams [146]. This section shows how travels performed by gender and socioeconomic groups are organised with respect to time.

We first analyse in detail the temporal characteristics of mobility for women and men in `SAO`. We group the travels into work-related (`work`), not work-related (`nonwork`) and overall (`all`) travels and analyse the fraction of travels per hour. Figure 3.18 shows the fraction of travels per each gender $X = \{\texttt{men}\ (M), \texttt{women}\ (W)\}$, and the differences between the fraction of travels between gender groups $F_t^M - F_t^W$. We observe that women, in general, have a higher fraction of travels at hours close to midday, and men tend to have higher mobility than women at the beginning and end of the day. This might be related to the fact that women feel safer at this time of the day in Latin-American countries [28]. Moreover, both genders are more likely to travel around three specific hours (peaks in the Figure 3.18): 6h, 12h and 17h with one hour of deviation. Interestingly, these hours tend to be the same across countries suggesting similar mobility regularities.

When we move our attention to commuting (`work` travels), we identify in Figure 3.3 that the temporal signature of the fraction of travels changes to one accentuated peak. We see that men are more likely to reach this peak before women. We also notice a second small peak after the time that may be devoted to lunch or to pick up (and leave) kids at school in Latin-American countries.

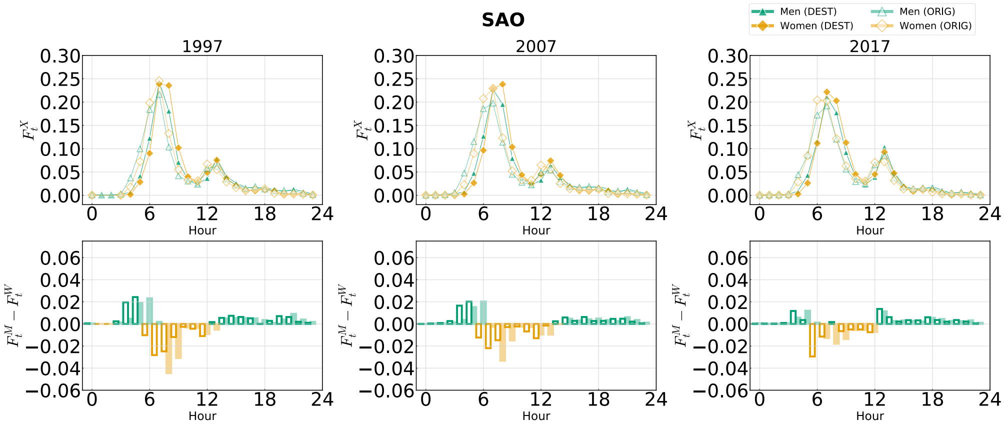

Figure 3.3: Fraction of travels F_t by each gender X ={`men` (M), `women` (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the `work` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

Finally, when we look at the `nonwork` travels in Figure 3.4, we see that the three peaks from Figure 3.18 come mainly from the `nonwork` travels because it combines a range of activities: health, going home, shopping, leisure, and studying. For trips related to studying, the temporal signature is also characterised by three peaks, and the gender differences become slightly higher at specific times of the day than for other purposes of travel in Figure 3.5. For the purpose of going home, the temporal signature shapes two accentuated peaks in the midday and evening in Figure 3.6, and the gender differences look like the ones from the overall mobility, even though the temporal signatures are different.

Regardless of the sociodemographic component, the differences in the temporal signature for each travel purpose show that mobility differences vary depending on the context of individuals' journeys. More relevant to this thesis, the purposes of travels affect the magnitude of gender differences in how women and men temporally organise their travels. At the same time, we also observe that specific purposes of travel tend to have a higher fraction of women travelling around midday, similar to the overall mobility.

We now analyse how systematic women and men temporally organise their travel across cities. Figures 3.7 and 3.8 show the fraction of `work` travels over time for several regions in the four countries. Indeed, we see a systematic tendency of

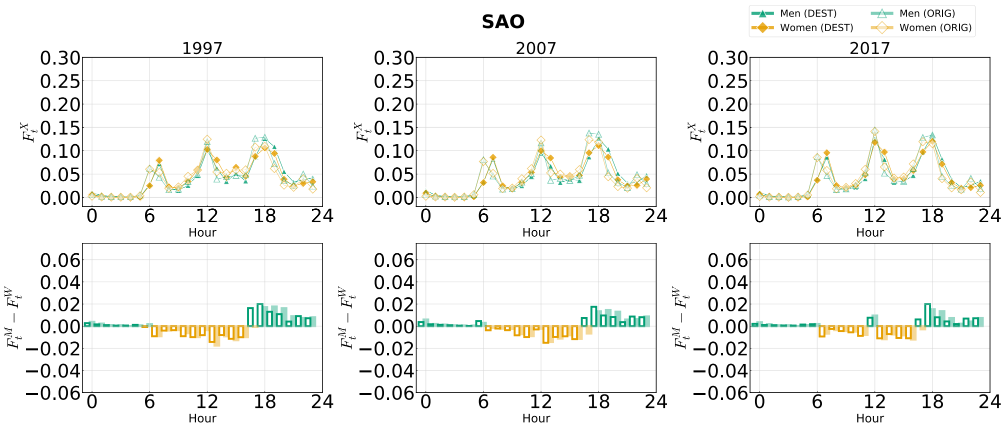

Figure 3.4: Fraction of travels F_t by each gender X ={men (M), women (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the nonwork travels in SAO and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

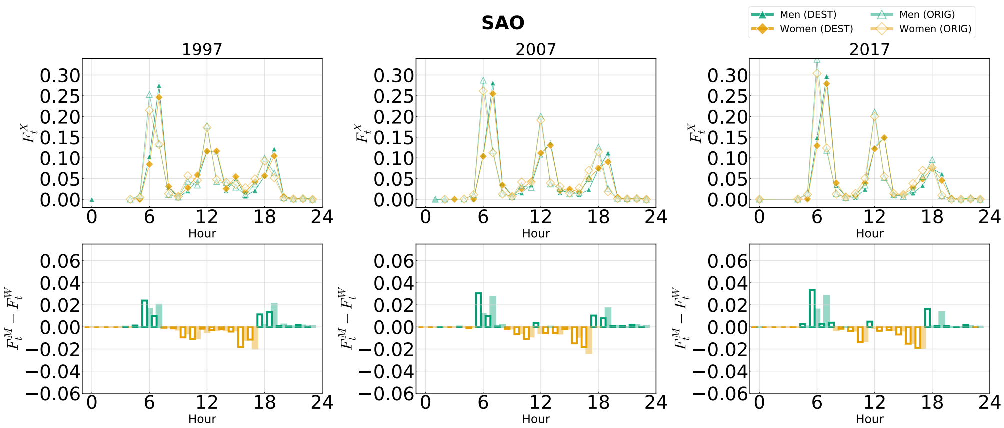

Figure 3.5: Fraction of travels F_t by each gender X ={men (M), women (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the study travels in SAO and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

men leaving for work earlier than women, and consequently, arriving at work earlier.

Turning now to the temporal signatures of the mobility from travellers belonging to different socioeconomic groups. We observe in Figure 3.11 that the upper income group tends to have a higher fraction of travels from 7h to 10h and 13h to 16h than the other groups. However, lower and middle income groups tend to have a higher fraction of travels than the upper income group from 4h-6h, 11h-12h, and 16h-18h. When we compare the lower income group with the middle income group, we see that the lower income group is more likely to travel earlier, and the middle income group is more likely to travel at night. In general, the three-peaked shape is what describes the temporal fraction of travels in SAO.

Looking at the commuting patterns in Figure 3.12, we see a presence of one

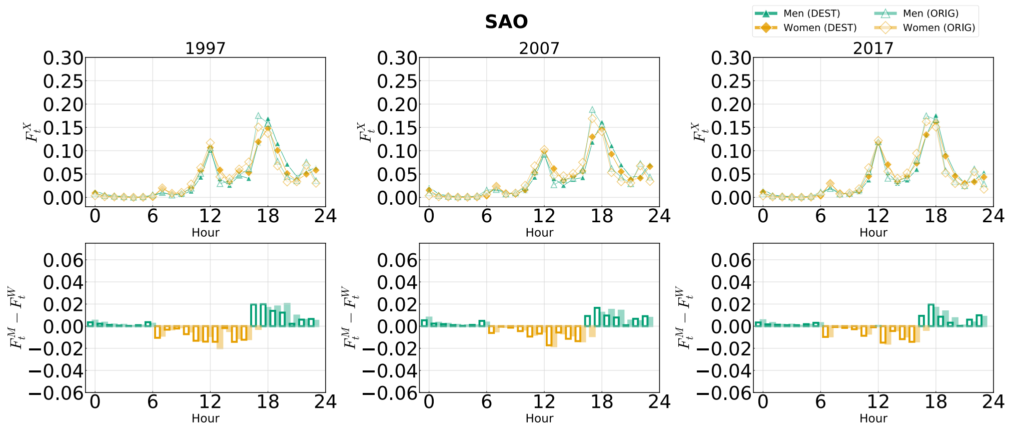

Figure 3.6: Fraction of travels F_t by each gender X ={`men` (M), `women` (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the travels related to going home in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

higher peak as the temporal signature. The lower income group is more likely to go to work earlier than the other groups, and individuals in the upper income group are more likely to concentrate their travels from 7h to 15h. However, we observe that the temporal signatures for other cities, visualised in Figure 3.16, are not as consistent in the socioeconomic dimension as they are in the gender dimension.

For `nonwork` travels, in Figure 3.13, we observe three peaks that are less-accentuated than the overall mobility. Middle and upper income groups are more likely to travel at night than the lower income group, and a higher fraction of individuals from the lower income group seem to travel in the evening. Looking at specific purposes of travel such as going home and studying (Figures 3.15 and 3.14), we see similar trends in the socioeconomic differences as we did in the overall mobility, but with variances in the magnitude. When controlling for the socioeconomic dimension, the gender differences in the fraction of travels persist, though the magnitudes of such differences are not consistently affected by the socioeconomic dimension. Taking into account the socioeconomic dimension might not impact the gender differences of the fraction of travels though it impacts slightly the magnitude of these differences (Figure 3.17).

å

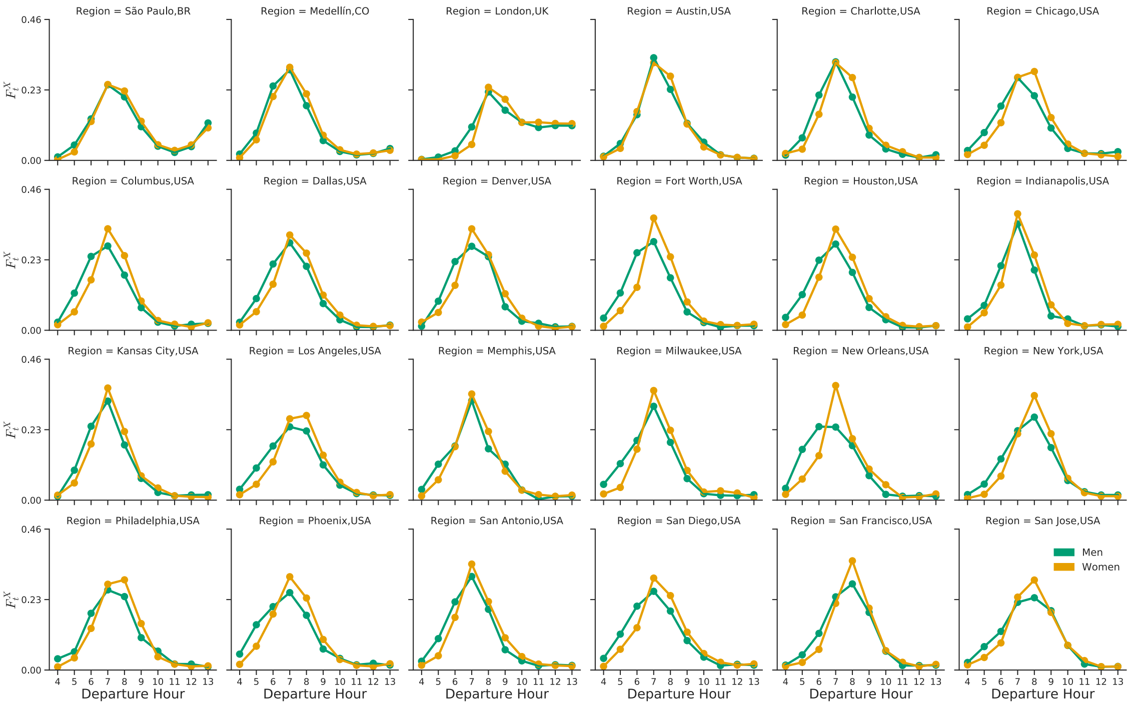

Figure 3.7: Fraction of travels F_t by each gender X ={`men` (M), `women` (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the `work` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

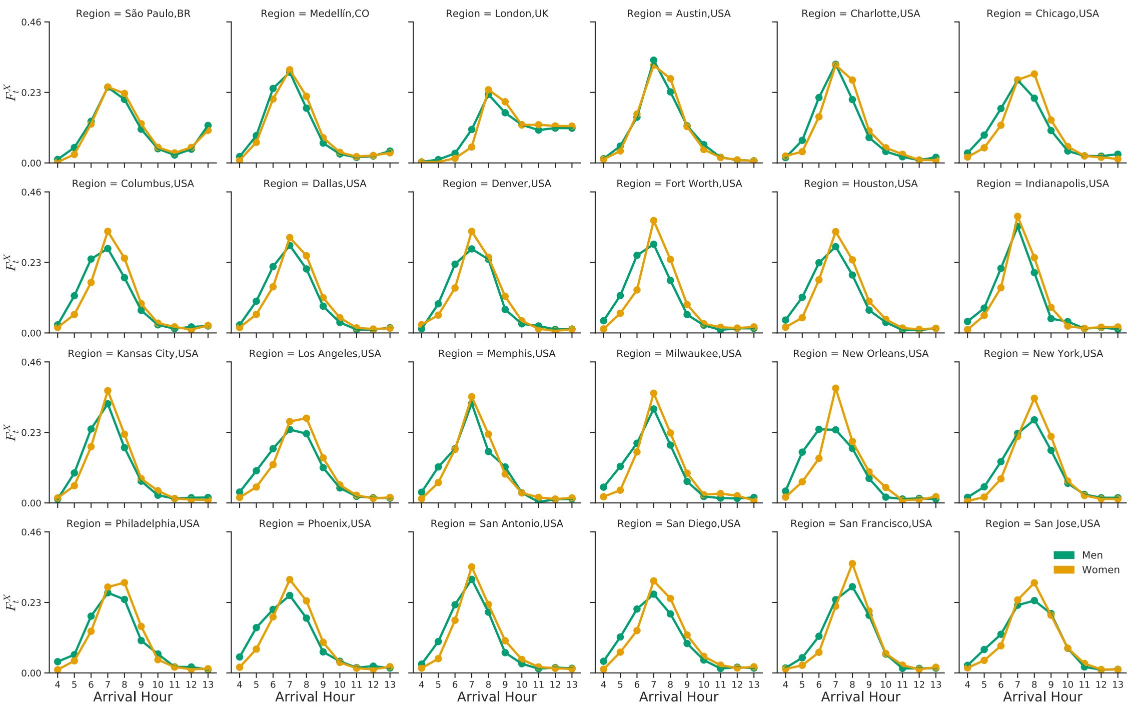

Figure 3.8: Fraction of travels F_t by each gender X ={`men` (M), `women` (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the `work` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

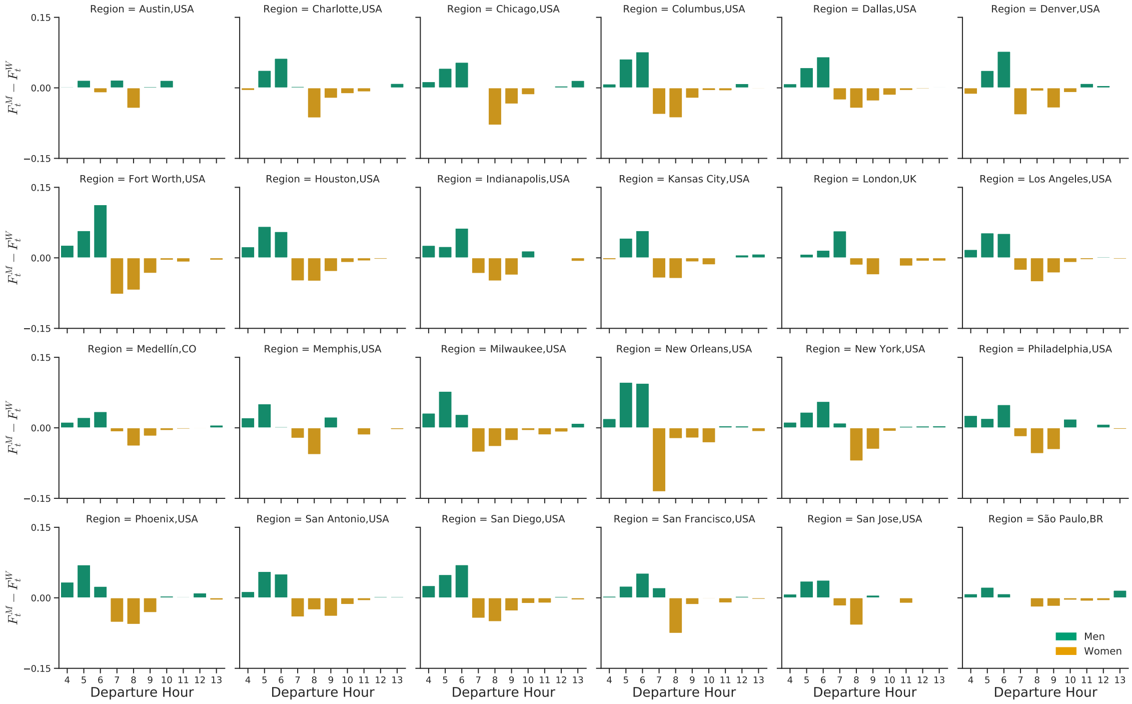

Figure 3.9: Gender differences $F_t^M - F_t^W$ (bar plots) of the fraction of travels considering the departure time of the `work` travels in `SAO`.

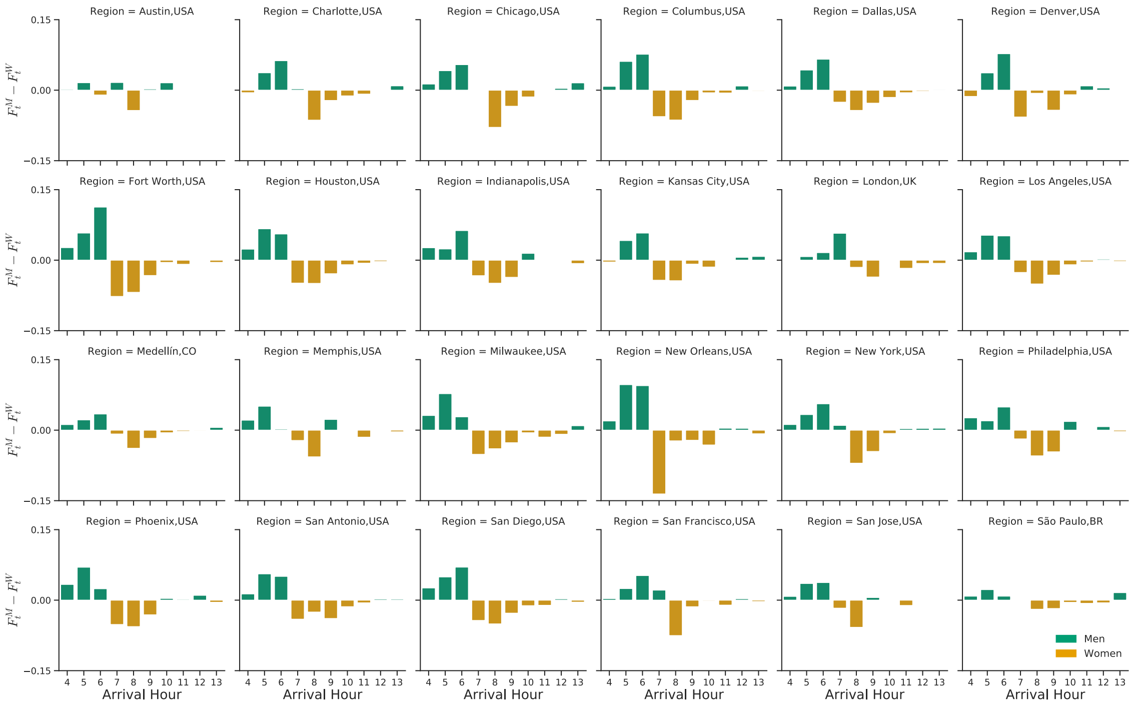

Figure 3.10: Gender differences $F_t^M - F_t^W$ (bar plots) of the fraction of travels considering the arrival time of the `work` travels in `SAO`.

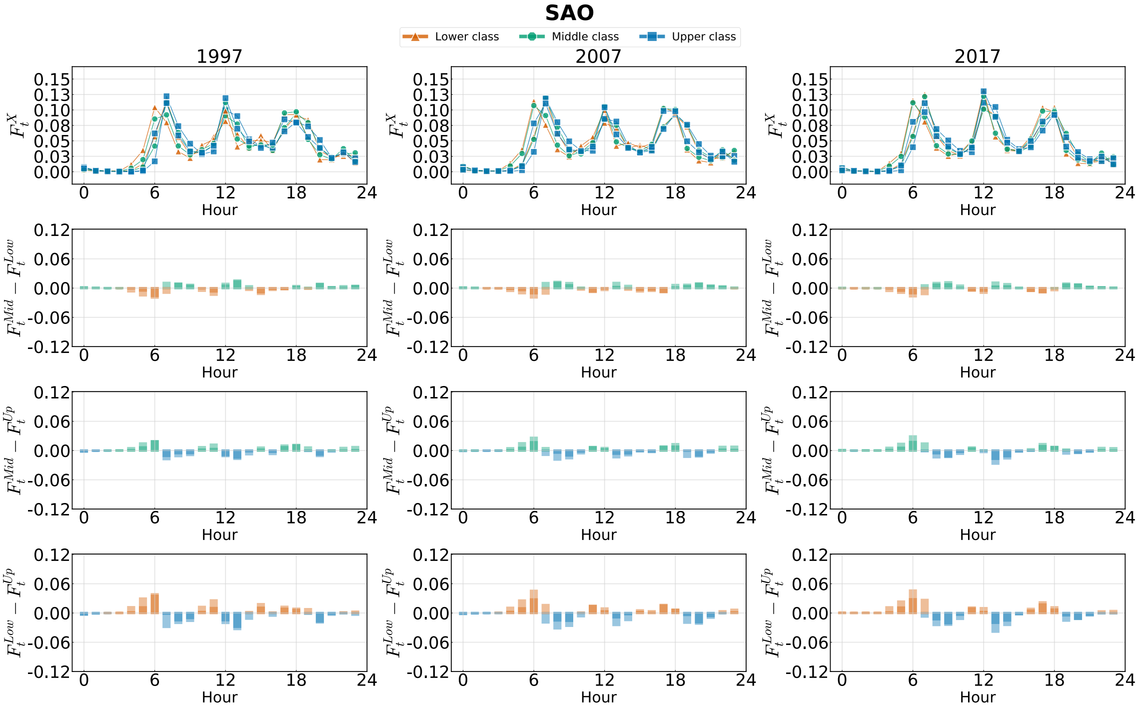

Figure 3.11: Fraction of travels F_t by each socioeconomic group X (line plots), and the socioeconomic differences $F_t^{X1} - F_t^{X2}$ (bar plots), of the `all` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

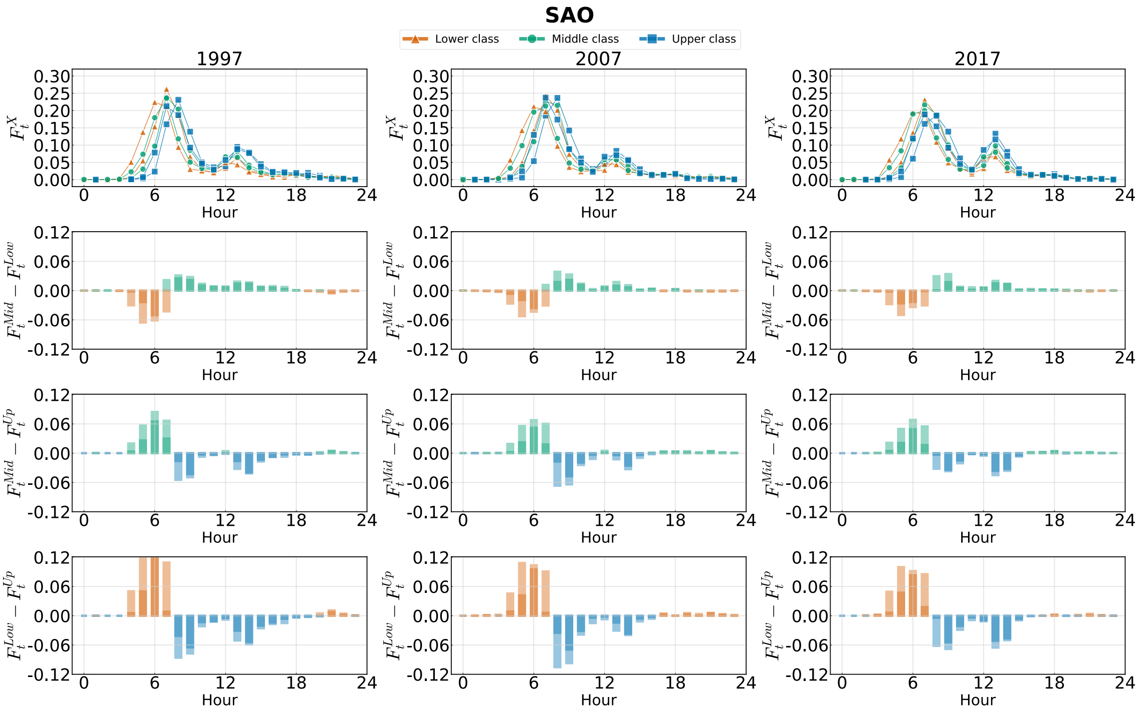

Figure 3.12: Fraction of travels F_t by each socioeconomic group X (line plots), and the socioeconomic differences $F_t^{X1} - F_t^{X2}$ (bar plots), of the `work` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

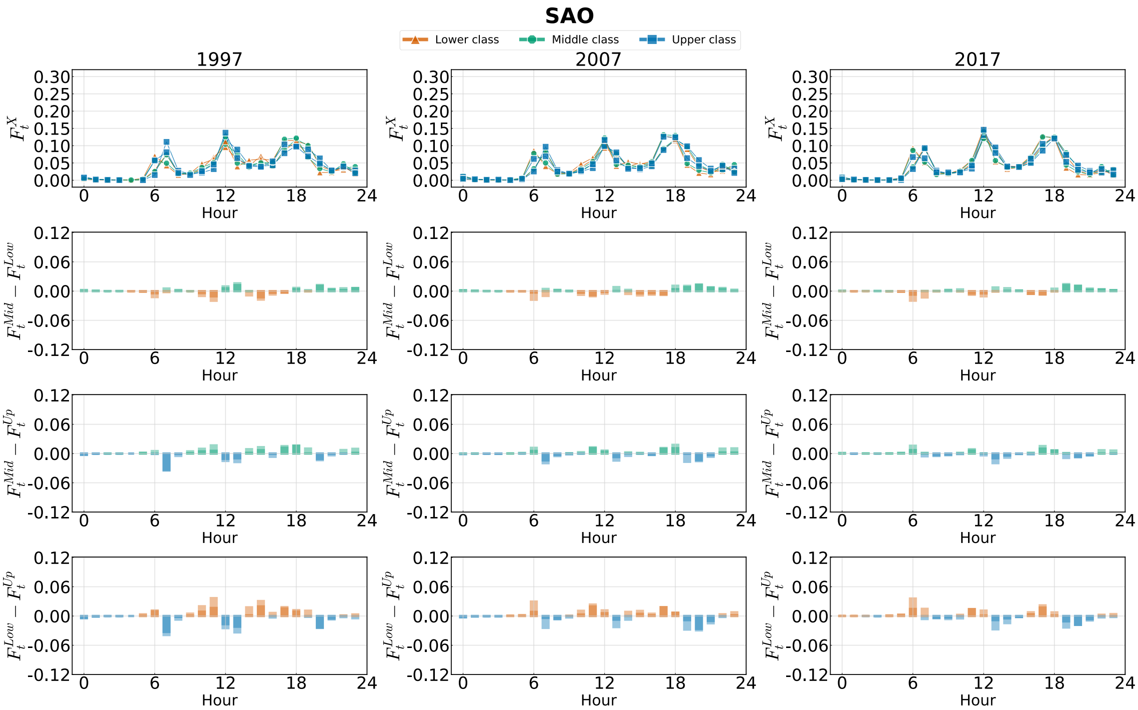

Figure 3.13: Fraction of travels F_t by each socioeconomic group X (line plots), and the socioeconomic differences $F_t^{X1} - F_t^{X2}$ (bar plots), of the `nonwork` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

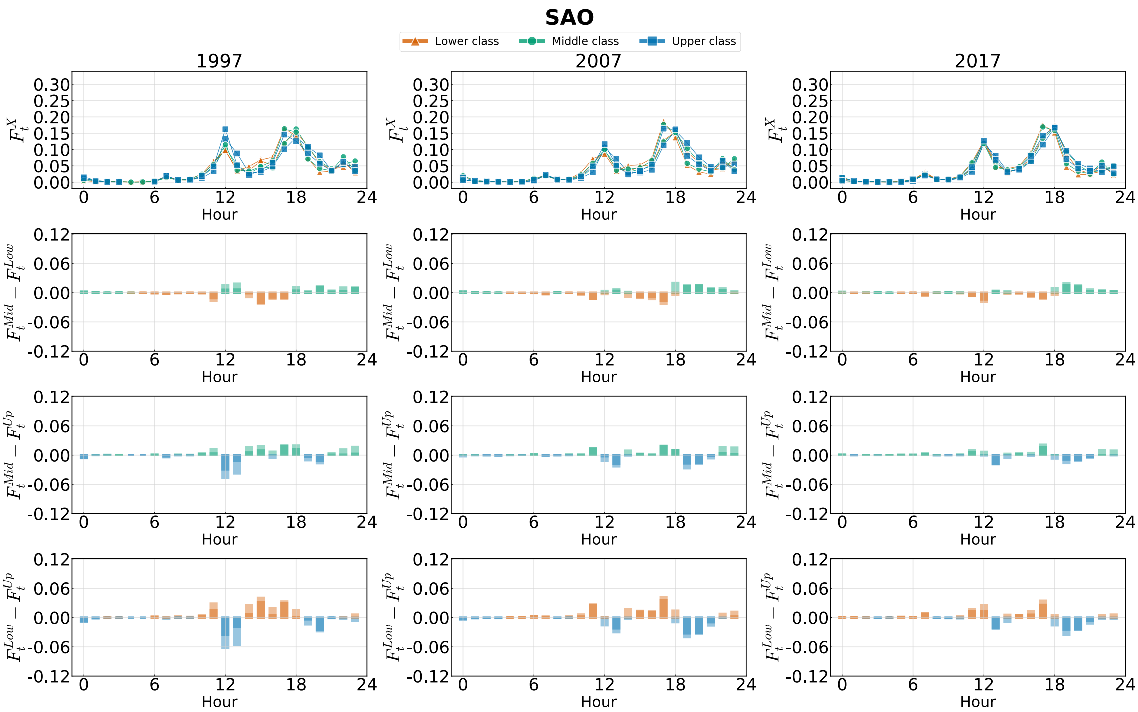

Figure 3.14: Fraction of travels F_t by each socioeconomic group X (line plots), and the socioeconomic differences $F_t^{X1} - F_t^{X2}$ (bar plots), of the travels going to home in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

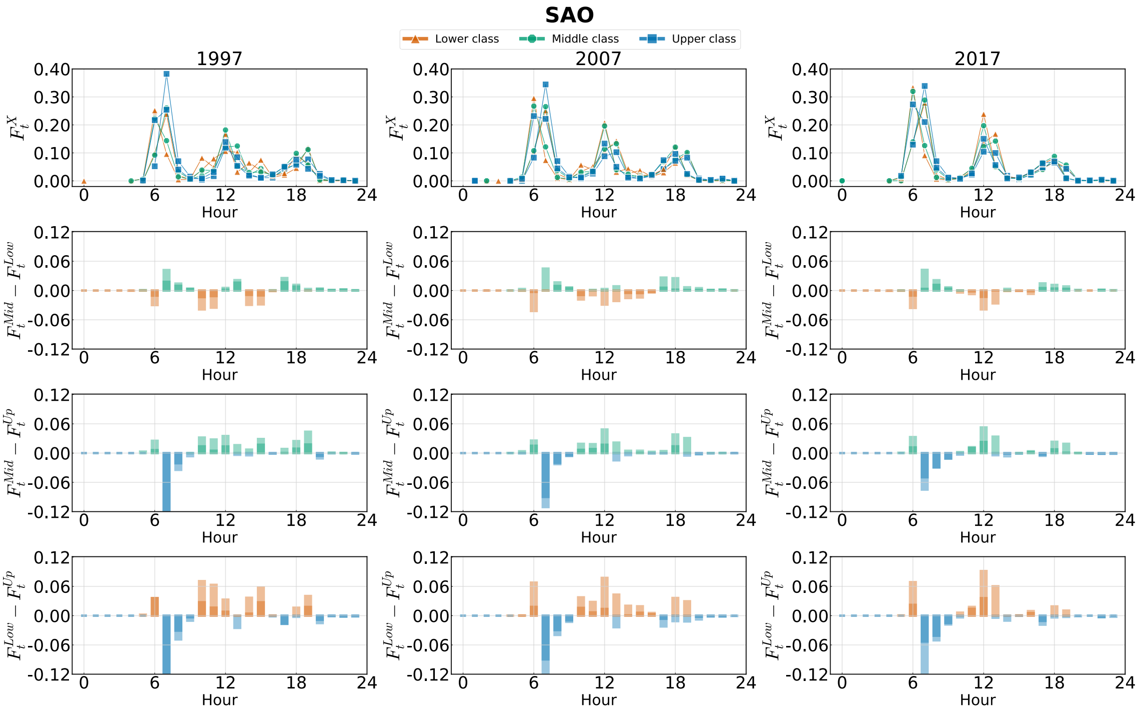

Figure 3.15: Fraction of travels F_t by each socioeconomic group X (line plots), and the socioeconomic differences $F_t^{X1} - F_t^{X2}$ (bar plots), of the study travels in SAO and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

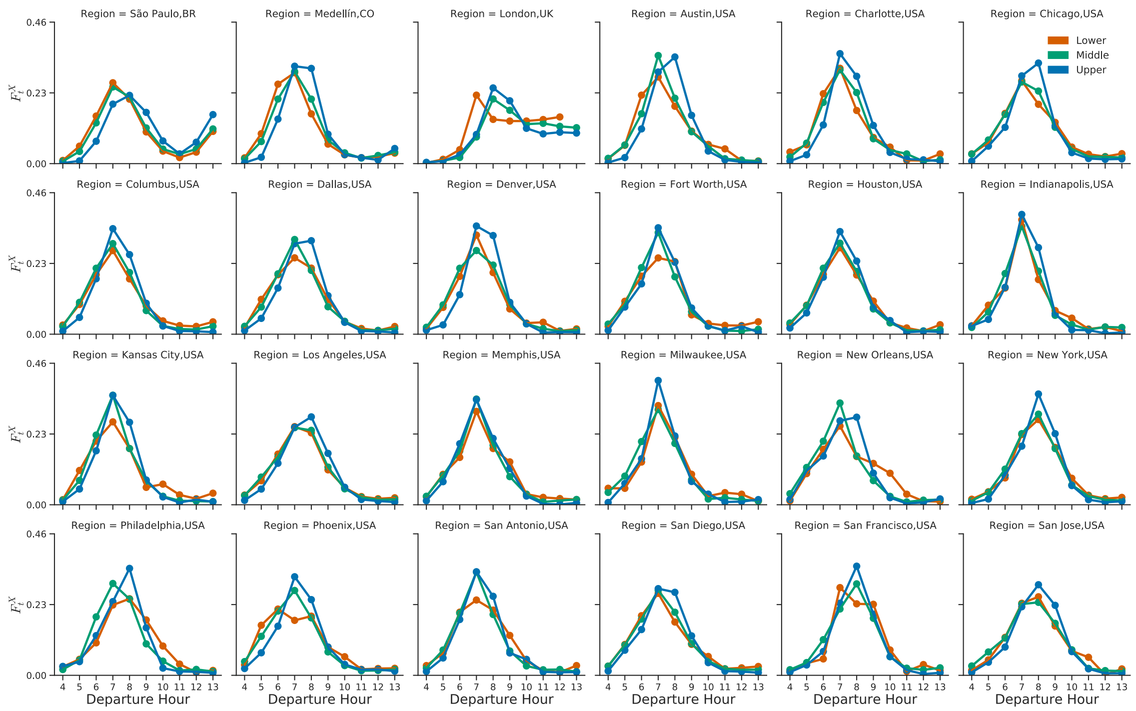

Figure 3.16: Fraction of travels F_t by each gender X (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the work travels in SAO and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

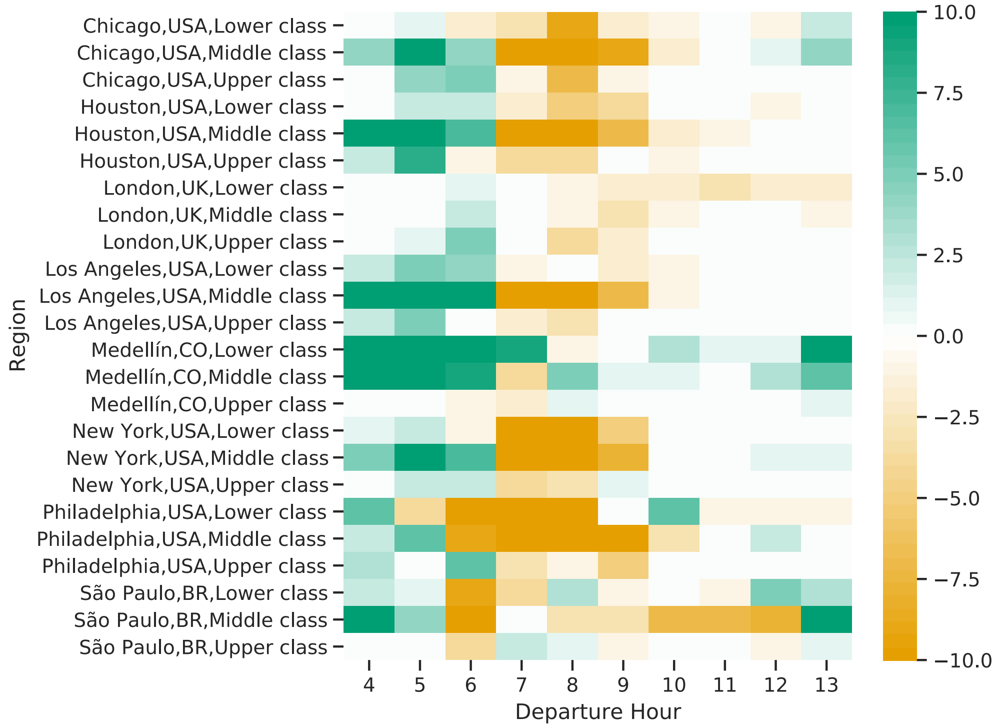

Figure 3.17: Gender differences from the fraction of travels, $F_t^M - F_t^W$, of the `work` travels and their aggregated case with respect to the departure time in hours.

3.5 Spatio-Temporal patterns

We then analyse how mobility is shaped over space and time simultaneously and how both space and time contribute to amplifying mobility's gender and socioeconomic differences. For the case of urban mobility networks, we notice small differences in macro-level metrics, such as density and clustering, for the overall networks of different gender groups. However, these small differences are amplified when we separate the network into morning, midday and afternoon. From Figure 3.18, we see that there are three peaks in the mobility. Thus, using `MDE` data, we look at the mobility network for each peak, detailed in Figure 3.19. Some areas that men highly visit during different periods of the day are not the places that women commonly visit. The distribution of visited areas is not the same for women and men (using Kolmogorov Smirnov (KS) test with a confidence interval of 99%), and almost 70% of the zones are more likely to have a higher fraction of travels of men than women. In this way, we post the question: do women and men explore urban areas differently? Chapter 4 will address this question in detail.

We then compare the network structures for gender and socioeconomic groups

Table 3.8: Portrait Divergence between the mobility network disaggregated by morning, midday and evening from travellers of each gender ($\texttt{women}, \texttt{men}$) and socioeconomic ($\texttt{lower}$,$\texttt{middle}$,$\texttt{upper}$) group, $PD(X_{i=1}, X_{i=2})$, where $X_i \epsilon \{\texttt{women}, \texttt{men}, \texttt{lower}, \texttt{middle}, \texttt{upper}\}$. The values extracted from the null models are in parentheses. The null models shuffle the gender and socioeconomic categories.

City	Year	Period	PD(women,men)	PD(lower,middle)	PD(lower,upper)	PD(middle,upper)
MDE	2005	Morning	0.8694 (0.4803)	0.9343 (0.4926)	0.9866 (0.6996)	0.9898 (0.6975)
		Midday	0.8455 (0.4526)	0.9511 (0.4618)	0.9870 (0.6630)	0.9861 (0.6649)
		Evening	0.8418 (0.4852)	0.9395 (0.4945)	0.9887 (0.7183)	0.9871 (0.7192)
	2017	Morning	0.9891 (0.4764)	0.9900 (0.49283)	0.9976 (0.6213)	0.9846 (0.6299)
		Midday	0.9889 (0.4488)	0.9643 (0.4665)	0.9980 (0.6176)	0.9880 (0.6293)
		Evening	0.9611 (0.5630)	0.9739 (0.5790)	0.9974 (0.7293)	0.9886 (0.7402)
BGT	2012	Morning	0.9428 (0.5616)	0.9918 (0.5698)	0.9937 (0.7235)	0.9953 (0.7259)
		Midday	0.9487 (0.5059)	0.9889 (0.5142)	0.9985 (0.7038)	0.9962 (0.7091)
		Evening	0.9730 (0.6161)	0.9920 (0.6256)	0.9982 (0.7733)	0.9955 (0.7771)
	2019	Morning	0.7350 (0.2204)	0.8261 (0.2283)	0.9898 (0.5616)	0.9961 (0.5661)
		Midday	0.9300 (0.2449)	0.9005 (0.2535)	0.9960 (0.6011)	0.9950 (0.6079)
		Evening	0.9929 (0.2571)	0.9962 (0.2674)	0.9966 (0.6032)	0.9983 (0.6096)
SAO	1997	Morning	0.8100 (0.3240)	0.8550 (0.3464)	0.9870 (0.5047)	0.9767 (0.4855)
		Midday	0.6298 (0.2442)	0.8198 (0.2607)	0.9808 (0.4112)	0.9515 (0.3958)
		Evening	0.7525 (0.3172)	0.9151 (0.3369)	0.9801 (0.5031)	0.9588 (0.4867)
	2007	Morning	0.7425 (0.3585)	0.7900 (0.4208)	0.9877 (0.4922)	0.9702 (0.4368)
		Midday	0.8360 (0.2812)	0.8488 (0.3260)	0.9653 (0.4020)	0.9463 (0.3591)
		Evening	0.8341 (0.3521)	0.8162 (0.4084)	0.9695 (0.4832)	0.9276 (0.4334)
	2017	Morning	0.6460 (0.3761)	0.9772 (0.4638)	0.9401 (0.5341)	0.9738 (0.4559)
		Midday	0.6899 (0.2732)	0.7825 (0.3443)	0.9715 (0.4142)	0.9452 (0.3473)
		Evening	0.7266 (0.3746)	0.9772 (0.4594)	0.9530 (0.5375)	0.9693 (0.4626)

using the Portrait Divergence. We see that the values of portrait divergence are all closer to 1 than 0, indicating that the network structures over the periods of time are different across gender and socioeconomic groups. We see that by disaggregating the network over time, we find higher values of portrait divergence than we did for the overall network. This indicates that time has a major effect on amplifying the differences in how travellers from various sociodemographic groups visit urban areas. Moreover, we see that the values computed for the null model — that shuffles the gender and socioeconomic label — are generally smaller than those computed from the correct labelling. This suggests that gender and socioeconomic factors contribute to the identified differences in network structure.

Even though the network structures across groups are different, we further analyse whether the highly visited areas (denoted hubs in network science) are similar across groups. To compare the highly visited areas across gender and socioeconomic groups, we rank the areas considering the fraction of travels per group and select the 50 most visited per group. Then, we compare the number of top 50 most visited areas that are shared between gender and socioeconomic groups, shown in Table 3.10.

We see a high number of highly visited areas in common between women and men, having 84% percentage of common highly visited areas. The time period only slightly affects the number of common highly visited areas, indicating that evening is the period with the highest values across cities and years.

When we look at the socioeconomic groups, we see that the number of common highly visited areas reaches a maximum of 66% (33 common areas out of a total of 50 areas). The percentage of common highly visited areas between `lower` and `upper` income groups are in general the lowest (maximum of 40%, or 20 areas), showing that, regardless of time, these two groups do not share a majority of highly visited areas. The period of time slightly affects the number of highly visited areas that are common across socioeconomic groups. In contrast with the comparison of gender groups, socioeconomic groups have the lowest values during the evening across the cities and years. We observe that socioeconomic groups have a few common highly visited areas in line with the finding that they spatially structure their travels differently over the urban areas (see Table 3.5).

To make sure that the results were not only impacted by residential segregation, we remove the travels that have the purpose of going home, and with the filtered travels, we redefine the top 50 most visited areas for each sociodemographic group. Table 3.10 presents the frequently visited areas that the different sociodemographic groups have in common. We see an effect on the values of common areas for both directions (increasing or decreasing) and generally the effect is in small magnitudes, indicating that highly visited areas are not common only because of the travels related to home. Removing travels to home seems to increase the number of common highly visited areas for socioeconomic groups, indicating that mobility in the three regions for other purposes concentrate more than the residential distribution. In fact, Table 3.12 shows that most travels do not have the destination and residential area as the same. For the case of gender, the number of common highly visited areas increase when we take into account mobility without the travels to home. This might be true because women are more likely to work closer to home than men [111]. In conclusion, the highly visited areas might be affected by residential segregation,

Table 3.9: Portrait Divergence between the mobility network for work-related purpose disaggregated by morning, midday and evening from travellers of each gender (`women`,`men`) and socioeconomic (`lower`,`middle`,`upper`) group, $PD(X_{i=1}, X_{i=2})$, where $X_i \epsilon${`women`,`men`, `lower`, `middle`, `upper` }. The values extracted from the null models are in parentheses. The null models shuffle the gender and socioeconomic categories.

City	Year	Period	PD(`women`,`men`)	PD(`lower`,`middle`)	PD(`lower`,`upper`)	PD(`middle`,`upper`)
`MDE`	2005	Morning	0.9543 (0.6261)	0.9601 (0.6350)	0.9948 (0.8262)	0.9901 (0.8231)
		Midday	0.9909 (0.7949)	0.9786 (0.8079)	0.9951 (0.8845)	0.9966 (0.8804)
		Evening	0.9936 (0.8732)	0.9786 (0.8764)	0.9976 (0.9165)	0.9948 (0.9194)
	2017	Morning	0.9923 (0.6663)	0.9873 (0.6781)	0.9972 (0.7862)	0.9854 (0.7924)
		Midday	0.9938 (0.8324)	0.9959 (0.8363)	0.9997 (0.9038)	0.9949 (0.9077)
		Evening	0.9846 (0.9467)	0.9650 (0.9433)	1.0000 (0.9869)	1.0000 (0.9867)
`BGT`	2012	Morning	0.9921 (0.8064)	0.9966 (0.8136)	0.9992 (0.8771)	0.9963 (0.8756)
		Midday	0.9974 (0.8375)	0.9992 (0.8450)	0.9983 (0.9040)	1.0000 (0.9009)
		Evening	0.9994 (0.9295)	0.9869 (0.9283)	1.0000 (0.9697)	0.9988 (0.9695)
	2019	Morning	0.9747 (0.4274)	0.9853 (0.4388)	0.9996 (0.7183)	0.9983 (0.7189)
		Midday	0.9985 (0.7227)	0.9887 (0.7333)	0.9996 (0.9014)	1.0000 (0.9014)
		Evening	0.9994 (0.8696)	0.9989 (0.8687)	0.9999 (0.9672)	0.9984 (0.9682)
`SAO`	1997	Morning	0.7982 (0.5045)	0.9815 (0.5316)	0.9904 (0.6997)	0.9849 (0.6770)
		Midday	0.9315 (0.5504)	0.9921 (0.5892)	0.9884 (0.7201)	0.9496 (0.6820)
		Evening	0.9957 (0.8171)	0.9769 (0.8240)	0.9988 (0.9129)	0.9936 (0.8955)
	2007	Morning	0.7916 (0.5367)	0.9028 (0.6198)	0.9866 (0.6888)	0.9818 (0.6168)
		Midday	0.9364 (0.6225)	0.9650 (0.7160)	0.9694 (0.7744)	0.9567 (0.6904)
		Evening	0.9775 (0.8478)	0.9914 (0.8844)	0.9893 (0.9357)	0.9956 (0.9031)
	2017	Morning	0.8576 (0.5624)	0.9938 (0.6753)	0.9852 (0.7365)	0.9842 (0.6403)
		Midday	0.9715 (0.5077)	0.9980 (0.6299)	0.9974 (0.6691)	0.9597 (0.5527)
		Evening	0.9946 (0.8801)	0.9982 (0.9278)	0.9968 (0.9602)	0.9947 (0.9211)

but this is not the only factor for making areas highly visited across gender and socioeconomic groups.

We see similar gender differences regardless of whether we account for socioeconomic status (similar to Table 3.10). We also observe that, regardless of gender, travellers from different socioeconomic statuses frequently visit different areas.

Instead of analysing `all` travels, we now analyse the `work` travels (Table 3.11). We draw the same conclusions that women and men tend to have a similar set of highly visited areas and that socioeconomic groups display a less similar set. However, we observe that for `work` travels, the number of common areas is higher than for `all` travels, in line with our finding in Chapter 5 that `work` travels concentrate in fewer areas than `all` travels.

In summary, we see consistent differences in the network structure of gender and socioeconomic groups over the three periods of time: morning, midday and evening. However, we observe slight differences in what are the most visited areas per gender. We argue that this is the case because the highly visited areas might reflect the concentration of urban amenities (e.g., residential, workplace, business

Table 3.10: Number of common top 50 most visited areas between gender (W, M) and socioeconomic (Low, Mid, Up) groups, A^X, where $X \epsilon \{W, M, Low, Mid, Up\}$. The values in parentheses display the effect of computing the same results for removing the mobility related to going home from the overall travels.

City	Year	Period	$\cap(A^W, A^M)$	$\cap(A^{Low}, A^{Mid})$	$\cap(A^{Low}, A^{Up})$	$\cap(A^{Mid}, A^{Up})$
MDE	2005	Morning	42 (+01)	33 (+00)	20 (+04)	29 (+03)
		Midday	37 (+01)	09 (+15)	03 (+10)	08 (+14)
		Evening	41 (–05)	05 (+24)	00 (+12)	03 (+15)
		All day	42 (+01)	33 (+00)	20 (+00)	29 (–01)
	2017	Morning	34 (+02)	18 (+00)	08 (+02)	21 (+03)
		Midday	37 (–08)	05 (+07)	01 (+01)	04 (+05)
		Evening	38 (–13)	04 (+11)	00 (+01)	02 (+01)
		All day	38 (–06)	18 (–02)	08 (–01)	21 (–02)
BGT	2012	Morning	31 (–01)	08 (+6)	02 (+03)	09 (+01)
		Midday	29 (–07)	06 (+00)	01 (+00)	02 (+07)
		Evening	33 (–19)	06 (+2)	01 (+01)	03 (+06)
		All day	36 (–06)	08 (+4)	02 (+08)	09 (+01)
	2019	Morning	36 (–03)	26 (–03)	12 (–01)	13 (+00)
		Midday	40 (–09)	25 (–03)	10 (+02)	11 (+01)
		Evening	33 (+07)	18 (–06)	13 (–01)	08 (+14)
		All day	41 (+02)	26 (+01)	13 (–01)	13 (+02)
SAO	1997	Morning	36 (+01)	22 (+2)	13 (+01)	27 (+00)
		Midday	41 (–01)	04 (+19)	13 (–05)	23 (–01)
		Evening	40 (–04)	23 (+04)	02 (+02)	12 (+08)
		All day	41 (–03)	23 (+01)	13 (–01)	27 (–02)
	2007	Morning	39 (–01)	29 (+00)	10 (+03)	21 (+05)
		Midday	39 (–10)	25 (–03)	06 (+00)	12 (+06)
		Evening	39 (–08)	31 (–08)	05 (+02)	12 (+07)
		All day	42 (–05)	31 (+06)	10 (–01)	21 (+00)
	2017	Morning	32 (–01)	24 (–01)	06 (+03)	18(+04)
		Midday	31 (+02)	21 (–06)	01 (+03)	10 (+07)
		Evening	36 (–07)	24 (–07)	00 (+08)	6 (+12)
		All day	36 (–01)	24 (+07)	06 (+00)	18 (+08)

and industry). Therefore, areas that have more opportunities will attract more people. However, we show in Chapter 4 that gender and socioeconomic status plays an important role in the mobility concentration over the urban areas.

3.6 Household arrangements

Two factors appear consistently in the literature as reinforcements of inequalities in gender: marital status and parenthood [111, 147–150]. These two factors seem to impact gender roles and exacerbate inequalities in the labour market [111, 147–150]. Therefore, we argue that these two factors might also impact the mobility of women and men.

This section explores whether household arrangements (marital status and parenthood) impact the urban mobility network for women and men. Specifically,

Table 3.11: Number of common top 50 most visited areas from the mobility related to work between gender (W, M) and socioeconomic (Low, Mid, Up) groups, A^X, where $X \epsilon \{W, M, Low, Mid, Up\}$.

City	Year	Period	$\cap(A^W, A^M)$	$\cap(A^{Low}, A^{Mid})$	$\cap(A^{Low}, A^{Up})$	$\cap(A^{Mid}, A^{Up})$
MDE	2005	Morning	41	39	24	34
		Midday	39	33	21	27
		Evening	27	18	13	20
		All day	42	42	24	34
	2017	Morning	27	24	19	33
		Midday	22	16	13	18
		Evening	07	08	02	03
		All day	27	24	19	33
BGT	2012	Morning	25	20	10	14
		Midday	14	10	04	12
		Evening	04	10	05	06
		All day	32	24	13	19
	2019	Morning	30	27	10	7
		Midday	24	21	12	10
		Evening	17	17	06	09
		All day	35	27	12	13
SAO	1997	Morning	35	30	23	27
		Midday	34	20	13	21
		Evening	21	17	10	18
		All day	35	31	23	27
	2007	Morning	34	31	22	32
		Midday	25	19	12	22
		Evening	20	13	10	15
		All day	34	33	22	32
	2017	Morning	34	27	14	27
		Midday	29	18	12	22
		Evening	14	09	11	09
		All day	36	28	16	29

we build a network from the mobility of travellers that are: (i) `single`; (ii) `married`; (iii) `parent`; and (iv) `without children`. We then use the portrait divergence to compare the network structures and compute the average portrait divergence from the network built from the shuffled gender labels in the dataset.

We see in Table 3.13 that the network structures exhibit gender differences for those that are `married`, `single`, `parent`, and `married parent`. Besides, we observe that the values of PD are higher for the network structure between women and men that are married parents. We see that when we disaggregated the data by household arrangements, the gender effect is more clear than comparing the network structure from the travels made by all men and women (Table 3.5). We also observed that the values for the null models shown in Table 3.13 are smaller than the values computed for our data, indicating that the combination of gender and household arrangement impacts the mobility network structure. We are not able to identify any further temporal impact within the social dimensions because the PD values are

Table 3.12: Percentages of the travels made for *all* purposes performed by traveller of type X (i.e. `all` (A), `men` (M) and `women` (W)) having as destination zone the same zone where the traveller lives, $P^{X}_{dest=live}$. Column $P^{X}_{live=work}$ denotes the same quantity computed for *work* travels.

City	Year	$P^{A}_{dest=live}(\%)$	$P^{M}_{dest=live}(\%)$	$P^{W}_{dest=live}(\%)$	$P^{A}_{live=work}(\%)$	$P^{M}_{live=work}(\%)$	$P^{W}_{live=work}(\%)$
MDE	2005	16.31	15.58	17.11	07.76	08.02	07.38
	2017	17.07	15.01	19.54	22.04	26.41	17.32
BGT	2012	18.85	16.99	20.37	13.72	14.37	12.92
	2019	04.84	04.35	05.34	10.81	10.54	11.15
SAO	1997	34.75	32.06	37.77	23.31	21.83	25.73
	2007	31.48	29.57	33.45	20.23	18.66	22.32
	2017	32.31	34.62	33.98	20.63	20.23	21.11

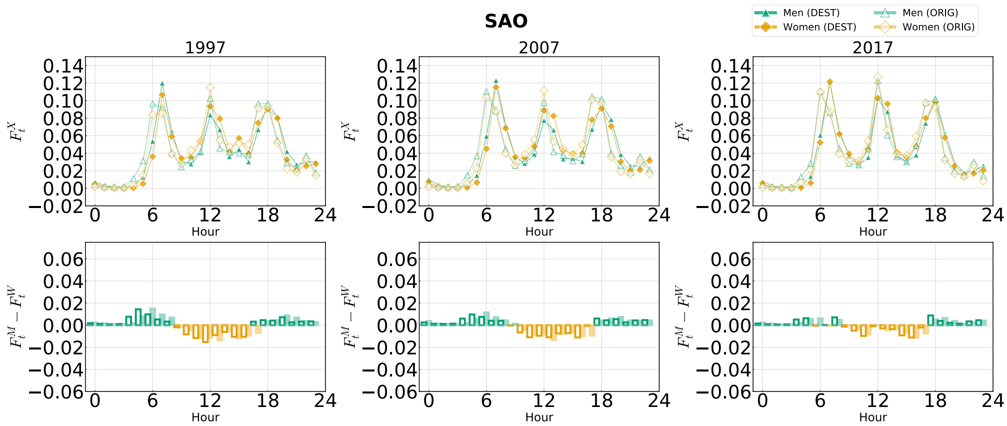

Figure 3.18: Fraction of travels F_t by each gender X ={`men` (M), `women` (W)} (line plots), and the gender differences $F_t^M - F_t^W$ (bar plots), of the `all` travels in `SAO` and their aggregated case with respect to the departure (empty symbols) and arrival (filled symbols) time in hours.

already extremely high when we just regard one of the dimensions (we cover this in Chapter 4).

3.7 Discussion

Cultural constructs and daily demands impose different roles on people [12, 124–127]. This chapter presented evidence that women and men reveal different spatio-temporal characteristics that emerge from their mobility across years and regions.

Specifically, we studied the main properties of the structural organisation of urban mobility for gender and socioeconomic groups. We showed that the network structure between women and men tend to be different. The differences in the network structure are amplified, taking into account the temporal and household arrangements. Therefore, the spatio-temporal organisation of the mobility of women

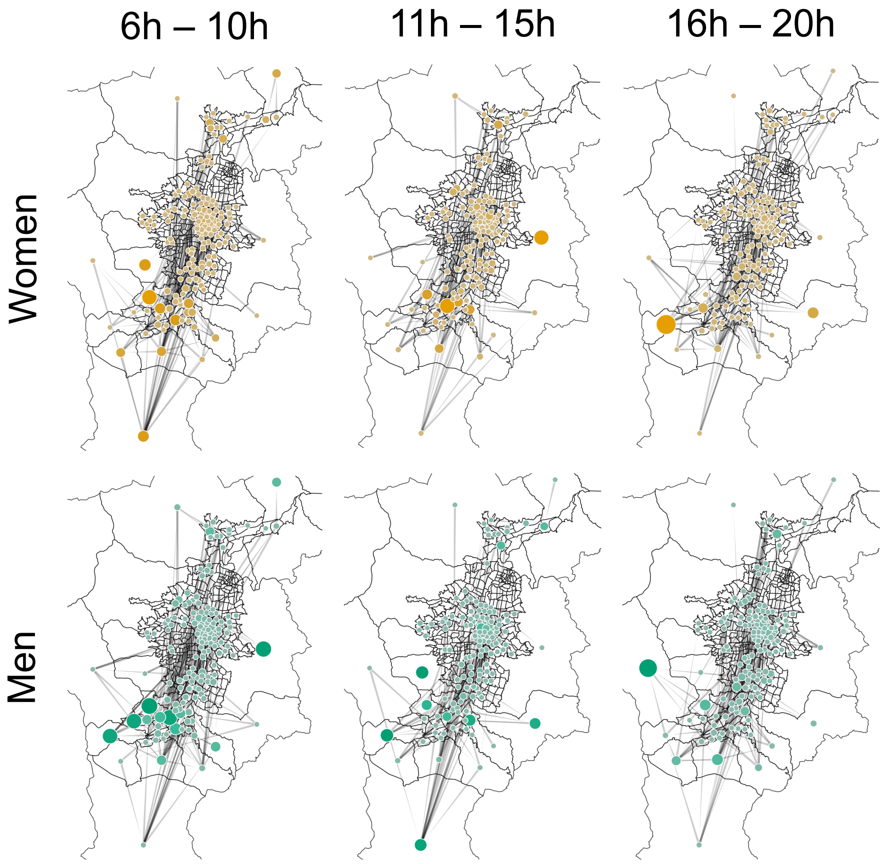

Figure 3.19: Network representation of gender mobility *flows* (i.e. number of travels) occurring during the morning (left column), midday (central column), and afternoon (right column). The nodes' size accounts for the in-strength of a zone (i.e. the sum of the weights of all edges entering a zone), while the edge thickness and colour accounts for the number of travels made between two zones. Data refer to the 2017's surveys.

and men are different and, as the literature points out, household arrangements impact the mobility of women and men differently [43, 45, 113]. We also presented that socioeconomic status alone plays a major role in the spatial organisation of urban mobility. Most of the highly visited areas tend to be the same for women and men, but this is not true for the socioeconomic groups. In particular, the lower and upper-income groups tend to have few highly visited areas in common. This finding might reflect the socioeconomic segregation of the studied cities — a well-known issue in urban areas [138, 151]. Moreover, in general, men tend to have a higher fraction of travels in the highly visited areas, which is in line with men being more likely to prefer larger cities than smaller ones, but in an urban context [152].

We demonstrated that the temporal patterns in the Latin American countries are similar to those observed in regions within the USA and the UK. However, we point out that some regions present more minor gender differences on the temporal

Table 3.13: Portrait Divergence between the mobility network from travellers of each gender (`women`,`men`) grouped by the following household arrangements: `married`, `single`, `parent` and `married parent`. We highlight in bold the maximum values reached for each row. The values extracted from the null models are in parentheses. The null models shuffle the gender category.

City	Year	PD(`married`)	PD(`single`)	PD(`parent`)	PD(`married parent`)
`MDE`	2005	0.8815 (0.5933)	0.7877 (0.3867)	0.6916 (0.4414)	**0.9404** (0.7175)
	2017	**0.9887** (0.4038)	0.8545 (0.4259)	0.9297 (0.4308)	0.8901 (0.7419)
`BGT`	2012	0.9612 (0.6533)	0.9759 (0.4936)	0.9595 (0.4730)	**0.9876** (0.7174)
	2019	0.9646 (0.2625)	0.1520 (0.1141)	0.8437 (0.2490)	**0.9874** (0.5178)
`SAO`	1997	**0.9911** (0.6856)	0.6209 (0.2150)	0.6240 (0.2295)	0.8988 (0.6940)
	2007	0.9152 (0.7312)	0.7059 (0.2497)	0.6497 (0.2778)	**0.9521** (0.8319)
	2017	0.9122 (0.7603)	0.6537 (0.2659)	0.5705 (0.2759)	**0.9561** (0.8952)

distribution of travels than others, which make us question whether cities that have similar urban structures and characteristics may present similar mobility differences across gender (see in Chapter 6).

We contribute to the literature showing that differences and inequalities are hidden in the complexity of what gender means and how gender can be interpreted. We argue that gender is a dynamic cultural construct that evolves over time and that changes in cultural constructs may reflect in mobility patterns changes. We presented new evidence that gender differences that result from marital status and parenthood are reflected in mobility patterns [14, 15, 18, 19, 24, 153]. Families in which women are the breadwinner indicate those gender roles are not inherent and can be reconfigured [154]. However, we do not observe this being translated in mobility — at least for population-based analysis.

Finally, to the best of our knowledge, the analyses using network science to demonstrate differences in how women and men distribute their travels over space and time is novel to the literature. We showed that the network from the mobility of men/women travellers has a global/local organisation. We present that individuals are more likely to visit other zones than those they reside in and that mobility tends to be more spatially concentrated than the residential spatial distribution. Areas highly visited by the upper-income group are usually not the ones visited by the lower income group showing socioeconomic segregation that may impact inequalities in the mobility of our studied countries.

4. Estimating the mobility diversity

There is unequal participation of different gender and socioeconomic groups across job sectors [14, 15, 17–19, 114, 115]. For example, the percentage of women in STEM-related jobs is around 30% in several countries such as Japan, Australia, and Canada [155]. In this way, we argue that if these job sectors are concentrated in a specific urban location, the likelihood of having more men visiting this location is probably higher than women. Thus, an uneven distribution of mobility within urban areas for different gender and socioeconomic groups might result from an uneven distribution of job opportunities in urban areas. The spatial concentration of mobility can be interpreted as a proxy for understanding the spatial landscape of opportunities. Furthermore, it can be used to explore the extent to which inequalities in job sectors influence commuting patterns.

In Chapter 3, we observe indicators that women and men organise their mobility differently through space and time. In this chapter, we quantify the spatial concentration of mobility and fairly compare them across gender and socioeconomic groups. Therefore, this chapter presents the second contribution of this thesis, which is the adaptation of the metrics *mobility diversity* to measure the spatial concentration of mobility across sociodemographic groups. This metric is an adaptation of the Shannon entropy [156] to compute the concentration of mobility within an urban area. Even though Shannon entropy is widely used in the literature [80, 157], we approached this measure from a new point of view, which allowed us to extend its applicability to indicate how the spatial landscape of mobility is different for sociodemographic groups. Therefore, *mobility diversity* presents a novel mobility pattern in which we capture differences between the mobility of women and men and socioeconomic groups. Our results are limited to the urban areas of Medellín, Bogotá and São Paulo, but we argue that it can be replicated for other regions if similar datasets are available. This chapter does not use data from the United States of America and the United Kingdom because we can not ensure similar robustness of our methodology in the spatial partitioning of both datasets.

We acknowledge that this chapter is a rewritten version of the reference: "**M. Macedo**, L. Lotero, A. Cardillo, R. Menezes and H. Barbosa, 'Differences in the spatial landscape of urban mobility: gender and socioeconomic perspectives', pp. 1–19, PLOS ONE, 2022.". The contribution of this work is summarised in the following points: (i) women and men differ in how they distribute their travels over urban areas; (ii) socioeconomic background amplify gender differences in mobility diversity; and (iii) work travels tend to be more concentrated than overall and non-work travels. These findings indicate that inequalities in the labour market and cultural constructs might be translated to differences in mobility.

4.1 Data

The data used in this chapter are from the national travel and household surveys from Medellín (`MDE`) [73], Bogotá (`BGT`) [129] and São Paulo (`SAO`) [72]. These data are further detailed in Section 3.1. As a summary, these data specify individual and household information such as gender, age and household income. They also detail the travels performed for different purposes/demands (e.g., work, study, go home and shop). Each travel has origin and destination zones, the departure and arrival times, purpose of travel and transportation mode used. Moreover, the governments and organisations of each country provide *expansion factors* for the travels. These *expansion factors* take into account the sociodemographic composition for each year of their partitioning (areas or zones) that projects the sample onto the real number of residents within the urban areas.

In order to be consistent in our analyses across time and to compare cities, we grouped the data into three socioeconomic strata: Lower class (`lower`), Middle class (`middle`), and Upper class (`upper`) (see details of the data in Chapter 3). In Table 4.1, we show the main characteristics of the raw data with respect to the number of people available in the data. In Table 4.2, we show the main characteristics of the raw data with respect to the number of travels available in the data. The population distributions obtained from the socioeconomic rearrangement using the expansion factors (Tables A.2 and A.3) were similar to what is frequently observed in modern societies [158, 159]. We point out that for the `MDE` survey of 2017, there is no available *expansion factor* of the travels, so we kept our analyses using the sample data. We show that the expansion factor can impact the magnitude of our measure — *mobility diversity*—, but differences across sociodemographic groups persist. Therefore, the absence of *expansion factor* can impact temporal analyses, and the usage of *expansion factor* is necessary for a more robust analysis.

Another data concern is that the number of zones and their boundaries change over time. In general, zones in our data become smaller over time as the population density and sociodemographic composition change. In this way, we use spatial partitioning for the first year available of each region, merging smaller areas into bigger ones for the following years. To ensure that our partitioning was reliable over the years, we ensure that the overall distributions of travel time, travel distance, and the fraction of travels would maintain consistent. Finally, the spatial division of the data corresponds to the area divisions of 2005 (`MDE`), 2012 (`BGT`), and 1997 (`SAO`), respectively. Two examples of spatial partitioning can be seen for the case of `BGT` in 2019 in Figures 4.1 and 4.4 where we show the gender and socioeconomic distribution of work travel over the metropolitan region.

4.2 Mobility Diversity

Mobility diversity is defined as the concentration/distribution of travels within a set of zones in an urban area. The term *mobility diversity* was proposed by Pappalardo et al. [157] based on the probability of an individual performing a travel from a zone x to a zone y. Whereas Lenormand et al. [80] compute a similar metrics considering the attractiveness of a location taking into account where people live and visit. Both papers use the Shannon entropy to compute the concentration of mobility but from different perspectives.

Here, we compute the *mobility diversity* based on the population or collective mobility from sociodemographic groups in urban areas. We are not necessarily interested in the zone where people live, so we do not constraint travel origins to consider the trip-chaining characteristics in mobility. From the literature, trip-chaining is more likely to be performed by women than by men [25, 26, 111], and disregarding this fact would potentially hide inequalities across genders. Therefore, *mobility diversity* in this chapter is the likelihood of a group of people X visiting a zone i for a specific demand d.

Mathematically, we describe the *mobility diversity*, H_d^X, of a group of travellers X based on their travels to fulfil purpose d as Equation (4.1). A group X can be a set of individuals that share one or more characteristics in common, such as gender and socioeconomic status. The probability of an individual of the group X visiting a given zone i to fulfil purpose d is $p_d^X(i)$, represented in Equation (4.2). $N_d^X(i)$ denotes the number of travels made by a group X to fulfil purpose d whose destination is zone i, and N_d^X denotes the total number of travels made by a group X to fulfil purpose d. The total number of zones is represented as N_Z.

$$H_d^X = -\frac{1}{\log_2 N_Z} \sum_{i=1}^{N_Z} p_d^X(i)\, \log_2 p_d^X(i)\,, \tag{4.1}$$

where

$$p_d^X(i) = \frac{N_d^X(i)}{N_d^X}\,. \tag{4.2}$$

Equation (4.1), H_d^X, outputs a value between zero and one ($H_d^X \in [0,1]$). The two extreme cases correspond to (i) a complete concentration of travels in one zone ($H_d^X = 0$) or (ii) a uniform distribution of travels over all the zones ($H_d^X = 1$). The boundaries of *mobility diversity* do not depend on the group of travellers or the purpose of travel in consideration. Under the assumption that all the travellers travel to one zone ($i = \tilde{i}$), the Equation (4.2) becomes:

$$p_d^X(i) = \begin{cases} 1 & \text{for } i = \tilde{i} \\ 0 & \text{otherwise} \end{cases}\,. \tag{4.3}$$

Then, we replace p in Equation (4.1), and rewrite H_d^X:

$$H_d^X = -\frac{1}{\log_2 N_Z}\left[(1\ \log_2 1) + \sum_{\substack{i=1\\ i\neq \tilde{i}}}^{N_Z} 0\ \log_2 0\right] = -\frac{1}{\log_2 N_Z}\ (0+0) = 0\,. \tag{4.4}$$

Otherwise, if we assume that the travellers visit all the zones with similar probabilities, Equation (4.2) becomes:

$$p_d^X(i) = \frac{N_d^X(i)}{N_d^X} = \frac{N_d^X/N_Z}{N_d^X} = \frac{N_d^X}{N_Z}\frac{1}{N_d^X} = \frac{1}{N_Z}\ \forall i\,, \tag{4.5}$$

where $N_d^X(i)$ is the total number of travels made by a group X with a purpose d to a destination area i and $N_d^X = \sum_i^{N_Z} N_d^X(i)$. We then replace Equation (4.5) in Equation (4.1):

$$H_d^X = -\frac{1}{\log_2 N_Z}\sum_{i=1}^{N_Z}\frac{1}{N_Z}\ \log_2\frac{1}{N_Z}\,, \tag{4.6}$$

as the sum does not depend on i, we solve:

$$H_d^X = -\frac{1}{\log_2 N_Z}N_Z\left[\frac{1}{N_Z}\ \log_2\frac{1}{N_Z}\right] = -\frac{1}{\log_2 N_Z}\Big(-\log_2 N_Z\Big) = 1\,. \tag{4.7}$$

The number of zones and travels can impact the magnitude of the *mobility diversity* (Equation 4.1), so we take extra precautions to ensure consistency. To account for the effect of sample and population sizes, we propose to compute *mobility diversity* using a bootstrapping strategy and estimate the H values from random samples of the data. We compute the *mobility diversity* (e.g. the value of H_d^X 1000 times using Equation (4.1)) and a random sample set of 60% travels made by a certain group of travellers, X, fulfilling a given purpose, d. With a set of 60% of travels, we ensure the consistency of the *mobility diversity* values while avoiding statistical fluctuations biased by the data (e.g. outliers). We describe the impact of varying the bootstrapping percentages of travels in Section 4.5. We also analysed the evolution of the values of H with respect to specific sizes (e.g., 5000 travels) of the bootstrap sample. We found that the value of H saturates as the sample size increases (see Section 4.5). Finally, we argue that the effects of the sample composition (e.g., the presence of more poor travellers than rich ones) on the value of H is part of the intrinsic nature of the Colombian and Brazilian societies and, therefore, does not constitute a bias in our results.

The variation in the area, length and population density of the zones can affect the values of H. To address this spatial partitioning issue, we compared the empirical values of H with those obtained using five null models accounting for both the non-homogeneous size of the areas and the non-uniform density of inhabitants (Section 4.6). We observe that the values of H are not completely described even taking into account the main factors of mobility, indicating that gender and socioeconomic dimensions potentially shape the *mobility diversity*.

In summary, we ensure that our analyses of *mobility diversity* are valid when considering multiple concerns related to sample size, spatial partitioning and population distribution. In line with Section 2.5 and the literature [3], the following factors are essential to mobility: travel distances following a truncated power-law distribution and non-homogeneous residential distribution. Nevertheless, these factors are insufficient to explain the differences in *mobility diversity* across gender and socioeconomic groups.

Table 4.1: Main properties of the raw datasets analysed in our study. For each region, we have the total area covered $\mathcal{A}$, and the number of zones into which it is divided, N_Z. Then, for each year we have the number of travellers N_P, the fraction of `men` (`women`) travellers f^M (f^W), and the fraction of Lower class, Middle class and Upper class travellers ($f^{\texttt{lower}}$, $f^{\texttt{middle}}$ and $f^{\texttt{upper}}$).

Region	$\mathcal{A}(km^2)$	N_Z	Year	N_P	$f^{\texttt{men}}$	$f^{\texttt{women}}$	$f^{\texttt{lower}}$	$f^{\texttt{middle}}$	$f^{\texttt{upper}}$
`MDE`	1,167	215	2005	22,840	0.48	0.52	0.48	0.46	0.06
			2017	30,290	0.49	0.51	0.54	0.38	0.08
`BGT`	24,477	400	2012	37,189	0.46	0.54	0.52	0.42	0.06
			2019	47,149	0.48	0.52	0.52	0.43	0.05
`SAO`	9,486	248	1997	37,316	0.48	0.52	0.30	0.63	0.07
			2007	51,103	0.49	0.51	0.23	0.62	0.15
			2017	48,085	0.50	0.50	0.23	0.63	0.14

Table 4.2: Main properties of the raw datasets analysed in our study. For each region, we have the total area covered $\mathcal{A}$, and the number of zones into which it is divided, N_Z. Then, for each year we have the number of travels N_T, the fraction of travels made by men (women) travellers $f^{\texttt{men}}$ ($f^{\texttt{women}}$), and the fraction of travels made by Lower class, Middle class and Upper class ($f^{\texttt{lower}}$, $f^{\texttt{middle}}$ and $f^{\texttt{upper}}$).

Region	$\mathcal{A}(km^2)$	N_Z	Year	N_T	$f_T^{\texttt{men}}$	$f_T^{\texttt{women}}$	$f_T^{\texttt{lower}}$	$f_T^{\texttt{middle}}$	$f_T^{\texttt{upper}}$
`MDE`	1,167	215	2005	70,773	0.48	0.52	0.47	0.46	0.07
			2017	65,228	0.49	0.51	0.15	0.66	0.19
`BGT`	24,477	400	2012	100,009	0.45	0.55	0.50	0.43	0.07
			2019	164,931	0.48	0.52	0.51	0.44	0.05
`SAO`	9,486	248	1997	93,376	0.48	0.52	0.29	0.63	0.08
			2007	137,411	0.49	0.51	0.20	0.62	0.18
			2017	125,544	0.50	0.50	0.20	0.63	0.17

4.3 Results

We explore the differences in the *mobility diversity* that are consistent and persistent over the years and metropolitan areas that can potentially indicate disadvantages and inequalities. Our hypothesis is that there is a structural disadvantage of certain groups in mobility, so inequalities in the labour market can manifest in the commuting patterns. In Chapter 3, we saw that both gender and socioeconomic groups display different network structures from their mobility, so now we estimate to what extent each group tend to concentrate their travels in a few areas.

We present the existence of effects played by gender and socioeconomic factors in urban mobility by computing the *mobility diversity* of the overall travels (`all`), and their differences with work (`work`) and non-work (`nonwork`) travels. Then, we compare the *mobility diversity* coming from the mobility across gender and socioeconomic strata (separately and combined). We point out that longitudinal changes of H for `MDE` are mostly presented in the discussion section as we do not have expansion factors available for 2017.

Looking at Figures 4.1- 4.6, we argue that there is already a visual perception of the role of gender and socioeconomic status in the mobility of `MDE`, `BGT` and `SAO`. Most of the areas are visited for the purpose of work by men and by the middle class, and we observe that the spatial distribution of mobility over the areas has some peculiarities for each city. In this way, we now use *mobility diversity* to capture this concentration of mobility.

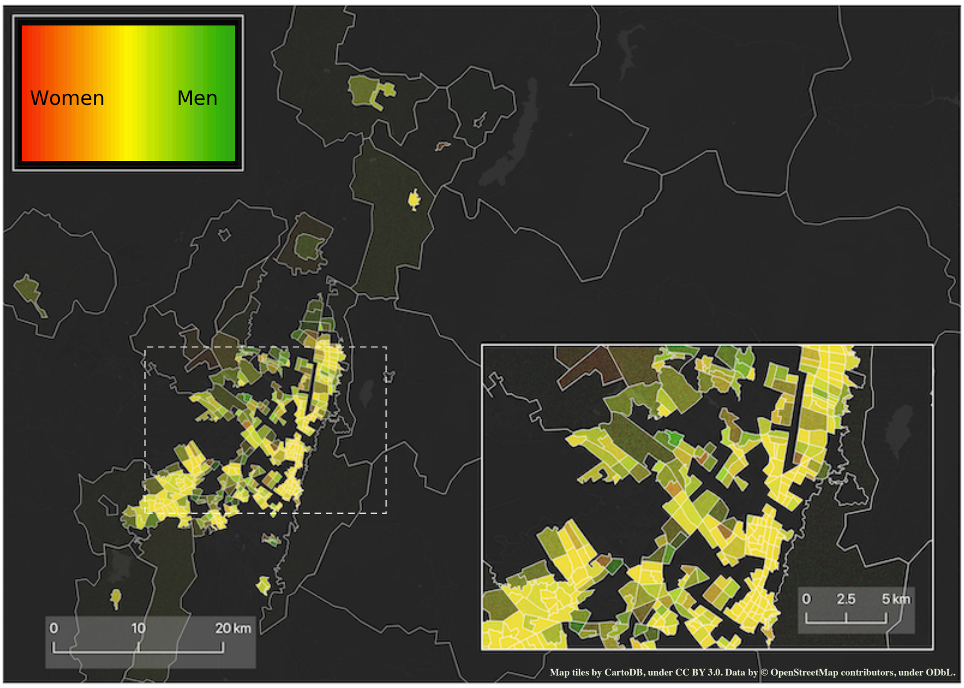

Figure 4.1: Density map of `work` travels made in `BGT` during the year 2019. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by women (red), men (green), or by both (yellow). The inset portrays a zoom of the city centre.

4.3.1 Analysis of the travel's purpose

We compute the *mobility diversity*, H, of travels belonging to three groups: related to work (`work`), related to any purpose except for work (`nonwork`) and regardless of their purpose (`all`). Figure 4.7 exhibits the distributions of the values of H for each group, city and year. Pairwise Welch's t-tests revealed that all the distributions are statistically different (p-value < 0.001). The values of H are all located above 0.80, indicating that the travels are — more or less — evenly distributed across all the zones available regardless of the purpose, city, or year considered.

We observe that `work` travels display smaller values of H than from other types of travels, which means that `work` travels are slightly more concentrated. In this way, we argue that job opportunities might be more spatially concentrated in urban areas than other purposes such as education and leisure. We also see that there is a decrease of H over time for the travels in `MDE` and `BGT`. This change might come from other opportunities that become available in other zones, so the travels became

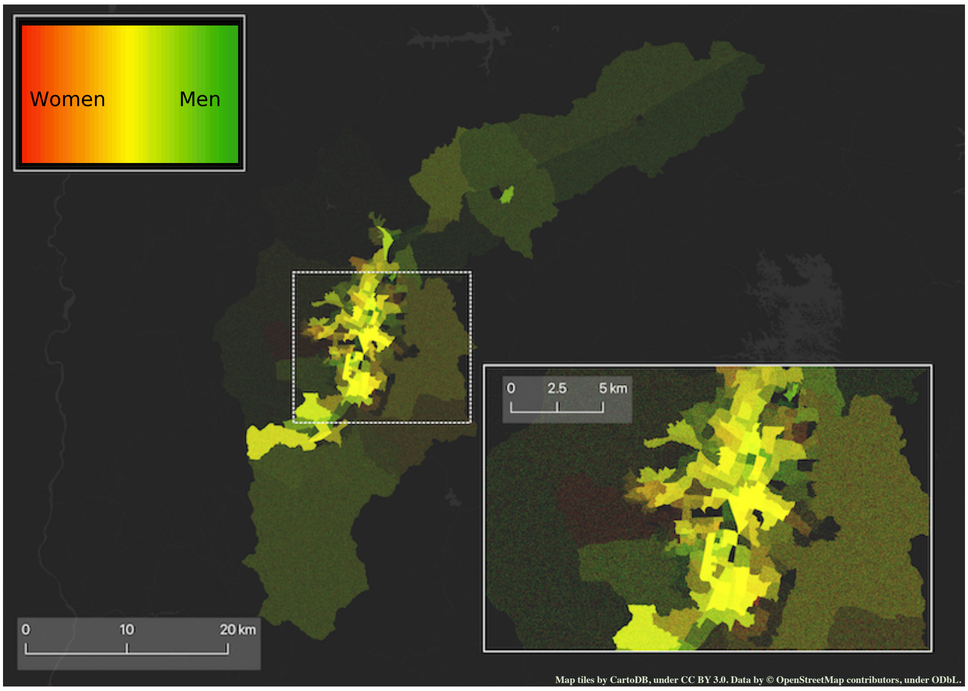

Figure 4.2: Density map of `work` travels made in `MDE` during the year 2017. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by women (red), men (green), or by both (yellow). The inset portrays a zoom of the city centre.

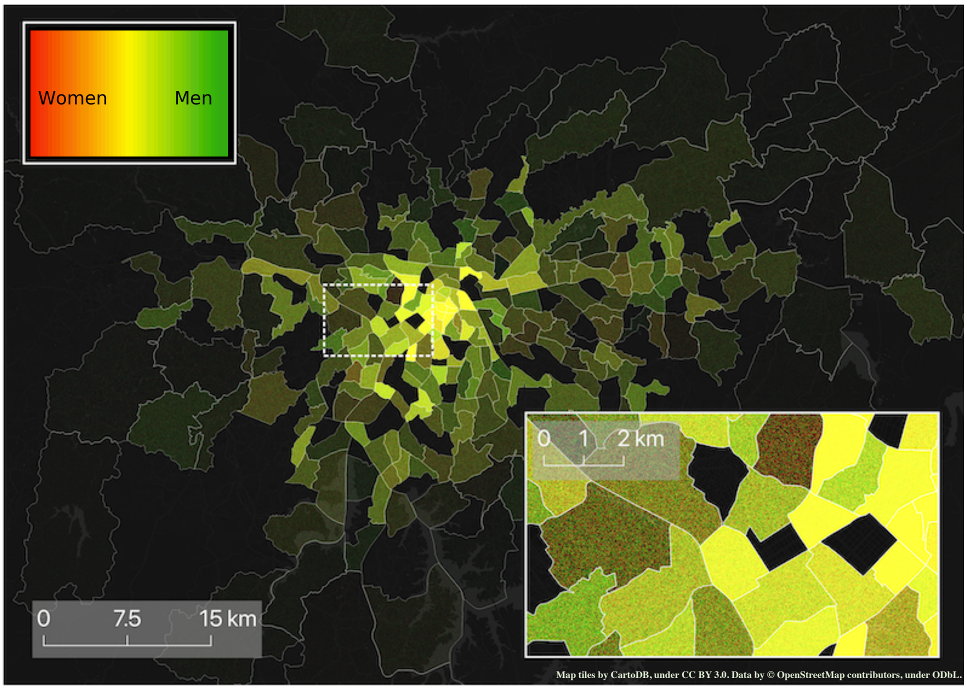

Figure 4.3: Density map of `work` travels made in `SAO` during the year 2017. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by women (red), men (green), or by both (yellow). The inset portrays a zoom of the city centre.

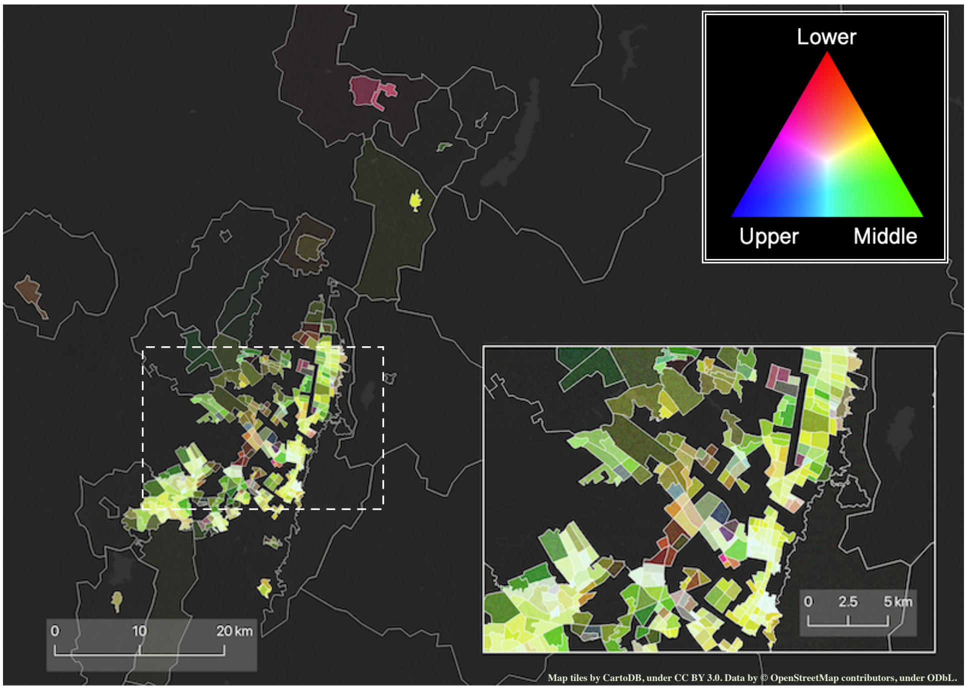

Figure 4.4: Density map of `work` travels made in `BGT` during the year 2019. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by travellers belonging to the `lower` (red), `middle` (green), `upper` (blue) or all three socioeconomic status. The inset portrays a zoom of the city centre.

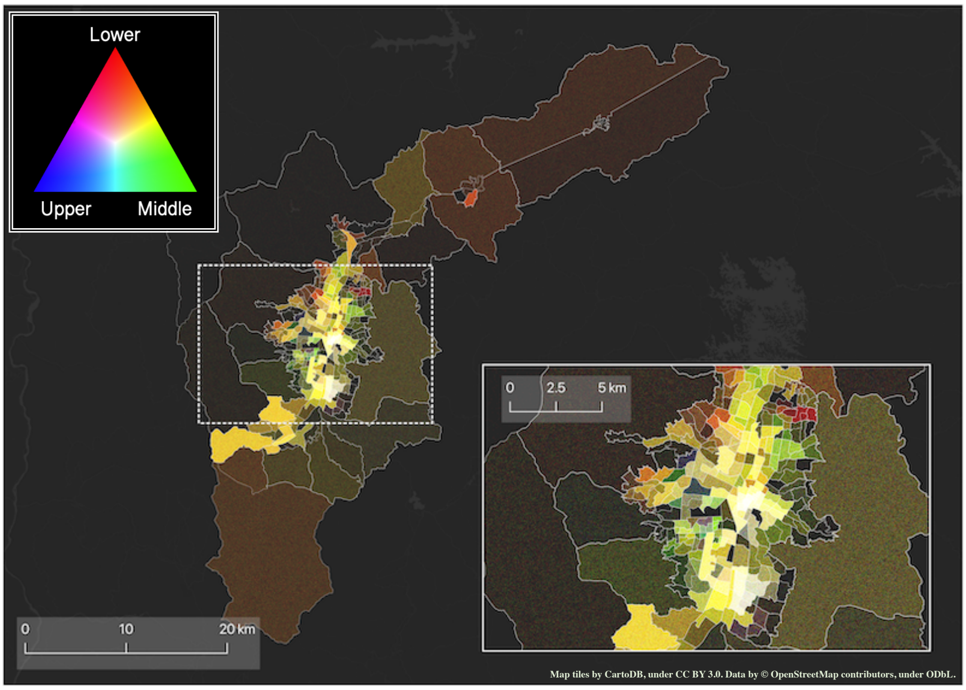

Figure 4.5: Density map of `work` travels made in `MDE` during the year 2017. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by travellers belonging to the `lower` (red), `middle` (green), `upper` (blue) or all three socioeconomic status. The inset portrays a zoom of the city centre.

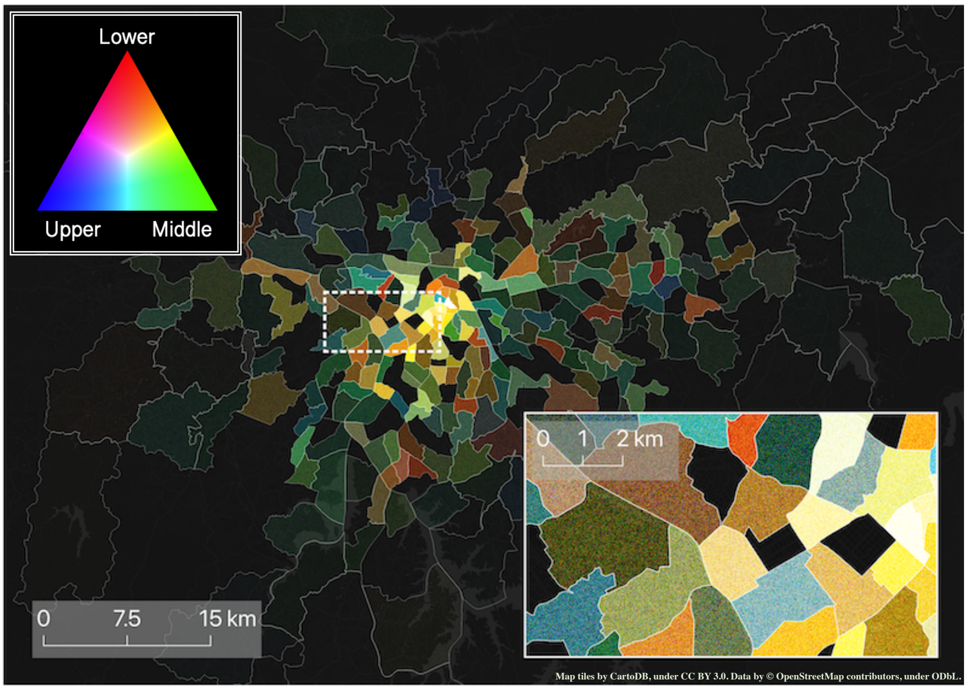

Figure 4.6: Density map of `work` travels made in `SAO` during the year 2017. Brighter colours represent a higher density of travels to work. The hue denotes whether for a given zone the majority of travels were made by travellers belonging to the `lower` (red), `middle` (green), `upper` (blue) or all three socioeconomic status. The inset portrays a zoom of the city centre.

more concentrated.

For the case of `SAO`, we observe that the increase of H only happened between 2007 and 2017 indicating a decrease on the concentration of travels. From 1997 to 2007, we observe a decrease of H that indicates an increase in the concentration of travels. We argue that `work` travels seem to be less stable over the years than `all` travels demonstrating the importance of analysing mobility not solely as an overall or average.

4.3.2 Effects of gender on mobility's diversity

We uncover in this section that there is a statistical difference between the distributions of *mobility diversity* across gender. We point out that men and women display different patterns in mobility such as average travel time, preferences on the mode of transportation, and commuting travel distance [25, 28, 45–48, 134], and *mobility diversity* is a new pattern in mobility in which they differ. We argue that as the participation of women in some job sectors is scarce and that some job sectors are

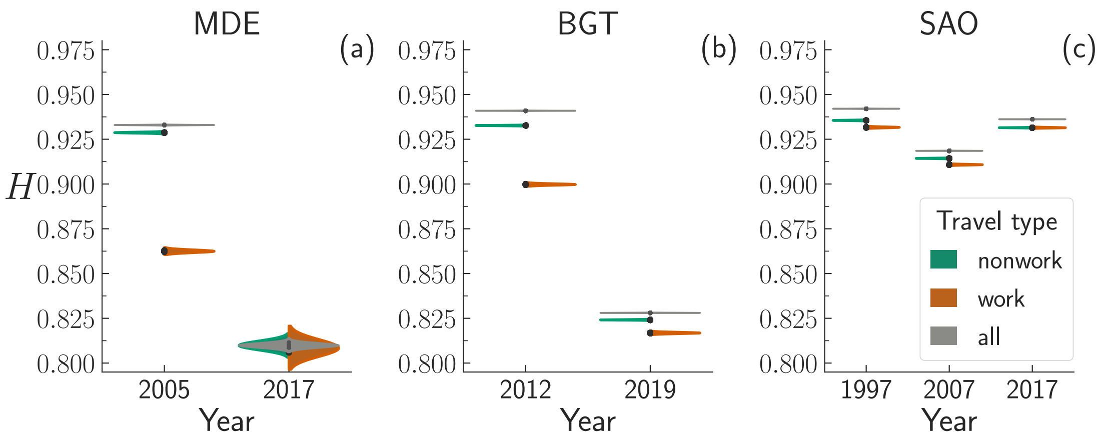

Figure 4.7: Violin plots of the bootstrapped *mobility diversity*, H, for `all`, `work` and `nonwork` travels made in each region and year. To better visualise the overlap (or not) between the distributions of `all`, `work`, and `nonwork` travels, we show the distributions for `all` travels duplicated (entire grey violins instead of half-violins).

concentrated in a set of areas, women and men explore the urban areas in different ways. Therefore, we use *mobility diversity* to measure this difference.

We first detail our analysis for `BGT`, then we summarise the analysis for the other cities. Figure 4.8 shows the Kernel Density Estimation (KDE) of the *mobility diversity*, H, of travels made by men (M), women (W), and all (A) travellers for all travels (`all`), and for work travels only (`work`). Statistical differences are found between the distributions of H across gender groups computed by Welch's t-test between all the possible pairs of distributions (p-value < 0.001).

The envelopes of the KDEs are more peaked for the most recent year, and they display a smaller peak-to-peak distance between the envelopes. Regardless of gender, the travels in 2019 become more concentrated than in 2012, indicating that some areas might start attracting more visits. Considering gender, we see that the average values of H from travels performed by women are always smaller than the ones performed by men. This suggests that men tend to explore more in urban areas than women.

We then repeat the same analysis for `MDE` and `SAO` (shown in detail in Figures 4.9 and 4.10). We provide an overview of the effects of gender on H for all the urban areas together over all the available years in Figure 4.11. We ensure that Welch's t-test confirmed that the distributions are statistically different (p-value < 0.001), except for the case between men and all travellers of `MDE` in 2017.

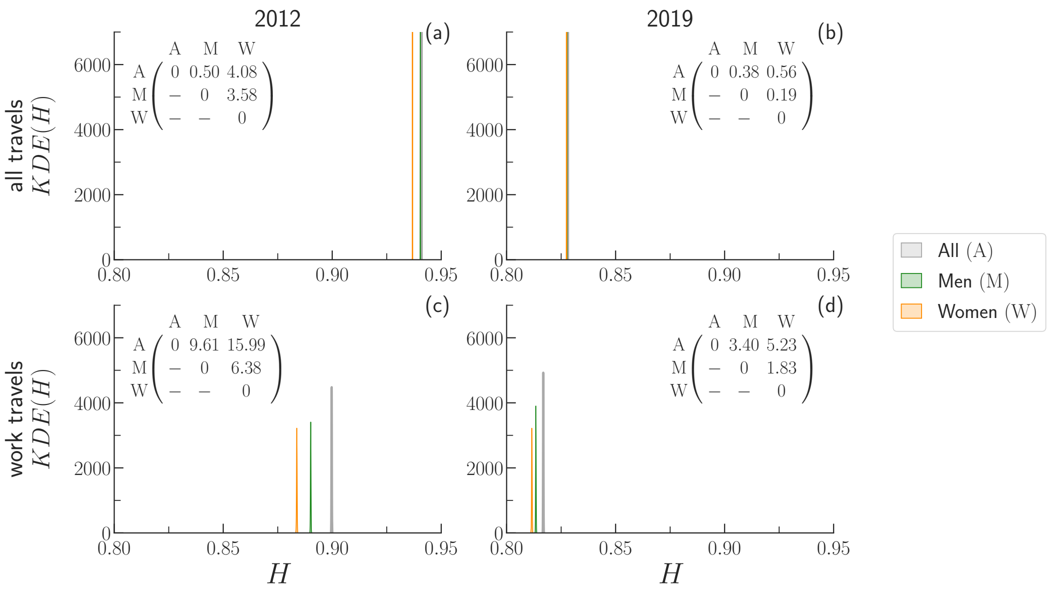

Figure 4.8: Kernel Density Estimation (KDE) plots of the *mobility diversity*, H, for `all` travels (panels a and b) and `work` travels (panels c and d) in the urban area of `BGT`. For each travel purpose, we plot KDE(H) for travels made by men (M), women (W), and all travellers (A). The matrix appearing within each graphic summarises the distances between the medians of the distributions (peak-to-peak distances multiplied by 10^{-3}).

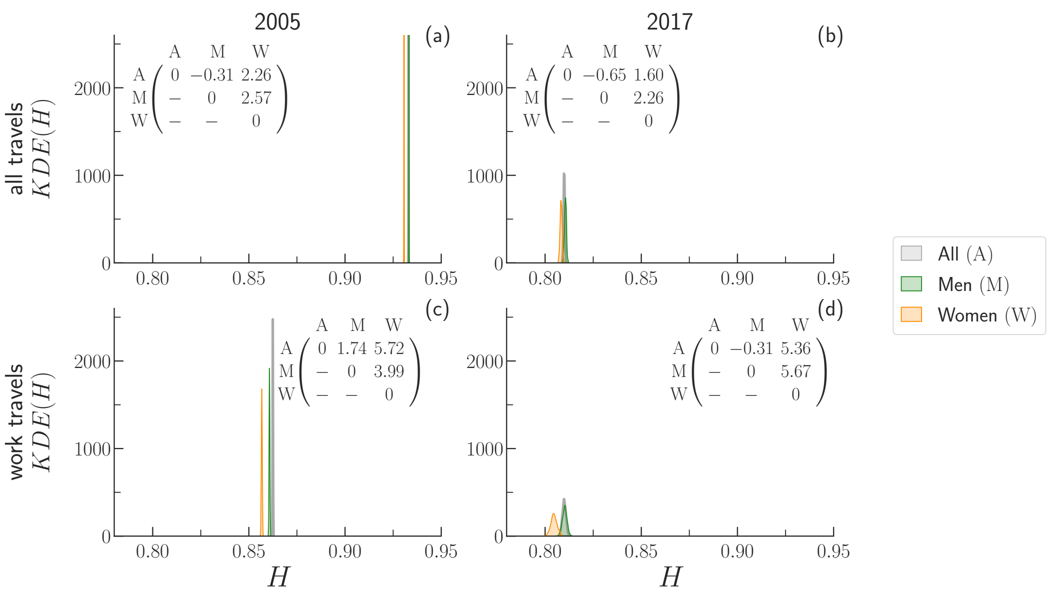

Figure 4.9: Kernel Density Estimation (KDE) plots of the *mobility diversity*, H, for `all` travels (panels a,b) and `work` travels (panels c,d) in `MDE`. For each travel purpose, we plot the KDE(H) for travels made by men (M), women (W), and all (A) travellers. The matrix appearing in the top left corner of each panel reports the peak-to-peak distance (i.e. the distance between the median of the distributions) multiplied by a factor of 10^{-3}. The KDEs are computed from a distribution of H obtained by bootstrapping 1,000 times 60% of the available travel records.

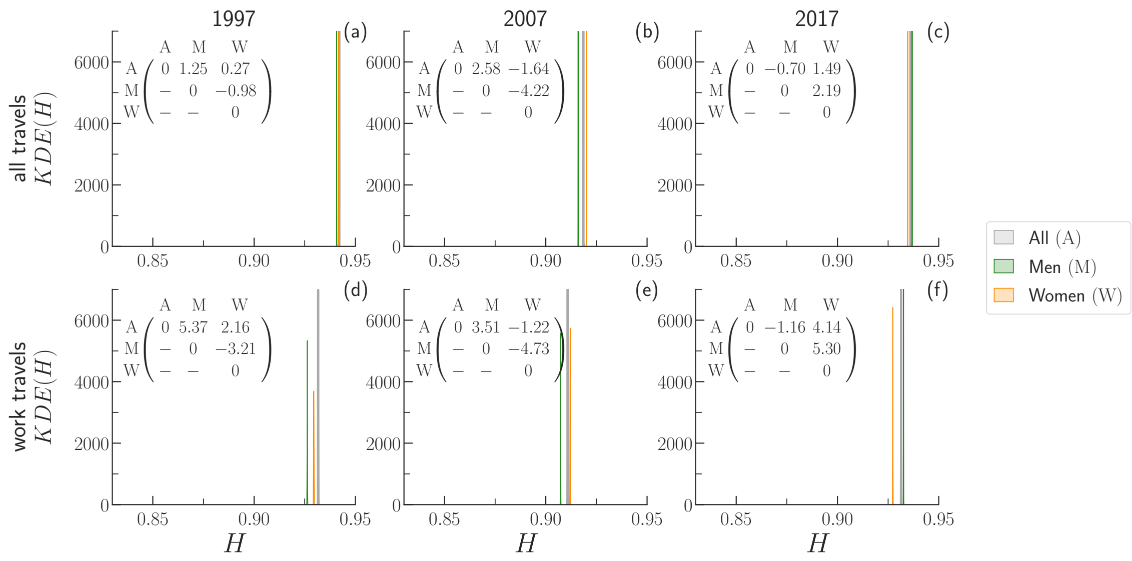

Figure 4.10: Kernel Density Estimation (KDE) plots of the *mobility diversity*, H, for `all` travels (panels a - c) and `work` travels (panels d-f) made in `SAO`. For each travel purpose, we plot the KDE(H) for travels made by men (M), women (W), and all travellers (A). The matrix appearing in the top left corner of each panel reports the peak-to-peak distance (i.e. the distance between the median of the distributions) multiplied by a factor of 10^{-3}. The KDEs are computed from a distribution of H obtained by bootstrapping 1,000 times 60% of the available records.

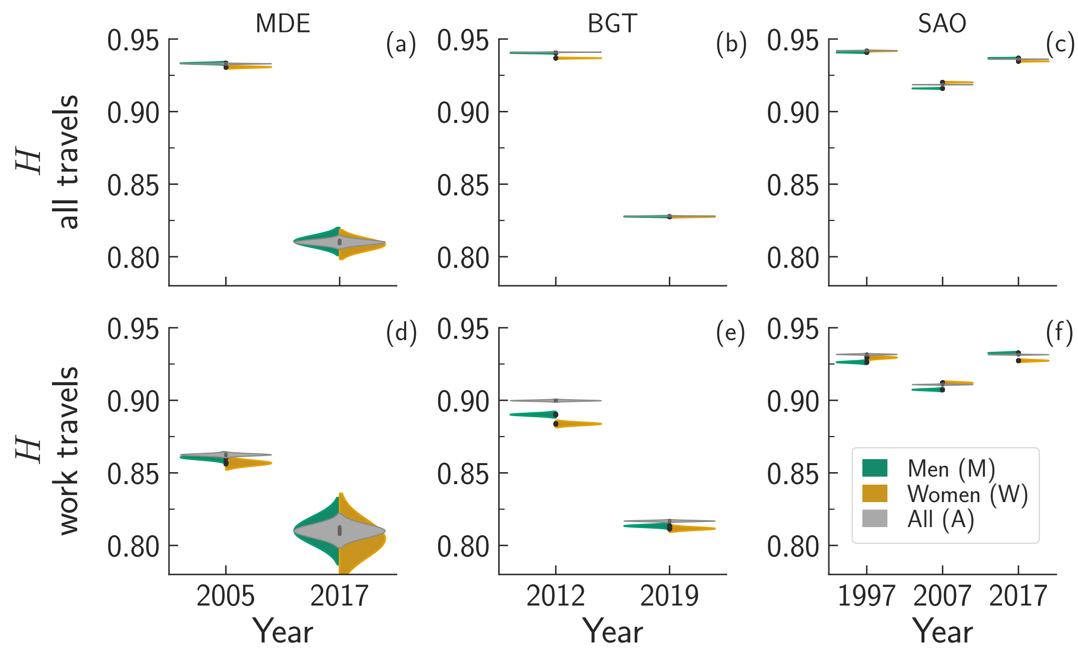

Figure 4.11: Violin plots of the bootstrapped *mobility diversity*, H, for `all` travels (top row, panels a-c), and `work` travels (bottom row, panels d-f). Each column refers to a different region: `MDE` (panels a and d), `BGT` (panels b and e), and `SAO` (panels c and f). For each region, we display the distribution of the values of H in each year. We show the distributions for `all` travels duplicated (entire violins), and the distributions for `men` and `women` travels in half-violins.

Similar to the results for BGT, the value of H for the travels performed by men tend to be higher than the ones computed for women's mobility for MDE and SAO. We identify one exception for the case of SAO in 2007 for which $H^W > H^M$. The violin plots show also that, in general, $\Delta H^{MA} < \Delta H^{WA}$, where $\Delta H^{XY} = \left|\langle H^X\rangle - \langle H^Y\rangle\right|$ and $X, Y \in \{A, M, W\}$. We conclude that no qualitative difference can be seen between the distributions of $H_{\texttt{all}}$ and $H_{\texttt{work}}$.

We highlight that travels performed by both genders can have values of *mobility diversity* greater than only computed from women or men travels. This is true because the distribution of probabilities of individuals, regardless of gender, visiting a location can become more uniform than when considering gender. The opposite can also happen. Even though the sample size varies, the variance on the probabilities of visiting a location plays the main role in establishing *mobility diversity*. Further details on the sample size effect can be seen in Section 4.5. Regardless of the sample size selected for the number of travels, we consistently see the values of *mobility diversity* from men travellers higher than the ones computed for women travellers.

Looking at the temporal evolution of H in Figure 4.11, we observe a decrease of H between years, except for the city of SAO in 2017 that shows the opposite trend. We in fact observe the same V-shaped patterns as in Figure 4.7. We argue that, in general, the value of H associated with men's mobility tend to be higher than women's, regardless of the travel's purpose.

We also point out that the absence of expansion factors in MDE 2017 indeed cause some variation in the *mobility diversity* for MDE. Looking at Figures 4.7 and 4.12, there is an evidence that *mobility diversity* decreases over the years. The gender groups are not affected by the expansion factor, but the magnitude is slightly impacted.

As in the literature, we also observe in our data that on average women are more likely to perform shorter travels than men (Figure 4.13, and Tables 4.3 and 4.4). We compute the travel distance, l, between the centroids of the origin and destination zones. Even though we see differences in the travel distances distributions, men could have longer travel distances while concentrating their travels in a few zones, which is not the case. Therefore, the differences in the *mobility diversity* do not necessarily

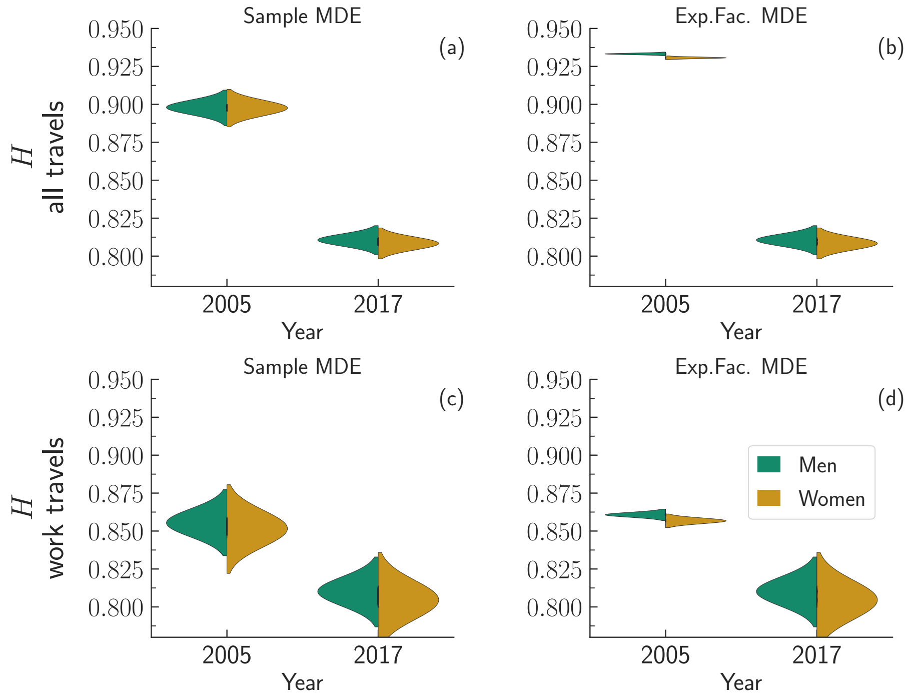

Figure 4.12: Comparing the distributions of the *mobility diversity* (H) for `all` travels (panels **a,b**) and `work` travels (panels **c,d**) within the `MDE` area.
Panels **a** and **c** display the case of raw travel records, whereas panels **b** and **d** display the case of travel records obtained using the expansion factors for year 2005.

emerge because of the travel distance distributions.

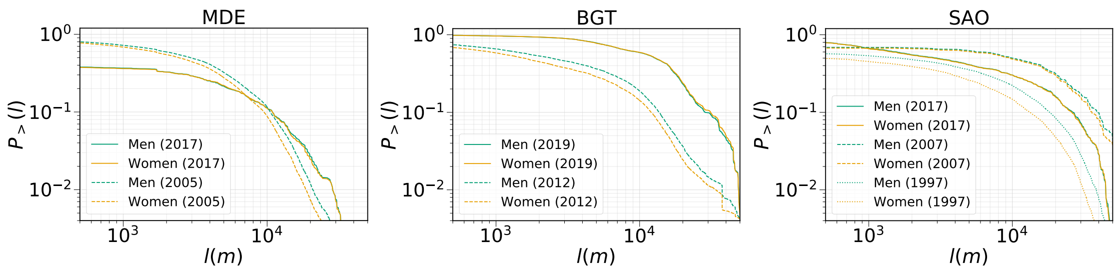

Figure 4.13: Complementary cumulative probability distribution function, $P_>(l)$, of the probability of making travel with a distance between origin and destination zones equal to or greater than l. Each panel refers to a different metropolitan area.

Nevertheless, the percentage of women or men working in the same zone that they live in can impact the *mobility diversity*, as endogenous mobility decreases the percentage of travels exploring other urban areas. In our data, women and men display a similar fraction of travels regardless of the travel's purpose (all or to work), and the fact that the origin and destination zones are the same (travels inside a zone) (Table 3.7). The fraction of travels that travellers live and work in the same zone are similar for women and men (Table 3.7). Furthermore, women and men work in general in only one zone (Figure 4.14), but the latter is slightly more likely to work in more than one zone. Thus, we argue that travels inside zones and the number of workplaces are not impacting differences in the *mobility diversity* of women and men.

Then, we check whether the majority of the zones are more likely to be visited by men than women for different travel purposes. Table 3.6 shows the percentage of zones for which there are more travels performed by men than by women for `all`, `work` and `nonwork` purposes. For `all` travels, `MDE` and `SAO` show most areas being visited by men, and `BGT` shows a majority of areas being visited by women. For `work` travels, regardless of the region, the majority of the areas are mostly visited by men, and the opposite happens to `nonwork` travel. We conclude that women are more likely to be concentrated in a few areas to work, and women are also a minority in most areas. We see that the travel's destination for women and men follows different distributions regardless of the purpose of travel except for `MDE` in 2017 (tested by Student t-test and Kolmogorov–Smirnov test with p-value < 0.01). This might be the case as we do not have the expansion factor available for `MDE` in 2017.

Table 4.3: Minimum ($\min(l)$), maximum ($\max(l)$), median ($med(l)$), average ($\langle l \rangle$), and standard error of the mean (ε_l) of the travel distance l (measured in m) made by men and women in each region and year.

City	Year	Gender	$\min(l)$	$\max(l)$	$med(l)$	$\langle l \rangle$	ε_l
`MDE`	2005	`men`	102.54	59149.04	3775.57	5025.26	3.37
		`women`		58728.50	3256.55	4463.16	3.17
	2017	`men`	104.58	39867.49	5349.23	7705.08	52.37
		`women`		39867.49	5631.93	7734.20	53.62
`BGT`	2012	`men`	101.35	115452.21	4207.63	6666.98	3.49
		`women`		81793.49	3008.38	5746.65	2.99
	2019	`men`	119.02	89115.59	12883.82	14561.42	4.94
		`women`		89115.59	12991.28	14761.55	4.74
`SAO`	1997	`men`	103.91	99384.35	7081.65	10162.40	4.05
		`women`		99384.35	5297.32	8136.43	3.88
	2007	`men`	281.25	85235.50	17888.80	21623.26	4.91
		`women`		85235.50	17627.78	20844.79	4.74
	2017	`men`	130.41	49996.82	4727.93	10015.40	3.11
		`women`		62504.24	4630.86	9952.94	3.10

Table 4.4: The p-values of the Kolmogorov–Smirnov ($KSTest$) and Student t ($TTest$) tests comparing the travel distance performed by `men` (M), `women` (W) and `all` travellers (A). The symbol *** represents that the p-value is smaller than 0.001.

City	Year	$KSTest(MW)$	$KSTest(AM)$	$KSTest(AW)$	$TTest(MW)$	$TTest(AM)$	$TTest(AW)$
MDE	2005	***	***	***	***	***	***
	2017	***	0.17	0.12	0.69	0.83	0.82
BGT	2012	***	***	***	***	***	***
	2019	***	***	***	***	***	***
SAO	1997	***	***	***	***	***	***
	2007	***	***	***	***	***	***
	2017	***	***	***	***	***	***

Table 4.5: Percentage of areas for which the fraction of travels performed by men is higher than the same quantity computed for women for `all`, `work` and `nonwork` travels ($P^M_{all,area} > P^W_{all,area}$, $P^M_{work,area} > P^W_{work,area}$, $P^M_{nonwork,area} > P^W_{nonwork,area}$).

City	Year	$P^M_{all,area} > P^W_{all,area}$	$P^M_{work,area} > P^W_{work,area}$	$P^M_{nonwork,area} > P^W_{nonwork,area}$
MDE	2005	61.46%	79.67%	43.20%
	2017	62.67%	91.12%	36.40%
BGT	2012	28.21%	63.79%	21.55%
	2019	26.84%	77.09%	15.02%
SAO	1997	63.88%	87.20%	39.84%
	2007	46.85%	86.12%	26.77%
	2017	50.00%	86.69%	34.00%

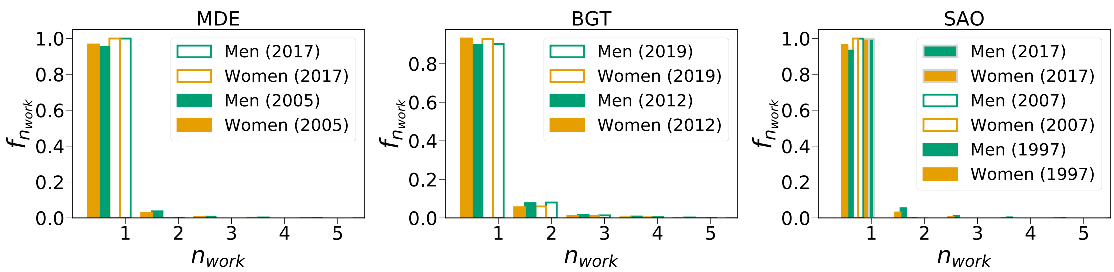

Figure 4.14: Fraction of the number of locations in which an individual works, $f_{n_{work}}$, disaggregated according to the gender of the travellers.

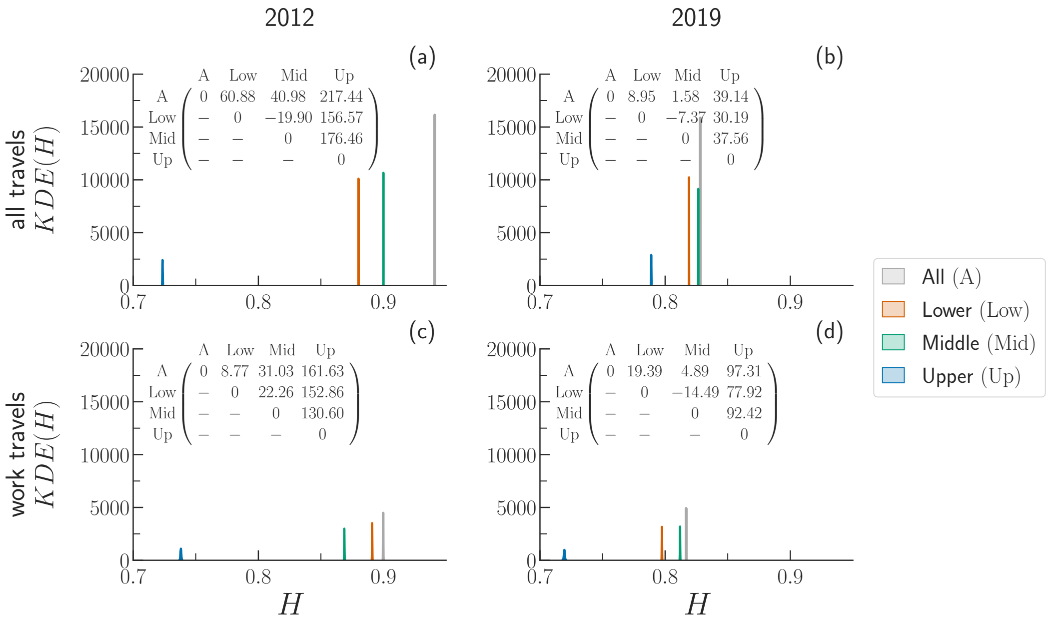

Figure 4.15: Kernel Density Estimation (KDE) plots of the *mobility diversity*, H, for travels made by travellers belonging to different socioeconomic status in the `BGT` area of 2012 (panels a and c) and 2019 (panels b and d). The top row (panels a and b) displays the values obtained considering `all` travels, whereas the bottom row (panels c and d) displays the values obtained considering only travels associated with the `work` purpose. We show the KDE(H) for travels made by all travellers (A), as well as for those belonging to the `lower` (Low), `middle` (Mid), or `upper` (Up) socioeconomic groups. The matrix appearing within each plot encapsulates the distance between the median of the distributions (peak-to-peak distances multiplied by 10^{-3}).

4.3.3 Effects of gender & socioeconomic status

In this section, we unveil an effect of socioeconomic status alone and combined with gender in the *mobility diversity* (H). In fact, we show that gender differences in H are amplified when we take the socioeconomic factor into account. We first analyse the role of socioeconomic status alone for travels made for `all` or `work` purposes. We compute the values of H of travels made by travellers belonging to the three socioeconomic groups as defined in Section 4.1 (i.e. `lower` Low, `middle` Mid, and `upper` Up). We also analyse the *mobility diversity* of the travels made by all travellers combined `all` (A).

We detail the analyses of *mobility diversity* for `BGT`, and then we move to other urban areas. Figure 4.15 displays the KDE(H) for the `BGT` area for the years 2012 and 2019, respectively. We find the same decreasing trend of *mobility diversity* over time considering both purpose of travels as in Sections 4.3.1 and 4.3.2.

The travels performed by the people belonging to the `upper` income group exhibit the lowest values of H, indicating that they concentrate more on their travels in a few urban areas. In contrast, `middle` income group travellers in general display the highest values of H, suggesting that individuals in the `middle` income group cover space more uniformly than other classes. However, we argue that the magnitude of the *mobility diversity* varies across the socioeconomic groups for different reasons.

For the studied cities, people belonging to `upper` income group might move to fewer zones because they are better located in relationship to opportunities. On the other hand, people belonging to lower-income class might not be able to afford to live close to a high volume of opportunities. Besides, the mobility of `lower` income group can rely more on the public transportation system, which can be insufficient or inadequate depending on the city and the travel destination. Finally, we observe that the peak-to-peak distances between the KDE become smaller over the years, indicating a possible decrease of socioeconomic inequalities in `BGT`. We highlight here that this decrease is also seen across gender groups.

Analysing the other urban areas, we observe that the KDE plots (Figures 4.16 and 4.17) confirm that travellers belonging to the `upper` group tend to display the lowest values of H, whereas those belonging to the `middle` group tend to cover more uniformly the urban region. There is no qualitative difference between the values of H for different purposes of travel, but we do see some peculiarities across cities. In `SAO` (Figure 4.17), the *mobility diversity* of lower and middle-income travellers display similar changes over the years, while the upper group shows the opposite trend. We argue here that the impact on the mobility of lower and middle income groups in 2007 could have been largely influenced by the profound economic changes Brazil underwent during that period [160, 161].

We now focus on the combined effect of gender and socioeconomic status, so we compute the *mobility diversity* of travels made by travellers having a certain social status (e.g. `middle`) and gender (e.g. W). We summarise the results of H in Figure 4.18 for travels made for `all` purposes by all the combinations of gender and socioeconomic status. What stands out in Figure 4.18 is that socioeconomic status shapes the mobility of people considerably, whereas gender shapes a smaller effect.

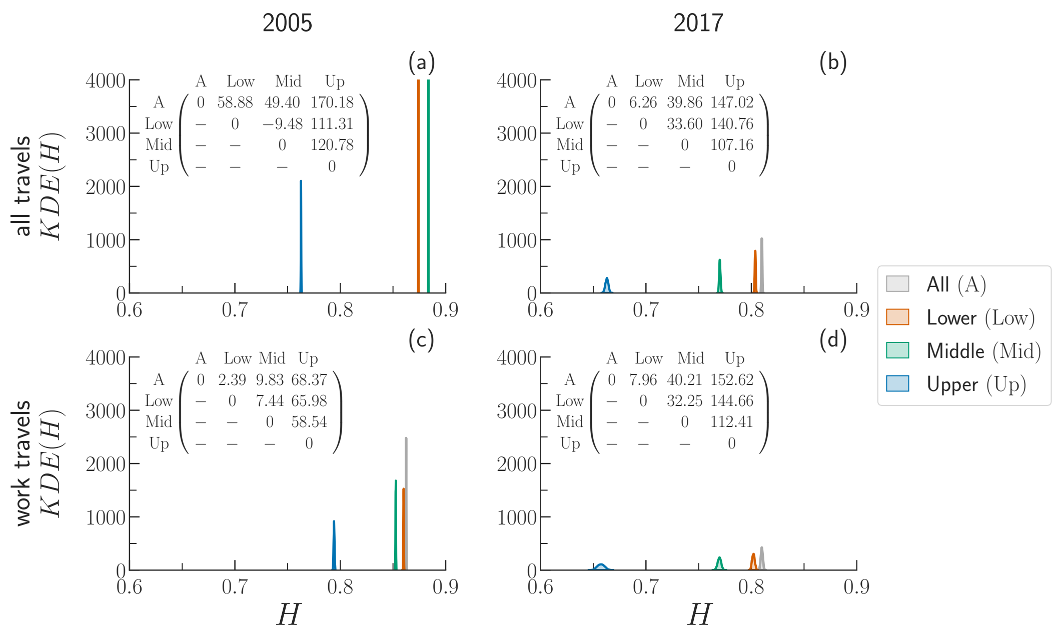

Figure 4.16: Kernel Density Estimation (KDE) plots of the *mobility diversity* H for `all` travels (panels **a** and **b**), and `work` travels (panels **c** and **d**) in Medellín. The matrix in the top left corner of each graph reports the peak-to-peak distance between the median of the distribution, multiplied by a factor of 10^{-3}.

We see a consistent gender distinction with men displaying higher values of H than women within the same socioeconomic group. We also continue to observe that peak-to-peak distances are smaller over time. Indeed, on average, the gender differences within each socioeconomic group tend to decrease over time, suggesting that the gender gap might be getting smaller (Table 4.6).

We notice that the greater differences between genders occur for travellers belonging to the `upper` group (Table 4.6). Similar conclusions are drawn from the *mobility diversity* computed for travels made for `work` purposes by all combinations of gender and socioeconomic status (Figure 4.19). We highlight that we can not reject the null hypothesis that the distributions of the `work` travels are statistically similar (tested by Welch's t-test with p-value < 0.01) in two cases for `MDE` in 2017: (i) all and men travellers, and (ii) men of the upper income group. We also notice that the values of *mobility diversity* are different from those obtained from the null models (see Section 4.6). On average, the gender-centred differences within each socioeconomic group tend to decrease over time, suggesting that a possible gender-level difference in mobility is, indeed, reducing (Table 4.6).

We assess the contribution of the gender and socioeconomic attributes (alone

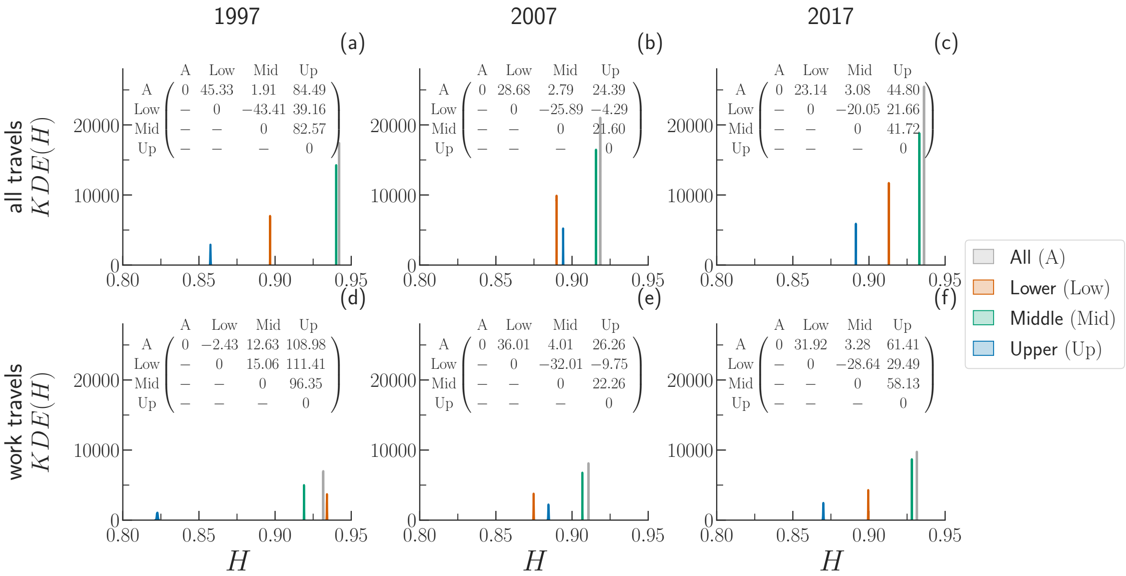

Figure 4.17: Kernel Density Estimation (KDE) plots of the *mobility diversity* H for `all` travels (panels **a**, **b**, and **c**), and `work` travels (panels **d**, **e**, and **f**) in São Paulo. The matrix in the top left corner of each graph reports the peak-to-peak distance between the median of the distribution, multiplied by a factor of 10^{-3}.

and combined) in the *mobility diversity* H applying three statistical tests. First, we use the ANOVA one-way test to compare if the averages of the *mobility diversity* distributions computed separately by either the gender or socioeconomic status groups are statistically different. Then, we apply the ANOVA two-way test to investigate if the averages of the *mobility diversity* distributions computed by gender and socioeconomic status together are statistically different. Finally, we apply Tukey's HSD post hoc test to identify within attributes which groups display statistically different average values of H. All the detailed explanations and specific values of F and p-values from ANOVA and Tukey's HSD post hoc tests are detailed in Section 4.4. Based on these tests, we can reject the hypothesis that the mean values of the *mobility diversity*, H, from the travels performed by gender (men, women and all travellers) and socioeconomic groups (`lower`, `middle`, `upper` and all travellers) are similar. When considering only gender or socioeconomic status, there is no exception in the statistical tests.

Lastly, we compare the *mobility diversity* distributions of gender and socioeconomic status taken together. We can also reject the null hypothesis that the mean values of the *mobility diversity*, H, distributions are similar from the majority of pairwise comparisons, except in `MDE` for the `work` travels in 2017 performed by (i) all and men travellers; (ii) all and men for the upper income group. Moreover, the value of H computed for the entire population is higher than that of the travels made by travellers of a given gender and socioeconomic group. Such a difference is due to the fact that gender and socioeconomic status play a role in concentrating the travels from many areas to fewer areas. Sections 4.5, 4.6 and 4.7 detail the effect of sample size, residential distribution, spatial tessellation and endogenous mobility.

Summing up, in general, the patterns in *mobility diversity* for different groups of gender and socioeconomic status taken separately or together are statistically different. Without exception, we claim that the socioeconomic group consistently accounts for the highest gap in *mobility diversity*. Thus, depending on the gender and socioeconomic group, individuals can be more likely to concentrate their travels in fewer areas indicating that the job distribution might also be more concentrated for certain sociodemographic groups.

Table 4.6: Gender differences, $median(H_S^M) - median(H_S^W)$, of the *mobility diversity* H of travels made for all and work purposes by travellers grouped according to their socioeconomic status, S, and gender, $X \in \{M, W\}$. The values report the peak-to-peak distance between the median of the distribution of H, multiplied by a factor of 10^{-3}. The values in bold represent the cases in which $median(H_S^W) > median(H_S^M)$.

City	Year	Purpose	lower	middle	upper
MDE	2005	all	13.0	3.3	26.4
		work	20.9	3.3	11.9
	2017	all	0.4	5.4	18.2
		work	0.9	8.4	18.2
BGT	2012	all	14.6	19.0	19.6
		work	16.3	7.8	36.1
	2019	all	3.3	0.1	**-1.8**
		work	7.4	4.9	**-0.6**
SAO	1997	all	13.7	0.5	19.5
		work	5.9	0.0	46.2
	2007	all	**-0.2**	**-8.0**	6.9
		work	0.7	**-7.3**	3.6
	2017	all	2.7	0.9	4.1
		work	12.3	2.4	10.7

4.4 Statistical verification of the *mobility diversity* distributions

To account for variations in sample sizes, we applied a bootstrapping technique that estimates the distribution of *mobility diversity* ensuring that fluctuations and outliers do not bias our analyses. We, then, apply statistical tests to compare the distributions of *mobility diversity* across groups. We used three tests: the Welch's t-test [162], the ANOVA [163], and the Tukey's HSD post hoc test [164]. Welch's t-test establishes whether two distributions - in our case, the distributions are the *mobility diversity* (H) values - are similar. Using the ANOVA test, we compare whether the average values of *mobility diversity* of two or more groups are statistically similar. Lastly, Tukey's HSD post hoc test points out the groups with different average values of mobility diversity when the ANOVA test rejects the hypothesis null of similar average values of the distributions.

We have shown in Section 4.3 that all the pairwise comparisons using Welch's t-test rejected the null hypothesis except for `MDE` in 2017 between (i) `all` and `men` travellers and (ii) `all` and `men` travellers from upper income group. Therefore, most of the pairwise comparisons between the *mobility diversity* distributions are statistically different.

Now, we apply the ANOVA one-way and two-way tests to assess if the distributions indicate similar average values of H. We present in Tables A.6 and A.7 all the values of F-statistic and p-value of the ANOVA test computed from the *mobility diversity*, H, of travels made for `all` and `work` purposes by travellers aggregated by gender and socioeconomic status.

Starting with the gender groups, we can reject the null hypothesis that the mean values of H from men, women, and all travellers regardless of the purpose of travels are statistically the same because the p-values are smaller than 0.01 and the F-values are not small. Next, we apply Tukey's HSD post hoc test to discover the specific groups that the distributions are not statistically similar. We observe in Tables A.8 - A.14 that the p-values from the multi-group means comparisons of the `women`, `men` and `all` distributions of the *mobility diversity* between the different purpose of travels are smaller than 0.01. The only exception is for `MDE` in 2017 when comparing the values of H from the `work` travels grouped by `all` and `men` travellers.

We then repeat the same statistical analyses considering the socioeconomic groups. The socioeconomic groups presented statistically different distributions and average values of H using the ANOVA test and Tukey's HSD post hoc test.

Finally, we apply the two-way ANOVA test to analyse the relationship between gender and socioeconomic status in measuring *mobility diversity*. In Tables A.6 and A.7, the values of F are small; hence the outcome of the ANOVA test does not exclude that the *mobility diversity*, H, of travellers belonging to different socioeconomic and gender groups could belong to the same distribution. To exclude such possibility, we decided to apply Tukey's HSD test. From all the multi-group means comparisons of the distributions of the *mobility diversity* between different sets of travels, Tables A.8 - A.14, we can reject that the values of H of the following distributions are from populations with the same mean values.

In summary, we conclude that the gender and socioeconomic groups, alone and combined, show different distributions of *mobility diversity*. This result indicates gender and socioeconomic effect on how each group distribute their travels throughout metropolitan areas.

4.5 Effects of changing the sample size

Because sociodemographic groups are not uniformly distributed, we have a higher percentage of travels, for example, from men and middle income group. Therefore, these groups with a higher percentage of travel have higher chances of exploring more different areas than groups with a smaller percentage of travel. However, a minimal amount of data is sufficient to well-represent the groups using a similar number of travels or percentage of travels per group. Thus, in this section, we explore the possible effect that differences in the sample size can bias our results.

We start by fixing the number of travels across groups and looking at the impact of the *mobility diversity* as we increase the number of travels for these groups.

We consider the range from 1,000 travels up to the maximum possible size for each region.

In respect to gender, Figures 4.20- 4.22 display the values of H as we increase the size of the sample for `BGT`, `SAO`, and `MDE`. There is a saturation of the values of H generally before the 60% number of travels per group (black symbols on Figures 4.20- 4.22). We continue to see the same qualitative results in which men consistently show higher values of H than women, and the differences between groups are smaller for the most recent year. Even though the magnitude is impacted, the gender effect maintains present in our results.

Now that we ensure that the effect of gender did not depend on the sample size, we move to the same analysis for the socioeconomic factor. Figures 4.23- 4.25 show the effects of changing the sample size for the three socioeconomic groups considered. We conclude that even though the sample size is important for a better picture from the differences in the *mobility diversity* across socioeconomic groups, the sample size does not dictate or change the robustness of our analysis.

Next, we look at the effects of changing the sample size for travels grouped according to gender and socioeconomic status simultaneously. Figures 4.26 and 4.27 do not highlight any qualitative difference with the hierarchies displayed in Figures 4.18 and 4.19. Notwithstanding, the values of H displayed in Figures 4.26 and 4.27 are not the same. Such a difference is due to the fact that the real population is not made of groups of equal size. Therefore, we conclude that the dissimilarity of group sizes should be taken into account when computing H. Nevertheless, we argue that our results are robust when we ensure that the group sample size is the same.

Finally, as our data do not have a similar sample size across groups, we use `BGT` data to argue that using 60% of the data is enough to capture the saturated *mobility diversity* values. Figure 4.28 shows the values of H as we increase the size of the sample for `BGT` area considering the percentage values per group. Using 10% of each data sample is already enough to indicate the differences across groups. However, we argue that at 60%, we have enough stability in the values of H per group. In summary, we conclude that sample size does not affect our conclusions or the validity of our results.

4.6 Null models

In order to ensure that our results do not reflect a simple physical and human mobility behaviour, we compare our results of H with computed H from five null models (NMx with $x \in \{1, \ldots, 5\}$) in respect to three aspects: *i*) the partitioning of the urban area into zones of different size, *ii*) the heterogeneity of the distance of travels, and *iii*) the population that live in a zone. NM1 being the least realistic model, and NM5 being the most realistic one. Figure 4.29 summarises the characteristics of all the null models.

We generate for all the models a similar number of travels (e.g. for `SAO` we generate 248,000 travels), which is proportional to 1,000 multiplied for the number of zones for each region. For the case of taking into account where people live, the number of people going out from a zone z is proportional to the density of people residing in the zone z. The distributions of travel distance are from a uniform random distribution and truncated power-law distribution [3]. Such distributions range between 100 and 60,000 meters, corresponding to the minimum and maximum distances estimated in our data. If a destination point of a generated travel falls outside the urban area, we extract a new point until its position falls within the urban area. After generating the travels, we compute first the probability that their destinations fall within a certain zone, i, using Equation (4.2), and then the value of H using Equation (4.1). Finally, we average the results over 1,000 realisations.

Starting from the most naive model, NM1 accounts only for the tessellation of the urban area by spreading random points (i.e. travels' destinations) over the urban area. Therefore, the probability that a travel ends in a given zone is only proportional to the zone's area. NM2 expands slightly NM1 by taking into account the travel distance limits (100 meters until 60,000 meters). NM3 is similar to NM2 with the exception that the travel's distance is extracted from a truncated power-law function. Such a difference translates into the presence of more short range travels which, in turn, corresponds to more travels towards neighbour zones if the origin zone is small, and more travels within the same zone if the origin zone is big, instead. NM4 is similar to NM2 but it accounts for the heterogeneity in the density of people living in each zone. This means that the number of travels starting from a given zone is proportional to the number of people living in it. Finally, NM5 is the most realistic model and differentiates from NM4 because the travel distance follows a truncated power-law distribution.

Tables 4.7 and 4.8 display the values of H computed for the spatial distributions of destinations in all the areas, years, purpose, and by the groups considered in our study. In addition, the tables also contain the values of H corresponding to the destinations' distributions generated by each null model. We performed both the Kolmogorov-Smirnov and Welch's t-test on each combination (pair) of empirical and synthetic distributions of H and, we have found that all of the combinations reject the null hypothesis of the aforementioned tests with a p-value smaller than 0.001.

In Section 4.5, Figures 4.26 and 4.27 display the violin plots of H computed using the highest possible value of sample size across groups. In this section, instead, we observe in Table 4.8 how much the median values of H for each group differ from the same quantity computed via the null models. The difference between these values indicates that taken together gender and socioeconomic status exert a remarkable effect on the value of the *mobility diversity*.

Table 4.7: Summary of the values of the *mobility diversity*, H, of empirical data and null models. For each area, year, and purpose of travel we report the values of H computed for all travels ($H_{\texttt{all}}$), gender ($H_{\texttt{men}}$, $H_{\texttt{women}}$), and socioeconomic status ($H_{\texttt{lower}}$, $H_{\texttt{middle}}$, $H_{\texttt{upper}}$). We report also the value of H computed using different null models (NMx with $x \in \{1, \dots, 5\}$) averaged over 1,000 realisations.

Area	Year	Travel Type	$H_{\texttt{all}}$	$H_{\texttt{men}}$	$H_{\texttt{women}}$	$H_{\texttt{lower}}$	$H_{\texttt{middle}}$	$H_{\texttt{upper}}$	$\langle H \rangle$				
									NM1	NM2	NM3	NM4	NM5
MDE	2005	all	0.9158	0.9160	0.9127	0.8235	0.8783	0.7281	0.6008	0.7379	0.9643	0.7020	0.8847
		work	0.8797	0.8757	0.8704	0.8648	0.8701	0.7548					
	2017	all	0.9030	0.9039	0.9013	0.8965	0.8594	0.7356	0.6008	0.7379	0.9643	0.7085	0.9230
		work	0.9042	0.9045	0.8994	0.8959	0.8601	0.7344					
BGT	2012	all	0.9075	0.9127	0.8950	0.7923	0.8826	0.6862	0.3902	0.6123	0.9680	0.9261	0.9485
		work	0.9077	0.8960	0.8910	0.8937	0.8782	0.7388					
	2019	all	0.9366	0.9368	0.9356	0.9264	0.9339	0.8968	0.3902	0.6123	0.9680	0.5988	0.9243
		work	0.9285	0.9270	0.9245	0.9081	0.9220	0.8458					
SAO	1997	all	0.9443	0.9441	0.9420	0.9014	0.9402	0.8623	0.7674	0.8885	0.9930	0.8760	0.9483
		work	0.9359	0.9312	0.9318	0.9363	0.9234	0.8382					
	2007	all	0.9184	0.9155	0.9198	0.8810	0.9143	0.8901	0.7674	0.8885	0.9930	0.8862	0.9446
		work	0.9179	0.9143	0.9184	0.8819	0.9125	0.8864					
	2017	all	0.9357	0.9359	0.9341	0.9060	0.9321	0.8778	0.7674	0.8885	0.9930	0.8839	0.9533
		work	0.9354	0.9360	0.9313	0.9001	0.9320	0.8744					

Table 4.8: Summary of the values of the *mobility diversity*, H, of empirical data and null models. For each area, year, and purpose of travel we report the values of H computed for the travels performed by each group considering the gender and socioeconomic status together: Men Lower class ($H_{\mathrm{M,L}}$), Men Middle class ($H_{\mathrm{M,M}}$), Men Upper class ($H_{\mathrm{M,U}}$), Men Lower class ($H_{\mathrm{W,L}}$), Women Middle class ($H_{\mathrm{W,M}}$) and Women Upper class ($H_{\mathrm{W,U}}$). We report also the value of H computed using different null models (NMx with $x \in \{1, \ldots, 5\}$) averaged over 1,000 realisations.

Area	Year	Travel Type	$H_{\mathrm{M,L}}$	$H_{\mathrm{M,M}}$	$H_{\mathrm{M,U}}$	$H_{\mathrm{W,L}}$	$H_{\mathrm{W,M}}$	$H_{\mathrm{W,U}}$	$\langle H \rangle$				
									NM1	NM2	NM3	NM4	NM5
MDE	2005	all	0.8275	0.8777	0.7309	0.8145	0.8744	0.7045	0.6008	0.7379	0.9643	0.7020	0.8847
		work	0.8608	0.8610	0.7234	0.8400	0.8577	0.7115					
	2017	all	0.8959	0.8612	0.7425	0.8955	0.8558	0.7243	0.6008	0.7379	0.9643	0.7085	0.9230
		work	0.8931	0.8603	0.7360	0.8921	0.8518	0.7177					
BGT	2012	all	0.7939	0.8858	0.6835	0.7793	0.8668	0.6640	0.3902	0.6123	0.9680	0.9261	0.9485
		work	0.8752	0.8579	0.7128	0.8588	0.8501	0.6767					
	2019	all	0.9272	0.9332	0.8904	0.9239	0.9330	0.8922	0.3902	0.6123	0.9680	0.5988	0.9243
		work	0.9055	0.9184	0.8175	0.8980	0.9135	0.8180					
SAO	1997	all	0.9053	0.9388	0.8623	0.8916	0.9383	0.8428	0.7674	0.8885	0.9930	0.8760	0.9483
		work	0.9278	0.9175	0.8283	0.9219	0.9175	0.7821					
	2007	all	0.8789	0.9092	0.8886	0.8791	0.9173	0.8817	0.7674	0.8885	0.9930	0.8862	0.9446
		work	0.8763	0.9072	0.8785	0.8756	0.9145	0.8749					
	2017	all	0.9041	0.9314	0.8742	0.9013	0.9305	0.8701	0.7674	0.8885	0.9930	0.8839	0.9533
		work	0.8973	0.9310	0.8705	0.8850	0.9286	0.8598					

4.7 Effects of endogenous and residential based travels on *mobility diversity*

In this section, we investigate whether the differences across *mobility diversity* disappears when we disregard the mobility performed (i) inside areas (endogenous) or (ii) to the areas where people live. Analysing these two scenarios help us identify whether the differences found in *mobility diversity* are mainly driven by residential distribution. We acknowledge that these two scenarios do not take into account travels that their origin is from home. Our goal is not to show that residential distribution does not play a role in mobility, but that destination area distribution differs from the residential landscape.

Removing endogenous mobility (first scenario) represents a removal of 27%/8% from `all`/`work` travels (Table 3.7). For the second scenario, residential-based travels, we remove between 4% and 38% of the travels depending on the purpose of the travel, region and year (Table 3.12). Thus, we compute the *mobility diversity* based on the mobility for the two scenarios considering at least 62% of the sample size data.

Figure 4.30 presents the violin plots of H computed for the sets of non-endogenous travels made either for `all` (panels a-c), or `work` (panels d-e) purpose. We compare the average values of H displayed in Figures 4.11 and 4.30. In general, we observe a decrease of H when disregarding endogenous travels, especially for `SAO`, and a greater gender difference. Besides, the removal of endogenous travels affects the diversity of `work` travels more than for the `all` travels.

We then repeat the comparison for the case of travels excluding the residential-based travels (Figure 4.31). As for the non-endogenous travels case, we also observe a general decrease of the values of H together with an amplification of the gender differences. Therefore, the gender differences persist even when we disregard endogenous and residential-based travels.

Next, we investigate how the gender differences will be affected when considering also the socioeconomic dimension. Figures 4.32 and 4.33 are the non-endogenous scenario of Figures 4.18 and 4.19, whereas Figures 4.34 and 4.35 account for the non-residential scenario.

We again see that there is an effect in the magnitude of *mobility diversity* for both scenarios. We observe a generalised decrease in $\langle H \rangle$ as well as, for a given socioeconomic status, an amplification of the gender differences, $\Delta H = \left| \langle H^W \rangle - \langle H^M \rangle \right|$ (with indices W and M denoting `women` and `men`). Endogenous and residential mobility clearly affects the *mobility diversity*, but the extent of both the decrease in $\langle H \rangle$ and increase of ΔH is not constant neither between regions nor across years or travel's purpose.

To quantify the impact on the values of H, we measure the peak-to-peak difference between pairs of $KDE(H)$, obtained for travels made by men (M) and women (W) belonging to socioeconomic status S, $\Delta g = median(H_S^M) - median(H_S^W)$ in Tables 4.9 and 4.10. The definition of Δg is the sign of the peak-to-peak difference determining the hierarchy between men and women. We observe how in most cases, the values of Δg are bigger than the same quantity computed, including also endogenous and residential travels, regardless of the travel's purpose. In some cases (e.g. `SAO`), the values of Δg are even bigger than the maximum difference computed considering all the travels, $\max(\Delta g^\star)$. Moreover, `work` travels differences appear to be those more affected by the removal of endogenous and residential mobility. Finally, we notice that sometimes the hierarchy between men and women gets inverted (denoted by highlighted cells).

In conclusion, gender differences persist when we disregard endogenous and residential-based mobility. Therefore, we argue that our analyses persist in being robust to the residential segregation, and differences in the way women and men move also come from other factors of mobility and society.

Table 4.9: Gender based differences of the Kernel Density Estimator $KDE(H)$ of the *mobility diversity*, Δg, for travels made for `all` and `work` purposes having different origin and destination zones made by travellers grouped according to their socioeconomic status, $S \in \{$`lower`, `middle`, `upper`$\}$, and gender. The values of Δg are multiplied by a factor of 10^3. Highlighted cells represent the cases in which $\Delta g < 0$ (i.e. $median(H_S^W) > median(H_S^M)$). Column G denotes the gender of the more diverse travellers (M for men and W for women). The symbol $>$ ($<$) in column T denotes differences whose values are — in absolute value — bigger (smaller) than the one obtained taking into account all travels. The value of $\max(\Delta g^\star)$ is computed from Table 4.6.

City	Year	Purpose	$\max(\Delta g^\star)$	`all`			`lower`			`middle`			`upper`		
				Δg	G	T	Δg	G	T	Δg	G	T	Δg	G	T
MDE	2005	`all`	12.10	4.21	M	$<$	6.30	M	$>$	1.45	M	$<$	9.47	M	$<$
		`work`	-8.07	32.15	M	$>$	54.41	M	$>$	51.83	M	$<$	1.72	W	$<$
	2017	`all`	16.15	3.11	M	$>$	1.55	W	$>$	7.21	M	$>$	15.33	M	$<$
		`work`	19.67	324.73	M	$>$	201.37	M	$>$	137.85	M	$>$	50.04	M	$>$
BGT	2012	`all`	15.39	2.97	W	$<$	1.24	M	$<$	0.06	W	$<$	21.13	M	$>$
		`work`	33.89	10.36	M	$>$	16.05	M	$>$	16.13	M	$>$	46.09	M	$>$
	2019	`all`	2.05	2.10	M	$<$	5.28	M	$>$	0.33	M	$<$	2.22	W	$<$
		`work`	5.15	2.41	M	$<$	9.46	M	$>$	0.76	W	$>$	13.76	M	$>$
SAO	1997	`all`	16.05	7.52	M	$>$	14.55	M	$>$	7.73	M	$>$	17.20	M	$>$
		`work`	30.75	13.18	M	$>$	22.92	M	$>$	15.57	M	$>$	50.80	M	$>$
	2007	`all`	-5.97	39.92	M	$>$	36.43	W	$>$	62.20	W	$>$	5.53	M	$<$
		`work`	9.95	47.08	M	$>$	66.70	W	$>$	94.36	M	$>$	116.40	M	$>$
	2017	`all`	6.51	37.53	M	$>$	62.91	M	$>$	8.25	M	$>$	19.63	W	$>$
		`work`	15.34	110.94	M	$>$	176.97	M	$>$	119.60	M	$>$	36.16	W	$>$

Table 4.10: Gender based differences of the Kernel Density Estimator $KDE(H)$ of the *mobility diversity*, Δg, for travels made for `all` and `work` purposes having not the residential zone of the traveller as destination zone made by travellers grouped according to their socioeconomic status, $S \in \{$`lower`, `middle`, `upper`$\}$, and gender. See the caption of Table 4.9 for notations and definitions.

City	Year	Purpose	$\max(\Delta g^\star)$	`all`			`lower`			`middle`			`upper`		
				Δg	G	T	Δg	G	T	Δg	G	T	Δg	G	T
MDE	2005	`all`	12.10	4.44	M	$>$	9.11	M	$>$	1.74	M	$<$	8.51	M	$<$
		`work`	-8.07	36.50	M	$>$	58.29	M	$>$	53.54	M	$>$	9.84	M	$>$
	2017	`all`	16.15	1.71	M	$<$	1.49	W	$>$	5.87	M	$>$	13.23	M	$<$
		`work`	19.67	290.01	M	$>$	157.93	M	$>$	166.62	M	$>$	52.31	M	$>$
BGT	2012	`all`	15.39	2.78	W	$<$	1.25	M	$<$	0.23	M	$<$	20.61	M	$>$
		`work`	33.89	9.76	M	$>$	15.50	M	$>$	15.43	M	$>$	43.84	M	$>$
	2019	`all`	2.05	2.16	M	$>$	5.03	M	$>$	0.32	M	$<$	0.43	W	$<$
		`work`	5.15	2.79	M	$>$	9.48	M	$>$	0.34	W	$>$	15.93	M	$>$
SAO	1997	`all`	16.05	6.98	M	$>$	14.55	M	$>$	7.73	M	$>$	17.20	M	$>$
		`work`	30.75	12.15	M	$>$	15.37	M	$>$	7.09	M	$>$	10.90	M	-
	2007	`all`	-5.97	18.92	M	$>$	35.16	W	$>$	25.35	M	$>$	16.93	W	$>$
		`work`	9.95	34.27	M	$>$	52.87	W	$>$	55.11	M	$>$	79.61	M	$>$
	2017	`all`	6.51	20.15	M	$>$	67.04	M	$>$	5.90	M	$>$	24.00	W	$>$
		`work`	15.34	93.56	M	$>$	161.93	M	$>$	98.13	M	$>$	30.85	W	$>$

4.8 Discussion

The literature extensively analyses and models the general patterns in human mobility [3, 38, 40–42, 165]. However, little is known about the differences between sociodemographic groups in respect to mobility. This chapter demonstrated that for three cities, Medellín, Bogotá and São Paulo, gender and socioeconomic status alone and combined impact mobility diversity.

We measure the spatial diversity of the travels performed by different groups of people (e.g., gender and socioeconomic groups) modifying well-known metrics from information theory, Shannon entropy [156]. *Mobility diversity* computes specifically the probability that a zone is a destination from the mobility of a particular group for a specific purpose of travel. Entropy can also be used as a proxy of the predictability of a group's mobility as low entropy can indicate higher predictability [39, 157]

Purpose of travels (e.g. going to work and home) shape mobility such as the probability of returning to the last visited location [40], the fraction of travels over time [166], the most frequent visited locations [3], and the amount of money spent related to the distance travelled [88]. We unveiled the role played by the following *purpose* of the travel: one made of travels due to work activities (`work`), one made of travels due to any purpose except work (`nonwork`), and another made by all travels together regardless of their purpose (`all`). We presented that the travels related to the purpose of working are more concentrated in a few areas than the other types of travels. This is the case because some areas offer a higher volume of job opportunities than others attracting more people to certain areas. We also observed that each urban area has a different evolution over time, suggesting that the economic context might be strongly intertwined with mobility. Therefore, we should further analyse other countries to understand the universality of our results.

From the perspective of gender, our analysis revealed the existence of a distinction between the mobility of men and women, with the men being more entropic/diverse than women. The conclusion is independent of the region, time, and purpose of the travel. This difference in mobility diversity between genders can be related to other differences already found in mobility, such as women tend to make shorter travels than men and avoid travelling to certain destinations, particularly during late hours [26, 43, 45–47, 167]. We have also observed that the gender differences in mobility diversity get smaller over time, which can be a result of policies, actions and regulations to reduce gender inequality in several sectors such as transportation and security [25, 27, 28, 108–110].

Turning now to the role of the socioeconomic status, travellers are divided into three socioeconomic groups: lower, middle and upper classes, and a greater effect is shown in mobility diversity compared to gender effect. We see a clear distinction among socioeconomic groups, with `upper` being the group exploring the space in the least diverse way, and `middle` travellers displaying the most diverse mobility patterns [31, 44, 119, 120]. We argue that for the `upper` income group, individuals appear to be more selective in their destinations, where they live and also move less possibly because they can afford a broad range of options (i.e. buying a car or living in expensive areas closer to where they work) [43, 159, 168, 169].

Finally, we showed that gender differences are amplified when the socioeconomic dimension is considered. Nevertheless, men continue to show higher values of H than women. The highest gender difference is found from the travellers belonging to the `upper` income group, and this might be in line with the fact that there is higher gender inequality in highly qualified jobs - women being less likely to be hired or promoted [170–172].

We replicated our results for five null models to ensure that differences from mobility diversity were not solely a byproduct of random processes considering spatial partitioning, travel distance distribution, and residential distribution. We also ensure that the results of mobility diversity are consistent regardless of the sample size. Finally, we performed extra analyses on the effect of endogenous and residential-based mobility, indicating that the gender differences persist. Thus, we conclude that our results hold robust to all the mentioned concerns.

Works presented in the past that gender of the travellers do not play a remarkable role in the predictability of mobility behaviour [39, 60]. However, recent works, instead, have shown that gender indeed plays a role in the discrimination of travellers [46, 108, 173]. This might be true because differences (peculiarities) in the behaviour of women and men are in the opposite direction of universalities.

Recently, literature addresses the role of gender and socioeconomic status of travellers separately in mobility [26, 43, 134, 174], but to the best of our knowledge, their combined effect is a novelty in the literature. We presented here a systematic new pattern in the mobility of how gender and socioeconomic status are intertwined. Women report disadvantages in several aspects of their life such as income, free time, and career's progression [110], and this chapter shows that mobility is another aspect in which women suffer [25, 26, 46, 175].

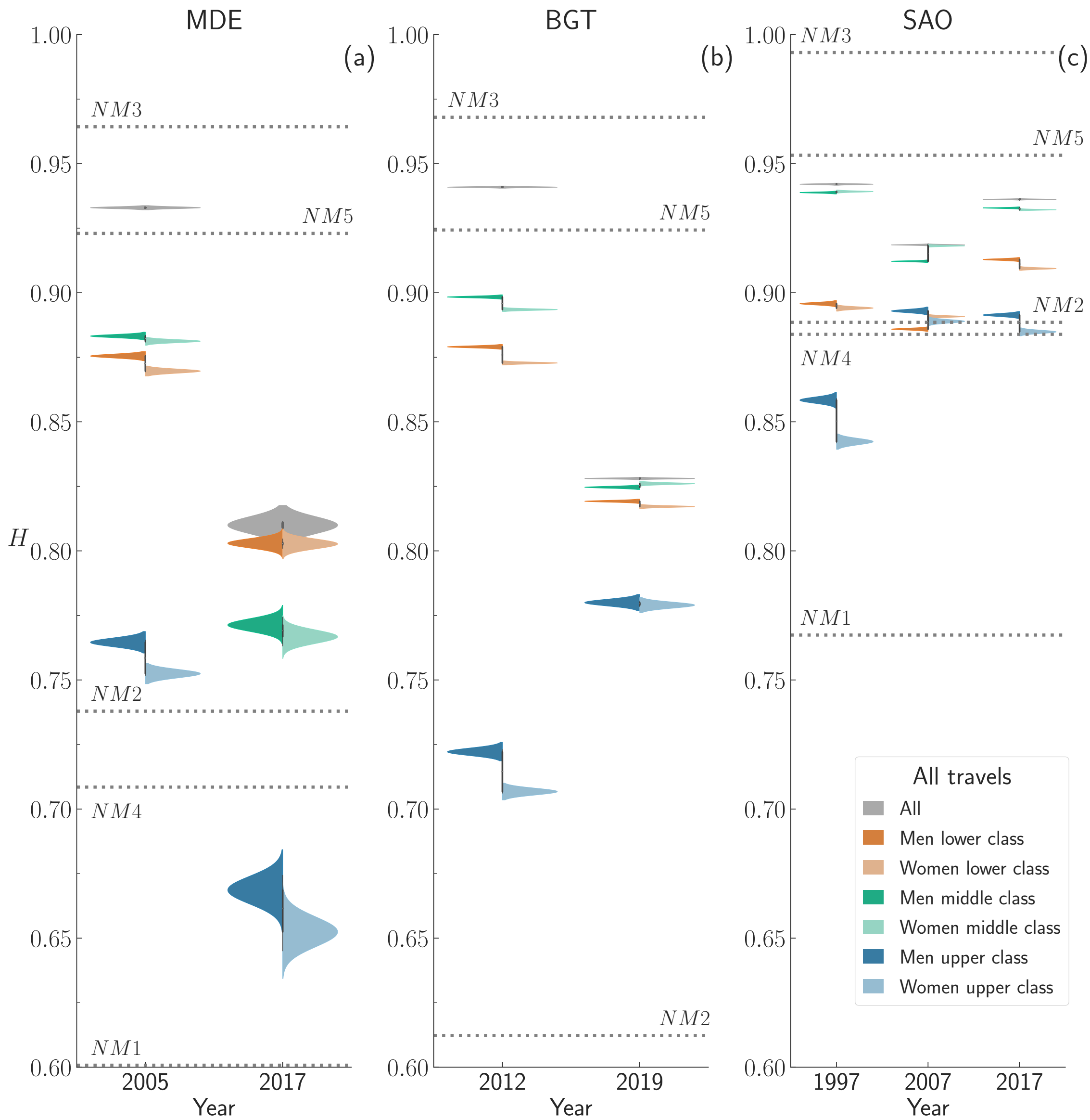

Figure 4.18: Violin plots of the *mobility diversity*, H, of travels made for `all` purposes by travellers grouped according to their socioeconomic status and gender. Each column refers to a different region, and for each region, we consider all the available years. For each socioeconomic status (`upper`, `middle`, and `lower`) a darker hue denotes men travellers, whereas lighter hue denotes women ones. Dotted lines in grey denote the values of H computed from travels generated using null models (See details in Section 4.6).

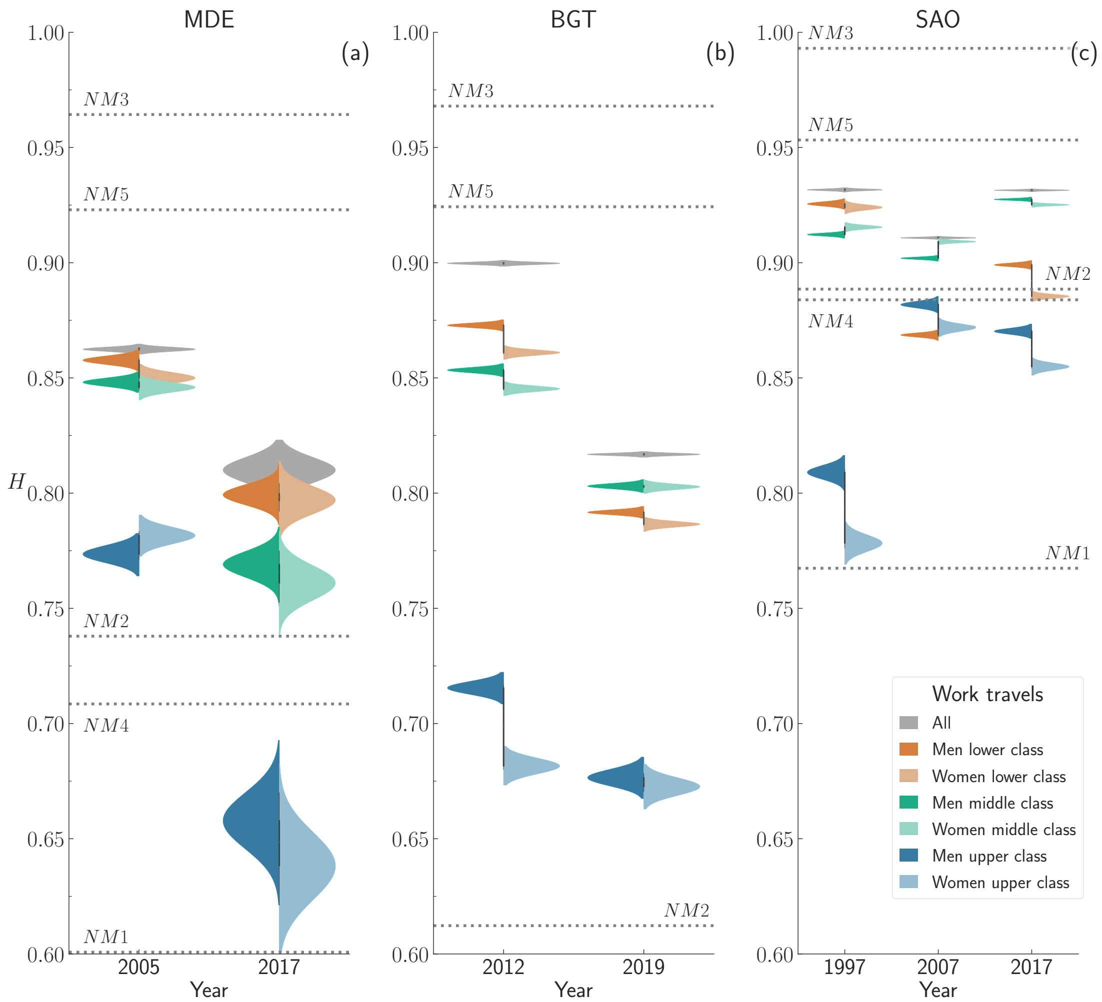

Figure 4.19: Distribution of the *mobility diversity*, H, for travels made by `work` purposes by travellers grouped according to their socioeconomic status and gender. Each column refers to a different region, and for each region, we consider all the available years. For each socioeconomic status (`upper`, `middle`, and `lower`) a darker hue denotes men travellers, whereas a lighter hue denotes women ones. Dotted grey lines display the values of *mobility diversity* for each null model (see Section 4.6 for the details).

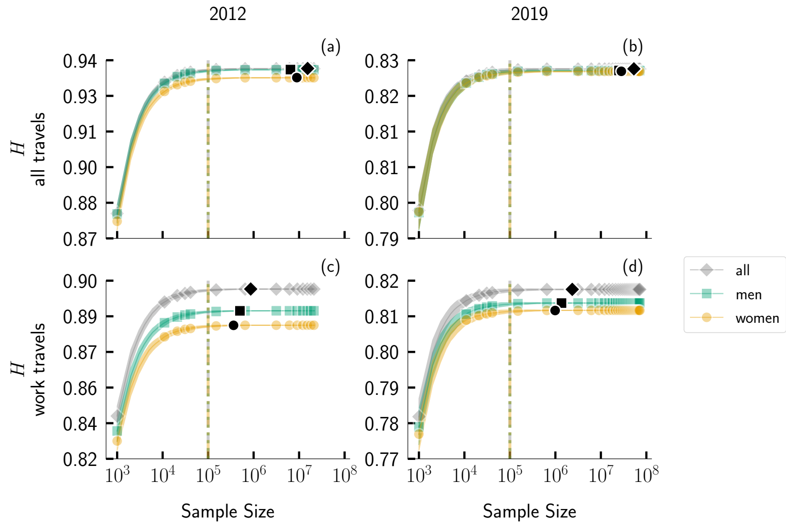

Figure 4.20: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by gender in `BGT`. We consider either `all` travels (panels **a** and **b**), or `work` travels (panels **c** and **d**) only. The shaded area accounts for the standard deviation of the values obtained from averaging the results over 1,000 realisations. Each column accounts for a different year. The vertical lines denote the size from which the values of *mobility diversity* stabilise. The black symbols correspond to the same quantity obtained using a sample size equal to 60% of all the travels available.

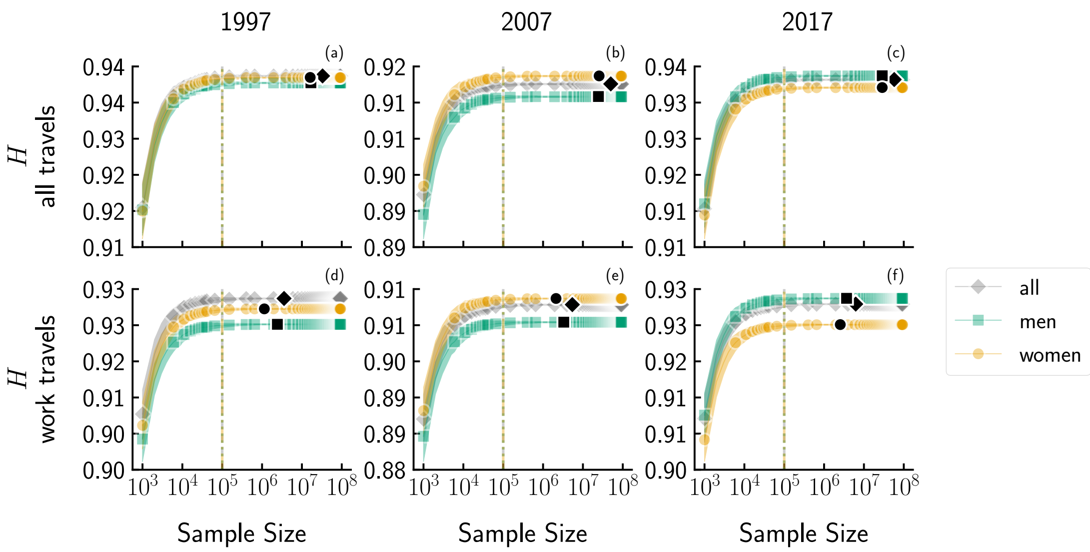

Figure 4.21: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by gender in `SAO`. We consider either `all` travels (panels **a**, **b**, and **c**), or `work` travels (panels **d**, **e**, and **f**) only. See the caption of Figure 4.20 for the description of the notation.

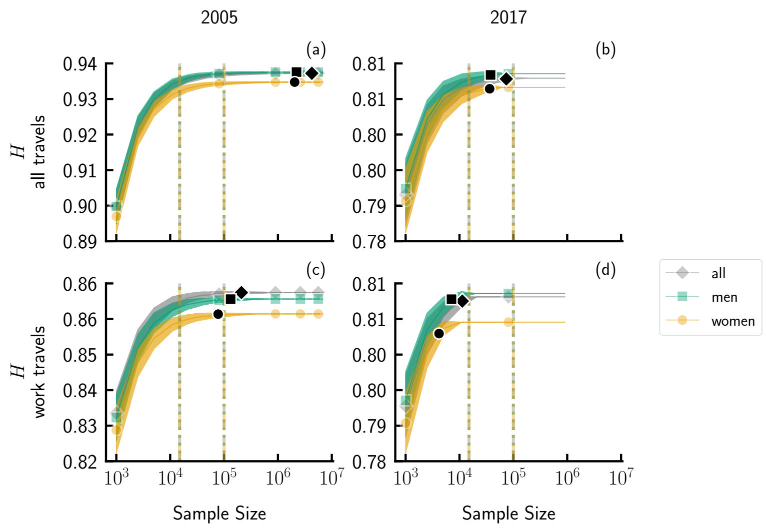

Figure 4.22: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by gender in `MDE`. We consider either `all` travels (panels **a** and **b**), or `work` travels (panels **c** and **d**) only. See the caption of Figure 4.20 for the description of the notation.

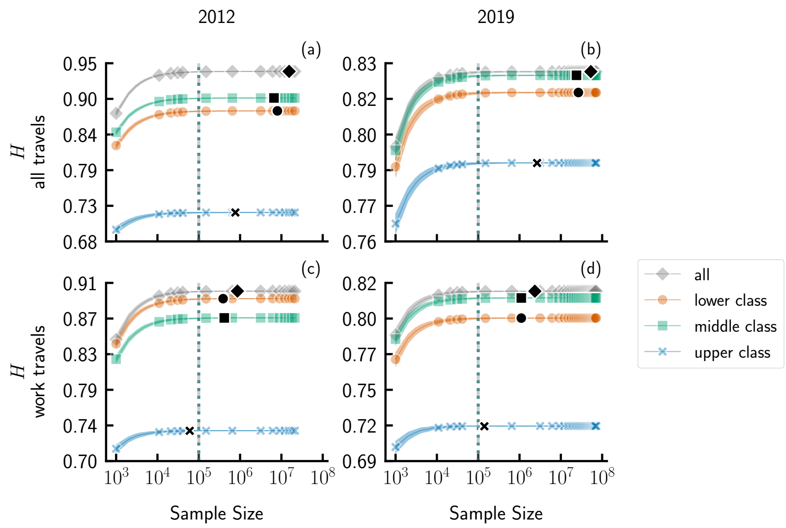

Figure 4.23: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by socioeconomic groups in `BGT`. We consider either `all` travels (panels **a** and **b**), or `work` travels (panels **c** and **d**) only. See the caption of Figure 4.20 for the description of the notation.

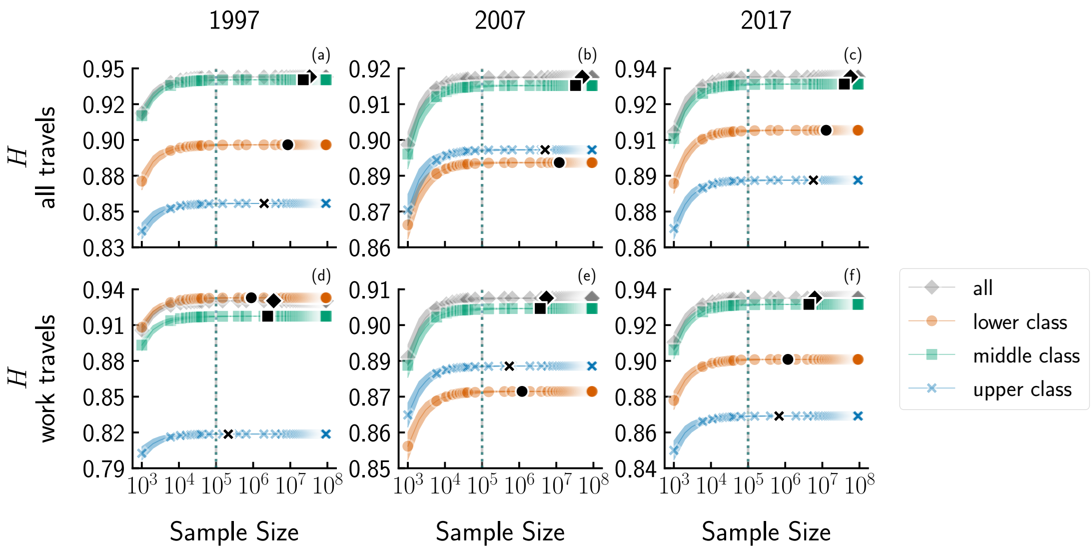

Figure 4.24: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by socioeconomic groups in `SAO`. We consider either `all` travels (panels **a**, **b**, and **c**), or `work` travels (panels **d**, **e**, and **f**) only. See the caption of Figure 4.20 for the description of the notation.

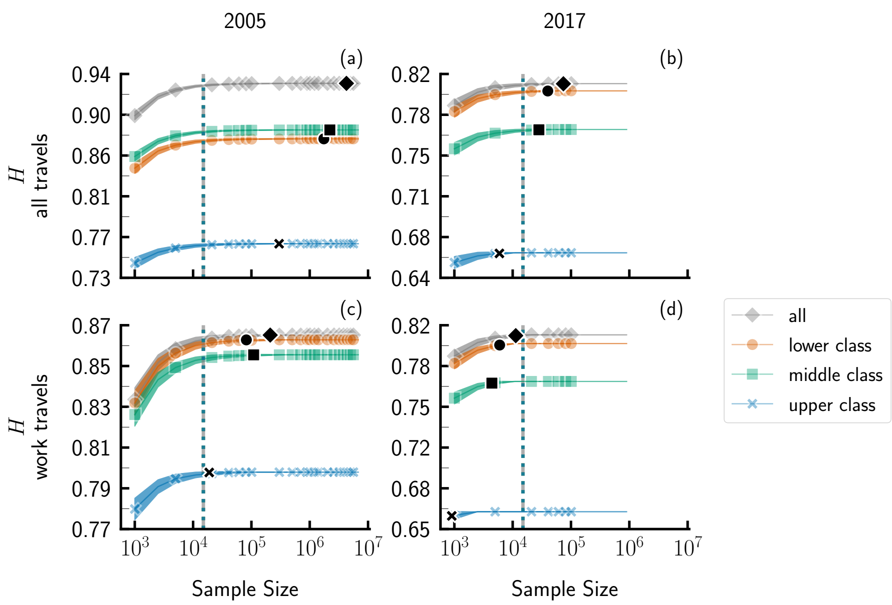

Figure 4.25: Values of the *mobility diversity*, H, for different sample sizes for travels made by travellers grouped by socioeconomic groups in `MDE`. We consider either `all` travels (panels **a** and **b**), or `work` travels (panels **c** and **d**) only. See the caption of Figure 4.20 for the description of the notation.

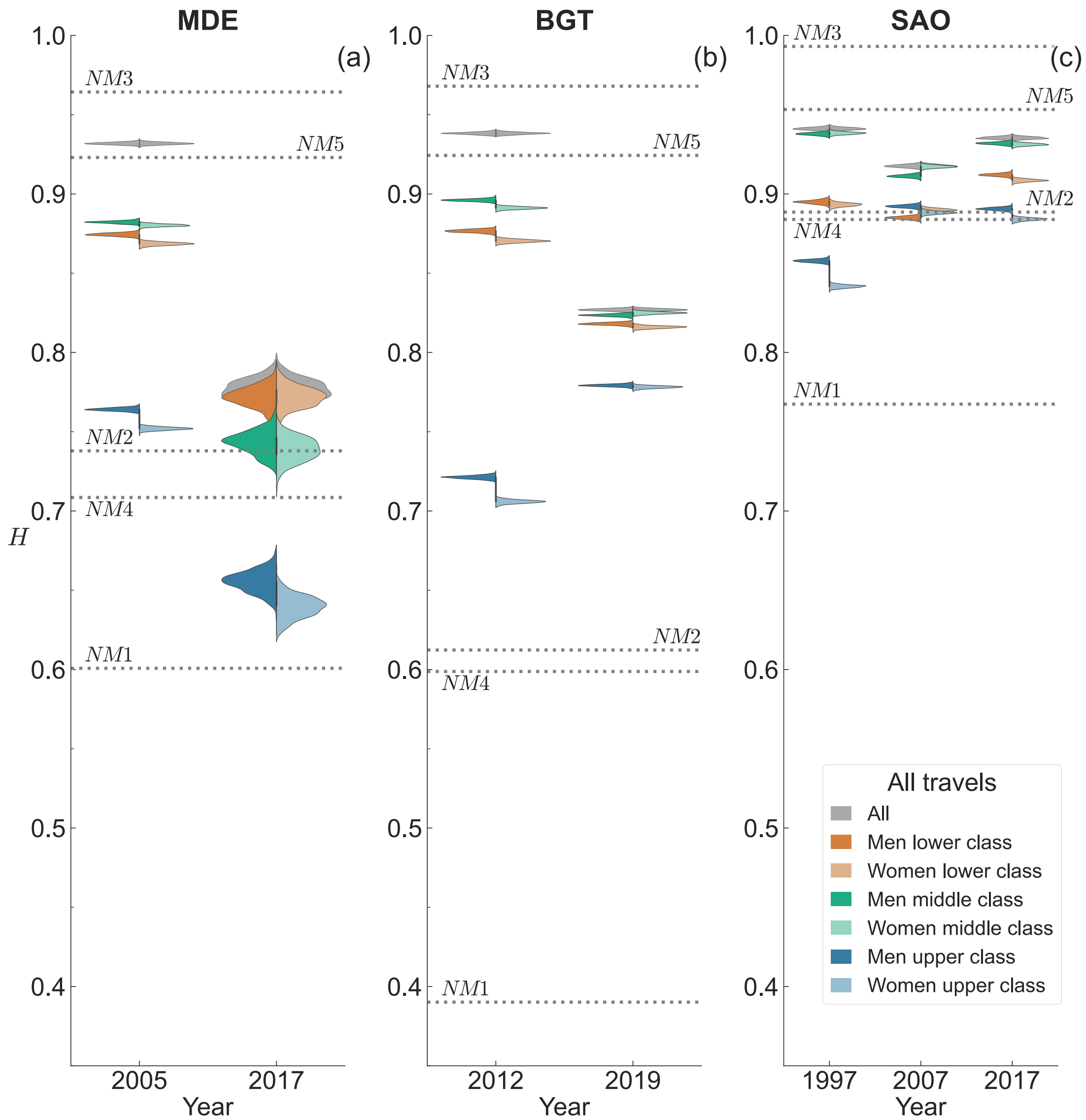

Figure 4.26: Distribution of the *mobility diversity*, H, for 25,000 travels (except `MDE` for 2017 that is 550 travels) in the sample size range made by `all` purposes by travellers grouped according to their socioeconomic status and gender. Each column refers to a different region, and for each region, we consider all the available years. For each socioeconomic status (`upper`, `middle`, and `lower`) darker hue denotes men travellers, whereas lighter hue denotes women ones. The dotted lines denote the values of H computed using travels generated by each null model further explained in Section 4.6.

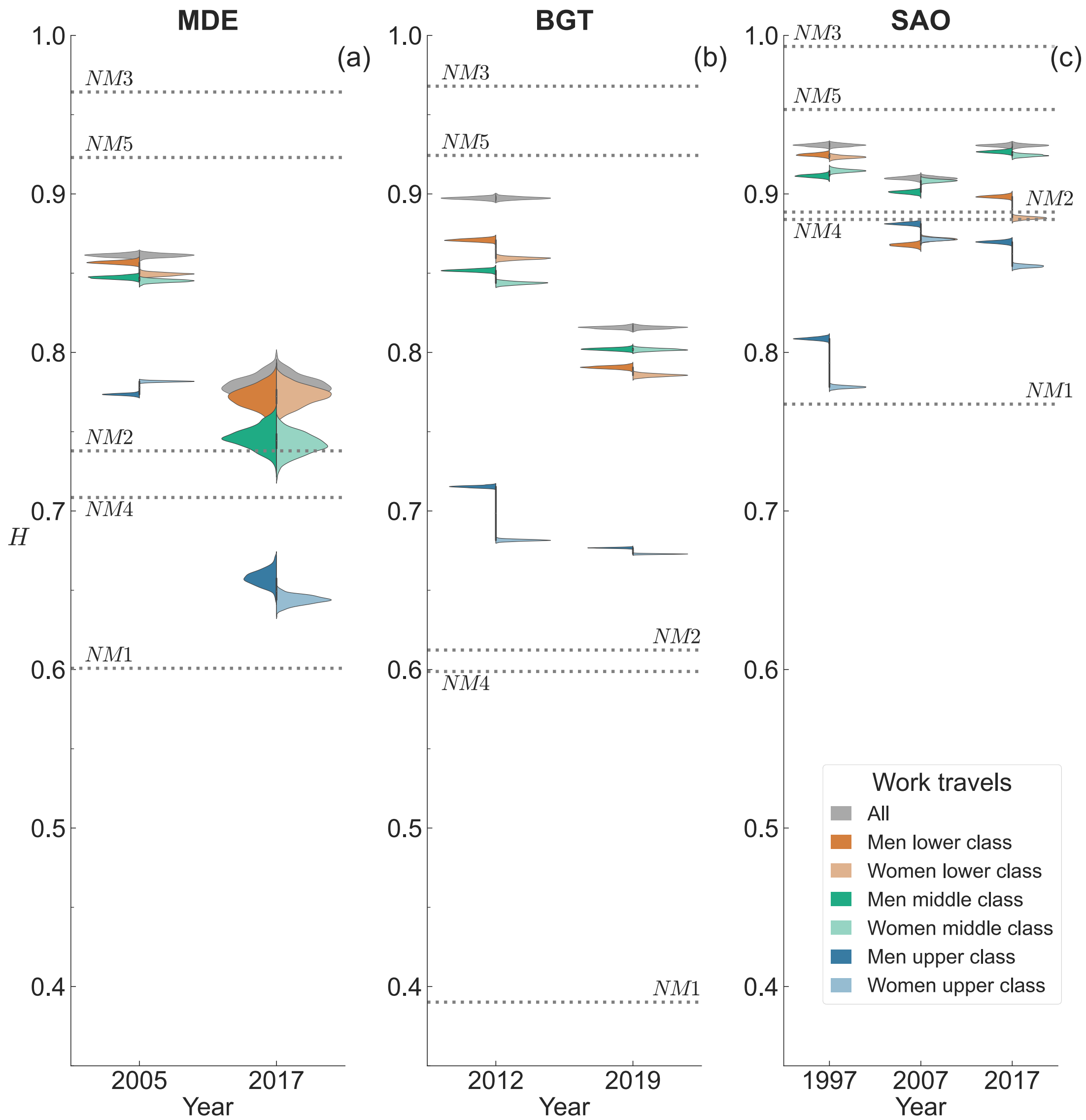

Figure 4.27: Distribution of the *mobility diversity*, H, for the maximum number of travels in the sample size range made by `work` purposes by travellers grouped according to their socioeconomic status and gender. Each column refers to a different region, and for each region, we consider all the available years. For each socioeconomic status (`upper`, `middle`, and `lower`) darker hue denotes men travellers, whereas lighter hue denotes women ones. The dotted lines denote the values of H computed using travels generated by each null model further explained in Section 4.6.

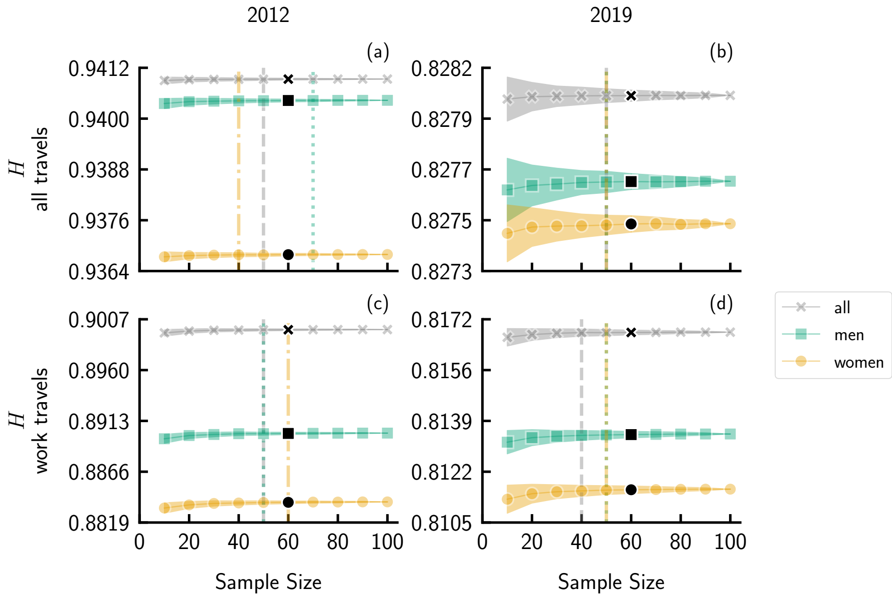

Figure 4.28: Values of the *mobility diversity*, H, for different sample sizes (in percentage) of travels made by travellers grouped by gender in `BGT`. We consider either `all` travels (panels **a** and **b**), or `work` travels (panels **c** and **d**) only. The shaded area accounts for the standard deviation of the values obtained from averaging the results over 1,000 realisations. Each column accounts for a different year, and the dashed lines represent the saturation of the values of the *mobility diversity*.

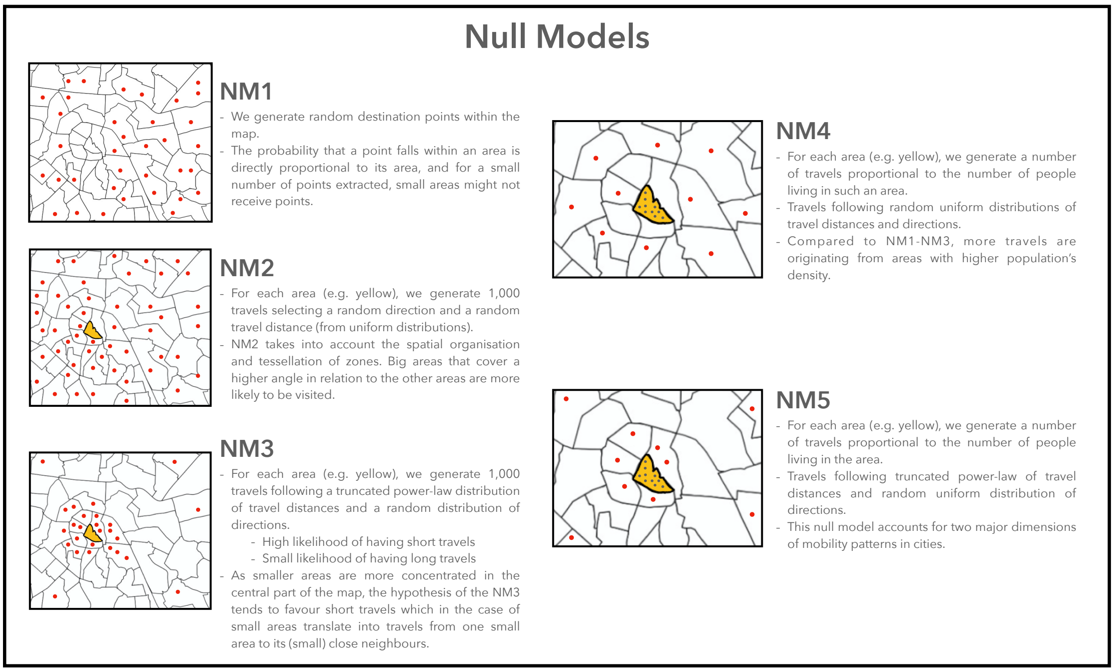

Figure 4.29: Schematic summary of the main features of the null models considered. For each null model, we list its main properties. Red dots appearing in the maps denote the travels' destinations.

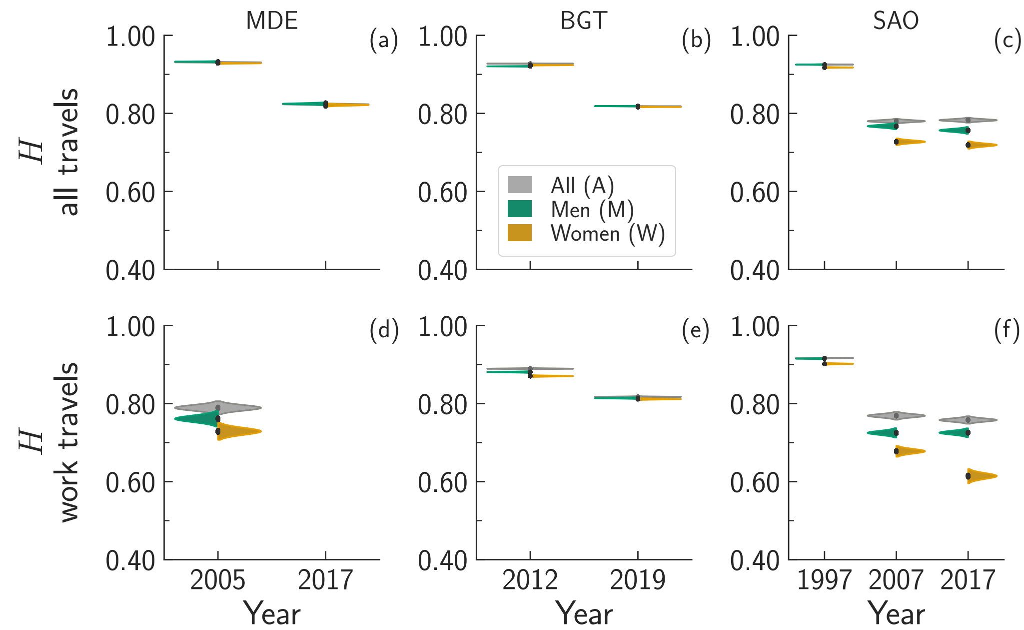

Figure 4.30: Violin plots of the bootstrapped *mobility diversity*, H, for travels having different origin and destination zones. The top row (panels a-c) accounts for travels made for `all` purposes, whereas the bottom row (panels d-e) displays the results for `work` travels. The data for year 2017 in the `MDE` area are missing as they are too scant.

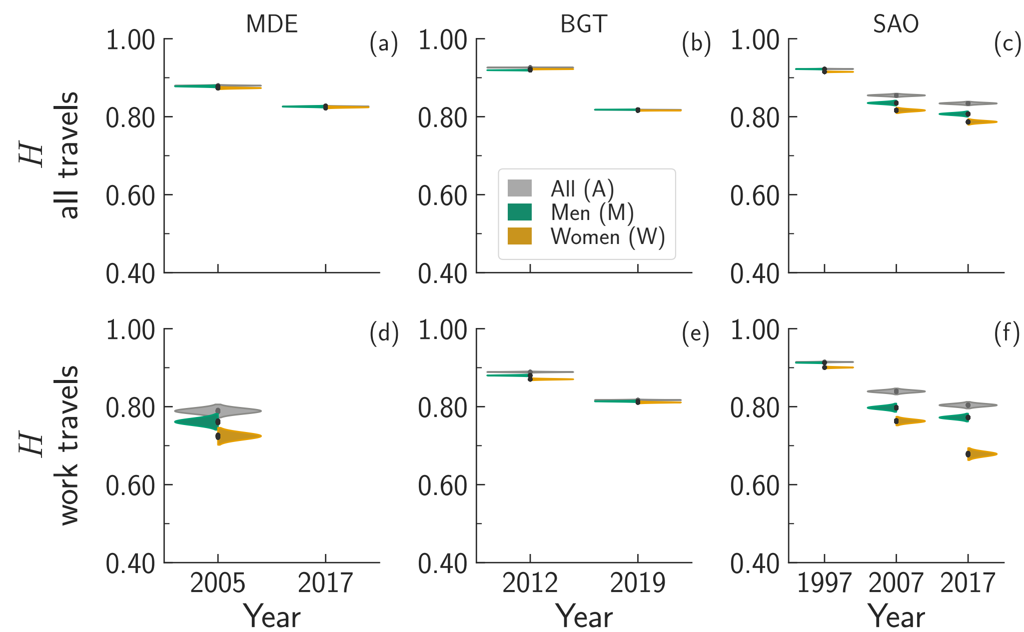

Figure 4.31: Violin plots of the bootstrapped *mobility diversity*, H, for travels whose destination zone does not coincide with the zone where the traveller lives. See the caption of Figure 4.30 for the notation's details and other information.

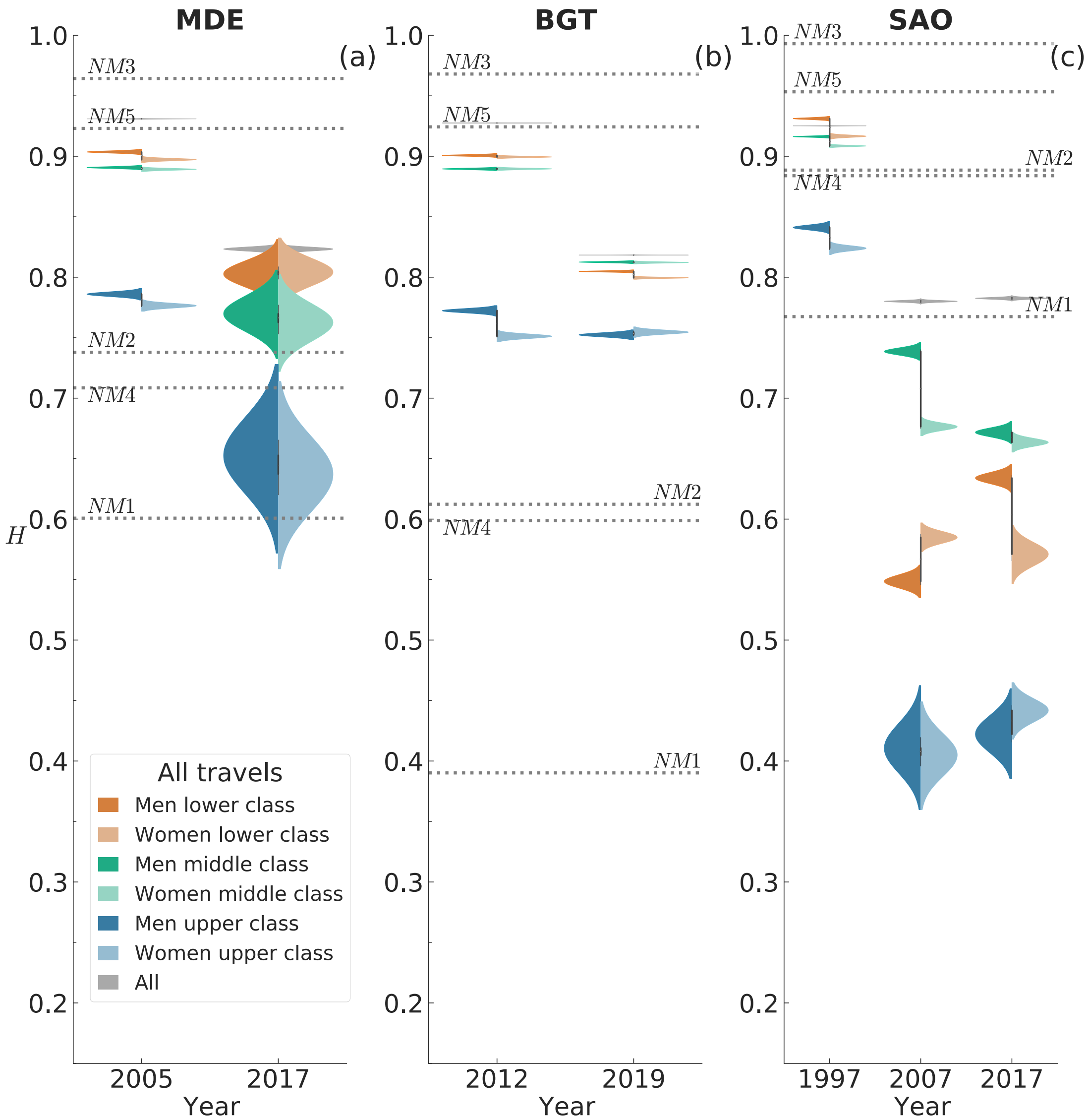

Figure 4.32: Violin plots of the *mobility diversity*, H, of travels made for `all` purposes having different origin and destination zones made by travellers grouped according to their socioeconomic status and gender. Each plot refers to a different region and, for each region, we consider all the available years. For each socioeconomic status (`upper`, `middle`, and `lower`) a darker hue denotes men travellers, whereas lighter hue denotes women ones. Dotted lines in grey denote the values of H computed from travels generated using null models NMx with $x \in \{1, \ldots, 5\}$ (see Section 4.6).

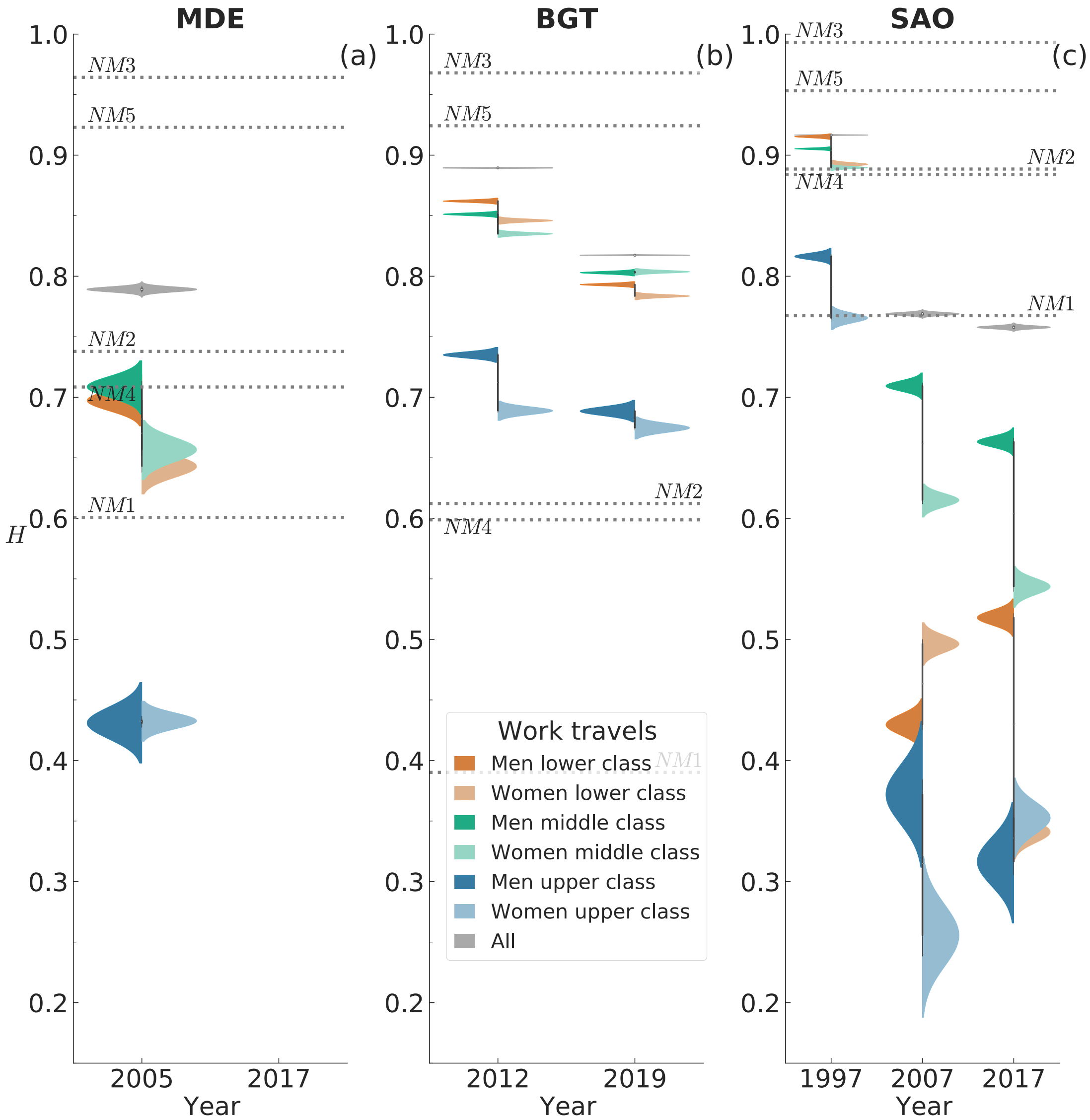

Figure 4.33: Violin plots of the *mobility diversity*, H, of travels made for `work` purposes that the origin and destination are different by travellers grouped according to their socioeconomic status and gender. See the caption of Figure 4.32 for further details.

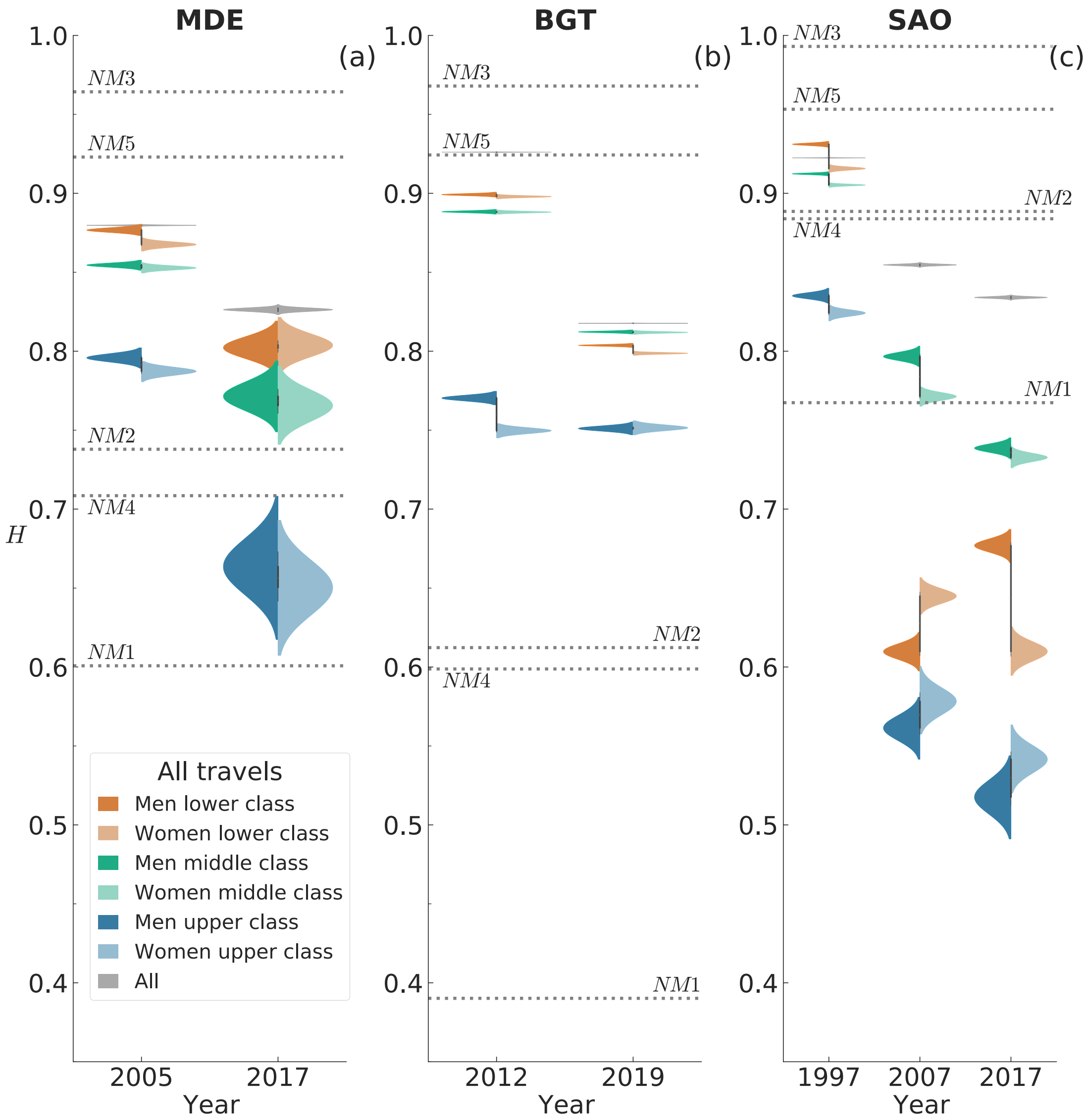

Figure 4.34: Violin plots of the *mobility diversity*, H, of travels made for `all` purposes whose destination zone is different from the traveller's home zone and made by travellers grouped according to their socioeconomic status and gender. See the caption of Figure 4.32 for further details.

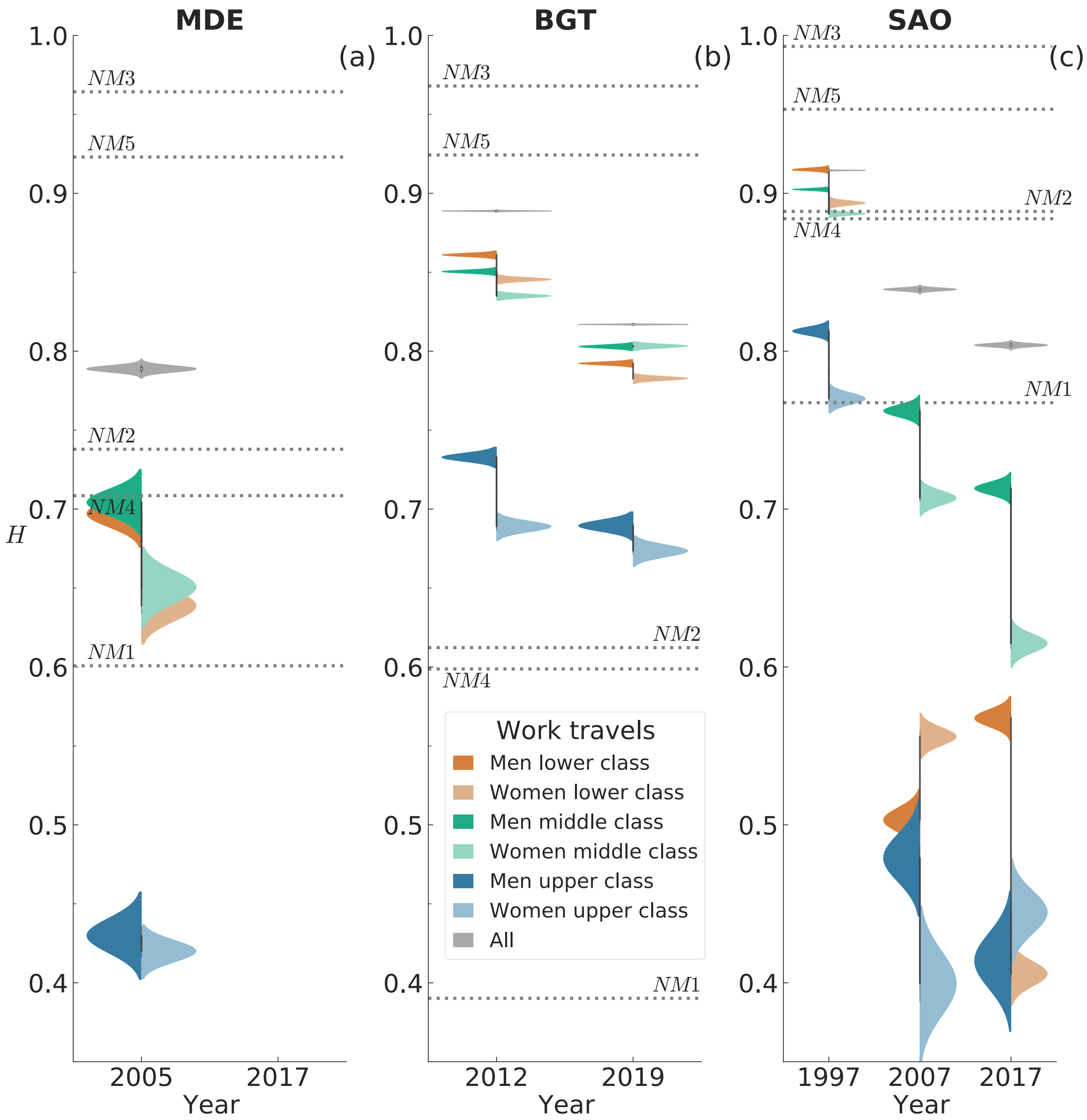

Figure 4.35: Violin plots of the *mobility diversity*, H, of travels made for `all` purposes whose destination zone is different from the traveller's home zone and made by travellers grouped according to their socioeconomic status and gender. See the caption of Figure 4.32 for further details.

5. Estimating mobility cost and reward

Urban areas are getting more crowded by people and amenities (e.g., restaurants, buildings and offices) [32]. Urban development together with a high density of people and amenities may lead to several issues such as traffic jams [146], and high real estate prices [176]. In some cities, the transportation system cannot accommodate such a high volume of people and fails to address the needs across sociodemographic groups [46]. Thus, the interaction between population, amenities and transportation systems might be important to the urban development in the direction of fairness, equality and sustainability.

Walking, cycling or using public transportation are still not preferable choices for those that have a high-income level or higher education degree [177]. In fact, women tend to opt for private transportation modes when they can afford to do so. However, given the gender differences in pay scale [178] and job promotions [172], women and men may have different budgets to access transportation modes. These gender gaps in the labour market often translate into mobility characteristics such as travel time. This is an example of how different income or education levels may lead to differences in the mobility patterns of women and men. Besides, the literature shows that mode of transportation and mobility perception can affect the well-being of women and men. For instance, walking and cycling tend to make women less stressed [179].

As we observed in Chapter 3 and 4, travellers from different gender and socioeconomic groups exhibit distinct patterns of mobility with respect to how they distribute their travels in space and time. Therefore, their mobility needs may differ.

For instance, as we observed in Chapter 3, in multiple cities, men tend to go to work earlier than women, then, it might be the case that the addition of buses and trains in the early morning would benefit men more than women. Studying the needs of different sociodemographic groups can clarify how interventions and policies can be introduced to improve urban issues, such as traffic jams and crowded buses. Thus, opportunities may be more accessible for some sociodemographic groups, causing these groups to be exposed to a different set of opportunities [8, 180].

We argue that the urban landscape distribution of amenities, transportation and facilities may be used differently by certain sociodemographic groups. For instance, areas that have a high population volume are generally more visited by men than women (showed in Chapter 3). We define in this chapter two mobility patterns: cost and reward. These two patterns are major factors in how mobility is modelled, as seen in Section 2.5, and how mobility differs between women and men.

Mobility cost is related in this thesis to how long it takes to travel between locations. Literature found that women are more inclined to have shorter travel times than men [13, 27], and in this chapter, we revisit this finding for many cities. Specifically, Craig and van Tienoven [111] argue that women tend to have shorter travels because they dedicate more hours to household and care responsibilities, requiring them to be near to where they live. Therefore, gender roles likely affect their commuting travel time differently, but it also depends on the household arrangement; hence we also explore the effect of household arrangements on the gender differences of the commuting travel time.

Mobility reward is defined in this thesis as the volume and diversity of amenities that people are exposed throughout their journey. We argue that having access to amenities (opportunities) is beneficial to anyone [112, 181, 182]. We show evidence that higher gender differences in the diversity of amenities may indicate higher gender differences in the commuting travel time. We also show that urban landscape (city organisation) affects gender differences in mobility. Therefore, changes in the urban landscape can potentially impact the gender differences in mobility.

5.1 Data

In this chapter, we use data from São Paulo (`SAO`) [72], Medellín (`MDE`) [73], Bogotá (`BGT`) [129], the metropolitan region of London in the United Kingdom [77] and our selected cities of the USA using the American Community Survey (ACS) [76]. Further details of these data are described in Section 3.1.

We also use data from OpenStreetMap [78] extracted from the categories: amenities, highways and buildings. These three categories list locations that have establishments and infrastructures related to several daily demands such as: eating (restaurants), going to work by bus (bus stops) and going to the doctor (hospitals). This thesis uses the term amenities to refer to all these establishments and infrastructures extracted from OpenStreetMap.

To avoid inconsistency across cities and regions, we group all these amenities into the following categories: education, food, health, leisure, services, transport and work. We highlight that some categories have more amenities than others, and some cities have more amenities than others. For example, bus stops present high volume and spatial concentration for some cities than others [183] impacting the number of amenities in the transport category. Moreover, the uneven distribution of amenities is consistent with the centralization of amenities in cities [184].

To compute the number of amenities in a zone, we use two approaches: (i) the number of amenities spatially inside a zone and (ii) the number of amenities within a certain geodesic distance from the centroid of a zone. We use a radius varying from 0.5 to 5.0 kilometres. For the first case, we estimate the composition of amenities in a zone, and in the second case, we estimate, for a given distance, the number of amenities that individuals can reach from the zone centre.

5.2 Mobility cost

Cultural norms impose women to spend more time on household and care responsibilities than men, potentially affecting how women choose where to work [12]. Women are more likely to use public transportation than men [46, 48], and women tend to go

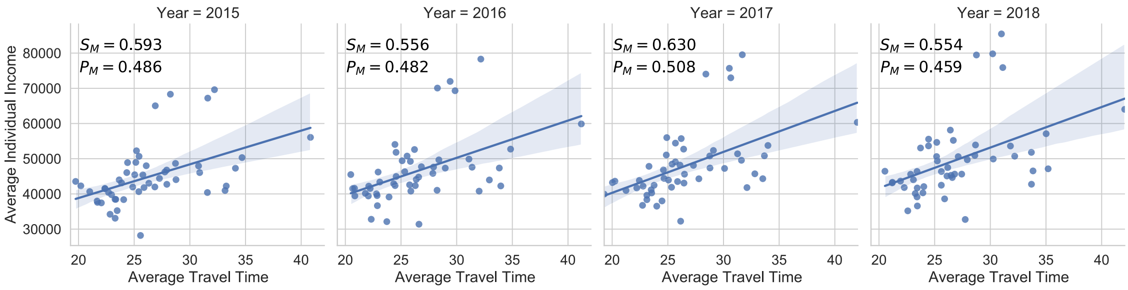

Figure 5.1: Average travel time versus average income level for the cities in the USA. We display the Spearman Correlation, S_M, and the Pearson Correlation, P_M, between the average travel time and average income level.

to work later than men (see Chapter 3). Hence, women and men likely have different mobility costs because they make distinct decisions.

We define the mobility cost based on the time taken to travel between locations. There are other ways to define the mobility cost, but here we argue that travel time is a key component for mobility, as people usually plan their mobility considering the arrival time and the mode of transportation, and in line with the literature [3]. Moreover, our hypothesis is that travel time can be associated with other dimensions such as income level. We observe that travel time is indeed associated with individual income for the cities in the USA, as shown in Figure 5.1. In this way, people with higher income also have higher travel time, indicating that travel time might be associated with the mobility cost individuals can afford. Gustafson [45] presents that higher income levels are related to a higher likelihood of travel for women and men, and we observe that it also increase the magnitude of the travel time.

First, we explore the average commuting travel time for women and men. Figure 5.2 shows the complementary cumulative distribution function of the commuting travel time, t, in minutes for women and men for some cities in five North American states (New York, California, Texas, Colorado and Florida). We see a trend that men tend to have a higher commuting travel time than women. To analyse whether this trend is consistent for all the cities across Brazil, Colombia, the United Kingdom and the United States of America, we perform a bootstrapping methodology to ensure that fluctuations and outliers in the data do not bias our conclusions. This methodology is similar to the one performed in Chapter 4.

For the bootstrapping, we randomly draw 80% of men and women travellers

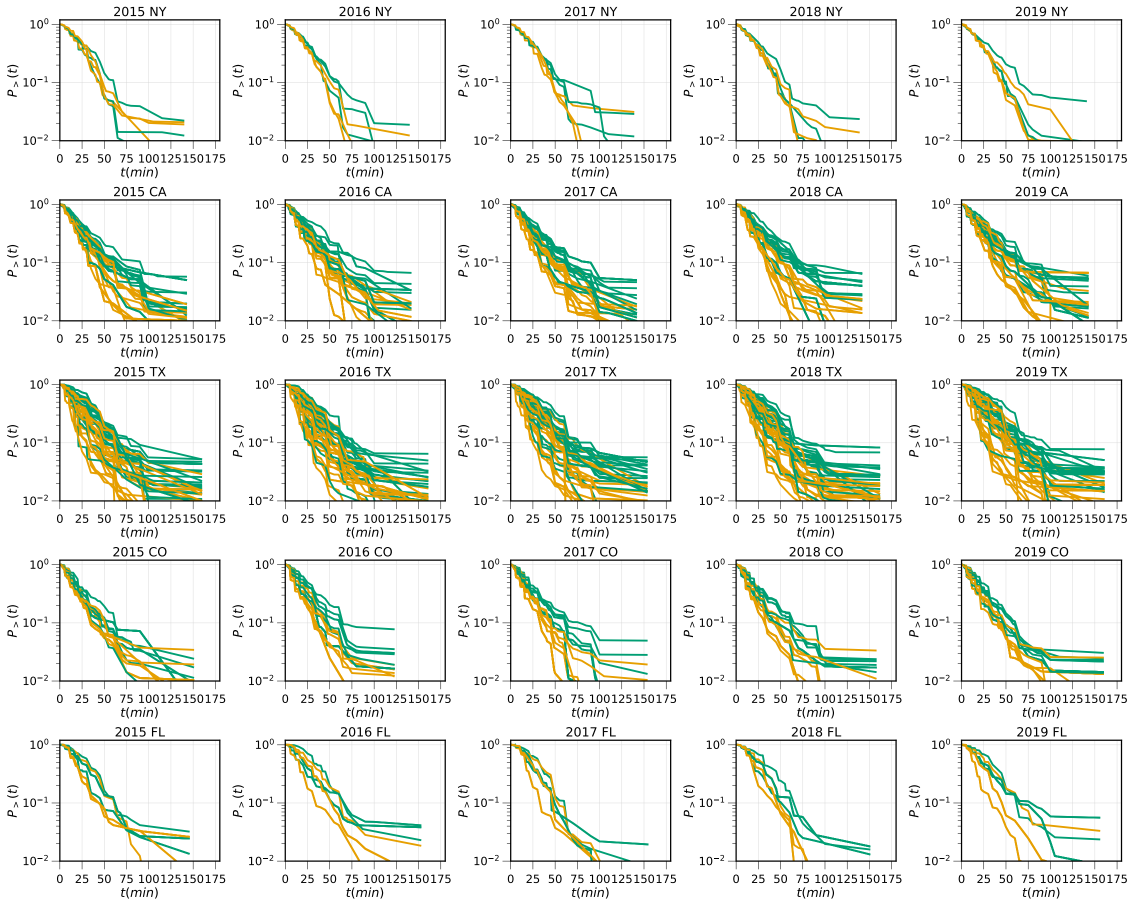

Figure 5.2: Complementary cumulative distribution function of the commuting travel time, $P_>(t)$, for cities inside five states: New York (NY), California (CA), Texas (TX), Colorado (CO) and Florida (FL).

from each city and year, making sure that each pair (random men and women) is from the same residential area. We compute the average travel time for these random sets and repeat these processes 1000 times. We then compare the average travel time distribution of women and men using Welch's t-test to check whether the mean of these distributions is statistically different. In the end, we compute the difference between the average travel time for women and men, with respect to each city and year, to ultimately compare it with the statistical test.

We group the travellers in the following household arrangements: (i) `single`; (ii) `married`; (iii) `parent`; and (iv) `without children`. We also combine more than one household arrangement, for example, `single parent` travellers.

The final results of the bootstrapping is shown in Figure 5.3. We observe that `women` travellers' average commuting travel time is smaller than the one from `men` travellers in most cities and years. Gender, alone, seems to play a major role

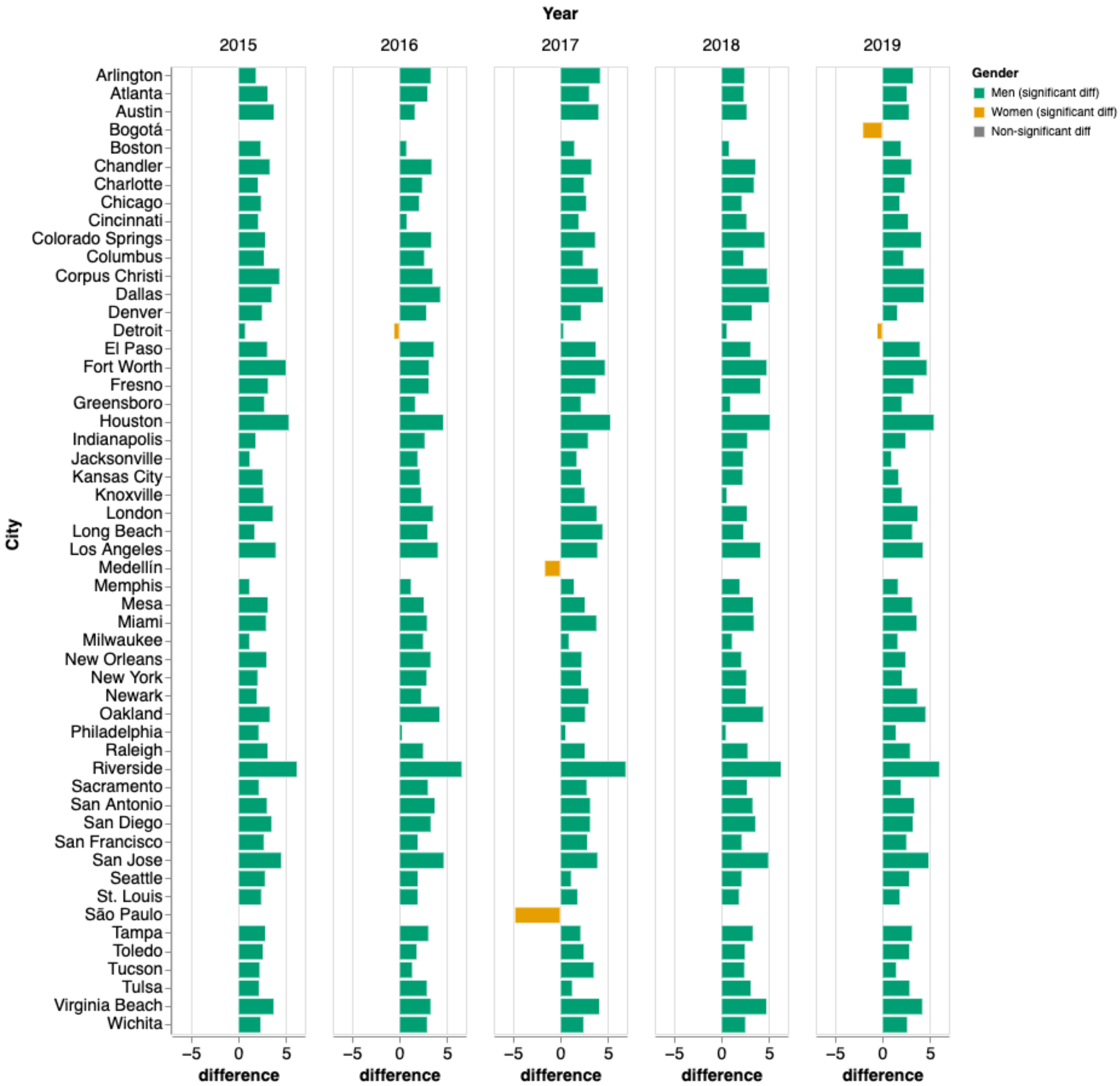

Figure 5.3: Gender differences in the average commuting travel time of the urban mobility from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

in the differences in travel time. However, we observe four exceptions Bogotá, Detroit, Medellín and São Paulo. These four cities show women having a longer commuting travel time than men. As in the United States we see a major consistent pattern, but for latin american cities we see another pattern, we argue that the urban landscape and cultural constructs might differ across these two continents. Our findings contribute to the literature by showing that gender differences in commuting travel time may not be a consequence solely from cultural constructs, but urban landscape may also play a role.

Next, we replicate the bootstrapping technique, taking into account the household arrangements of individuals. For instance, in the literature, smaller gender

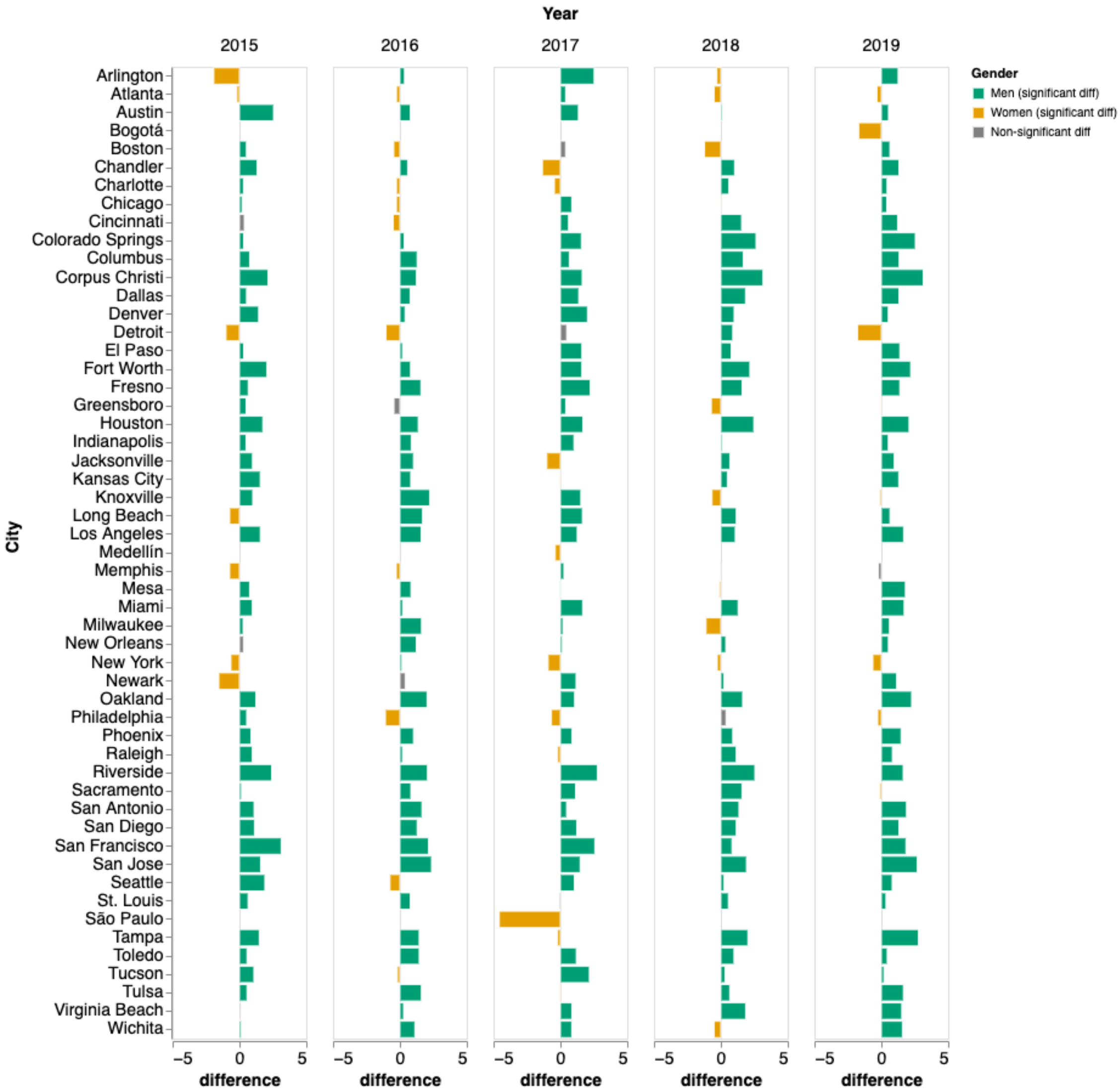

Figure 5.4: Gender differences in the average commuting travel time of the urban mobility from *single* travellers in the time period from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

differences in the commuting travel distance are more likely to be seen when considering people that are not married and do not have children [13].

For `single` travellers in our datasets, we observe in Figure 5.4 that the gender differences in the commuting travel time appear to be smaller than the ones seen in Figure 5.3, for most cities. We also plot, for the most recent year, the gender differences between household arrangements in Figure 5.6. We observe that for some cities such as Boston and New York, the commuting travel time of `single women` is higher than `single men`. When we remove the travellers that have children from the `single` group, we observe similar gender differences trends in Figure 5.5, but `single without children` travellers show higher gender differences than `single`

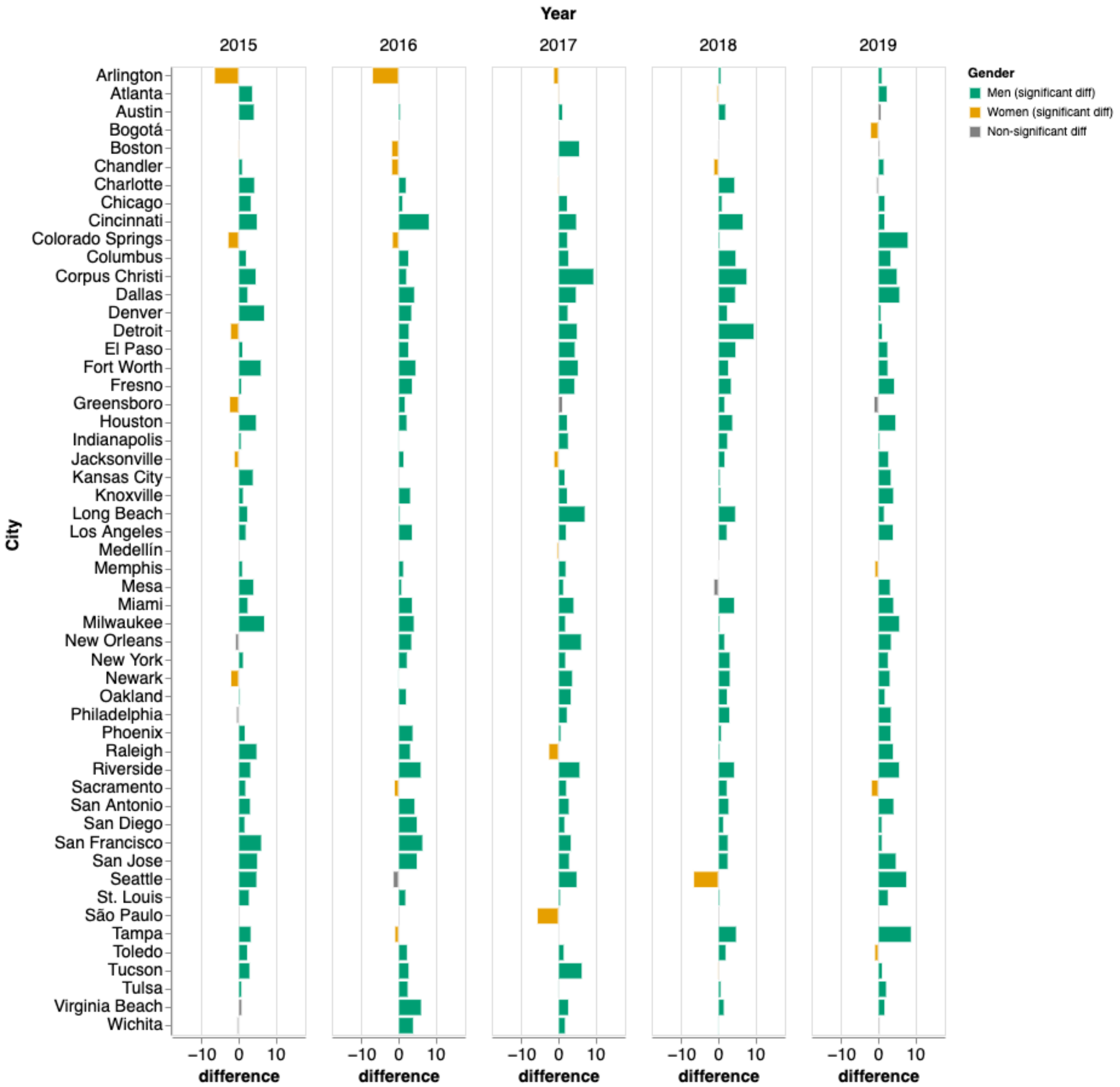

Figure 5.5: Gender differences in the average commuting travel time of the urban mobility from `single without children` travellers in the time period from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

and `single parent` travellers (Figure 5.6). Therefore, we argue that single travellers might be less affected by gender roles than overall travellers, and gender differences in the `single without children` group might come from the gender gaps in the labour market. Our findings are in contrast with the results of Fan [185], and Marcén and Morales [186] that found that differences in travel time are only statistically different between women and men for households that have children.

The gender differences between the commuting travel time from `married` travellers (Figure 5.7) are higher than for single and overall travellers (Figures 5.4 and 5.3) across cities and years (see also the comparison between household arrangements

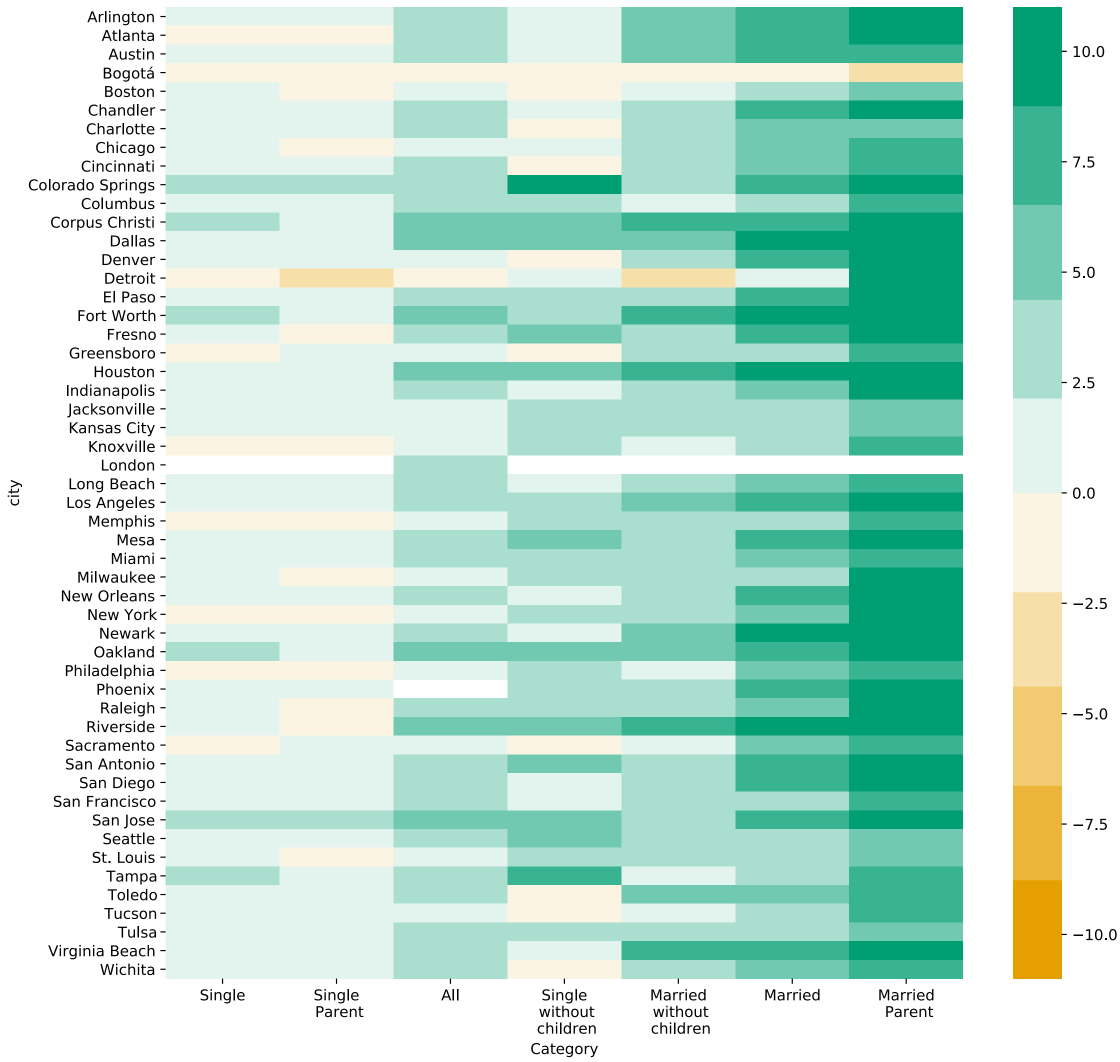

Figure 5.6: Gender differences from the mobility of travellers grouped by household arrangements in 2019.

in Figure 5.6). `Married men` tend to have a higher commuting travel time than `married` women, and these differences are smaller than for single travellers. We identify a positive correlation (around 0.2 using Spearman) between the commuting travel time of husbands and wives (pairwise comparison), indicating that higher travel time for one partner might be associated with the other partner in the same household. Moreover, the positive correlation disappears (less than 0.01) if we remove this household pairwise comparison between husband and wife (shuffling couples). We then conclude that marriage might play a role in travel time, which might be more accentuated in some cities such as Dallas, Fort Worth, Houston, Newark and Riverside. We highlight, again, that we identified that higher travel time indicates higher income levels within the United States. This implies that cities such as

Boston and New York might be starting a process of decreasing gender gaps in the labour market as they show different trends. For the case of latin american cities (in developing countries), we argue that their urban landscape might emerge with distinct constraints in the mobility of women and men.

Turning to the analyses of commuting travel time between `married` travellers with and without children, we observe in Figures 5.8 and 5.9 that men appear to have higher commuting travel times in most of the cities. Gender differences in the commuting patterns of married travellers are in line with the literature [13, 185, 186]. The gender differences of `married parent` travellers are the highest across all household arrangements shown in Figure 5.6. Therefore, we conclude that gender differences are present even for `single` travellers, and marital status, and having children tend to amplify the gender differences in commuting travel time.

Cities have different public systems, hierarchical organisations, job opportunity landscapes and political divisions. However, we consistently see that travel time for men is higher than for women in most of the cities in the USA, also indicating income disparity across genders. If differences in travel time are related to other dimensions — related to what gender is as a cultural construct —, cities that have implemented more regulations to mitigate gender inequality, in general, might have fewer gender differences in the travel time. However, we only see these gender differences being impacted for some cities when we compare `single` individuals. This can be an indicator that some cities are slowly getting closer to gender equality in the travel time differences.

Having children and getting married can then amplify gender differences in travel time. However, we argue that there is a gender effect in the commuting travel time that exists regardless of household arrangements. Our findings are relevant for the literature and policymakers as interventions related to household arrangements might not reduce these gender differences ultimately.

Moreover, shorter travel times may be related to jobs that pay low wages. This may be the case because upper income groups are more likely to be working in a small set of areas, a finding that is presented in Chapter 4. Thus, we argue that

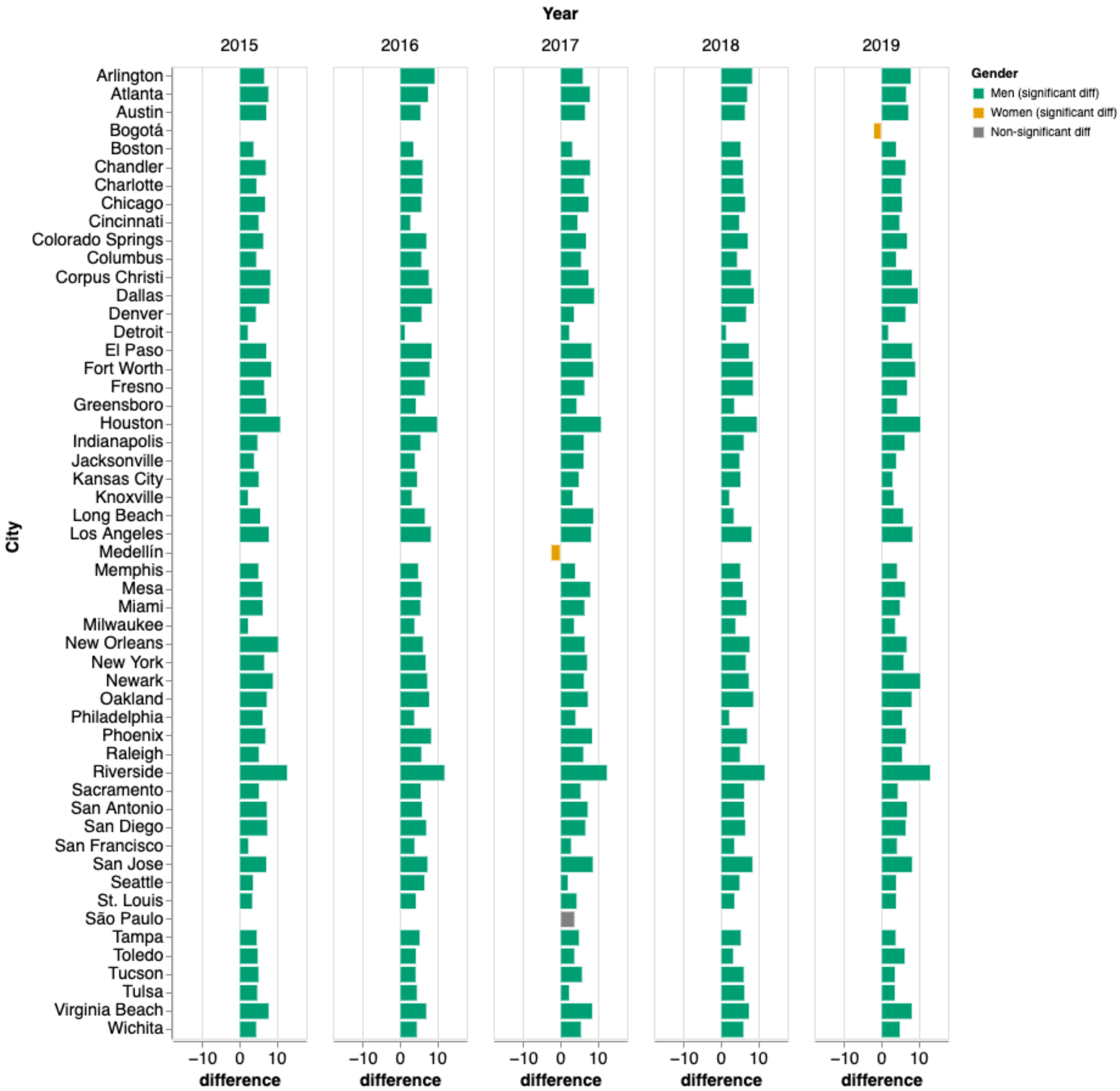

Figure 5.7: Gender differences in the average commuting travel time of the urban mobility from `married` travellers in the time period from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

the gender gap in pay [178] and the spatial landscape of job opportunities might be major reinforcements of differences in the commuting travel time over the years. As mobility cost seems to be associated with individual income, we argue that society should strive for gender equality in this dimension as a 50-50% balance.

We also found that the Spearman correlation between population volume and travel time or the gender differences are close to 0 (for example, -0.029). Therefore, it might be that population volume is not alone an indicator of differences in travel time, and differences in travel time are hidden in a complex set of urban and cultural characteristics. In this way, we will explore the spatial landscape of amenities among cities to understand how women and men can benefit from the urban space.

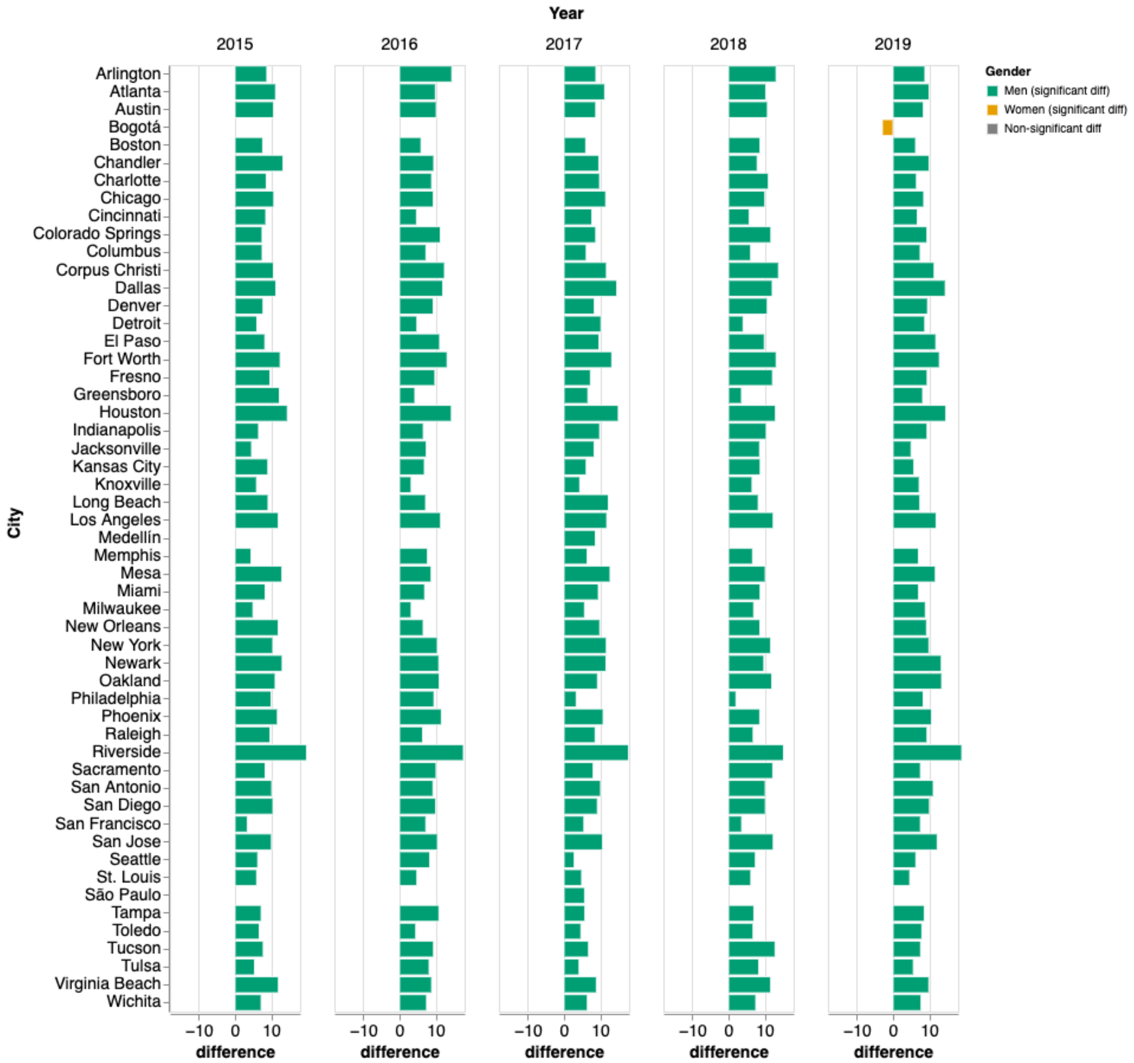

Figure 5.8: Gender differences in the average commuting travel time of the urban mobility from `married parent` travellers in the time period from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

5.3 Mobility reward

Being exposed (or nearer) to more opportunities can increase the likelihood of people accessing them via mobility [112, 182]. Areas that offer a high number of job and education opportunities coupled with a reasonable public transportation system may attract a high number of people. As one of the major dimensions of mobility, distance plays a unique role in mobility [3]. Here, we study the number and diversity of opportunities near to where women and men live, as the residential location is important for access to opportunities [16].

Differences in the residential distribution reflect the quantity and diversity of amenities that women and men are exposed. We establish mobility reward as the

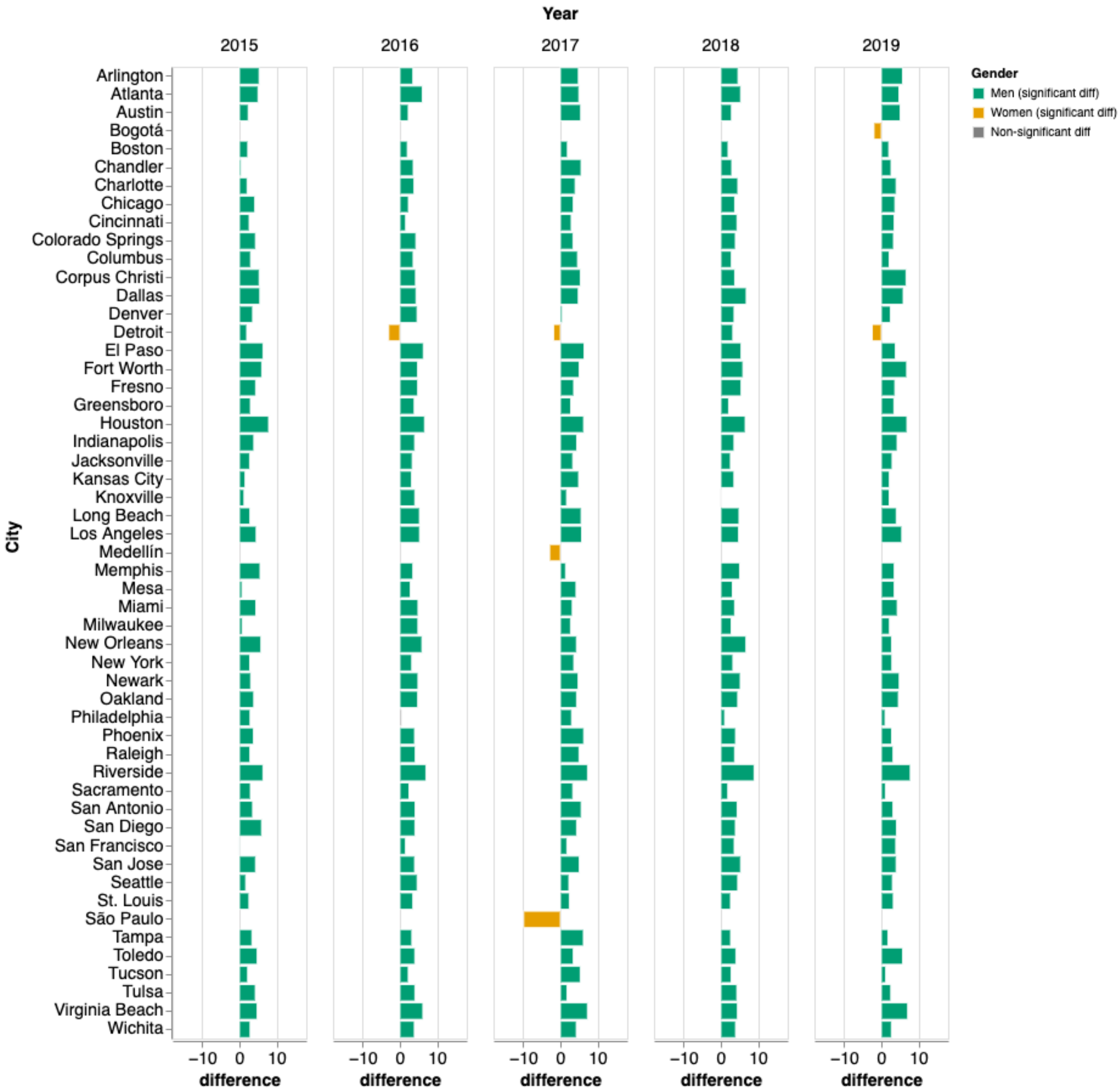

Figure 5.9: Gender differences in the average commuting travel time of the urban mobility from `married without children` travellers in the time period from 2015 to 2019. We apply Welch's t-test (p-value < 0.001) to ensure that distributions between women and men have different means. For the comparisons that presented similar means, we coloured the bars by grey.

magnitude of exposure to amenities that people can access near to where they live — short mobility. Exposure in this thesis means that people would be able to reach a certain number and diversity of amenities. We use OpenStreetMap data to categorise and locate the amenities, and we use the mobility data to estimate the likelihood of a group X being exposed to amenities.

We argue that as women stay near to where they live [111] than men, distance to amenities play an important role in women's lives. We compute the quantity and diversity of amenities that women and men are exposed to (i) inside the zone, and (ii) as they move from 0.5 km to 5 km radius that originates in the centroid of the

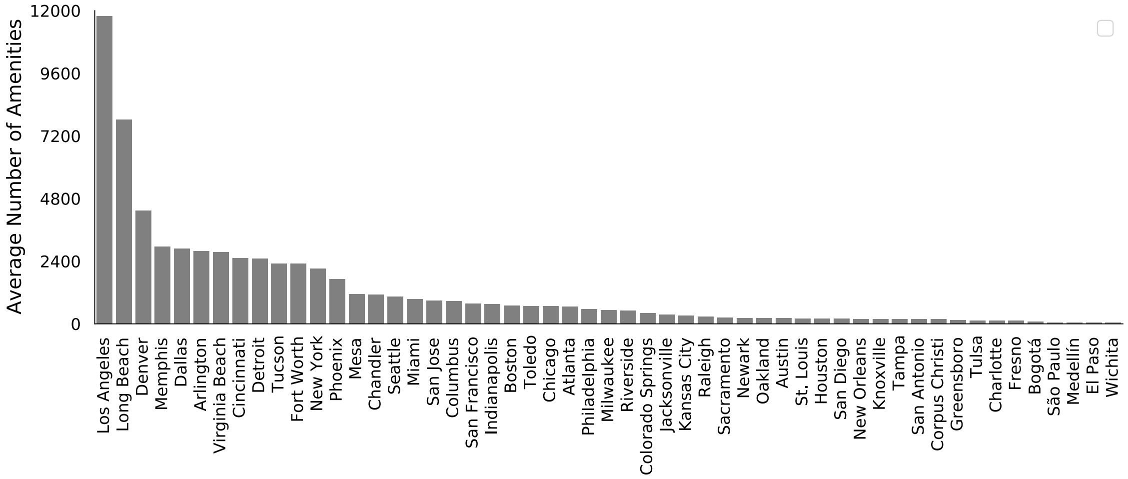

Figure 5.10: Ordered cities based on the average number of amenities extracted from OpenStreetMap [78]

.

zone in which they reside. Both metrics will indicate how far women and men travel to access a certain quantity and diversity of amenities. We apply Welch's t-test (p-value < 0.001) to ensure that distributions for each metrics have different means.

We start plotting in Figure 5.10 the average number of amenities estimated from the zones of each city. In the case of the United States, each zone is populated by around 100,000 people. For the case of the other countries, the zones can be populated by few people, and areas near to the city centre are usually smaller in extension as they are more populated. Thus, spatial partitioning considers the population and sociodemographic distribution more than the area of the zone.

We can also have an imbalance of work travels across genders. In general, men have a higher number of work travels than women (as shown in Chapter 3). Therefore, we normalise the number of amenities by city and by gender to fairly compare the exposure of women and men to amenities. For instance, we compute the fraction of amenities in a city that each person can access inside a zone or at a specific distance.

To enable the comparison between women and men, we compute the gender differences on the fraction of amenities, depending first on where they live (Figure 5.11). The scale in Figure 5.11 goes from -0.08 to 0.08 ($\pm$8%) that corresponds to a higher exposure of up to 1,200 amenities, which can vary across category. In most cities, the gender differences remain consistent across all categories. The same happens to the comparison between distances.

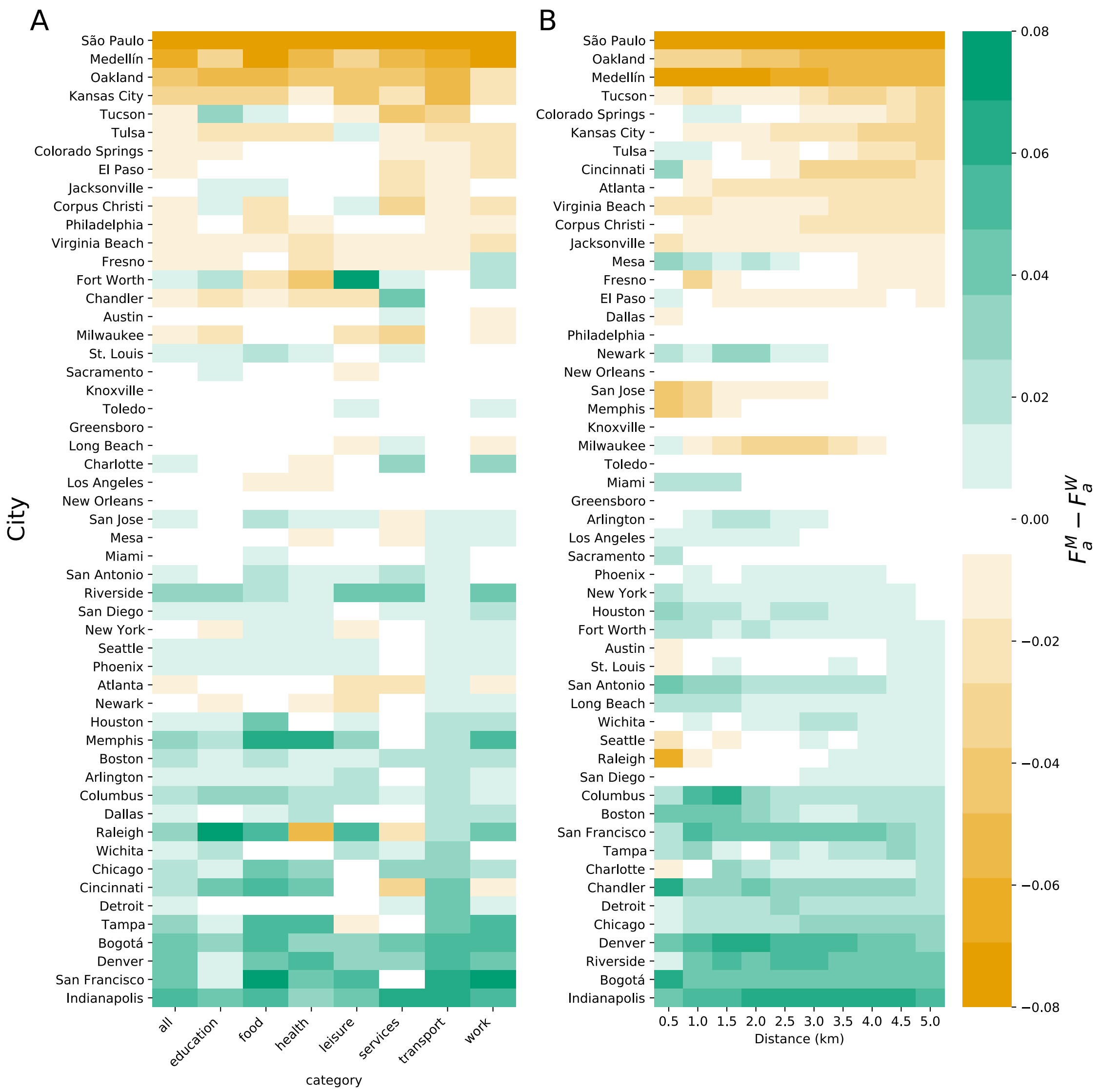

Figure 5.11: Gender differences of the fraction of amenities grouped by category (A) and by distance (B) estimated from the mobility of travellers in each city for 2019. For the cities of São Paulo and Medellín, we are using the mobility data of 2017 that is the most recent data that we have available.

We observe in Figure 5.11 that three groups of cities are identified when considering amenity exposure. First, cities such as São Paulo, Medellín, Oakland and Kansas City provide a higher fraction of amenities for women than for men. As women tend to have shorter travels, it might be the case that a higher number of amenities are nearly available in these cities, which can influence women to stay within the zone that they live. In fact, for the cities of São Paulo and Medellín, we observed in Chapter 4 that women tend to concentrate their work travels in fewer areas than men, and in this chapter, we see that these cities also show a higher number of amenities related to work near to where women live.

Second, cities such as Greensboro, Knoxville and Sacramento exhibit similar fractions of amenity exposure for women and men. We can argue two possible scenarios, (i) amenities are typically located far from where people live (more than 5 km) or (ii) women and men are more similarly distributed across all the areas making the amenities similarly accessible for both.

Lastly, cities such as Bogotá, Boston, Detroit, Denver and Wichita provide a higher fraction of amenities that are reachable for men. These cities also show women having higher travel time than men in Figure 5.6 which indicates that women go farther to reach job destinations than in other cities. Thus, we argue that as women are not as near to a higher number of amenities than men, women might need to go farther to access them in these cities.

We now investigate how socioeconomic status can impact the exposure to amenities for women and men. In Figure 5.12, we plot a comparison between the fraction of amenities that women and men are exposed to for each socioeconomic group. We notice that the gender differences in the upper income group are not usually the same as the ones seen in the remaining income groups. Moreover, the gender differences tend to be smaller for the upper income group in comparison to the other groups indicating that upper income women and men are more likely to be exposed to similar fractions of amenities in some cities.

We, then, can conclude that women and men are exposed to different fractions of amenities, and upper income groups, in general, show smaller gender differences. We see an amplification of the differences when we use mode of transportation and travel time as a proxy for gender constraints. While the fluctuations lead to less robust results, these proxies can be applied to research in mobility to make the estimation more accurate.

It is important to highlight that we found more cities that display higher fractions of amenities that are reachable by men than by women. Moreover, we found that there are cities where women and men have similar exposure to amenities through mobility. Thus, we argue that it may be the case that some characteristics of the urban landscape can provide fairer exposure to the quantity of accessible amenities. Although differences in travel time are not associated with the differences in the fractions of amenities, we argue that it might be the case for some cities such as Boston, Bogotá, Oakland, and Kansas City.

Although having a higher number of accessible amenities is important, it is also important to have diversity, as they fulfil different needs. We can summarise some needs into the following categories: food, education, work, health, transport, leisure and services. Then, we compute the diversity of amenities that women and men are exposed to depending on where they live. The diversity of amenities is calculated similarly to the mobility diversity (Section 4.2) considering the probability of a group X being exposed to the amenity type a for a given zone i.

Figure 5.13 shows the gender differences in the diversity of amenities across cities. We see that three groups of cities can be seen in the image: (i) the cities that women are exposed to a more diverse set of amenities, (ii) the cities that women and men are exposed to a similarly diverse set of amenities, and (iii) the cities that men are exposed to a more diverse set of amenities. Moreover, we can see a relationship between the quantity and diversity of amenities for few cities.

In the first case, the cities of Oakland, Virginia Beach and Philadelphia tend to have a higher quantity and diversity of amenities higher for women than for men, so women have a higher mobility reward than men. For the second case, cities such as Toledo, New Orleans and Knoxville tend to display similar values of quantity and diversity of amenities across gender groups. The last case shows cities such as Indianapolis, Tampa and Columbus having a higher quantity and diversity of amenities computed for men travellers than for women.

However, there are some cities that have gender differences reversed for quantity and diversity of amenities. Therefore, in relation to amenities and exposure to them, we see several peculiarities among cities. In Chapter 6, we group cities considering their mobility patterns to study possible mechanisms that can impact inequalities and differences on the exposure to amenities.

We also used Spearman correlation to compare the gender differences for travel time, fraction of amenities and diversity of amenities. We found a positive correlation of 0.22 between travel time and diversity of amenities. We then argue that higher gender differences in the travel time can be associated with higher gender differences in the diversity of amenities.

Finally, we can analyse the gender differences of the diversity of amenities between women and men across socioeconomic groups in Figure 5.14. We identify that most of the cities have lower gender differences for lower income group, and that the differences between middle and upper income groups tend to be higher for different genders. In relation to the diversity, the socioeconomic status plays a role in how exposed individuals are to amenities.

We can also observe that there are no statistical differences in the diversity of amenities in Bogotá, Medellín and São Paulo. In this way, we argue that in these cities the differences in the distribution of travels for women and men seen in Chapter 4 might be coming from the number or absence of amenities near to where they live. The amenities near to where people live might not offer enough job and other opportunities as individuals in these cities tend to visit more other zones than the zone they reside.

In summary, our findings indicate that women and men are exposed to a different quantity and diversity of amenities across cities. Therefore, the mobility reward for women and men vary across cities. We found that people from upper income groups show different gender differences than people from the other income groups. Finally, we present that gender differences become smaller in relation to diversity as we increase the distance individuals travel from their place of residence. This is in line with the idea that improving accessibility to locations may be a good solution to achieve fairness in mobility [182].

5.4 Discussion

Women and men tend to display different patterns in mobility. In this chapter, we built on existing literature to show that, for most cities, commuting travel time — mobility cost — continues to be a major difference when comparing the mobility of women and men [27, 111]. We also found that people with high income tend to have higher commuting travel times. Then, aligned with this finding, men that are usually considered the breadwinners [23, 24] tend to show higher commuting travel time than women in most of the cities, indicating that gender roles play a role in work-related mobility.

In contrast, we also presented that women are more likely to have average commuting travel time than men for the Latin-American cities. In this way, differences in the cultural constructs and urban landscape may play a role in the gender differences for travel time. We argue that the public system in these Latin-American cities are not as efficient as in the USA [187], so the higher number of public system travels by women may impact the average travel time. This can also be related to the fear and perception of crime in these cities, as women are more likely to take precautions in their mobility [188]. We argue that longer travels might be chosen by women to avoid certain locations that are dark, unpopulated or have high rates of crime.

For most cities, we observed that gender differences in the travel time for the `married parent` group are higher than for the other household arrangements such as `single` and `single parent` groups. Therefore, we argue that gender roles in household arrangements might amplify the gender differences in mobility, which aligns with the literature [13, 113]. However, contrary to the literature, our results indicate that gender plays a role in the commuting travel time even for individuals that are `single` and `without children`.

For the mobility reward, we investigated the fraction and diversity of amenities that women and men are exposed to, depending on where they live and the distance from their residence. Mobility can be seen as an indicator of power because it expands the accessibility to opportunities [112]. We found that men have more exposure to a higher fraction and variety of amenities in most cities. Furthermore, as the distance from where people live increases, the gender differences for amenity diversity decreases.

We finally conclude in this chapter that mobility may have a higher cost for men, but men are generally compensated with high income. Shorter travel time indicates a greater difference in the diversity of amenities between women and men. Moreover, although we found cities where the mobility reward is higher for women, many cities show a higher mobility reward for men.

Little is known in the literature regarding how women and men are exposed to opportunities or have access to them. Most of the works cover the need for public transportation [113], new policies and regulations [182] and changes in the labour market [24]. More relevantly, Hail et al. [181] indicate that public transportation is important to the accessibility to facilities.

Based on the literature and our findings, we conclude that there is a gender, household and urban landscape effect in the mobility patterns. Changing the urban landscape likely decreases some inequalities in mobility, but changes in the cultural constructs might also be necessary for a stronger effect on gender inequality. Finally, we acknowledge that the literature emphasises the public transportation system as the most important dimension for decreasing gender inequality in mobility, but the constraints might also be related to historical gender gaps in the labour market. We argue that amenity exposure is a crucial dimension to consider since gender gaps in income may translate to where people live. Therefore, where people (afford to) live may also be a factor for gender inequalities in mobility that impose trends on the mobility of women.

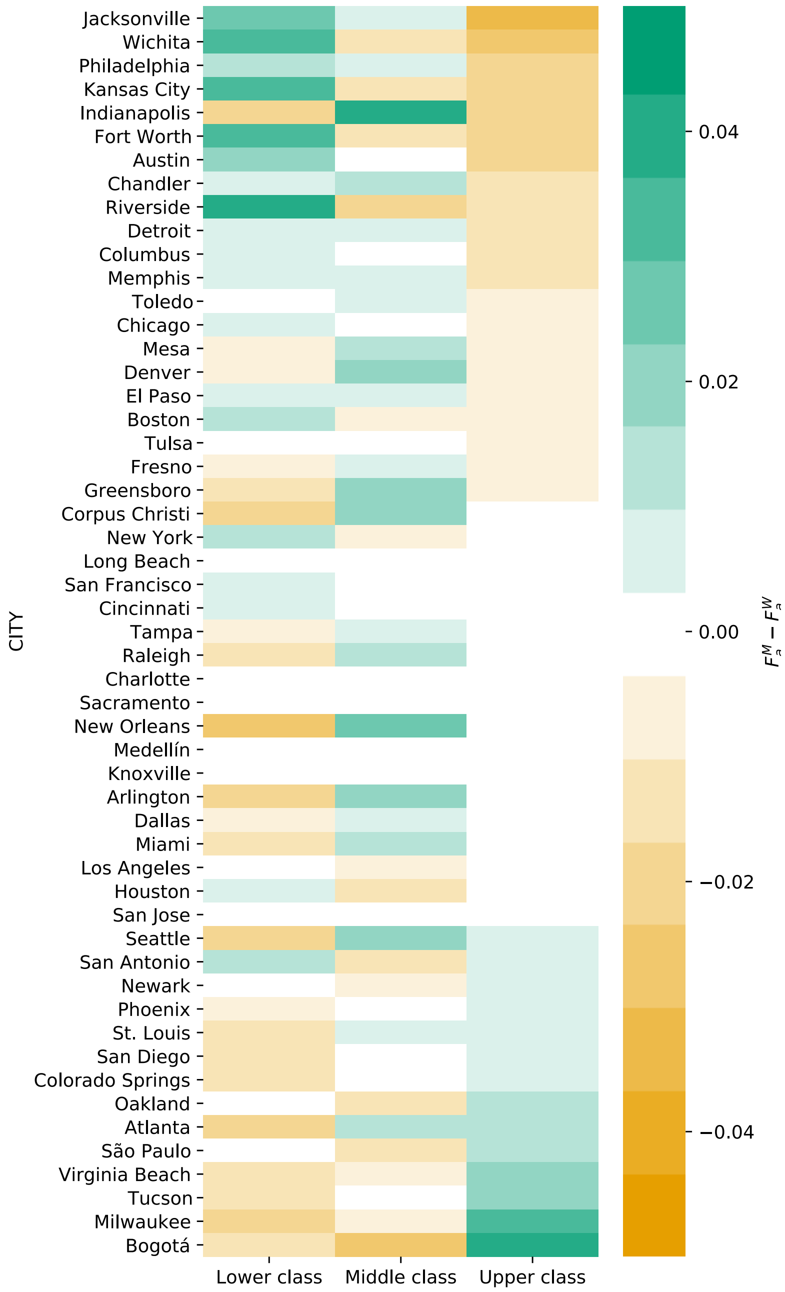

Figure 5.12: Gender differences of the fraction of amenities grouped by socioeconomic groups estimated from the mobility of travellers in each city for 2019. For the cities of São Paulo and Medellín, we are using the mobility data of 2017 that is the most recent data that we have available.

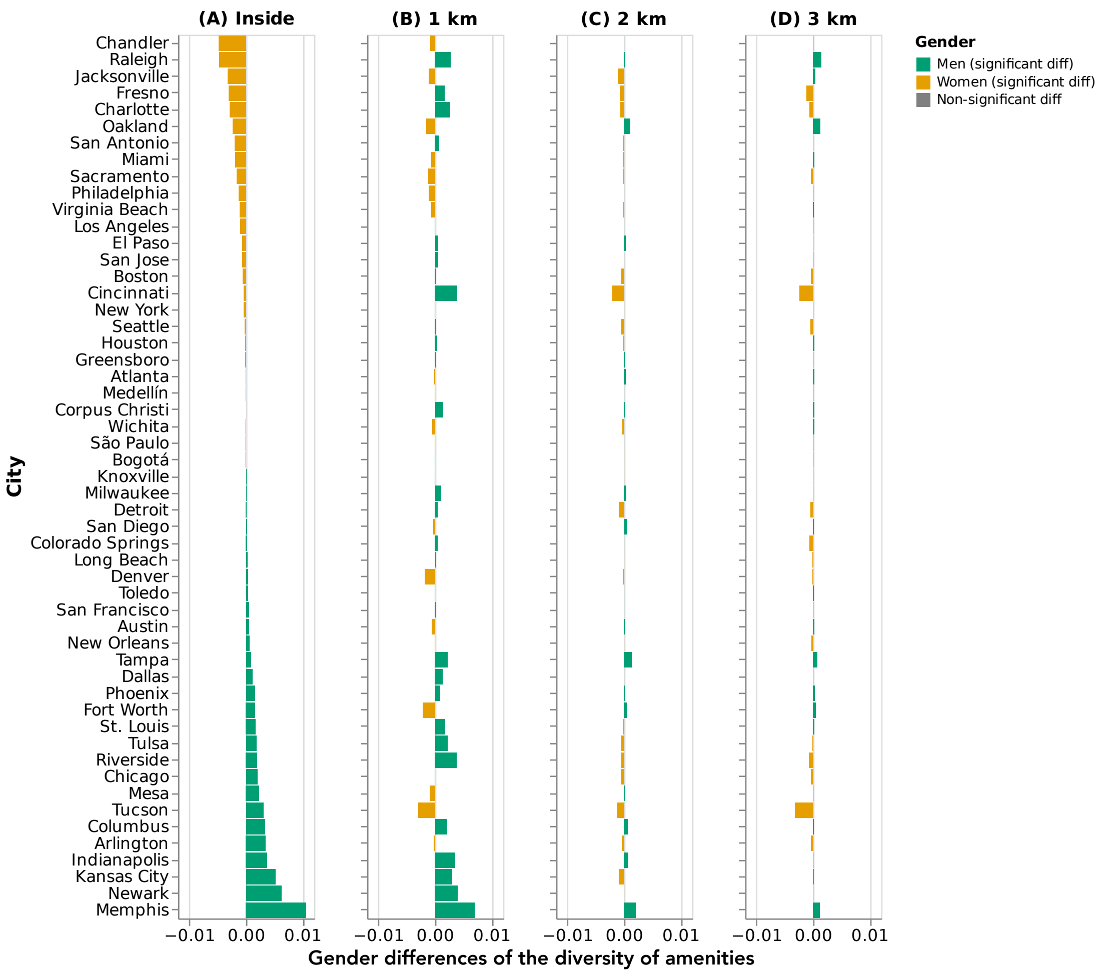

Figure 5.13: Gender differences of the diversity of amenities (A) from the zones where individuals live, (B) from 1 km, (C) from 2 km and (D) from 3 km grouped by cities in 2019.

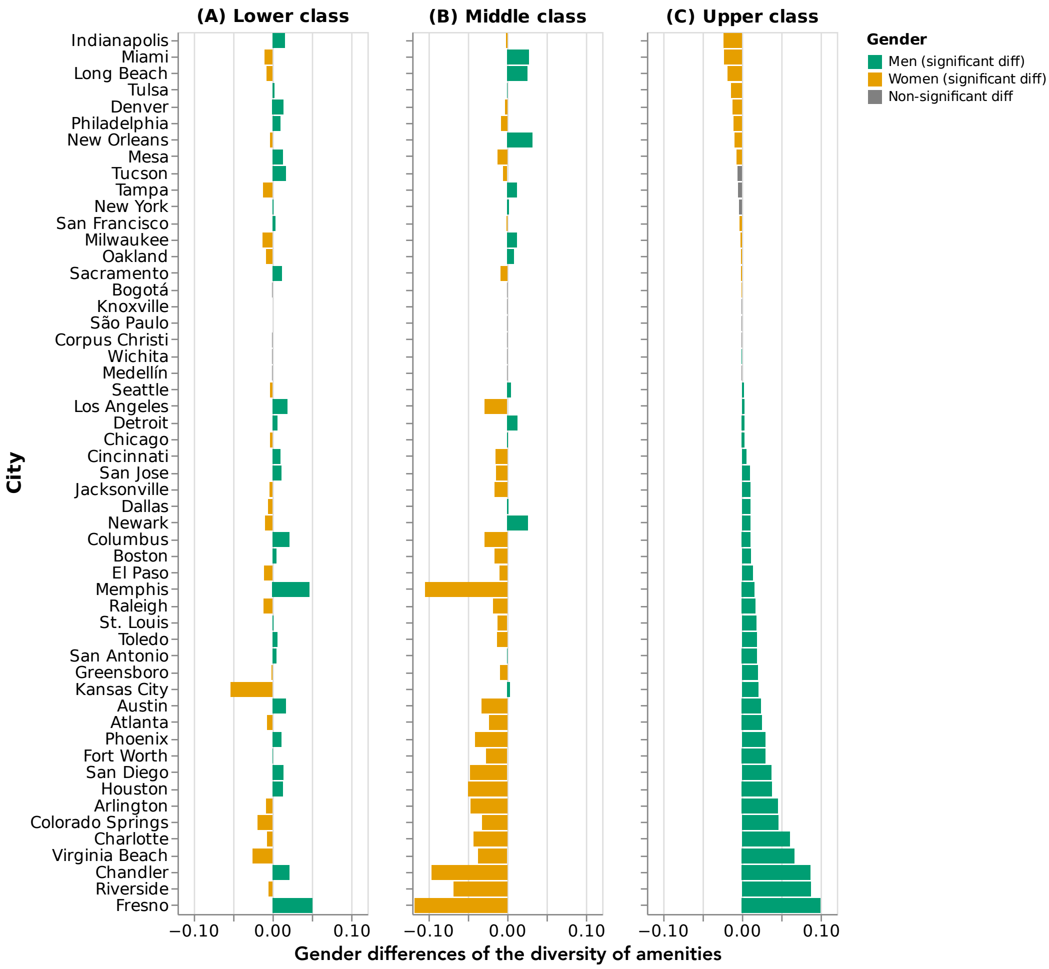

Figure 5.14: Gender differences for each socioeconomic group (A) Lower class, (B) Middle class and (C) Upper class of the diversity of amenities from the mobility of travellers grouped by cities in 2019.

6. Profiling cities based on mobility characteristics

The development of sustainable and inclusive cities is a complex task. As urban areas are becoming denser in population, buildings, roads and facilities, it is becoming more complex to intervene in the public and private systems [189] due to possible unintended consequences. Literature shows that improving the transportation system can provide higher accessibility to all areas, and consequently, provide a better environment for development [190]. However, creating an efficient transportation system, for instance, is a challenging task that usually demands changes in the environment and legislation [191].

Cities with similar urban infrastructure may have a mutual benefit from related interventions that aim to improve the urban landscape and well-being of their populations. Hence, the understanding of how cities share mobility patterns and urban characteristics increases the likelihood of successful interventions and solutions from the past.

The city's infrastructure, development, and design can affect how people belonging to different sociodemographic groups behave. Urbanisation and population density can be associated with smaller gender gaps in 12 low and middle-income countries [16]. For instance, in urban areas, education seems to be beneficial to the improvement of female labour force participation (FLFP) [16], and improving the accessibility to job and education opportunities via mobility might help decrease the gender gaps in the labour market. Here, we study the similarities of the work landscape that emerge from the mobility in several cities in the USA to investigate whether this job landscape can tell us important information about possible mobility inequalities.

In this chapter, we group cities based on three mobility characteristics that describe the urban landscape of work-related mobility. The first mobility characteristic, hotspot cutoff level 1, estimates the fraction of areas that are highly-visited by work travel [32]. The second one, flow hierarchy, estimates the mobility flow between hotspots that emerge from work travels [32]. Last, the third one, mobility diversity, estimates how work travels are distributed within all the areas (proposed in Chapter 4).

Based on these three mobility characteristics, we identify the existence of two major clusters that differ from public transportation usage, population volume, number of amenities, diversity of amenities, and average travel time. Therefore, cities that share similar mobility characteristics also share similarities in the urban landscape. Our findings indicate that mobility patterns are related to city development and that Latin American cities show similar mobility characteristics as cities in the USA. Last, we demonstrate that gender differences in average travel time are higher for cities with low overall public transportation usage and a low number of amenities, suggesting that the urban landscape can affect the mobility of women and men differently.

6.1 Data

In this chapter, we use the data of LODES [74] from the USA to compute the mobility features because of its high spatial resolution for work-related mobility. Then, we use the ACS [76] data, which has rich metadata about the individual and household characteristics, to analyse the sociodemographic characteristics of cities. Section 3.1 explains the details of both datasets that are used in this chapter. We also use the quantity and diversity of amenities computed from the OpenStreetMap [78] in Chapter 4.

The data from the Latin-American cities: Medellín, Bogotá and São Paulo are used as case studies in comparison to the USA. The data from Colombia and Brazil have a different resolution than the data from the USA. Moreover, the cultural constructs and federal regulations are also different across these countries. With this in mind, we argue that adding these three cities to the main analyses would bias the results. Moreover, with the data of these three Latin-American cities, we compute the mobility features disaggregated by gender.

6.2 Mobility features

We group the cities based on three characteristics/features of urban mobility: hotspot cutoff level 1, flow hierarchy and mobility diversity. These three features indicate how work-related mobility is concentrated over areas, what the highly visited areas (called hotspots) are, and how the flow is structured between hotspots [32, 189]. Thus, these three features summarise how concentrated the work travels are in each city.

A hotspot, in this chapter, is a set of areas that concentrates mobility flow. The hotspot can have n levels, where level 1 comprises the areas with the highest flows on each area, and the last level n represents the areas where individuals have the smallest flows. We define the hotspot cutoff level 1 as the percentage of zones that are not grouped in the hotspot level 1 [32]. Highly uneven distributions of trips within areas result in high values of the hotspot cutoff level 1. Galotti et al. [192] showed that large polycentric cities have many hotspots, and they appear to be more segregated and less integrated than smaller and monocentric cities.

We compute the hotspot cutoff level 1 similar to Bassolas et al. [32] using the Loubar method. We measure the Lorenz curve represented as the blue curved line in Figure 6.1 based on the sorted cumulative distribution of inflows in ascending order. This results in the relationship between the normalized cumulative number of areas in axis x, and the fraction of total inflow in axis y. Then, the hotspot cutoff level 1 is calculated as the point that a segment (represented by the red line in Figure 6.1) touches the x-axis from the derivative of the Lorenz curve at the extreme point where $x = 1$ and $y = 1$. This hotspot cutoff level 1 that intersects the x-axis defines the zones that will be grouped at the first level based on the ascending order distribution. This methodology can be used iteratively to compute the other levels of hotspots

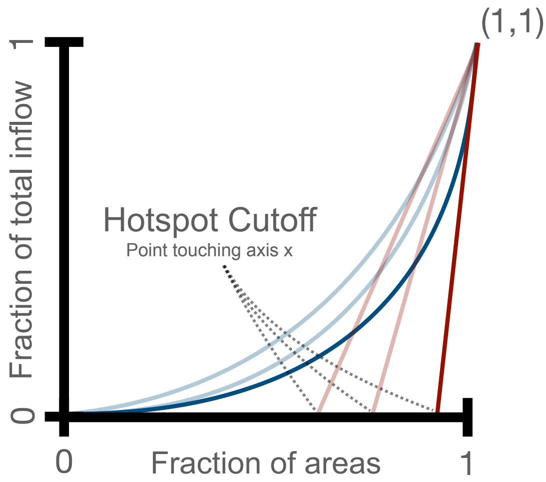

Figure 6.1: Representation of the hotspot cutoff points using the Loubar method. Curved blue lines represent the Lorenz curve and straight red lines represent the segment that intersects the x-axis from the point $x = 1$ and $y = 1$.

using the remaining zones that were not grouped in previous levels of hotspots.

The flow hierarchy is the sum of the fraction of travels between the hotspots of a similar level, one level above and one level below; therefore, it is the strength of mobility between similar hotspot levels [32]. Mathematically, the flow hierarchy is defined as:

$$\phi = \sum_{i,j=1}^{L} F_{ij}(\delta_{ij} + \delta_{i(j-1)} + \delta_{(i-1)j}), \tag{6.1}$$

where L is the maximum number of hotspot levels, F_{ij} is the fraction of travels between the hotspot levels i and j, and δ is the Kronecker delta. High values of ϕ (close to 1) indicate a strong hierarchical organization, and small values of ϕ (close to 0) indicate a more uniform distribution of travels across the hotspots levels.

Bassolas et al. [32] compute the hotspots based on the outflow of the mobility. The high values of flow hierarchy are associated with higher rates of public transport usage and walk modal share, fewer rates of pollution emissions and traffic fatal injuries. In this way, high values of flow hierarchy indicate desirable urban indicators in terms of health, liveability and accessibility.

Different to Bassolas et al. [32], we compute hotspots based on the inflow of the mobility related to commuting. Therefore, the hotspots represent the areas that individually have higher fraction of travels related to work; concentrate job destinations. To be robust across cities, we compute a total of 5 levels of hotspots per city, and measure the flow hierarchy based on the flow between these 5 levels.

Lastly, we use mobility diversity as proposed in this thesis to estimate how concentrated the work travels are in different cities. High values (close to 1) of mobility diversity indicate an even job distribution over all the areas, indicating similarly visited areas for work purposes. The other extreme (values close to 0) indicates a high concentration of travels in a few areas.

In summary, hotspot cutoff level 1 represents the fraction of areas outside the first hotspot, flow hierarchy represents the flow between hotspots, and mobility diversity represents how the work destinations are concentrated considering all the areas. We estimate these three mobility features for the 50 cities selected from the LODES data (detailed in Section 3.1), and the values range in the following intervals:

- Hotspot cutoff level 1: values from 0.80 to 0.98
- Flow Hierarchy: values from 0.47 to 0.88
- Mobility diversity: values from 0.71 to 0.89

The high values of hotspot cutoff level 1 in all the cities indicate that only a maximum of 20% of the areas in any city is highly being visited for work purposes. The wide range of flow hierarchy shows that some cities are more hierarchically organised between hotspots levels than others. Finally, the values of mobility diversity indicate low levels of the overall concentration of job opportunities in a few areas, as the values are closer to 1 than 0. These mobility diversity values are in line with the ones found for the Latin-American cities in Chapter 4. We also observe that high levels of flow hierarchy are associated with high levels of mobility diversity and a high percentage of highly-visited areas.

Another interesting finding is that the flow hierarchy computed based on the hotspots of work destinations are smaller than the ones based on the mobility outflow in Bassolas et al. [32]. Therefore, as we showed in Chapter 4, we argue that, based on mobility data, workplaces are more concentrated than the places related to other purposes, and the hierarchy between hotspots is also stronger based on the residential and overall landscape.

6.3 City profiles

Little is known about the types of urban structures that reflect or amplify mobility inequalities. Here, we study several cities to investigate whether the three selected mobility features can indicate possible sociodemographic inequalities reflected in mobility. Therefore, we start applying the clustering technique to find groups of cities that share mobility similarities, and after, we analyse these groups' composition.

Specifically, we apply a hierarchical clustering on the feature vector using the standard normalisation, which scales each feature separately by subtracting the mean and dividing by the standard deviation. The distribution will have a mean equal to zero and a standard deviation equal to one. This normalisation ensures that one feature will not entirely bias the definition of clusters.

We then use the hierarchical clustering with the complete linkage method and the euclidean distance [193, 194]. Using the complete method, the groups should have all the elements, the most distant possible between different groups. Euclidean distance computes the distance between the elements as a segment. In our context, mobility in the cities in one group should be different from the other cities. We argue that as we only have three dimensions, similar accuracy is drawn using other methods and distances, but we chose the most appropriate for our goal that is creating groups that are the most distinct from each other.

Two groups of cities are identified while clustering the feature vectors for each year. We show that one of the outstanding differences between clusters (`C1` and `C2`) is public transportation usage. The mobility in cluster `C2` has a higher usage of public transportation compared to `C1`. When we look at the number of amenities from the OpenStreetMap, we ensure that `C2` is the cluster that provides a wider range of public transportation amenities, probably attracting people to use the system more. Moreover, in cities like New York, people might either opt for public transportation or private transportation, which is more costly with respect to time (traffic) and money (expensive parking lots or taxis). We also notice that the cluster `C2` tends to have the biggest and most dense cities compared to the other cluster.

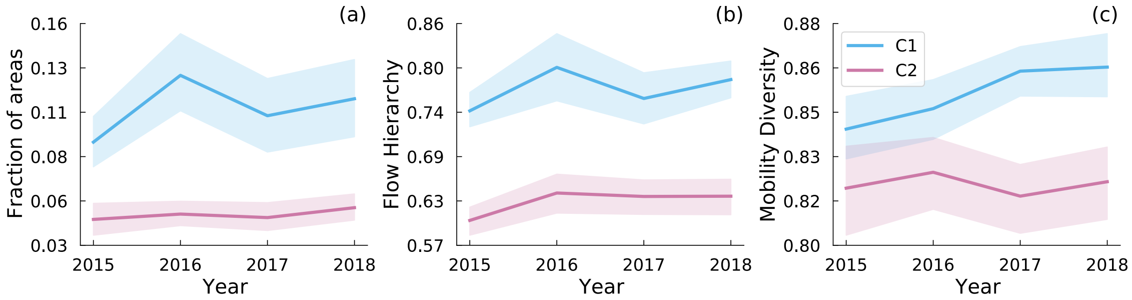

Figure 6.2: Variance of the mobility features in each cluster over the years.

6.3.1 Grouping evolution over the years

There are some variations on which cluster a city is grouped for each year (2015-2018), but only 4 cities are not consistently grouped in the same clusters over the years. We can see 3 scenarios:

- Cities that are most of the time in cluster `C1`: Long Beach, St. Louis, Virginia Beach, Kansas City, Miami, Oakland, Riverside, San Jose, Tulsa, Wichita, Charlotte, Corpus Christi, Fresno, Raleigh, Greensboro

- Cities that are most of the time in cluster `C2`: Arlington, Atlanta, Austin, Boston, Chandler, Chicago, Cincinnati, Columbus, Dallas, Denver, Detroit, El Paso, Fort Worth, Houston, Indianapolis, Jacksonville, Knoxville, Los Angeles, Memphis, Mesa, Milwaukee, New Orleans, New York, Newark, Philadelphia, Phoenix, Sacramento, San Antonio, San Diego, San Francisco, Seattle, Tampa, Toledo, Tucson, Colorado Springs

- Cities that are 50-50% of the time in each cluster: Chandler, Newark, Toledo, Wichita

Figure 6.2 shows that cities in the cluster `C1` display higher values for the three mobility features than cluster `C2` for all the years. Therefore, `C1` represents the cluster in which the work travels are more distributed across all the areas that have a higher fraction of travels representing the hotspot level 1, and that most of the mobility is performed across similar hotspot levels. We also observe that the population volume in the cities of cluster `C1` tend to be smaller than the ones from cluster `C2` (Figure 6.3).

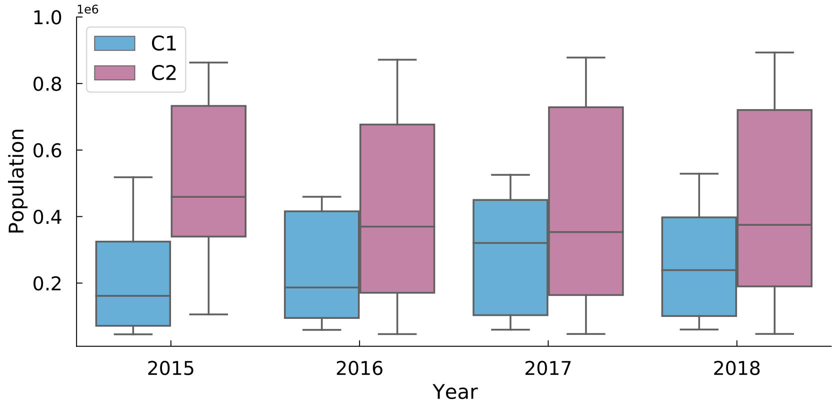

Figure 6.3: Population volume variance in the cities of each cluster over the years.

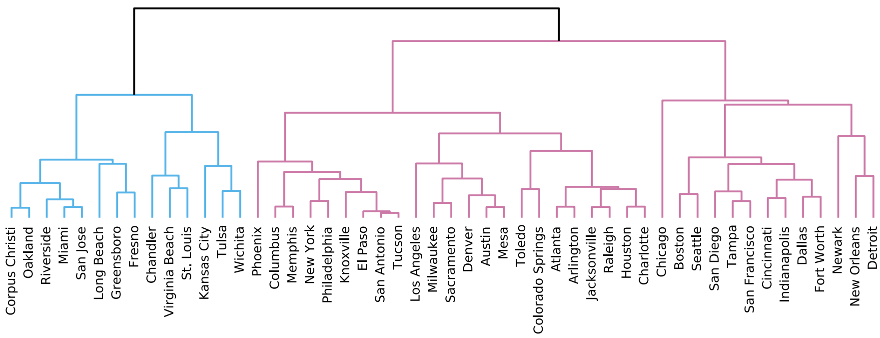

Figure 6.4: Cities grouped by the mobility features extracted from the data in 2018.

In line with our observation that most of the cities are consistently grouped in the same cluster, we choose the last year available, 2018, to further analyse the characteristics of each cluster. These characteristics can indicate a certain division of mobility characteristics between clusters. As we use data from other sources such as OpenStreetMap, the most recent year is the best option for data integration. Figure 6.4 displays the cities in their respective cluster for 2018 in which `C1` is coloured by blue and `C2` is coloured by pink throughout this entire chapter.

Using the data from 2018, the next section analyses the differences found between the two clusters. We will not explore the income level in this chapter because we did not find any difference coming from the median individual income across clusters. For instance, there are differences in the median individual income between women and men in all the cities regardless of the cluster, so men tend to have higher individual incomes than women.

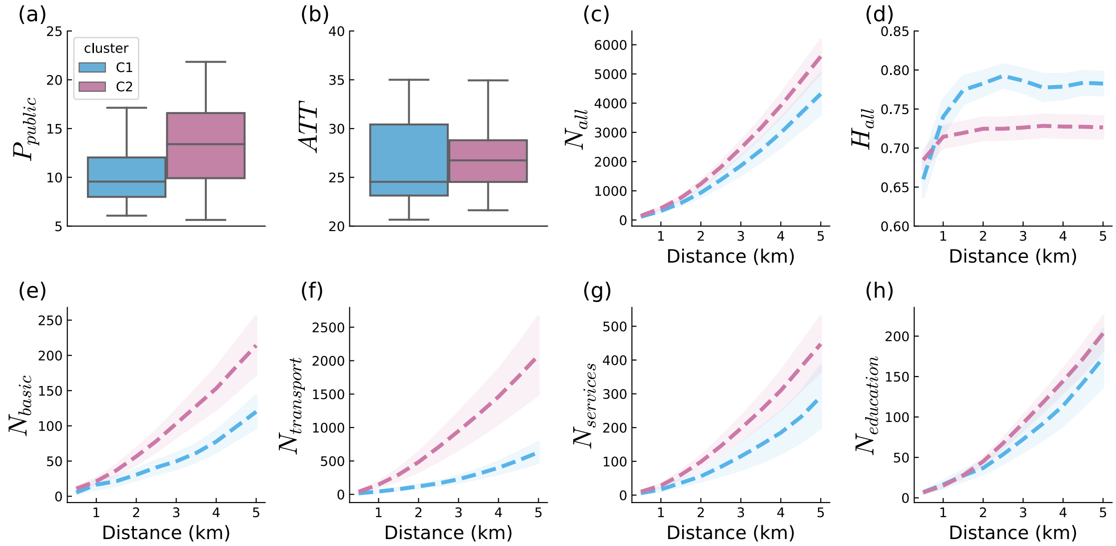

Figure 6.5: Characteristics of the cities grouped by the clusters from the data in 2018. (a) Percentage of public transportation usage, P_{public}; (b) Average Travel time, ATT; (c) Number of amenities, N_{all}; (d) Diversity of amenities, H_{all}; (e) Number of amenities related to food and health, N_{basic}; (f) Number of amenities related to transport, $N_{transport}$; (g) Number of amenities related to services, $N_{services}$; (h) Number of amenities related to education, $N_{education}$.

6.4 Cluster analyses

Figure 6.5 displays a summary of characteristics of each cluster. Firstly, we see that the usage of public transportation, P_{public}, is higher for cluster `C2` than for cluster `C1`. We then order the cities based on the usage of public transportation to understand how cities deviate from this usage (Figure 6.6). We observe that cities such as New York, Boston, and San Francisco grouped in the cluster `C2` have much more public transportation usage than the cities in the cluster `C1`.

It might be the case that the cities in which people use more public transportation also provide better public systems for women to reach a higher number of job opportunities, as women can not afford to have private transportation as much as men can [177]. In this way, we order the cities based on decreasing gender differences in public transportation use. This allows us to investigate whether high gender differences are seen in a particular cluster. Figure 6.7 display that most of the cities, regardless of the cluster, have a higher fraction of women using the public system than men, in line with the literature [25, 46, 48, 177]. Moreover, in most of the cities, men are more likely to travel using private transportation (e.g., car,

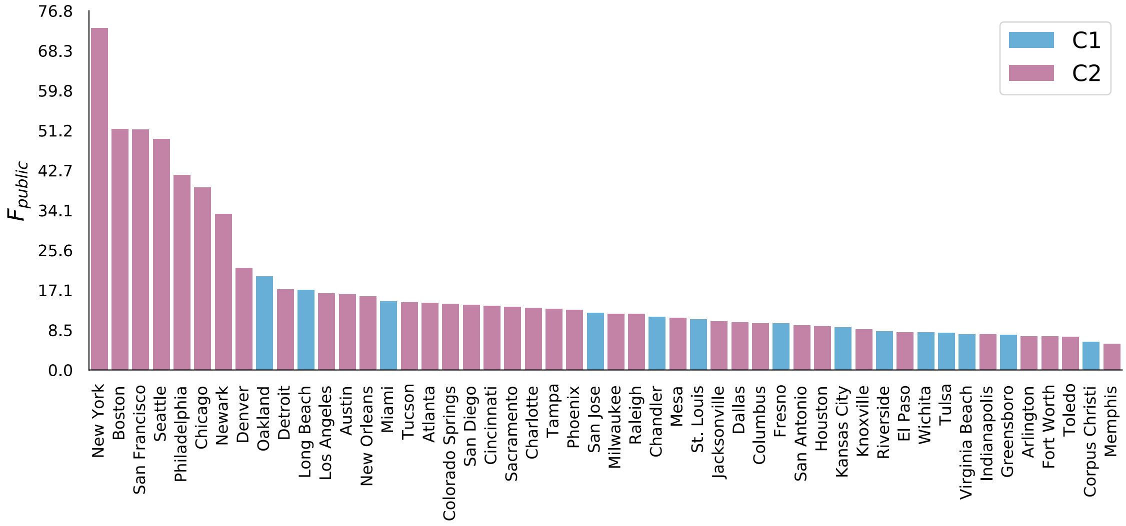

Figure 6.6: Cities ordered by the public transportation usage extracted from the data in 2018.

truck or van). Thus, we are not able to conclude that a particular cluster has higher gender differences than the other regarding the usage of the public system.

Considering transportation modes are related to travel time, we compare the average travel time across clusters in Figure 6.5 (b). On average, the commuting travel time from the urban mobility of cities in `C2` tends to be higher than `C1`. However, the cities present high variance on the average travel time, and we question whether the gender differences found in Chapter 4 are grouped in magnitude across clusters.

Ordering the cities with respect to the magnitude of gender differences in the commuting travel time (Figure 6.8), we identify that cities in the cluster `C1` tend to have higher gender differences than the cities in the cluster `C2`. This finding, along with the gender differences in public system usage, reveals that longer travel times may not be solely influenced by the usage of the public system but may be impacted by the distribution of job opportunities (that resulted in the clusters).

We then analyse the quantity and diversity of amenities grouped by clusters in Figure 6.5 (c-h). We notice that the cluster `C1` tend to have a higher overall number of amenities, N_{all}, as the distance increases, and this is true for the categories of (i) basic amenities composed of amenities related to food and health related, N_{basic} (ii) amenities related to transport, $N_{transport}$, (iii) amenities related to services, $N_{services}$, and (iv) amenities related to education, $N_{education}$. However, we see that the diversity

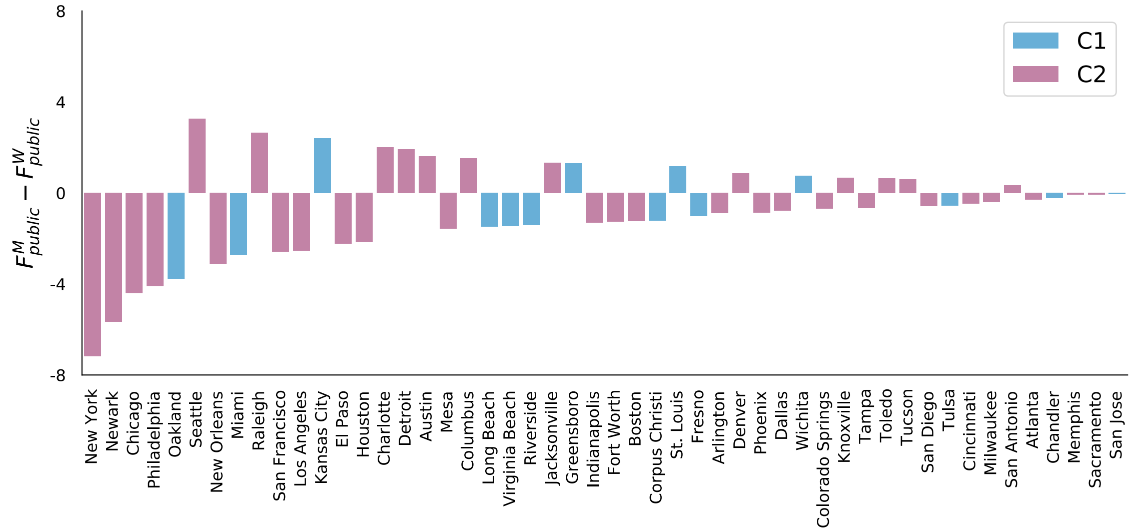

Figure 6.7: Cities ordered by the gender differences on the public transportation usage ($F^M_{public} - F^W_{public}$) extracted from the data in 2018.

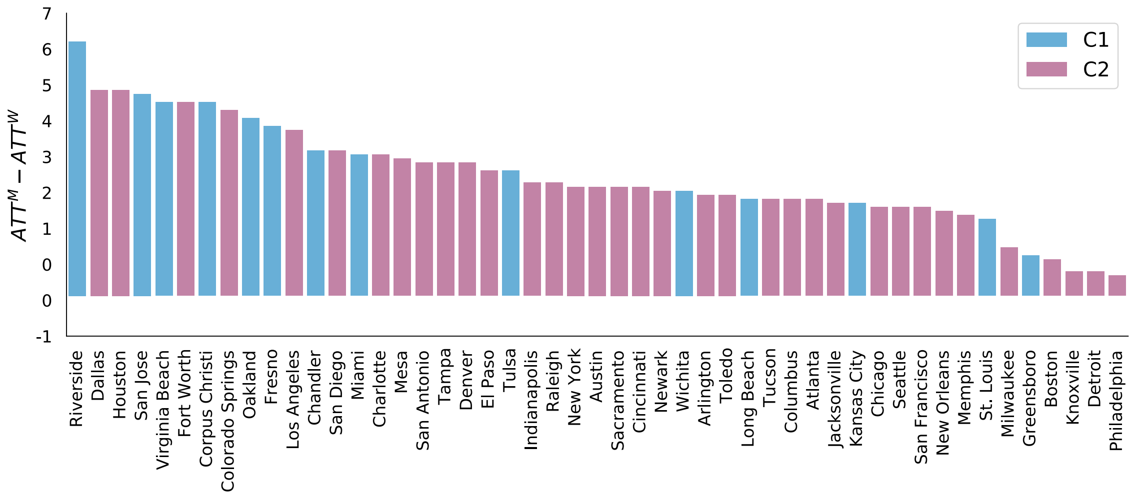

Figure 6.8: Cities ordered by the gender differences in the commuting travel time ($ATT^M - ATT^W$) extracted from the data in 2018.

of amenities, H_{all}, is higher for C1 than C2 as we increase the distance. This may be the case because the cities in C2 tend to concentrate amenities belonging to certain categories. As we observe for the number of amenities related to transport, the cluster C2 has a much higher number of amenities than C1.

In summary, we conclude that cities with smaller public transportation usage indicate a more even distribution of mobility related to work over all the areas and a high mobility flow related to work between similar hotspots. Higher gender differences in commuting travel time are also seen in these cities. The number of amenities per category is not as high as the other cities, but diversity of amenities is high as we increase the distance from where people live.

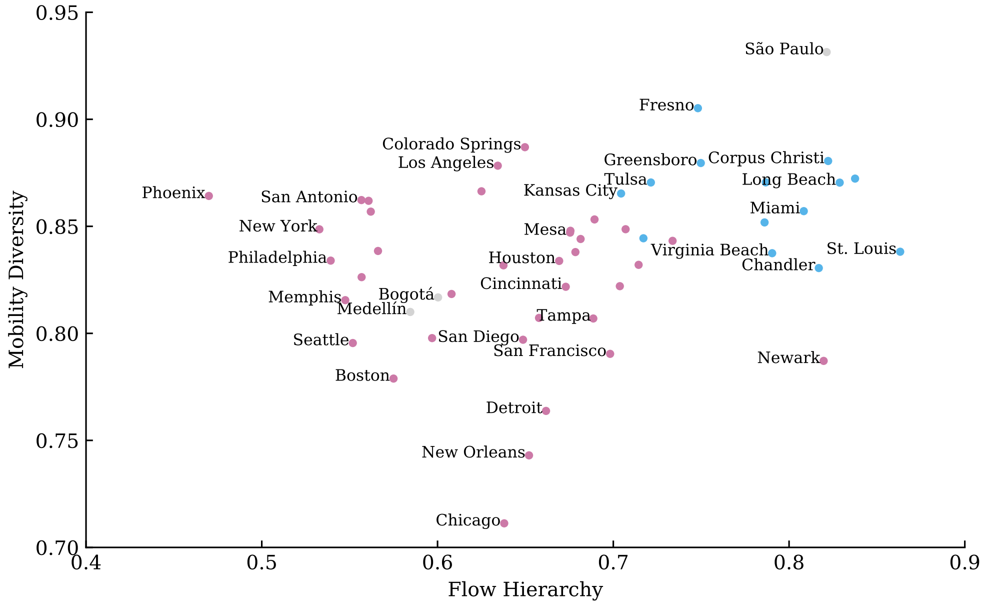

Figure 6.9: Mobility diversity and flow hierarchy of Latin American cities located between the same values from the cities of the USA. We label some cities to avoid overlapping labels.

6.4.1 Mobility features for the Latin American cities

Based on the results of the USA, we use the data of Medellín, Bogotá and São Paulo to estimate the cluster in which these three cities would be grouped. These three cities are considered to be big cities in Colombia and Brazil. However, both countries are still considered as developing countries. The mobility features are estimated using their mobility, which has a different spatial granularity as the mobility data from the USA.

However, we can observe in Figure 6.9 that the values from the Latin-American cities are similar to other cities in the USA. The values from Bogotá and Medellín are closer to cities grouped in cluster `C2`, such as San Diego, Memphis and Cincinnati. In contrast, the city of São Paulo is closer to the cities in the cluster `C1`, such as Fresno, Corpus Christi and Long Beach. This finding suggests that cities share urban similarities regardless of the cultural constructs.

For the case of `MDE`, `BGT` and `SAO`, we also compute the three mobility features in Table 6.1 from the mobility of women and men travellers. We observe that men have a higher flow hierarchy than women, and women tend to have higher hotspot cutoff level 1. As we presented in Chapter 5, men tend to have higher mobility diversity than women.

Similar to the division in Figure 6.5, high values of one feature display high values for the other feature. Moreover, we highlight that high values of hotspot cutoff level 1 mean low values of the fraction of zones in hotspot level 1.

The work travels from men travellers seem to concentrate within the hotspots more than from women, which can be related to our finding in Chapter 3 that indicated men are more likely to go to areas that are highly visited. The percentages of flow between the first and second hotspots are usually the highest compared to the other pairwise comparisons. The percentage of hotspots that women and men have in common (around 60%) is in line with the top 50 most visited areas in Chapter 3. However, the number of areas are higher than 50 inside the hotspots, reaching, for example, 87 areas in SAO 2017.

Table 6.1: Mobility features computed from the mobility of `all`, `men` and `women` travels.

City	Year	Flow Hierarchy			Hotspot Cutoff Level 1			Mobility Diversity		
		all	men	women	all	men	women	all	men	women
MDE	200s5	0.7010	0.7065	0.6927	0.8931	0.8604	0.8930	0.8624	0.8607	0.8567
	2017	0.5845	0.6119	0.5387	0.9366	0.9213	0.9520	0.8100	0.8101	0.8045
BGT	2012	0.6306	0.6425	0.6148	0.9002	0.8925	0.9232	0.8997	0.8901	0.8837
	2019	0.6003	0.5957	0.6064	0.9065	0.9362	0.9048	0.8168	0.8134	0.8115
SAO	1997	0.8014	0.8072	0.7909	0.8740	0.8663	0.8843	0.9316	0.9262	0.9294
	2007	0.8071	0.8106	0.8023	0.8889	0.8540	0.8910	0.9107	0.9072	0.9120
	2017	0.8214	0.8295	0.8106	0.8627	0.8336	0.8858	0.9314	0.9325	0.9272

6.5 Discussion

We observed in Chapter 3 that men are more likely to leave home for work earlier than women. In Chapter 4, we presented that men are more likely to have a higher travel time than women. This chapter shows that men are more likely to travel using private transportation than women, and women are more likely to travel using public transportation than men. All these findings are aligned with other works [25, 46, 111, 177].

All these differences in mobility may be a consequence of cultural constructs in the cities. As the city develops, the cultural constructs evolve with the city. Then, as the city develops, people from different sociodemographic groups can use the urban landscape differently. It might be the case that as the city evolves, the population starts to use more public than private transportation because the city also starts to provide better public systems.

The high population density might also require a better public system to serve the population, so if the city does not evolve, the city might not function optimally. In fact, urban and dense areas face more challenging issues because creating or changing the public system requires more creative solutions to reuse the environment [191]. Better public systems tend to encourage people to not opt for private transportation [195], and the unequal distribution of transport systems potentially heighten inequalities pertaining to the accessibility of basic amenities and products [187]. Therefore, transportation may be a key component to fairness and justice in mobility because individuals should be able to access locations as needed [182]. Moreover, we argue that improvements on the public transportation systems should not only take into account population volume but also aim to reach all neighbourhoods [196].

In this chapter, we found that less dense areas presented a low rate of usage of the public system and a low number of amenities close to where the individuals live. People might have to go farther in these less dense areas to access a certain amount of amenities, but the diversity of amenities increases with the distance. In contrast, for the denser areas, we observed that the usage of the public system is high, the number of available amenities is also high, but the diversity of amenities tends to be stable as we increase the distance from where people live.

Kansas City is a city that historically developed neighbourhoods and public systems with racial/ethnic segregation, urban inequality, and uneven distribution of wealth [197]. This city is an example of the cluster in which we identified low public transportation usage and high gender differences in commuting travel time. Therefore, we argue that this cluster might reveal cities that are still under severe segregation problems and inequality.

We also argue that our clustering may have revealed clusters that tend to have different types of co-located amenities. Complementary and competitive co-located amenities can benefit from each other in different ways [198]. Areas with complementary amenities offer a more diverse set of amenities, and individuals tend to fulfil more than one need in a small distant radius (e.g., eat in a restaurant, pay bills in the bank, and go to the doctor). For instance, communities that share similar sociodemographic backgrounds create environments with diverse types of amenities that people can fulfil their needs walking short distances [199].

In contrast, competitive co-located amenities can create an attractive environment to bring enough people that allows sharing. For instance, New York concentrates on similar amenities in Times Square, such as restaurants, coffee shops and stores. Because Times Square is an area that attracts numerous tourists, several amenities can compete for customers.

Our findings show that cities considered more developed are associated with more competitive co-located amenities within a short distance from home than with complementary co-located amenities. This indicates that competitive co-located amenities might be more beneficial to low gender inequality, but there is still a need to investigate further.

7. Discussion, conclusion & limitation

The increase in data granularity and availability coupled with the development of new methodologies and frameworks lead to a future where disaggregation of human patterns based on spatial, temporal, or social idiosyncrasies becomes not only possible but likely. Even though individuals are seemingly unique, with their also seemingly individual characteristics, one can often look for similar traits and group them, leading to the possibility of building predictable systems and a better understanding of human behaviour. To identify the proper idiosyncrasies, we focus this thesis on studying the patterns in the mobility of women and men.

Statistical analyses (e.g., network and data) enable us to represent and quantify mobility in an intertwined way between spatial, temporal and cultural aspects [200]. Aggregated knowledge of mobility is widely studied in the literature [3, 38, 39] without taking into account the cultural aspect that is fundamental to the understanding of inequalities and differences in mobility. The plethora of complex factors that can play a role in mobility makes us focus on one sociodemographic component: gender, combined with three factors that affect gender: socioeconomic status, marital status, and parenthood. As a city develops, its cultural constructs change, and consequently, changes in mobility are also identified. In fact, we identified in our results that cities can indeed share similar gender differences in mobility, but some singularities in the mobility patterns are also seen in few cities.

We contribute to the literature with a new methodology and pattern in mobility in which women and men differ — mobility diversity. This new pattern may be a proxy to unveil inequalities that emerge from the labour market, and to compute how opportunities are distributed over metropolitan areas. We found that socioeconomic status amplifies gender differences in mobility diversity, probably associated with the gender gaps in the labour market. For instance, we saw that higher gender differences in mobility diversity are found from individuals in the upper-income group that can be associated with higher gender gaps in the labour market for highly skilled jobs [170–172].

In line with the results of mobility diversity, we identified from the properties of the mobility networks of women and men that women tend to have a more local organisation (high average clustering), and men tend to have a more global organisation (high density). Besides, we showed that gender and socioeconomic level play a role in their mobility network structure. However, analyses considering only the properties of the network did not unveil systematic patterns that emerged from the combined dimensions of gender and socioeconomic status. Thus, these analyses highlight the complexity of unveiling disaggregated patterns in mobility.

We also found that socioeconomic status may not amplify the differences in the temporal distribution of travels made by women and men travellers indicating similar gender differences regardless of the socioeconomic group. In contrast, the commuting travel time is associated with individual income, so high travel time is correlated with high individual income in most cities of the USA. Therefore, we argue that women and men across socioeconomic groups might afford to reach different locations, impacting mobility diversity. In this way, cities with better public transportation might offer higher accessibility to opportunities for individuals across gender groups. We observed that cities with low gender differences in commuting travel time between single individuals tend to be the ones that have high usage of public transportation.

We also identified the existence of two cities profiles that share similar usage of public transportation systems and similar population volume. These profiles differ on the concentration of job distributions using mobility diversity, fraction of areas in hotspot level 1 and flow hierarchy. High dense cities with high public system usage are associated with smaller mobility diversity values than the other cities of the United States. Values of mobility diversity tend to vary between 0.71 to 0.97

for all the cities, indicating that values smaller than 0.84 are cities with more urban development than those with values higher than 0.84.

We observed that cities such as New York, Detroit, Boston, and Philadelphia that have a high fraction of travels using public transportation also show higher diversity of amenities exposed to women and low gender differences in the commuting travel time. Moreover, in these cities, considering short distances, women and men tend to be similarly exposed to a high diversity of amenities. In contrast, we found that cities such as Kansas City, Corpus Christi, Virginia Beach, and Tulsa with low public transportation usage also indicate high gender differences in mobility.

Interestingly, we found that Medellín and Bogotá in Colombia share similar mobility characteristics with cities in the United States with high development. We argue that in Colombia, the government made several interventions to benefit the mobility of their population, also taking into account gender and socioeconomic status [201, 202]. We were able to identify several changes in the mobility of Medellín and Bogotá over the years that can be a consequence of the interventions and improvements in the public system and regulations in the country. We observed that the divisions of women and men in the zones of Colombia are more balanced (50-50%) in the most recent year, which might indicate a more fair spatial inclusion of women and men in the labour market. We also showed that mobility diversity decreased over the years, indicating that job opportunities became more concentrated over the metropolitan areas. This might indicate that a complete evenly distribution of opportunities might not be related to high urban development. Our analysis is more robust for Bogotá than for Medellín as we do not have the expansion factor available for both years in Medellín.

In this thesis, we also explored the impact of household arrangements on the mobility patterns of women and men. We identified that marital status and parenthood change the network structure of the mobility of women and men and that high gender differences are seen for the commuting travel time. Thus, household arrangements continue to amplify the gender differences in mobility in the studied datasets in line with the literature [13, 185, 186]. However, gender differences are also identified for single individuals who do not have children, showing that gender roles play a role in mobility regardless of household arrangements, in contrast to the literature [185, 186].

We then in this thesis provided an overview of the characteristics and patterns of mobility that women and men differ. We indicated based on data of Brazil, Colombia, the United Kingdom and the United States of America that: (i) women and men tend to go to different locations for job opportunities, (ii) tend to cover urban areas differently, (iii) have different commuting travel time, (iv) their work travels have different peak hours, (v) use the transportation system differently and (vi) are exposed to different amount and diversity of amenities. Therefore, making opportunities accessible might require a gender-inclusive solution.

In this thesis, we do not indicate possible interventions. However, we state that our goal is not to make everyone the same but make the opportunities reachable/accessible to anyone. We have strong indicators that regulations and actions in Colombia are affecting the country positively. Besides, because we found that urban landscapes have different patterns in mobility across cities and gender groups, we argue that improving the urban landscape may help mitigate some gender inequalities.

We argue that to have a better understanding of inequalities and differences, we still need more data. We have been searching for data for three years, and the majority of data do not have available sociodemographic information, or are not enough to capture patterns of individuals in mobility. We also highlight the importance of using expansion factors to project the sample studied to the real population distribution. The spatial distribution of individuals from gender and socioeconomic groups can impact the analyses of mobility patterns, as individuals can be exposed to different opportunities and transportation modes. We also argue that it is necessary to have better models explaining the mobility for overall and specific groups of people such as gender and socioeconomic groups (see unsuccessful modelling using two major human mobility models in Appendix A.1).

Inequalities in gender, socioeconomic groups and any other type are a relevant topic to study. However, this topic requires several precautions. We do not want to make the inequalities more amplified by stereotyping individuals. In contrast, we need to be aware of the differences across sociodemographic groups to understand better the indicators that are re-feeding the inequalities. It is a complex and long journey but indeed necessary.

We are in the twenty-first century, and we still have barriers and burdens in mobility and the labour market [8]. There is a major role in the transportation system for helping people access opportunities, but not sufficient to solve the inequalities in mobility across sociodemographic groups [8, 190, 196]. We need more data collection in regard to gender, and we need more understanding about the needs and characteristics of women and men in mobility [203]. Gender gaps still exist regardless of economic and urban development, and only social and cultural interventions from governments and organisations will allow a change of paradigm that will permit sustainable and robust progress towards gender equality. Acting on the mobility dimension might be the easiest way to stop the reinforcement loop of inequalities between mobility, the labour market and cultural constructs, as mobility can help expanding individual opportunities.

7.1 Limitations

Our data analyses come from assumptions and definitions that are limited by the cities and years that we have available data. We assumed in our analyses that gender is binary classified and no physical and biological differences between any person play a role. We agree that gender is always changing its concept, and other studies should take into account other gender classifications, but our data do not allow us to take this step further. We argue that we still need more awareness and studies on better gender classification and that data such as national surveys should take into account more dimensions of this social construct.

Moreover, our data have the following limitations:

Representativeness: The dataset of each city was designed by a competent organisation to represent the universal patterns of mobility, but it can have demographic bias where some groups of people are not well represented people such as homeless and extreme rich people (income above millions). Moreover, our data represent mostly the patterns repeated with high frequency (routine), so unexpected or non-frequent behaviours are not capture on our data. Other types of data such as mobile phones or credit card records might be better data to capture non-frequent behaviours of individuals.

Subjectivity: Surveys are usually at best designed to avoid subjectivity of answers. Several precautions are done by the organisations such as asking questions in a very specific and objective manner that does not allow the person to answer creatively. However, the process is not completely safe from subjectivity. For instance, women can express their mobility with more detail than men, which we might see this expression translated on their patterns in mobility.

Spatial inhomogeneity: Urban and dense areas have better data resolution than rural and non-dense areas. As the areas are usually divided considering population density, dense areas (usually urban ones) are usually smaller areas. In this way, one non-dense area will have similar extent of multiple dense areas hidden peculiarities that happen inside a big non-dense area.

Missing data: All our datasets suffer from lack of data in many ways. For example, some areas in our data are only represented by men work travels. We argue that this is unlikely to be the true reality, but we argue that this missing data represent the small chance of women working in this particular area, not the absence of all women in it. The absence of women can be true for some countries where women are prohibited by religion and government of working outside their home, but for Latin-American countries, for example, this should be unlikely to happen. Also, our argument is valid for the spatial granularity studied. For instance, even though it is unlikely, it is possible that small neighbourhoods have only women or men working on it because of low participation of one gender in a particular job sector (e.g., nursery and army).

7.2 Ethical implications

Surveys and censuses collect data following multiple good practices [204, 205] especially because of privacy and ethical concerns. In this way, before being made available to organisations or to the public, these data require anonymisation and aggregation [206].

After the data is available, the use of this data also requires other precautions [206, 207]. First, any analysis made in the data is limited to the three major aspects: spatial, temporal and social. The data can tell the story of a particular population limited to a specific geographic area and a set of social connections. Therefore, generalisations should be made with caution.

Second, data should be aggregated to avoid privacy issues, and limited access to disaggregated data is relevant [208]. Having access to aggregated data help researchers to work on their projects, but it also constrains the data analysis. Results using aggregated data might have more uncertainties in comparison to using disaggregated data, so usefulness, relevance and privacy should be well-balanced to maintain the rigour of science and people's privacy. Moreover, choosing the proper safeguarding methodology is also crucial to data analysis as each data will allow researchers to answer a limited set of questions [207].

Third, limitations to conclusions, recommendations and interventions might be cautiously indicated [209, 210]. We should take into account that other people can use the research findings and statements in further works, so establishing the limitations and the extent to which the analyses and conclusions can be valid is indispensable.

Publications

Selected Publications

1. **M. Macedo**, L. Lotero, A. Cardillo, R. Menezes and H. Barbosa, 'Gender patterns of human mobility in Colombia: Reexamining Ravenstein's laws of migration', in *Complex Networks XI: Proceedings of the 11th Conference on Complex Networks CompleNet 2020*, pp. 269–281, Springer Nature, 2020.

2. **M. Macedo**, L. Lotero, A. Cardillo, R. Menezes and H. Barbosa, 'Differences in the spatial landscape of urban mobility: gender and socioeconomic perspectives', pp. 1–19, PLOS ONE, 2022.

3. A. Jaramillo, **M. Macedo** and R. Menezes, 'Reaching to the top: the gender effect in highly-ranked academics in computer science', *Advances in Complex Systems*, 2021.

Other Publications

1. **M. Macedo**, H. Siqueira, E. Figueiredo, C. Santana, R. Lira, A. Gokhale and C. Bastos-Filho, 'Overview on Binary Optimisation using Swarm-inspired Algorithms', *IEEE Access*, 2021.

2. R. Lira, **M. Macedo**, H. Siqueira, R. Menezes and C. Bastos-Filho, 'Modelling the social interactions in Grey Wolf Optimizer', in *2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, IEEE, 2021.

3. **M. Macedo**, L. Taw, N. Gurrapadi, R. Lira, D. Pinheiro, M. Oliveira, C. Bastos-Filho and R. Menezes, 'Fishing for interactions: a network science

approach to modeling fish school search', in *Proceedings of the Genetic and Evolutionary Computation Conference 2021*, pp. 40–48, 2021.

4. C. Bastos-Filho, F. de Lima-Neto, M. Lacerda, **M. Macedo**, C. Santana, H. Siqueira, R. Silva, A. Neto, B. Menezes and I. Albuquerque, 'Fish School Search: Account for the First Decade', in *Handbook of AI-based Metaheuristics*, CRC Press, pp. 21–42, 2021.

5. **M. Macedo**, M. Santana, W.P. dos Santos, R. Menezes and C. Bastos-Filho, 'Breast cancer diagnosis using thermal image analysis: A data-driven approach based on swarm intelligence and supervised learning for optimized feature selection', *Applied Soft Computing*, 2021.

6. H. Siqueira, C. Santana, **M. Macedo**, E. Figueiredo, A. Gokhale and C. Bastos-Filho, 'Simplified binary cat swarm optimization', *Integrated Computer-Aided Engineering*, pp. 35–50, 2021.

7. M. Oliveira, D. Pinheiro, **M. Macedo**, C. Bastos-Filho and R. Menezes, 'Uncovering the social interaction network in swarm intelligence algorithms', *Applied Network Science*. pp. 1–20, 2020.

8. H. Siqueira, **M. Macedo**, Y.D. Tadano, T.A. Alves, S.L. Stevan, D.S. Oliveira, M.H. Marinho, P.S. Neto, J.F. de Oliveira, I. Luna and L.A. Sarubbo, 'Selection of temporal lags for predicting riverflow series from hydroelectric plants using variable selection methods', *Energies*, no. 16, 2020.

9. N. Gurrapadi, L. Taw, **M. Macedo**, M. Oliveira, D. Pinheiro, C. Bastos-Filho and R. Menezes, 'Modelling the Social Interactions in Ant Colony Optimization', in *International Conference on Intelligent Data Engineering and Automated Learning 2019*, Springer, pp. 216–224, 2019.

10. **M. Macedo**, C. Santana, H. Siqueira, R.L. Rodrigues, J.L. Ramos, J. Silva, A. Maciel and C. Bastos-Filho, 'Investigation of college dropout with the fuzzy c-means algorithm', in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, IEEE, vol. 2161, pp. 187–189, 2019.

11. C. Santana, C. Bastos-Filho, **M. Macedo** and H. Siqueira, 'SBFSS: Simplified binary fish school search', in *2019 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, pp. 2595–2602, 2019.

12. L. Taw, N. Gurrapadi, **M. Macedo**, M. Oliveira, D. Pinheiro, C. Bastos-Filho and R. Menezes, 'Characterizing the social interactions in the artificial bee colony algorithm', in *2019 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, pp. 1243–1250, 2019.

13. E. Figueiredo, **M. Macedo**, H. Siqueira, C. Santana, A. Gokhale and C. Bastos-Filho, 'Swarm intelligence for clustering—A systematic review with new perspectives on data mining', *Engineering Applications of Artificial Intelligence*, Elsevier, pp. 313–329, 2019.

14. C. Santana, **M. Macedo**, H. Siqueira, A. Gokhale and C. Bastos-Filho, 'A novel binary artificial bee colony algorithm', *Future Generation Computer Systems*, Science Direct, pp. 180–96, 2019.

Bibliography

[1] C. Panigutti, M. Tizzoni, P. Bajardi, Z. Smoreda and V. Colizza, 'Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models', *Royal Society open science*, vol. 4, no. 5, p. 160 950, 2017.

[2] S. Rubrichi, Z. Smoreda and M. Musolesi, 'A comparison of spatial-based targeted disease mitigation strategies using mobile phone data', *EPJ Data Science*, vol. 7, no. 1, pp. 1–15, 2018, ISSN: 21931127. DOI: 10.1140/epjds/s13688-018-0145-9. arXiv: 1706.00690. [Online]. Available: http://dx.doi.org/10.1140/epjds/s13688-018-0145-9.

[3] H. Barbosa *et al.*, 'Human mobility: Models and applications', *Physics Reports*, 2018.

[4] P. J. Zufiria *et al.*, 'Identifying seasonal mobility profiles from anonymized and aggregated mobile phone data. Application in food security', *PLoS ONE*, vol. 13, no. 4, pp. 1–20, 2018. DOI: 10.1371/journal.pone.0195714.

[5] E. B. Peterson, K. Neels, N. Barczi and T. Graham, 'The economic cost of airline flight delay', *Journal of Transport Economics and Policy*, vol. 47, no. 1, pp. 107–121, 2013, ISSN: 00225258. [Online]. Available: http://www.jstor.org/stable/24396355.

[6] T. Cresswell, 'Towards a politics of mobility', *Environment and planning D: society and space*, vol. 28, no. 1, pp. 17–31, 2010.

[7] D. Peters, 'Gender and sustainable urban mobility', 2011.

[8] Y. A. S. Harumain *et al.*, 'The urban women travelling issue in the twenty-first century', *Journal of Regional and City Planning*, vol. 32, no. 1, pp. 1–14, 2021.

[9] C. B. Brettell, 'Marriage and migration', *Annual Review of Anthropology*, vol. 46, pp. 81–97, 2017.

[10] C. Teelken, Y. Taminiau and C. Rosenmöller, 'Career mobility from associate to full professor in academia: Micro-political practices and implicit gender stereotypes', *Studies in Higher Education*, vol. 46, no. 4, pp. 836–850, 2021.

[11] S. Ayeb-Karlsson, 'No power without knowledge: A discursive subjectivities approach to investigate climate-induced (im) mobility and wellbeing', *Social Sciences*, vol. 9, no. 6, p. 103, 2020.

[12] N. Havet, C. Bayart and P. Bonnel, 'Why do gender differences in daily mobility behaviours persist among workers?', *Transportation research part A: policy and practice*, vol. 145, pp. 34–48, 2021.

[13] L. Hu, 'Gender differences in commuting travel in the us: Interactive effects of race/ethnicity and household structure', *Transportation*, vol. 48, no. 2, pp. 909–929, 2021.

[14] S. Klasen, T. T. N. Le, J. Pieters and M. Santos Silva, 'What drives female labour force participation? comparable micro-level evidence from eight developing and emerging economies', *The Journal of Development Studies*, vol. 57, no. 3, pp. 417–442, 2021.

[15] D. Ç. Yildirim and H. Akinci, 'The dynamic relationships between the female labour force and the economic growth', *Journal of Economic Studies*, 2020.

[16] E.-M. Egger, A. Arslan and E. Zucchini, 'Does connectivity reduce gender gaps in off-farm employment?', *WIDER Working Paper 2021*, 2021.

[17] M. Ibáñez and E. Garcia-Mingo, 'Maping gendered social closure mechanisms through examination of seven male-dominated occupations', *Gender, Work & Organization*, 2021.

[18] M. Torre and J. A. Jacobs, 'The gender mobility paradox: Gender segregation and women's mobility across gender-type boundaries, 1970–2018', *Gender & Society*, p. 08 912 432 211 046 328, 2021.

[19] A. Menzel and C. Woodruff, 'Gender wage gaps and worker mobility: Evidence from the garment sector in bangladesh.', *Labour Economics*, p. 102 000, 2021.

[20] R. Simpson, 'Masculinity at work: The experiences of men in female dominated occupations', *Work, employment and society*, vol. 18, no. 2, pp. 349–368, 2004.

[21] J. McDowell, 'Masculinity and non-traditional occupations: Men's talk in women's work', *Gender, Work & Organization*, vol. 22, no. 3, pp. 273–291, 2015.

[22] J. McDonald, 'Conforming to and resisting dominant gender norms: How male and female nursing students do and undo gender', *Gender, Work & organization*, vol. 20, no. 5, pp. 561–579, 2013.

[23] M. T. Cardador, P. L. Hill and A. Salles, 'Unpacking the status-leveling burden for women in male-dominated occupations', *Administrative Science Quarterly*, p. 00 018 392 211 038 505, 2021.

[24] D. Bridges, E. Wulff and L. Bamberry, 'Resilience for gender inclusion: Developing a model for women in male-dominated occupations', *Gender, Work & Organization*, 2021.

[25] L. Mejía-Dorantes and P. Soto Villagrán, 'A review on the influence of barriers on gender equality to access the city: A synthesis approach of Mexico City and its Metropolitan Areaa', *Cities*, vol. 96, p. 102 439, 2020, ISSN: 02642751. DOI: 10.1016/j.cities.2019.102439. [Online]. Available: https://doi.org/10.1016/j.cities.2019.102439.

[26] L. Gauvin *et al.*, 'Gender gaps in urban mobility', *Humanities and Social Sciences Communications*, vol. 7, no. 1, pp. 1–13, 2020. DOI: 10.1057/s41599-020-0500-x.

[27] L. Farré, J. Jofre-Monseny and J. Torrecillas, *Commuting Time and the Gender Gap in Labor Market Participation*, IZA Discussion Papers, 13213. Available at: https://www.iza.org/publications/dp/13213/commuting-time-and-the-gender-gap-in-labor-market-participation, 2020.

[28] G. C. Moreira and V. A. Ceccato, 'Gendered mobility and violence in the São Paulo metro, Brazil', *Urban Studies*, vol. 58, pp. 1–20, 2020. DOI: 10.1177/0042098019885552.

[29] J.-J. Lin and Y.-C. Cheng, 'Access to jobs and apartment rents', *Journal of Transport Geography*, vol. 55, pp. 121–128, 2016.

[30] F. Andersson, J. C. Haltiwanger, M. J. Kutzbach, H. O. Pollakowski and D. H. Weinberg, 'Job displacement and the duration of joblessness: The role of spatial mismatch', *Review of Economics and Statistics*, vol. 100, no. 2, pp. 203–218, 2018.

[31] V. Frias-Martinez, J. Virseda-Jerez and E. Frias-Martinez, 'On the relation between socio-economic status and physical mobility', *Information Technology for Development*, vol. 18, no. 2, pp. 91–106, 2012. DOI: 10.1080/02681102.2011.630312.

[32] A. Bassolas *et al.*, 'Hierarchical organization of urban mobility and its connection with city livability', *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[33] T.-I. Gizelis, S. Pickering and H. Urdal, 'Conflict on the urban fringe: Urbanization, environmental stress, and urban unrest in africa', *Political Geography*, vol. 86, p. 102 357, 2021.

[34] Y. S. Chang, S. J. Jo, Y.-T. Lee and Y. Lee, 'Population density or populations size. which factor determines urban traffic congestion?', *Sustainability*, vol. 13, no. 8, p. 4280, 2021.

[35] J. Lu, B. Li, H. Li and A. Al-Barakani, 'Expansion of city scale, traffic modes, traffic congestion, and air pollution', *Cities*, vol. 108, p. 102 974, 2021.

[36] U. Nations, *Sustainable development goals (sdgs)*, Available at: https://sdgs.un.org/goals. (visited on 02/02/2021).

[37] D. Brockmann, L. Hufnagel and T. Geisel, 'The scaling laws of human travel', *Nature*, vol. 439, no. 7075, p. 462, 2006.

[38] M. C. Gonzalez, C. A. Hidalgo and A.-L. Barabási, 'Understanding individual human mobility patterns', *Nature*, vol. 453, no. 7196, p. 779, 2008. DOI: 10.1038/nature06958.

[39] C. Song, Z. Qu, N. Blumm and A.-L. Barabási, 'Limits of predictability in human mobility', *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010. DOI: 10.1126/science.1177170.

[40] H. Barbosa, F. B. de Lima-Neto, A. Evsukoff and R. Menezes, 'The effect of recency to human mobility', *EPJ Data Science*, vol. 4, no. 1, p. 21, 2015. DOI: epjds/s13688-015-0059-8.

[41] L. Alessandretti, U. Aslak and S. Lehmann, 'The scales of human mobility', *Nature*, vol. 587, no. 7834, pp. 402–407, 2020. DOI: 10.1038/s41586-020-2909-1.

[42] F. Simini, M. C. González, A. Maritan and A.-L. Barabási, 'A universal model for mobility and migration patterns', *Nature*, vol. 484, no. 7392, p. 96, 2012. DOI: 10.1038/nature10856.

[43] M. Lenormand *et al.*, 'Influence of sociodemographic characteristics on human mobility', *Scientific reports*, vol. 5, no. 1, pp. 1–15, 2015.

[44] G. Carra, I. Mulalic, M. Fosgerau and M. Barthelemy, 'Modelling the relation between income and commuting distance', *Journal of the Royal Society Interface*, vol. 13, no. 119, 2016. DOI: 10.1098/rsif.2016.0306.

[45] P. Gustafson, 'Work-related travel, gender and family obligations', *Work, employment and society*, vol. 20, no. 3, pp. 513–530, 2006. DOI: 10.1177/0950017006066999.

[46] Y. J. Singh, 'Is smart mobility also gender-smart?', *Journal of Gender Studies*, vol. 29, no. 7, pp. 832–846, 2020. DOI: 10.1080/09589236.2019.1650728.

[47] L. Rodríguez, J. Palanca, E. del Val and M. Rebollo, 'Analyzing urban mobility paths based on users' activity in social networks', *Future Generation Computer Systems*, vol. 102, pp. 333–346, 2020, ISSN: 0167739X. DOI: 10.1016/j.future.2019.07.072. [Online]. Available: https://doi.org/10.1016/j.future.2019.07.072.

[48] I. Psylla, P. Sapiezynski, E. Mones and S. Lehmann, 'The role of gender in social network organization', *PLoS ONE*, vol. 12, no. 12, pp. 1–21, 2017, ISSN: 19326203. DOI: 10.1371/journal.pone.0189873.

[49] M. Arntz, S. B. Yahmed and F. Berlingieri, 'Working from home and covid-19: The chances and risks for gender gaps', *Intereconomics*, vol. 55, no. 6, pp. 381–386, 2020.

[50] D. Kondor, S. Grauwin, Z. Kallus, I. Gódor, S. Sobolevsky and C. Ratti, 'Prediction limits of mobile phone activity modelling', *Royal Society Open Science*, vol. 4, no. 2, 2017, ISSN: 20545703. DOI: 10.1098/rsos.160900.

[51] T. Liu, Z. Yang, Y. Zhao, C. Wu, Z. Zhou and Y. Liu, 'Temporal understanding of human mobility: A multi-time scale analysis', *Plos One*, vol. 13, e0207697, 2018, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0207697.

[52] F. Alhasoun, M. Alhazzani, F. Aleissa, R. Alnasser and F. González, 'City scale next place prediction from sparse data through similar strangers', in *Proceedings of the ACM KDD Workshop, Halifax, NS, Canada*, vol. 14, 2017.

[53] H. Jiang, Y. Wu, Y. Zhang, S. Wang and Y. Zhang, 'From social community to spatio-temporal information: A new method for mobile data exploration', *Journal of Visual Languages and Computing*, vol. 41, pp. 1339–1351, 2017, ISSN: 1045926X. DOI: 10.1016/j.jvlc.2017.05.002.

[54] S. Jiang, J. Ferreira and M. C. Gonzalez, 'Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore', *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, 2017, ISSN: 2332-7790. DOI: 10.1109/TBDATA.2016.2631141. [Online]. Available: http://ieeexplore.ieee.org/document/7755745/.

[55] S. Luo, F. Morone, C. Sarraute, M. Travizano and H. A. Makse, 'Inferring personal economic status from social network location', *Nature Communications*, vol. 8, no. May, pp. 1–7, 2017, ISSN: 20411723. DOI: 10.1038/ncomms15227. arXiv: 1704.01572.

[56] R. Di Clemente, M. Luengo-Oroz, M. Travizano, S. Xu, B. Vaitla and M. C. González, 'Sequences of purchases in credit card data reveal lifestyles in urban populations', *Nature Communications*, vol. 9, no. 1, pp. 1–8, 2018, ISSN: 20411723. DOI: 10.1038/s41467-018-05690-8. arXiv: 1703.00409. [Online]. Available: http://dx.doi.org/10.1038/s41467-018-05690-8.

[57] K. Qian, C. Wu, Z. Yang, Y. Liu and K. Jamieson, 'Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi', *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing - Mobihoc '17*, pp. 1–10, 2017. DOI: 10.1145/3084041.3084067. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3084041.3084067.

[58] C. E. Kontokosta and N. Johnson, 'Urban phenology: Toward a real-time census of the city using Wi-Fi data', *Computers, Environment and Urban Systems*, vol. 64, pp. 144–153, 2017. DOI: 10.1016/j.compenvurbsys.2017.01.011. [Online]. Available: http://dx.doi.org/10.1016/j.compenvurbsys.2017.01.011.

[59] V. Kulkarni, A. Mahalunkar, B. Garbinato and J. D. Kelleher, 'On the Inability of Markov Models to Capture Criticality in Human Mobility', *arXiv preprint*, pp. 1–15, 2018. arXiv: 1807.11386. [Online]. Available: http://arxiv.org/abs/1807.11386.

[60] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann and A. Baronchelli, 'Evidence for a conserved quantity in human mobility', *Nature Human Behaviour*, vol. 2, no. 7, pp. 485–491, 2018, ISSN: 23973374. DOI: 10.1038/s41562-018-0364-x.

[61] W. Zheng, X. Huang and Y. Li, 'Understanding the tourist mobility using gps: Where is the next place?', *Tourism Management*, vol. 59, pp. 267–280, 2017. [Online]. Available: `http://dx.doi.org/10.1016/j.tourman.2016.08.009`.

[62] Q. Liu, Z. Wang and X. Ye, 'Comparing mobility patterns between residents and visitors using geo-tagged social media data', *Transactions in GIS*, no. November 2017, pp. 1372–1389, 2018, ISSN: 14679671. DOI: `10.1111/tgis.12478`.

[63] C. Ponte, H. P. M. Melo, C. Caminha, J. S. Andrade and V. Furtado, 'Traveling heterogeneity in public transportation', *EPJ Data Science*, vol. 7, no. 1, 2018, ISSN: 21931127. DOI: `10.1140/epjds/s13688-018-0172-6`. [Online]. Available: `http://dx.doi.org/10.1140/epjds/s13688-018-0172-6`.

[64] N. A. Rehman, K. Relia and R. Chunara, 'Creating full individual-level location timelines from sparse social media data', in *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2018, pp. 379–388.

[65] Z. Mao, Y. Jiang, G. Min, S. Leng, X. Jin and K. Yang, 'Mobile social networks: Design requirements, architecture, and state-of-the-art technology', *Computer Communications*, vol. 100, pp. 1–19, 2017, ISSN: 01403664. DOI: `10.1016/j.comcom.2016.11.006`.

[66] J. Hinds and A. N. Joinson, 'What demographic attributes do our digital footprints reveal? a systematic review', *PloS one*, vol. 13, no. 11, 2018.

[67] Y. Wang, Q. Wang and J. E. Taylor, 'Aggregated responses of human mobility to severe winter storms: An empirical study', *PLoS ONE*, vol. 12, no. 12, pp. 1–15, 2017, ISSN: 19326203. DOI: `10.1371/journal.pone.0188734`.

[68] M. Li, R. Westerholt, H. Fan and A. Zipf, 'Assessing spatiotemporal predictability of LBSN: a case study of three Foursquare datasets', *GeoInformatica*, vol. 22, no. 3, pp. 541–561, 2018, ISSN: 13846175. DOI: `10.1007/s10707-016-0279-5`.

[69] A. Dmowska, T. F. Stepinski and P. Netzel, 'Comprehensive framework for visualizing and analyzing spatio-temporal dynamics of racial diversity in the entire United States', *PLoS ONE*, vol. 12, no. 3, pp. 1–20, 2017, ISSN: 19326203. DOI: 10.1371/journal.pone.0174993.

[70] D. Piovani, E. Arcaute, G. Uchoa, A. Wilson and M. Batty, 'Measuring Accessibility using Gravity and Radiation Models', *arXiv preprint*, 2018, ISSN: 2054-5703. DOI: 10.1098/rsos.171668. arXiv: 1802.06421. [Online]. Available: http://arxiv.org/abs/1802.06421.

[71] J. G. Hoendervanger, A. F. Ernst, C. J. Albers, M. P. Mobach and N. W. Van Yperen, 'Individual differences in satisfaction with activity-based work environments', *PLoS ONE*, vol. 13, no. 3, pp. 1–15, 2018, ISSN: 19326203. DOI: 10.1371/journal.pone.0193878.

[72] *Metrô São Paulo — pesquisa origem e destino: 50 anos.* Available at: http://www.metro.sp.gov.br/pesquisa-od/pesquisa-od-50-anos.aspx. (visited on 02/02/2021).

[73] *Datos abiertos del area metropolitana de valle de aburrá*, Available at: https://datosabiertos.metropol.gov.co/search/field_topic/movilidad-y-transporte-2/type/dataset?sort_by=changed. (visited on 02/02/2021).

[74] U. C. Bureau, *Longitudinal employer-household dynamics*, Available at: https://lehd.ces.census.gov/data/. (visited on 02/02/2021).

[75] S. Ruggles *et al.*, *Ipums usa*, Available at: https://usa.ipums.org/usa/. (visited on 02/02/2021).

[76] U. C. Bureau, *American community survey*, Available at: https://www.census.gov/programs-surveys/acs.html. (visited on 02/02/2021).

[77] U. D. S. from the Department for Transport, *National travel survey*, Available at: http://doi.org/10.5255/UKDA-SN-5340-10. (visited on 02/02/2021).

[78] OpenStreetMap contributors, *Planet dump retrieved from https://planet.osm.org*, https://www.openstreetmap.org, 2017.

[79] S. Salomón, C. Tîrnauca, R. Duque and J. L. Montaña, 'User identification from mobility traces', *Journal of Ambient Intelligence and Humanized Computing*, 2018, ISSN: 1868-5137. DOI: 10.1007/s12652-018-1117-4. [Online]. Available: http://link.springer.com/10.1007/s12652-018-1117-4.

[80] M. Lenormand, H. Samaniego, J. C. Chaves, V. da Fonseca Vieira, M. A. H. B. d. Silva and A. G. Evsukoff, 'Entropy as a measure of attractiveness and socioeconomic complexity in rio de janeiro metropolitan area', *Entropy*, vol. 22, no. 3, p. 368, 2020. DOI: 10.3390/e22030368.

[81] K.-S. Lee, J. K. Eom, J. Lee and S. Ko, 'Analysis of the activity and travel patterns of the elderly using mobile phone-based hourly locational trajectory data: Case study of gangnam, korea', *Sustainability*, vol. 13, no. 6, p. 3025, 2021.

[82] Y. Wang, C. Breitinger, B. Sommer, F. Schreiber and H. Reiterer, 'Comparing Sequential and Temporal Patterns from Human Mobility Data for Next-Place Prediction', *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization - UMAP '18*, pp. 157–164, 2018, ISSN: 1365-3148. DOI: 10.1145/3213586.3226212. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3213586.3226212.

[83] W. Geng and G. Yang, 'Partial correlation between spatial and temporal regularities of human mobility', *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017. [Online]. Available: http://dx.doi.org/10.1038/s41598-017-06508-1.

[84] C. Yu *et al.*, 'Modeling User Activity Patterns for Next-Place Prediction', *IEEE Systems Journal*, vol. 11, no. 2, pp. 1060–1071, 2017, ISSN: 19379234. DOI: 10.1109/JSYST.2015.2445919.

[85] X. Y. Yan, W. X. Wang, Z. Y. Gao and Y. C. Lai, 'Universal model of individual and population mobility on diverse spatial scales', *Nature Communications*, vol. 8, no. 1, pp. 1–9, 2017, ISSN: 20411723. DOI: 10.1038/s41467-017-01892-8. [Online]. Available: http://dx.doi.org/10.1038/s41467-017-01892-8.

[86] J. Monteiro, B. Martins and J. M. Pires, 'A hybrid approach for the spatial disaggregation of socio-economic indicators', *International Journal of Data Science and Analytics*, vol. 5, no. 2, pp. 189–211, 2017, ISSN: 2364-415X. DOI: 10.1007/s41060-017-0080-z. [Online]. Available: http://link.springer.com/10.1007/s41060-017-0080-z.

[87] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti and A.-L. Barabási, 'Returners and explorers dichotomy in human mobility', *Nature communications*, vol. 6, p. 8166, 2015.

[88] M. Lenormand, J. M. Arias, M. San Miguel and J. J. Ramasco, 'On the importance of trip destination for modelling individual human mobility patterns', *Journal of The Royal Society Interface*, vol. 17, no. 171, p. 20 200 673, 2020. DOI: 10.1098/rsif.2020.0673. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2020.0673.

[89] Z. Chen, S. Kelty, B. F. Welles, J. P. Bagrow, R. Menezes and G. Ghoshal, 'Contrasting social and non-social sources of predictability in human mobility', *arXiv preprint arXiv:2104.13282*, 2021.

[90] C. Fan, Y. Liu, J. Huang, Z. Rong and T. Zhou, 'Correlation between social proximity and mobility similarity', *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017, ISSN: 20452322. DOI: 10.1038/s41598-017-12274-x. arXiv: arXiv:1609.02669v1. [Online]. Available: http://dx.doi.org/10.1038/s41598-017-12274-x.

[91] L. Alessandretti, S. Lehmann and A. Baronchelli, 'Understanding the interplay between social and spatial behaviour', *EPJ Data Science*, vol. 7, no. 1, p. 36, 2018.

[92] C. Song, T. Koren, P. Wang and A.-L. Barabási, 'Modelling the scaling properties of human mobility', *Nature Physics*, vol. 6, no. 10, p. 818, 2010.

[93] M. Zhao, L. Mason and W. Wang, 'Empirical study on human mobility for mobile wireless networks', in *MILCOM 2008-2008 IEEE Military Communications Conference*, IEEE, 2008, pp. 1–7.

[94] S. Erlander and N. F. Stewart, *The gravity model in transportation analysis: theory and extensions*. Vsp, 1990, vol. 3.

[95] X. Li, H. Tian, D. Lai and Z. Zhang, 'Validation of the gravity model in predicting the global spread of influenza', *International journal of environmental research and public health*, vol. 8, no. 8, pp. 3134–3143, 2011.

[96] R. Patuelli, A. Reggiani, S. P. Gorman, P. Nijkamp and F.-J. Bade, 'Network analysis of commuting flows: A comparative static approach to german data', *Networks and Spatial Economics*, vol. 7, no. 4, pp. 315–331, 2007.

[97] M. Lenormand, A. Bassolas and J. J. Ramasco, 'Systematic comparison of trip distribution laws and models', *Journal of Transport Geography*, vol. 51, pp. 158–169, 2016.

[98] Y. Yang, C. Herrera, N. Eagle and M. C. González, 'Limits of predictability in commuting flows in the absence of data for calibration', *Scientific reports*, vol. 4, p. 5662, 2014.

[99] S. A. Stouffer *et al.*, *Intervening opportunities: a theory relating mobility and distance*. JSTOR, 1940.

[100] M. L. Bright and D. S. Thomas, 'Interstate migration and intervening opportunities', *American Sociological Review*, vol. 6, no. 6, pp. 773–783, 1941.

[101] E. R. Ruiter, 'Toward a better understanding of the intervening opportunities model', *Transportation Research*, vol. 1, no. 1, pp. 47–56, 1967.

[102] G. Krings, F. Calabrese, C. Ratti and V. D. Blondel, 'Urban gravity: A model for inter-city telecommunication flows', *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 07, p. L07003, 2009.

[103] P. Kaluza, A. Kölzsch, M. T. Gastner and B. Blasius, 'The complex network of global cargo ship movements', *Journal of the Royal Society Interface*, vol. 7, no. 48, pp. 1093–1103, 2010.

[104] M. Barthélemy, 'Spatial networks', *Physics Reports*, vol. 499, no. 1-3, pp. 1–101, 2011.

[105] L. Lotero, A. Cardillo, R. Hurtado and J. Gómez-Gardeñes, 'Several multiplexes in the same city: The role of socioeconomic differences in urban mobility', in *Interconnected networks*, Springer, 2016, pp. 149–164.

[106] A. Aleta, S. Meloni and Y. Moreno, 'A multilayer perspective for the analysis of urban transportation systems', *Scientific reports*, vol. 7, p. 44 359, 2017.

[107] E. Strano, S. Shai, S. Dobson and M. Barthelemy, 'Multiplex networks in metropolitan areas: Generic features and local effects', *Journal of The Royal Society Interface*, vol. 12, no. 111, p. 20 150 651, 2015.

[108] B. Petrongolo and M. Ronchi, 'Gender gaps and the structure of local labor markets', *Labour Economics*, vol. 64, no. November 2019, p. 101 819, 2020, ISSN: 09275371. DOI: 10.1016/j.labeco.2020.101819. [Online]. Available: https://doi.org/10.1016/j.labeco.2020.101819.

[109] V. Montoya-Robledo *et al.*, 'Gender stereotypes affecting active mobility of care in Bogotá', *Transportation Research Part D: Transport and Environment*, vol. 86, p. 102 470, 2020. DOI: 10.1016/j.trd.2020.102470.

[110] S. Chant, 'Cities through a "gender lens": A golden "urban age" for women in the global South?', *Environment and Urbanization*, vol. 25, no. 1, pp. 9–29, 2013, ISSN: 09562478. DOI: 10.1177/0956247813477809.

[111] L. Craig and T. P. van Tienoven, 'Gender, mobility and parental shares of daily travel with and for children: A cross-national time use comparison', *Journal of transport geography*, vol. 76, pp. 93–102, 2019.

[112] S. Hanson, 'Gender and mobility: New approaches for informing sustainability', *Gender, Place & Culture*, vol. 17, no. 1, pp. 5–23, 2010.

[113] K. Nosal Hoy and S. Puławska-Obiedowska, 'The travel behaviour of polish women and adaptation of transport systems to their needs', *Sustainability*, vol. 13, no. 5, p. 2693, 2021.

[114] A. M. Agogino, 'Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering', *Lecture, College of Engineering, University of Oklahoma*, 2007.

[115] L. Holman, D. Stuart-Fox and C. E. Hauser, 'The gender gap in science: How long until women are equally represented?', *PLoS biology*, vol. 16, no. 4, e2004956, 2018.

[116] J. M. Sheltzer and J. C. Smith, 'Elite male faculty in the life sciences employ fewer women', *Proceedings of the National Academy of Sciences*, vol. 111, no. 28, pp. 10 107–10 112, 2014.

[117] L. Craig and J. E. Brown, 'Weekend work and leisure time with family and friends: Who misses out?', *Journal of Marriage and Family*, vol. 76, no. 4, pp. 710–727, 2014.

[118] C. Fan, Y. Liu, J. Huang, Z. Rong and T. Zhou, 'Correlation between social proximity and mobility similarity', *Scientific Reports*, vol. 7, p. 11 975, 2017. DOI: 10.1038/s41598-017-12274-x. [Online]. Available: http://dx.doi.org/10.1038/s41598-017-12274-x.

[119] W. Alonso, 'Location and land use. toward a general theory of land rent.', *Harvard University Press*, 1964.

[120] A. Carlsson-Kanyama and A.-L. Linden, 'Travel patterns and environmental effects now and in the future:: Implications of differences in energy consumption among socio-economic groups', *Ecological Economics*, vol. 30, no. 3, pp. 405–417, 1999. DOI: 10.1016/S0921-8009(99)00006-3.

[121] L. Camarero and J. Oliva, 'Thinking in rural gap: mobility and social inequalities', *Palgrave Communications*, vol. 5, no. 1, pp. 1–7, 2019, ISSN: 20551045. DOI: 10.1057/s41599-019-0306-x. [Online]. Available: http://dx.doi.org/10.1057/s41599-019-0306-x.

[122] B. F. Milan and F. Creutzig, 'Lifting peripheral fortunes: Upgrading transit improves spatial, income and gender equity in medellin', *Cities*, vol. 70, pp. 122–134, 2017.

[123] M. M. Ferree, 'Filling the glass: Gender perspectives on families', *Journal of Marriage and Family*, vol. 72, no. 3, pp. 420–439, 2010.

[124] I. Johnston-Anumonwo, ‘The influence of household type on gender differences in work trip distance’, *The Professional Geographer*, vol. 44, no. 2, pp. 161–169, 1992.

[125] S. F. Mennino and A. Brayfield, ‘Job-family trade-offs: The multidimensional effects of gender’, *Work and occupations*, vol. 29, no. 2, pp. 226–256, 2002.

[126] C. Bastiaansen, E. Verspeek and H. van Bakel, ‘Gender differences in the mitigating effect of co-parenting on parental burnout: The gender dimension applied to covid-19 restrictions and parental burnout levels’, *Social Sciences*, vol. 10, no. 4, p. 127, 2021.

[127] L. Bari, T. Turner and M. O’Sullivan, ‘Gender differences in solo self-employment: Gendered flexibility and the effects of parenthood’, *Gender, Work & Organization*, 2021.

[128] Y. Lee, M. Hickman and S. Washington, ‘Household type and structure, time-use pattern, and trip-chaining behavior’, *Transportation Research Part A: Policy and Practice*, vol. 41, no. 10, pp. 1004–1020, 2007.

[129] *Secretaría distrital de movilidad*, Available at: `https://www.simur.gov.co/`. (visited on 02/02/2021).

[130] L. Kish, *Statistical design for research.* John Wiley & Sons, 2005.

[131] O. R. N. laboratory funded by Federal Highway Administration, *National household travel survey*, Available at: `https://nhts.ornl.gov/`. (visited on 02/02/2021).

[132] *Instituto brasileiro de geografia e estatística (ibge)*, Available at: `https://www.ibge.gov.br/`. (visited on 02/02/2021).

[133] *Associação brasileira de empresas de pesquisa (abep)*, Available at: `https://www.abep.org/`. (visited on 02/02/2021).

[134] M. Macedo, L. Lotero, A. Cardillo, H. Barbosa and R. Menezes, ‘Gender patterns of human mobility in Colombia: Reexamining Ravenstein’s laws of migration’, in *Complex Networks XI: Proceedings of the 11th Conference on Complex Networks CompleNet 2020*, Springer Nature, 2020, pp. 269–

281. DOI: 10.1007/978-3-030-40943-2_23. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-40943-2%5C_23.

[135] A. Barrat, M. Barthelemy, R. Pastor-Satorras and A. Vespignani, 'The architecture of complex weighted networks', *Proceedings of the national academy of sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.

[136] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, 'Complex networks: Structure and dynamics', *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[137] T. Opsahl and P. Panzarasa, 'Clustering in weighted networks', *Social Networks*, vol. 31, no. 2, pp. 155–163, 2009, ISSN: 0378-8733. DOI: https://doi.org/10.1016/j.socnet.2009.02.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378873309000070.

[138] J. J. Mijs and E. L. Roe, 'Is america coming apart? socioeconomic segregation in neighborhoods, schools, workplaces, and social networks, 1970–2020', *Sociology Compass*, e12884, 2021.

[139] F. JFdS *et al.*, 'Inequality and income segregation in brazilian cities: A nationwide analysis', 2021.

[140] A. L. Martínez and O. E. C. Mina, 'Socioeconomic residential segregation and income inequality in bogotá: An analysis based on census data of 2005', *Urban Socio-Economic Segregation and Income Inequality*, p. 433, 2021.

[141] A. Comandon and P. Veneri, 'Residential segregation between income groups in international perspective', *Urban Socio-Economic Segregation and Income Inequality*, p. 27, 2021.

[142] J. P. Bagrow and E. M. Bollt, 'An information-theoretic, all-scales approach to comparing networks', en, *Applied Network Science*, vol. 4, no. 1, p. 45, Jul. 2019, ISSN: 2364-8228. DOI: 10.1007/s41109-019-0156-x. [Online]. Available: https://doi.org/10.1007/s41109-019-0156-x (visited on 12/06/2021).

[143] T. Osán, D. Bussandri and P. Lamberti, 'Quantum metrics based upon classical jensen–shannon divergence', *Physica A: Statistical Mechanics and its Applications*, p. 127 001, 2022.

[144] F. Pérez-Cruz, 'Kullback-leibler divergence estimation of continuous distributions', in *2008 IEEE international symposium on information theory*, IEEE, 2008, pp. 1666–1670.

[145] T. Reisch, G. Heiler, J. Hurt, P. Klimek, A. Hanbury and S. Thurner, 'Behavioral gender differences are reinforced during the covid-19 crisis', *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[146] D. Heinrichs and J. S. Bernet, 'Public transport and accessibility in informal settlements: Aerial cable cars in medellín, colombia', *Transportation research procedia*, vol. 4, pp. 55–67, 2014.

[147] L. Craig and A. Powell, 'Non-parental childcare, time pressure and the gendered division of paid work, domestic work and parental childcare', *Community, Work & Family*, vol. 16, no. 1, pp. 100–119, 2013.

[148] M. Colley and R. N. Buliung, 'Gender differences in school and work commuting mode through the life cycle: Exploring trends in the greater toronto and hamilton area, 1986 to 2011', *Transportation Research Record*, vol. 2598, no. 1, pp. 102–109, 2016.

[149] L. Ruppanner and J. Treas, 'Working weekends: Changing european time regimes and gender inequality in household labor', *Journal of Family Issues*, vol. 36, no. 13, pp. 1782–1809, 2015.

[150] T. Buchanan, A. McFarlane and A. Das, 'The gender gap in market work hours among canadians: Examining essential (ist) linkages to parenting time and household labour hours', *Journal of Interdisciplinary Economics*, vol. 33, no. 2, pp. 167–189, 2021.

[151] M. van Ham, T. Tammaru, R. Ubarevičienė and H. Janssen, *Urban socio-economic segregation and income inequality: A global perspective.* Springer Nature, 2021.

[152] D. Albouy, 'Are big cities bad places to live? estimating quality of life across metropolitan areas', National Bureau of Economic Research, Tech. Rep., 2008.

[153] U. Henz, 'Couples' daily childcare schedules: Gendered patterns and variations', *Families, Relationships and Societies*, 2021.

[154] G. Ranson, 'Who's (really) in charge? mothers and executive responsibility in'non-traditional'families', *Families, Relationships and Societies*, vol. 2, no. 1, pp. 79–95, 2013.

[155] Catalyst, *Women in science, technology, engineering, and mathematics (stem)*, 2020. [Online]. Available: `https://www.catalyst.org/research/women-in-science-technology-engineering-and-mathematics-stem/` (visited on 21/07/2021).

[156] C. E. Shannon, 'A mathematical theory of communication', *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[157] L. Pappalardo, D. Pedreschi, Z. Smoreda and F. Giannotti, 'Using big data to study the link between human mobility and socio-economic development', *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, pp. 871–878, 2015. DOI: `10.1109/BigData.2015.7363835`.

[158] Y. Leo, E. Fleury, J. I. Alvarez-Hamelin, C. Sarraute and M. Karsai, 'Socioeconomic correlations and stratification in social-communication networks', *Journal of The Royal Society Interface*, vol. 13, no. 125, p. 20 160 598, 2016. DOI: `10.1098/rsif.2016.0598`.

[159] J. Gao, Y. C. Zhang and T. Zhou, 'Computational socioeconomics', *Physics Reports*, vol. 817, pp. 1–104, 2019, ISSN: 03701573. DOI: `10.1016/j.physrep.2019.05.002`. [Online]. Available: `https://doi.org/10.1016/j.physrep.2019.05.002`.

[160] J. R. M. de Barros, 'The impact of the international financial crisis on Brazil (ARI)', *International Cooperation & Development*, 2010.

[161] F. Ferrari Filho, 'Brazil's response: How did financial regulation and monetary policy influence recovery?', *Brazilian Journal of Political Economy*, vol. 31, no. 5, pp. 880–888, 2011.

[162] M. Delacre, D. Lakens and C. Leys, 'Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test', *International Review of Social Psychology*, vol. 30, no. 1, pp. 92–101, 2017. DOI: `10.5334/irsp.82`.

[163] B. G. Tabachnick and L. S. Fidell, *Experimental designs using ANOVA*. Belmont (CA), USA: Thomson/Brooks/Cole, 2007.

[164] A. M. Brown, 'A new software for carrying out one-way ANOVA post hoc tests', *Computer methods and programs in biomedicine*, vol. 79, no. 1, pp. 89–95, 2005. DOI: `10.1016/j.cmpb.2005.02.007`.

[165] F. Asgari, V. Gauthier and M. Becker, 'A survey on Human Mobility and its applications', *arXiv*, p. 1307.0814, 2013. eprint: `1307.0814`. [Online]. Available: `http://arxiv.org/abs/1307.0814`.

[166] H. Ballis and L. Dimitriou, 'Revealing personal activities schedules from synthesizing multi-period origin-destination matrices', *Transportation research part B: methodological*, vol. 139, pp. 224–258, 2020. DOI: `10.1016/j.trb.2020.06.007`.

[167] W.-S. Ng and A. Acker, *Understanding urban travel behaviour by gender for efficient and equitable transport policies*, International Transport Forum Discussion Paper. Available at: `https://www.itf-oecd.org/understanding-urban-travel-behaviour-gender-efficient-and-equitable-transport-policies`, 2018.

[168] E. A. Vasconcellos, 'Urban transport policies in brazil: The creation of a discriminatory mobility system', *Journal of transport geography*, vol. 67, pp. 85–91, 2018. DOI: `10.1016/j.jtrangeo.2017.08.014`.

[169] E. A. De Vasconcellos, 'Urban change, mobility and transport in São Paulo: Three decades, three cities', *Transport Policy*, vol. 12, no. 2, pp. 91–104, 2005. DOI: `10.1016/j.tranpol.2004.12.001`.

[170] H. Stier and M. Yaish, 'Occupational segregation and gender inequality in job quality: A multi-level approach', *Work, employment and society*, vol. 28, no. 2, pp. 225–246, 2014. DOI: 10.1177/0950017013510758.

[171] H. M. Clarke, 'Gender stereotypes and gender-typed work', in *Handbook of Labor, Human Resources and Population Economics*, Springer, 2020, pp. 1–23. DOI: 10.1007/978-3-319-57365-6_21-1.

[172] R. da Rocha Grangeiro, L. E. N. Silva and C. Esnard, 'I broke the glass ceiling, now what? overview of metaphors to explain gender inequality in organizations', *International Journal of Organizational Analysis*, 2021.

[173] F. Martin and A. Dragojlovic, 'Gender, Mobility Regimes, and Social Transformation in Asia', *Journal of Intercultural Studies*, vol. 40, no. 3, pp. 275–286, 2019, ISSN: 14699540. DOI: 10.1080/07256868.2019.1599166.

[174] L. Lotero, R. G. Hurtado, L. M. Floría and J. Gómez-Gardeñes, 'Rich do not rise early: Spatio-temporal patterns in the mobility networks of different socio-economic classes', *Royal Society open science*, vol. 3, no. 10, p. 150 654, 2016. DOI: 10.1098/rsos.150654.

[175] L. Levin and K. Thoresson, 'Gender equality and "smart" mobility', *Gendering Smart Mobilities*, pp. 143–161, 2019. DOI: 10.4324/9780429466601-9.

[176] S. Althobaiti, S. Alghumayjan, M. R. Frank, E. Moro, A. Alabdulkareem and A. Pentland, 'Housing prices and the skills composition of neighborhoods', *Frontiers in big Data*, vol. 4, p. 25, 2021.

[177] T. Saigal, A. K. Vaish and N. Rao, 'Socio-economic-and gender-based differentials in choice of sustainable mode of transport: Evidence from jaipur, india', *Millennial Asia*, p. 0 976 399 621 993 432, 2021.

[178] J. Säve-Söderbergh, 'Gender gaps in salary negotiations: Salary requests and starting salaries in the field', *Journal of Economic Behavior & Organization*, vol. 161, pp. 35–51, 2019.

[179] K. M. Herman and R. Larouche, 'Active commuting to work or school: Associations with subjective well-being and work-life balance', *Journal of Transport & Health*, vol. 22, p. 101 118, 2021.

[180] L. Dobbs, 'Stuck in the slow lane: Reconceptualizing the links between gender, transport and employment', *Gender, Work & Organization*, vol. 14, no. 2, pp. 85–108, 2007.

[181] Y. Hail and R. McQuaid, 'The concept of fairness in relation to women transport users', *Sustainability*, vol. 13, no. 5, p. 2919, 2021.

[182] R. H. Pereira, T. Schwanen and D. Banister, 'Distributive justice and equity in transportation', *Transport reviews*, vol. 37, no. 2, pp. 170–191, 2017.

[183] J. Hansson, F. Pettersson-Löfstedt, H. Svensson and A. Wretstrand, 'Effects of rural bus stops on travel time and reliability', *Public Transport*, pp. 1–22, 2021.

[184] T. A. Bittencourt and M. Giannotti, 'The unequal impacts of time, cost and transfer accessibility on cities, classes and races', *Cities*, vol. 116, p. 103 257, 2021.

[185] Y. Fan, 'Household structure and gender differences in travel time: Spouse/partner presence, parenthood, and breadwinner status', *Transportation*, vol. 44, no. 2, pp. 271–291, 2017.

[186] M. Marcén and M. Morales, 'Culture and the cross-country differences in the gender commuting gap', *Journal of Transport Geography*, vol. 96, p. 103 184, 2021.

[187] A. P. Rudke *et al.*, 'Spatial and socio-economic analysis of public transport systems in large cities: A case study for belo horizonte, brazil', *Journal of Transport Geography*, vol. 91, p. 102 975, 2021.

[188] P. Coppola and F. Silvestri, 'Gender inequality in safety and security perceptions in railway stations', *Sustainability*, vol. 13, no. 7, p. 4007, 2021.

[189] V. Volpati and M. Barthelemy, 'The spatial organization of the population density in cities', *arXiv preprint arXiv:1804.00855*, 2018.

[190] D. Chacon-Hurtado, I. Kumar, K. Gkritza, J. D. Fricker and L. J. Beaulieu, 'The role of transportation accessibility in regional economic resilience', *Journal of Transport Geography*, vol. 84, p. 102 695, 2020.

[191] R. Hrelja, J. Khan and F. Pettersson, 'How to create efficient public transport systems? a systematic review of critical problems and approaches for addressing the problems', *Transport Policy*, vol. 98, pp. 186–196, 2020.

[192] R. Gallotti, G. Bertagnolli and M. De Domenico, 'Unraveling the hidden organisation of urban systems and their mobility flows', *EPJ Data Science*, vol. 10, no. 1, p. 3, 2021.

[193] D. Müllner, 'Modern hierarchical, agglomerative clustering algorithms', *arXiv preprint arXiv:1109.2378*, 2011.

[194] E. K. Tokuda, C. H. Comin and L. d. F. Costa, 'Revisiting agglomerative clustering', *Physica A: Statistical Mechanics and its Applications*, p. 126 433, 2021.

[195] J. M. Roos, F. Sprei and U. Holmberg, 'Sociodemography, geography, and personality as determinants of car driving and use of public transportation', *Behavioral Sciences*, vol. 10, no. 6, p. 93, 2020.

[196] K. F. Anderson and J. Galaskiewicz, 'Racial/ethnic residential segregation, socioeconomic inequality, and job accessibility by public transportation networks in the united states', *Spatial Demography*, pp. 1–33, 2021.

[197] J. M. González-Pérez, 'Racial/ethnic segregation and urban inequality in kansas city, missouri: A divided city', *City & Community*, p. 1 535 684 121 990 799, 2021.

[198] C. A. Hidalgo, E. Castañer and A. Sevtsuk, 'The amenity mix of urban neighborhoods', *Habitat International*, vol. 106, p. 102 205, 2020.

[199] A. N. Garay-Huamán and C. Irazábal-Zurita, 'Latinos in kansas city: The political economy of placemaking', *Journal of Planning Literature*, vol. 36, no. 2, pp. 131–154, 2021.

[200] K. Manderscheid, 'Quantifying mobilities? reflections on a neglected method in mobilities research', *Applied Mobilities*, vol. 1, no. 1, pp. 43–55, 2016.

[201] Colombia. 'Report of colombia on global leaders' meeting on gender equality and women's empowerment'. (), [Online]. Available: `https://www.unwomen.org/en/get-involved/step-it-up/commitments/colombia` (visited on 03/03/2020).

[202] UN Women. 'Status of women's rights in colombia'. (), [Online]. Available: `https://lac.unwomen.org/en/donde-estamos/colombia` (visited on 04/03/2020).

[203] C. C. Perez, *Invisible women: Exposing data bias in a world designed for men*. Random House, 2019.

[204] K. Kelley, B. Clark, V. Brown and J. Sitzia, 'Good practice in the conduct and reporting of survey research', *International Journal for Quality in health care*, vol. 15, no. 3, pp. 261–266, 2003.

[205] A. Fink, *How to conduct surveys: A step-by-step guide*. Sage Publications, 2015.

[206] A. Zwitter, 'Big data ethics', *Big Data & Society*, vol. 1, no. 2, p. 2 053 951 714 559 253, 2014.

[207] J. Kaur, 'Analysis of ethical issues in data science', *International Journal of Progressive Research in Science and Engineering*, vol. 2, no. 3, pp. 49–52, 2021.

[208] D. R. Karp *et al.*, 'Ethical and practical issues associated with aggregating databases', *PLoS medicine*, vol. 5, no. 9, e190, 2008.

[209] M. Sloane, 'Inequality is the name of the game: Thoughts on the emerging field of technology, ethics and social justice', in *Weizenbaum Conference*, DEU, 2019, p. 9.

[210] C. Delpierre and M. Kelly-Irving, 'Big data and the study of social inequalities in health: Expectations and issues', *Frontiers in Public Health*, vol. 6, p. 312, 2018.

A. Appendix

This appendix list all the extra tables and figures performed to support this thesis.

A.1 Fitting models for sociodemographic groups

There is still a need for better mobility models that explain mobility for the overall population as well as for sociodemographic groups. We observe in Figure A.1 that the gravity (singly constrained) and radiation models do not predict well the mobility of women and men and that they are also characterised differently. For instance, the distance parameter of the gravity model for women's mobility is smaller than for men's, in line with our results in Chapter 4. Further studies about fitting models for the mobility of specific sociodemographic groups are necessary to uncover the behaviour of the minorities.

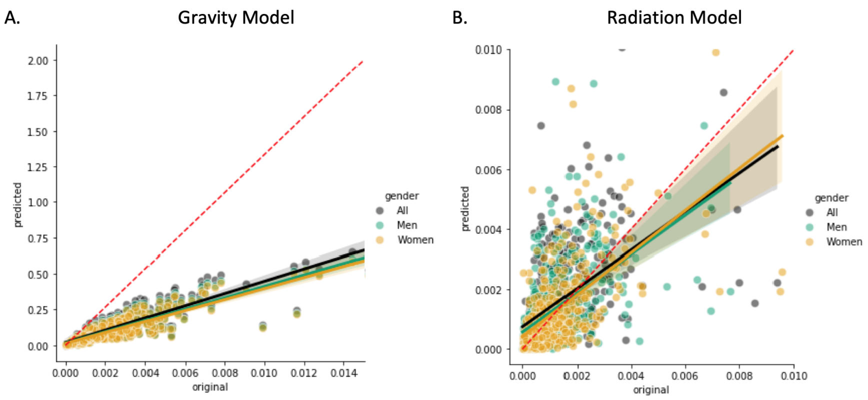

Figure A.1: Fitting (A.) Gravity model and (B.) Radiation model for the mobility of each gender.

Table A.1: Data description from the top 50 cities from the United States considering the spatial partitioning (ZL2) from the Longitudinal Employer-Household Dynamics survey [74]. The spatial partitioning from the American Community Survey (ZL1) is defined by the Public Use Microdata Areas (PUMAs) [76]. The top 50 cities are also the most populated cities in the USA as shown in the table by the estimated population.

Rank	State	City	ZL1	ZL2	Population Estimation				
					2015	2016	2017	2018	2019
1	NY	New York	55	1718	8463049	8469153	8437478	8390081	8336817
2	CA	Los Angeles	24	705	3938568	3963226	3975788	3977596	3979576
3	TX	Houston	37	530	2286908	2309544	2316750	2318573	2320268
4	IL	Chicago	17	504	2724344	2716723	2711069	2701423	2693976
5	PA	Philadelphia	11	355	1571065	1576051	1580601	1583592	1584064
6	AZ	Phoenix	13	326	1583690	1612199	1633560	1654675	1680992
7	TX	Dallas	25	302	1301329	1323916	1342479	1341802	1343573
8	TX	San Antonio	16	259	1464043	1487843	1511154	1530016	1547253
9	CA	San Diego	22	253	1387323	1402089	1412621	1421917	1423851
10	TX	Austin	33	195	921114	939447	951553	962469	978908
11	IN	Indianapolis	7	186	863319	868704	873179	880739	886220
12	NC	Charlotte	5	182	825668	843117	860002	872514	885708
13	OH	Columbus	5	178	854950	866894	881694	890869	898553
14	CA	San Francisco	7	175	863010	871512	878040	880696	881549
15	TX	Fort Worth	9	170	835356	856177	874809	893216	909585
16	TN	Memphis	6	159	654106	652548	650878	651104	651073
17	CA	San Jose	12	158	1025980	1030242	1032335	1028020	1021795
18	MO	Kansas City	4	157	475073	481670	488329	492012	495327
19	WI	Milwaukee	5	154	600477	596996	593725	591375	590157
20	LA	New Orleans	3	146	389742	391843	391493	391004	390144
21	CO	Denver	5	142	683285	696159	704961	716265	727211
22	MI	Detroit	8	139	679410	677143	674631	672977	670031
23	GA	Atlanta	4	139	468303	479174	491670	498183	506811
24	AZ	Tucson	6	139	535607	537528	541377	544858	548073
25	WA	Seattle	5	136	687386	709631	728661	742235	753675
26	FL	Jacksonville	7	136	865836	880520	892025	902437	911507
27	CA	Fresno	7	120	518203	521426	525373	528814	531576
28	AZ	Mesa	4	117	481811	490695	500021	507945	518012
29	OK	Tulsa	4	114	403491	404007	402060	400414	401190
30	TX	El Paso	5	113	676242	679955	681343	679875	681728
31	CO	Colorado Springs	6	105	449572	458714	465167	472567	478221
32	CA	Sacramento	13	104	488133	494127	500777	507737	513624
33	KS	Wichita	4	103	389412	390519	390277	389231	389938
34	FL	Tampa	3	101	371464	380344	391026	397232	399700
35	NC	Raleigh	3	98	449546	459469	465776	469314	474069
36	CA	Oakland	3	97	418211	420947	424382	429056	433031
37	CA	Long Beach	4	97	470128	468719	466646	465865	462628
38	FL	Miami	4	97	434738	449149	456617	462819	467963
39	MA	Boston	8	90	670491	679848	687788	691147	692600
40	VA	Virginia Beach	3	87	450304	450983	449896	449849	449974
41	OH	Cincinnati	3	85	299439	299748	301587	302277	303940
42	NJ	Newark	6	81	279636	280681	281237	281752	282011
43	MO	St. Louis	2	81	316010	312633	308233	303419	300576
44	NC	Greensboro	2	78	284610	289182	291537	294518	296710
45	TX	Arlington	2	77	388728	394384	397173	398123	398854
46	TN	Knoxville	3	75	184046	184986	186905	187362	187603
47	TX	Corpus Christi	4	72	324672	325786	325568	326307	326586
48	OH	Toledo	2	72	279898	278897	276688	274864	272779
49	AZ	Chandler	3	71	243679	247845	253599	257186	261165
50	CA	Riverside	15	70	320483	323684	326882	329654	331360

Table A.2: Summary of the composition of all the expanded data sets for travels made for `all` purposes. For a given location and year, we report: the number of travellers N_P, the number of travels N_T, the fraction of men (women) travellers f^M (f^W), and the fraction of travels made by men (women) f_T^M (f_T^W). We report also the fraction of travellers belonging to the `lower` ($f^{\texttt{lower}}$), `middle` ($f^{\texttt{middle}}$), and `upper` ($f^{\texttt{upper}}$) socioeconomic classes, and the same quantities discriminated by gender (e.g. $f^{\texttt{lower}W}$). Finally, we report the fraction of travels made by travellers with a given socioeconomic class and gender (e.g. $f_T^{\texttt{lower}W}$). The data sets are obtained applying the expansion factors to the raw data from the surveys.

Location	Medellín (`MDE`)		Bogotá (`BGT`)		São Paulo (`SAO`)		
Year	2005	2017	2012	2019	1997	2007	2017
N_P	22,702	38,048	11,672	47,149	37,316	54,745	48,085
N_T	7,102,052	123,449	25,628,970	88,620,670	54,939,650	83,313,240	95,948,930
f^M	0.52	0.51	0.46	0.48	0.52	0.49	0.50
f^W	0.48	0.49	0.54	0.52	0.48	0.51	0.50
f_T^M	0.52	0.51	0.42	0.47	0.51	0.49	0.50
f_T^W	0.48	0.49	0.58	0.53	0.49	0.51	0.50
$f^{\texttt{lower}}$	0.50	0.55	0.46	0.49	0.30	0.21	0.20
$f^{\texttt{middle}}$	0.46	0.38	0.48	0.46	0.63	0.63	0.65
$f^{\texttt{upper}}$	0.04	0.07	0.06	0.05	0.07	0.16	0.15
$f_T^{\texttt{lower}}$	0.41	0.54	0.52	0.50	0.26	0.24	0.22
$f_T^{\texttt{middle}}$	0.52	0.38	0.43	0.45	0.68	0.66	0.68
$f_T^{\texttt{upper}}$	0.07	0.08	0.05	0.05	0.06	0.10	0.10
$f^{\texttt{lower}\,M}$	0.26	0.28	0.21	0.24	0.16	0.10	0.09
$f^{\texttt{middle}\,M}$	0.23	0.19	0.22	0.22	0.32	0.31	0.32
$f^{\texttt{upper}\,M}$	0.02	0.04	0.03	0.02	0.04	0.08	0.08
$f^{\texttt{lower}\,W}$	0.24	0.26	0.25	0.25	0.14	0.11	0.10
$f^{\texttt{middle}\,W}$	0.23	0.19	0.26	0.24	0.30	0.32	0.33
$f^{\texttt{upper}\,W}$	0.02	0.04	0.03	0.03	0.04	0.08	0.08
$f_T^{\texttt{lower}\,M}$	0.22	0.28	0.21	0.23	0.13	0.11	0.10
$f_T^{\texttt{middle}\,M}$	0.26	0.19	0.19	0.21	0.35	0.39	0.35
$f_T^{\texttt{upper}\,M}$	0.04	0.04	0.02	0.02	0.03	0.05	0.05
$f_T^{\texttt{lower}\,W}$	0.19	0.26	0.32	0.27	0.13	0.13	0.12
$f_T^{\texttt{middle}\,W}$	0.26	0.19	0.24	0.23	0.33	0.32	0.35
$f_T^{\texttt{upper}\,W}$	0.03	0.04	0.03	0.03	0.03	0.05	0.05

Table A.3: Summary of the composition of all the expanded data sets for travels made only for `work` purpose. See the caption of Table A.8 for the description of each row.

Location	Medellín (`MDE`)		Bogotá (`BGT`)		São Paulo (`SAO`)		
Year	2005	2017	2012	2019	1997	2007	2017
N_P	9,081	17,466	6,844	20,208	17,806	29,640	25,333
N_T	349,963	18,814	1,437,599	3,916,047	5,939,612	9,038,745	10,363,550
f^M	0.61	0.62	0.55	0.56	0.62	0.55	0.55
f^W	0.39	0.38	0.45	0.44	0.38	0.45	0.45
f_T^M	0.63	0.63	0.58	0.58	0.68	0.61	0.59
f_T^W	0.37	0.37	0.42	0.42	0.32	0.39	0.41
$f^{\texttt{lower}}$	0.50	0.54	0.47	0.48	0.30	0.19	0.17
$f^{\texttt{middle}}$	0.46	0.38	0.46	0.46	0.63	0.64	0.66
$f^{\texttt{upper}}$	0.04	0.08	0.07	0.06	0.07	0.17	0.17
$f_T^{\texttt{lower}}$	0.39	0.53	0.45	0.47	0.25	0.22	0.19
$f_T^{\texttt{middle}}$	0.52	0.39	0.48	0.47	0.69	0.68	0.70
$f_T^{\texttt{upper}}$	0.09	0.08	0.07	0.06	0.06	0.10	0.11
$f^{\texttt{lower}\,M}$	0.32	0.35	0.27	0.28	0.19	0.10	0.09
$f^{\texttt{middle}\,M}$	0.27	0.23	0.25	0.25	0.39	0.35	0.36
$f^{\texttt{upper}\,M}$	0.02	0.05	0.04	0.03	0.04	0.10	0.10
$f^{\texttt{lower}\,W}$	0.18	0.19	0.20	0.20	0.11	0.08	0.07
$f^{\texttt{middle}\,W}$	0.19	0.15	0.21	0.21	0.25	0.29	0.30
$f^{\texttt{upper}\,W}$	0.02	0.03	0.03	0.03	0.02	0.08	0.08
$f_T^{\texttt{lower}\,M}$	0.26	0.34	0.26	0.28	0.17	0.13	0.11
$f_T^{\texttt{middle}\,M}$	0.31	0.24	0.28	0.26	0.47	0.42	0.42
$f_T^{\texttt{upper}\,M}$	0.06	0.05	0.04	0.04	0.04	0.06	0.06
$f_T^{\texttt{lower}\,W}$	0.13	0.18	0.19	0.19	0.08	0.09	0.08
$f_T^{\texttt{middle}\,W}$	0.21	0.15	0.20	0.21	0.22	0.26	0.29
$f_T^{\texttt{upper}\,W}$	0.03	0.04	0.03	0.02	0.02	0.04	0.04

Table A.4: Network properties from the mobility of men and women across socioeconomic groups.

Metrics	Region	Year	women lower	women middle	women upper	men lower	men middle	men upper
Average inflow	MDE	2005	1179	1450	142	1370	1427	162
		2017	51	45	44	51	46	44
	BGT	2012	1728	1928	54	1561	1675	54
		2019	1966	1419	71	1887	1303	69
	SAO	1997	7677	20697	400	8599	22471	651
		2007	3649	11809	355	3384	11422	419
		2017	1955	10863	161	1734	10024	180
Indegree centrality	MDE	2005	0.7662	0.4153	0.7687	0.4354	0.4387	0.7687
		2017	0.0538	0.0924	0.1579	0.0629	0.0885	0.3077
	BGT	2012	0.1467	0.1338	0.2634	0.1356	0.1284	0.2634
		2019	0.1124	0.1480	0.3207	0.1165	0.1403	0.3207
	SAO	1997	0.1526	0.4153	0.2036	0.2092	0.4378	0.2206
		2007	0.1388	0.3989	0.1869	0.1469	0.4206	0.1944
		2017	0.0601	0.2380	0.1408	0.0585	0.2345	0.1552
Average clustering	MDE	2005	0.4860	0.3550	0.4860	0.3010	0.3340	0.4860
		2017	0.3720	0.3060	0.2760	0.3800	0.3120	0.1870
	BGT	2012	0.1474	0.1281	0.2057	0.1345	0.1263	0.1848
		2019	0.1079	0.1442	0.1686	0.1161	0.1407	0.1402
	SAO	1997	0.3450	0.4180	0.3500	0.2670	0.3820	0.3100
		2007	0.3980	0.3800	0.3010	0.4270	0.3920	0.3510
		2017	0.4800	0.3870	0.3160	0.4320	0.3770	0.3320
Density	MDE	2005	0.2041	0.0880	0.2047	0.0719	0.0942	0.2047
		2017	0.0110	0.0150	0.0763	0.0119	0.0168	0.1209
	BGT	2012	0.0152	0.0213	0.0498	0.0146	0.0208	0.0498
		2019	0.0166	0.0223	0.0569	0.0160	0.0215	0.0569
	SAO	1997	0.0464	0.0907	0.0395	0.0621	0.1128	0.0439
		2007	0.0217	0.0494	0.0317	0.0216	0.0543	0.0351
		2017	0.0120	0.0339	0.0235	0.0131	0.0362	0.0251
Average neighbour indegree	MDE	2005	113	58	122	54	62	125
		2017	81	84	83	82	84	83
	BGT	2012	28	29	73	25	29	73
		2019	35	75	109	46	69	123
	SAO	1997	19	39	17	26	50	19
		2007	16	20	95	11	42	105
		2017	32	14	38	8	17	67

Table A.5: Portrait Divergence between the mobility network from travellers of each gender (`women`,`men`) in a socioeconomic (`lower`,`middle`,`upper`) group, $PD(X_{i=1}, X_{i=2})$, where $X_i \epsilon \{$`women lower`,`women middle`,`women upper`, `men lower`,`men middle`,`men upper` $\}$. The values extracted from the null models are in parentheses. The null models shuffle the gender and socioeconomic categories.

Region	Year	`lower` PD(`women`,`men`)	`middle` PD(`women`,`men`)	`upper` PD(`women`,`men`)
`MDE`	2005	0.7930 (0.4861)	0.7357 (0.7922)	0.9199 (0.4889)
	2017	0.8458 (0.5463)	0.6096 (0.7492)	0.8925 (0.3352)
`BGT`	2012	0.9082 (0.5680)	0.9112 (0.5935)	0.9580 (0.8096)
	2019	0.9472 (0.1522)	0.9704 (0.1824)	0.9423 (0.6475)
`SAO`	1997	0.4622 (0.3352)	0.5545 (0.2657)	0.8476 (0.5659)
	2007	0.8147 (0.4259)	0.5808 (0.2967)	0.8217 (0.4829)
	2017	0.8861 (0.4920)	0.8949 (0.3060)	0.8501 (0.5119)

Table A.6: F Statistic of the ANOVA Test computed from the mobility diversity of all travels. All the p-values are smaller than 0.001.

Location	`MDE`		`BGT`		`SAO`		
Years	2005	2017	2012	2019	1997	2007	2017
Gender Groups	573569	4690	6023260	93689	537431	6456450	2296951
Socioeconomic Groups	391579690	4127072	1138264618	35079596	301501736	177736269	228956044
Combined Groups	8604	3527	5711	3060	5145	8739	5984

Table A.7: F Statistic of the ANOVA Test computed from the mobility diversity of work travels. All the p-values are smaller than 0.001.

Location	`MDE`		`BGT`		`SAO`		
Years	2005	2017	2012	2019	1997	2007	2017
Gender Groups	186411	4826	6164761	742155	1551884	1377470	2164971
Socioeconomic Groups	7272494	762876	66412851	20475418	36892298	37295014	75000823
Combined Groups	3420	3826	4705	3406	2522	9095	6249

Table A.8: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in `MDE` 2005 using Tukey's HSD test. The values presented are multiplied by 10^2. The *** symbol denotes a p-value smaller than 0.001. We highlight the cells of groups having p-values higher than 0.001.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
`all`	`(all)` × `(men)`	0.0308	0.0302	0.0314	***
	`(all)` × `(women)`	-0.2263	-0.2269	-0.2257	
	`(men)` × `(women)`	-0.2571	-0.2577	-0.2565	
	`(all)` × `(lower)`	86.472	86.4707	86.4733	
	`(all)` × `(middle)`	87.4194	87.4181	87.4207	
	`(all)` × `(upper)`	75.3408	75.3395	75.3421	
	`(lower)` × `(middle)`	0.9475	0.9462	0.9488	
	`(lower)` × `(upper)`	-11.1312	-11.1325	-11.1299	
	`(middle)` × `(upper)`	-12.0786	-12.0799	-12.0773	
	`(all)` × `(men-lower)`	-5.7453	-5.7475	-5.7431	
	`(all)` × `(men-middle)`	-4.9728	-4.975	-4.9706	
	`(all)` × `(men-upper)`	-16.8361	-16.8383	-16.8339	
	`(all)` × `(women-lower)`	-6.3292	-6.3314	-6.327	
	`(all)` × `(women-middle)`	-5.1697	-5.1719	-5.1675	
	`(all)` × `(women-upper)`	-18.047	-18.0492	-18.0448	
	`(men-lower)` × `(men-middle)`	0.7725	0.7703	0.7747	
	`(men-lower)` × `(men-upper)`	-11.0908	-11.093	-11.0886	
	`(men-lower)` × `(women-lower)`	-0.5839	-0.5861	-0.5817	
	`(men-lower)` × `(women-middle)`	0.5756	0.5734	0.5778	
	`(men-lower)` × `(women-upper)`	-12.3017	-12.3039	-12.2995	
	`(men-middle)` × `(men-upper)`	-11.8633	-11.8655	-11.8611	
	`(men-middle)` × `(women-lower)`	-1.3564	-1.3586	-1.3542	
	`(men-middle)` × `(women-middle)`	-0.1969	-0.1991	-0.1947	
	`(men-middle)` × `(women-upper)`	-13.0742	-13.0764	-13.072	
	`(men-upper)` × `(women-lower)`	10.5069	10.5047	10.5091	
	`(men-upper)` × `(women-middle)`	11.6664	11.6642	11.6686	
	`(men-upper)` × `(women-upper)`	-1.2109	-1.2131	-1.2087	
	`(women-lower)` × `(women-middle)`	1.1595	1.1573	1.1617	
	`(women-lower)` × `(women-upper)`	-11.7178	-11.72	-11.7156	
	`(women-middle)` × `(women-upper)`	-12.8773	-12.8795	-12.8751	
`work`	`(all)` × `(men)`	-0.1733	-0.1755	-0.1711	***
	`(all)` × `(women)`	-0.5707	-0.5729	-0.5685	
	`(men)` × `(women)`	-0.3974	-0.3996	-0.3952	
	`(all)` × `(lower)`	85.1415	85.1383	85.1447	
	`(all)` × `(middle)`	84.3991	84.3959	84.4024	
	`(all)` × `(upper)`	78.5477	78.5445	78.551	
	`(lower)` × `(middle)`	-0.7424	-0.7456	-0.7392	
	`(lower)` × `(upper)`	-6.5938	-6.597	-6.5905	
	`(middle)` × `(upper)`	-5.8514	-5.8546	-5.8482	
	`(all)` × `(men-lower)`	-0.4769	-0.4823	-0.4715	
	`(all)` × `(men-middle)`	-1.4315	-1.4369	-1.4261	
	`(all)` × `(men-upper)`	-8.8917	-8.8971	-8.8863	
	`(all)` × `(women-lower)`	-1.2447	-1.2501	-1.2393	
	`(all)` × `(women-middle)`	-1.6542	-1.6596	-1.6488	
	`(all)` × `(women-upper)`	-8.084	-8.0894	-8.0786	
	`(men-lower)` × `(men-middle)`	-0.9547	-0.9601	-0.9493	
	`(men-lower)` × `(men-upper)`	-8.4148	-8.4202	-8.4094	
	`(men-lower)` × `(women-lower)`	-0.7678	-0.7733	-0.7624	
	`(men-lower)` × `(women-middle)`	-1.1773	-1.1827	-1.1719	
	`(men-lower)` × `(women-upper)`	-7.6072	-7.6126	-7.6018	
	`(men-middle)` × `(men-upper)`	-7.4601	-7.4656	-7.4547	
	`(men-middle)` × `(women-lower)`	0.1868	0.1814	0.1922	
	`(men-middle)` × `(women-middle)`	-0.2226	-0.228	-0.2172	
	`(men-middle)` × `(women-upper)`	-6.6525	-6.6579	-6.6471	
	`(men-upper)` × `(women-lower)`	7.647	7.6416	7.6524	
	`(men-upper)` × `(women-middle)`	7.2375	7.2321	7.2429	
	`(men-upper)` × `(women-upper)`	0.8076	0.8022	0.813	
	`(women-lower)` × `(women-middle)`	-0.4095	-0.4149	-0.404	
	`(women-lower)` × `(women-upper)`	-6.8393	-6.8447	-6.8339	
	`(women-middle)` × `(women-upper)`	-6.4299	-6.4353	-6.4245	

Table A.9: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in MDE 2017 using the Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
all	(all) × (men)	0.0637	0.0586	0.0688	***
	(all) × (women)	-0.1613	-0.1664	-0.1562	
	(men) × (women)	-0.225	-0.2301	-0.2199	
	(all) × (lower)	79.5541	79.5448	79.5633	
	(all) × (middle)	76.1984	76.1891	76.2076	
	(all) × (upper)	65.4818	65.4725	65.491	
	(lower) × (middle)	-3.3557	-3.3649	-3.3465	
	(lower) × (upper)	-14.0723	-14.0815	-14.0631	
	(middle) × (upper)	-10.7166	-10.7258	-10.7074	
	(all) × (men lower)	-0.7026	-0.7195	-0.6858	
	(all) × (men-middle)	-3.8696	-3.8864	-3.8528	
	(all) × (men-upper)	-14.1376	-14.1544	-14.1208	
	(all) × (women lower)	-0.7279	-0.7447	-0.7111	
	(all) × (women-middle)	-4.3216	-4.3384	-4.3048	
	(all) × (women-upper)	-15.7535	-15.7703	-15.7367	
	(men lower) × (men-middle)	-3.1669	-3.1838	-3.1501	
	(men lower) × (men-upper)	-13.4349	-13.4517	-13.4181	
	(men lower) × (women lower)	-0.0252	-0.042	-0.0084	
	(men lower) × (women-middle)	-3.6189	-3.6357	-3.6021	
	(men lower) × (women-upper)	-15.0508	-15.0676	-15.034	
	(men-middle) × (men-upper)	-10.268	-10.2848	-10.2512	
	(men-middle) × (women lower)	3.1417	3.1249	3.1585	
	(men-middle) × (women-middle)	-0.452	-0.4688	-0.4352	
	(men-middle) × (women-upper)	-11.8839	-11.9007	-11.8671	
	(men-upper) × (women lower)	13.4097	13.3929	13.4265	
	(men-upper) × (women-middle)	9.816	9.7992	9.8328	
	(men-upper) × (women-upper)	-1.6159	-1.6327	-1.5991	
	(women lower) × (women-middle)	-3.5937	-3.6105	-3.5769	
	(women lower) × (women-upper)	-15.0256	-15.0424	-15.0088	
	(women-middle) × (women-upper)	-11.4319	-11.4487	-11.4151	
work	(all) × (men)	0.0109	-0.0025	0.0244	0.1356
	(all) × (women)	-0.5492	-0.5626	-0.5357	***
	(men) × (women)	-0.5601	-0.5735	-0.5467	
	(all) × (lower)	79.3821	79.3596	79.4046	
	(all) × (middle)	76.1565	76.134	76.179	
	(all) × (upper)	64.9359	64.9134	64.9584	
	(lower) × (middle)	-3.2256	-3.2481	-3.2031	
	(lower) × (upper)	-14.4462	-14.4686	-14.4237	
	(middle) × (upper)	-11.2206	-11.243	-11.1981	
	(all) × (men lower)	-1.0741	-1.1158	-1.0323	
	(all) × (men-middle)	-4.0935	-4.1352	-4.0517	
	(all) × (men-upper)	-15.223	-15.2647	-15.1813	
	(all) × (women lower)	-1.306	-1.3477	-1.2643	
	(all) × (women-middle)	-4.8986	-4.9403	-4.8569	
	(all) × (women-upper)	-17.1903	-17.232	-17.1486	
	(men lower) × (men-middle)	-3.0194	-3.0611	-2.9777	
	(men lower) × (men-upper)	-14.149	-14.1907	-14.1072	
	(men lower) × (women lower)	-0.2319	-0.2736	-0.1902	
	(men lower) × (women-middle)	-3.8245	-3.8662	-3.7828	
	(men lower) × (women-upper)	-16.1162	-16.158	-16.0745	
	(men-middle) × (men-upper)	-11.1296	-11.1713	-11.0878	
	(men-middle) × (women lower)	2.7875	2.7458	2.8292	
	(men-middle) × (women-middle)	-0.8051	-0.8468	-0.7634	
	(men-middle) × (women-upper)	-13.0968	-13.1386	-13.0551	
	(men-upper) × (women lower)	13.917	13.8753	13.9588	
	(men-upper) × (women-middle)	10.3244	10.2827	10.3662	
	(men-upper) × (women-upper)	-1.9673	-2.009	-1.9256	
	(women lower) × (women-middle)	-3.5926	-3.6343	-3.5509	
	(women lower) × (women-upper)	-15.8843	-15.926	-15.8426	
	(women-middle) × (women-upper)	-12.2917	-12.3335	-12.25	

Table A.10: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in `BGT` 2012 using Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval		Adjusted p-value
			Lower bound	Upper bound	
all	(all) × (men)	-0.0497	-0.05	-0.0494	***
	(all) × (women)	-0.4077	-0.408	-0.4074	
	(men) × (women)	-0.358	-0.3583	-0.3577	
	(all) × (lower)	87.0619	87.0609	87.0629	
	(all) × (middle)	89.0512	89.0502	89.0522	
	(all) × (upper)	71.4057	71.4046	71.4067	
	(lower) × (middle)	1.9893	1.9883	1.9903	
	(lower) × (upper)	-15.6563	-15.6573	-15.6552	
	(middle) × (upper)	-17.6455	-17.6466	-17.6445	
	(all) × (men lower)	-6.1861	-6.1878	-6.1844	
	(all) × (men-middle)	-4.2557	-4.2574	-4.254	
	(all) × (men-upper)	-21.8737	-21.8754	-21.872	
	(all) × (women lower)	-6.812	-6.8137	-6.8103	
	(all) × (women-middle)	-4.743	-4.7446	-4.7413	
	(all) × (women-upper)	-23.4137	-23.4154	-23.412	
	(men lower) × (men-middle)	1.9304	1.9287	1.9321	
	(men lower) × (men-upper)	-15.6875	-15.6892	-15.6858	
	(men lower) × (women lower)	-0.6258	-0.6275	-0.6241	
	(men lower) × (women-middle)	1.4432	1.4415	1.4449	
	(men lower) × (women-upper)	-17.2275	-17.2292	-17.2258	
	(men-middle) × (men-upper)	-17.618	-17.6196	-17.6163	
	(men-middle) × (women lower)	-2.5562	-2.5579	-2.5545	
	(men-middle) × (women-middle)	-0.4872	-0.4889	-0.4855	
	(men-middle) × (women-upper)	-19.1579	-19.1596	-19.1563	
	(men-upper) × (women lower)	15.0617	15.06	15.0634	
	(men-upper) × (women-middle)	17.1307	17.129	17.1324	
	(men-upper) × (women-upper)	-1.54	-1.5417	-1.5383	
	(women lower) × (women-middle)	2.069	2.0673	2.0707	
	(women lower) × (women-upper)	-16.6017	-16.6034	-16.6	
	(women-middle) × (women-upper)	-18.6707	-18.6724	-18.669	
work	(all) × (men)	-0.9601	-0.9612	-0.9589	***
	(all) × (women)	-1.5977	-1.5989	-1.5966	
	(men) × (women)	-0.6377	-0.6388	-0.6365	
	(all) × (lower)	88.1965	88.1941	88.1989	
	(all) × (middle)	85.972	85.9696	85.9743	
	(all) × (upper)	72.908	72.9056	72.9104	
	(lower) × (middle)	-2.2246	-2.227	-2.2222	
	(lower) × (upper)	-15.2885	-15.2909	-15.2861	
	(middle) × (upper)	-13.0639	-13.0663	-13.0615	
	(all) × (men lower)	-2.6924	-2.6965	-2.6884	
	(all) × (men-middle)	-4.6356	-4.6396	-4.6315	
	(all) × (men-upper)	-18.4256	-18.4297	-18.4216	
	(all) × (women lower)	-3.8732	-3.8773	-3.8692	
	(all) × (women-middle)	-5.4457	-5.4497	-5.4416	
	(all) × (women-upper)	-21.8152	-21.8192	-21.8111	
	(men lower) × (men-middle)	-1.9431	-1.9472	-1.9391	
	(men lower) × (men-upper)	-15.7332	-15.7373	-15.7292	
	(men lower) × (women lower)	-1.1808	-1.1849	-1.1768	
	(men lower) × (women-middle)	-2.7533	-2.7573	-2.7492	
	(men lower) × (women-upper)	-19.1227	-19.1268	-19.1187	
	(men-middle) × (men-upper)	-13.7901	-13.7941	-13.786	
	(men-middle) × (women lower)	0.7623	0.7583	0.7664	
	(men-middle) × (women-middle)	-0.8101	-0.8142	-0.8061	
	(men-middle) × (women-upper)	-17.1796	-17.1836	-17.1755	
	(men-upper) × (women lower)	14.5524	14.5483	14.5564	
	(men-upper) × (women-middle)	12.98	12.9759	12.984	
	(men-upper) × (women-upper)	-3.3895	-3.3936	-3.3855	
	(women lower) × (women-middle)	-1.5724	-1.5765	-1.5684	
	(women lower) × (women-upper)	-17.9419	-17.946	-17.9379	
	(women-middle) × (women-upper)	-16.3695	-16.3735	-16.3654	

Table A.11: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in `BGT` 2019 using Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
all	(all) × (men)	-0.0376	-0.038	-0.0373	***
	(all) × (women)	-0.0561	-0.0564	-0.0557	
	(men) × (women)	-0.0184	-0.0188	-0.0181	
	(all) × (lower)	81.0821	81.0813	81.083	
	(all) × (middle)	81.8183	81.8174	81.8191	
	(all) × (upper)	78.0621	78.0612	78.063	
	(lower) × (middle)	0.7361	0.7353	0.737	
	(lower) × (upper)	-3.0201	-3.0209	-3.0192	
	(middle) × (upper)	-3.7562	-3.757	-3.7553	
	(all) × (men lower)	-0.8819	-0.8835	-0.8804	
	(all) × (men-middle)	-0.3397	-0.3413	-0.3382	
	(all) × (men-upper)	-4.8074	-4.8089	-4.8059	
	(all) × (women lower)	-1.0872	-1.0888	-1.0857	
	(all) × (women-middle)	-0.1996	-0.2012	-0.1981	
	(all) × (women-upper)	-4.9056	-4.9071	-4.904	
	(men lower) × (men-middle)	0.5422	0.5407	0.5437	
	(men lower) × (men-upper)	-3.9255	-3.927	-3.9239	
	(men lower) × (women lower)	-0.2053	-0.2068	-0.2038	
	(men lower) × (women-middle)	0.6823	0.6808	0.6838	
	(men lower) × (women-upper)	-4.0236	-4.0252	-4.0221	
	(men-middle) × (men-upper)	-4.4677	-4.4692	-4.4661	
	(men-middle) × (women lower)	-0.7475	-0.749	-0.746	
	(men-middle) × (women-middle)	0.1401	0.1386	0.1416	
	(men-middle) × (women-upper)	-4.5658	-4.5674	-4.5643	
	(men-upper) × (women lower)	3.7202	3.7186	3.7217	
	(men-upper) × (women-middle)	4.6078	4.6062	4.6093	
	(men-upper) × (women-upper)	-0.0982	-0.0997	-0.0966	
	(women lower) × (women-middle)	0.8876	0.8861	0.8891	
	(women lower) × (women-upper)	-3.8183	-3.8199	-3.8168	
	(women-middle) × (women-upper)	-4.706	-4.7075	-4.7044	
work	(all) × (men)	-0.3398	-0.3409	-0.3387	***
	(all) × (women)	-0.523	-0.5242	-0.5219	
	(men) × (women)	-0.1833	-0.1844	-0.1822	
	(all) × (lower)	78.9254	78.9227	78.9281	
	(all) × (middle)	80.3745	80.3718	80.3772	
	(all) × (upper)	71.1352	71.1325	71.1379	
	(lower) × (middle)	1.4491	1.4465	1.4518	
	(lower) × (upper)	-7.7902	-7.7929	-7.7875	
	(middle) × (upper)	-9.2393	-9.242	-9.2366	
	(all) × (men lower)	-2.5151	-2.5197	-2.5104	
	(all) × (men-middle)	-1.3778	-1.3825	-1.3731	
	(all) × (men-upper)	-14.0293	-14.034	-14.0246	
	(all) × (women lower)	-3.0307	-3.0354	-3.026	
	(all) × (women-middle)	-1.4085	-1.4132	-1.4038	
	(all) × (women-upper)	-14.4171	-14.4218	-14.4125	
	(men lower) × (men-middle)	1.1372	1.1326	1.1419	
	(men lower) × (men-upper)	-11.5143	-11.5189	-11.5096	
	(men lower) × (women lower)	-0.5156	-0.5203	-0.5109	
	(men lower) × (women-middle)	1.1066	1.1019	1.1113	
	(men lower) × (women-upper)	-11.9021	-11.9068	-11.8974	
	(men-middle) × (men-upper)	-12.6515	-12.6562	-12.6468	
	(men-middle) × (women lower)	-1.6529	-1.6576	-1.6482	
	(men-middle) × (women-middle)	-0.0307	-0.0354	-0.026	
	(men-middle) × (women-upper)	-13.0393	-13.044	-13.0346	
	(men-upper) × (women lower)	10.9986	10.9939	11.0033	
	(men-upper) × (women-middle)	12.6208	12.6161	12.6255	
	(men-upper) × (women-upper)	-0.3878	-0.3925	-0.3831	
	(women lower) × (women-middle)	1.6222	1.6175	1.6269	
	(women lower) × (women-upper)	-11.3865	-11.3911	-11.3818	
	(women-middle) × (women-upper)	-13.0087	-13.0133	-13.004	

Table A.12: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in `SAO` 1997 using Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
`all`	(`all`) × (`men`)	-0.1254	-0.1257	-0.1251	***
	(`all`) × (`women`)	-0.0272	-0.0275	-0.0269	
	(`men`) × (`women`)	0.0981	0.0978	0.0984	
	(`all`) × (`lower`)	88.7281	88.7272	88.7289	
	(`all`) × (`middle`)	93.0693	93.0684	93.0702	
	(`all`) × (`upper`)	84.8127	84.8118	84.8136	
	(`lower`) × (`middle`)	4.3413	4.3404	4.3422	
	(`lower`) × (`upper`)	-3.9154	-3.9163	-3.9145	
	(`middle`) × (`upper`)	-8.2567	-8.2575	-8.2558	
	(`all`) × (`men lower`)	-4.627	-4.6286	-4.6255	
	(`all`) × (`men-middle`)	-0.3233	-0.3248	-0.3218	
	(`all`) × (`men-upper`)	-8.3645	-8.366	-8.3629	
	(`all`) × (`women lower`)	-4.7965	-4.7981	-4.795	
	(`all`) × (`women-middle`)	-0.2844	-0.286	-0.2829	
	(`all`) × (`women-upper`)	-9.9697	-9.9712	-9.9681	
	(`men lower`) × (`men-middle`)	4.3037	4.3022	4.3053	
	(`men lower`) × (`men-upper`)	-3.7374	-3.7389	-3.7359	
	(`men lower`) × (`women lower`)	-0.1695	-0.171	-0.168	
	(`men lower`) × (`women-middle`)	4.3426	4.3411	4.3441	
	(`men lower`) × (`women-upper`)	-5.3426	-5.3442	-5.3411	
	(`men-middle`) × (`men-upper`)	-8.0412	-8.0427	-8.0396	
	(`men-middle`) × (`women lower`)	-4.4732	-4.4748	-4.4717	
	(`men-middle`) × (`women-middle`)	0.0388	0.0373	0.0404	
	(`men-middle`) × (`women-upper`)	-9.6464	-9.6479	-9.6448	
	(`men-upper`) × (`women lower`)	3.5679	3.5664	3.5695	
	(`men-upper`) × (`women-middle`)	8.08	8.0785	8.0815	
	(`men-upper`) × (`women-upper`)	-1.6052	-1.6068	-1.6037	
	(`women lower`) × (`women-middle`)	4.5121	4.5105	4.5136	
	(`women lower`) × (`women-upper`)	-5.1731	-5.1747	-5.1716	
	(`women-middle`) × (`women-upper`)	-9.6852	-9.6868	-9.6837	
`work`	(`all`) × (`men`)	-0.5365	-0.5373	-0.5357	***
	(`all`) × (`women`)	-0.2166	-0.2175	-0.2158	
	(`men`) × (`women`)	0.3199	0.319	0.3207	
	(`all`) × (`lower`)	92.4728	92.4706	92.4749	
	(`all`) × (`middle`)	90.9687	90.9666	90.9709	
	(`all`) × (`upper`)	81.3344	81.3323	81.3366	
	(`lower`) × (`middle`)	-1.504	-1.5062	-1.5019	
	(`lower`) × (`upper`)	-11.1383	-11.1405	-11.1362	
	(`middle`) × (`upper`)	-9.6343	-9.6364	-9.6321	
	(`all`) × (`men lower`)	-0.6123	-0.6164	-0.6082	
	(`all`) × (`men-middle`)	-1.9421	-1.9462	-1.9381	
	(`all`) × (`men-upper`)	-12.2619	-12.2659	-12.2578	
	(`all`) × (`women lower`)	-0.7635	-0.7676	-0.7594	
	(`all`) × (`women-middle`)	-1.6211	-1.6252	-1.6171	
	(`all`) × (`women-upper`)	-15.3371	-15.3412	-15.333	
	(`men lower`) × (`men-middle`)	-1.3298	-1.3339	-1.3258	
	(`men lower`) × (`men-upper`)	-11.6496	-11.6536	-11.6455	
	(`men lower`) × (`women lower`)	-0.1512	-0.1553	-0.1471	
	(`men lower`) × (`women-middle`)	-1.0088	-1.0129	-1.0048	
	(`men lower`) × (`women-upper`)	-14.7248	-14.7289	-14.7207	
	(`men-middle`) × (`men-upper`)	-10.3197	-10.3238	-10.3156	
	(`men-middle`) × (`women lower`)	1.1786	1.1745	1.1827	
	(`men-middle`) × (`women-middle`)	0.321	0.3169	0.3251	
	(`men-middle`) × (`women-upper`)	-13.395	-13.399	-13.3909	
	(`men-upper`) × (`women lower`)	11.4983	11.4943	11.5024	
	(`men-upper`) × (`women-middle`)	10.6407	10.6366	10.6448	
	(`men-upper`) × (`women-upper`)	-3.0752	-3.0793	-3.0712	
	(`women lower`) × (`women-middle`)	-0.8576	-0.8617	-0.8535	
	(`women lower`) × (`women-upper`)	-14.5736	-14.5777	-14.5695	
	(`women-middle`) × (`women-upper`)	-13.716	-13.72	-13.7119	

Table A.13: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in SAO 2007 using Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
all	(all) × (men)	-0.2582	-0.2584	-0.2579	***
	(all) × (women)	0.164	0.1637	0.1642	
	(men) × (women)	0.4221	0.4219	0.4224	
	(all) × (lower)	88.067	88.0665	88.0675	
	(all) × (middle)	90.6562	90.6557	90.6568	
	(all) × (upper)	88.4965	88.496	88.497	
	(lower) × (middle)	2.5893	2.5887	2.5898	
	(lower) × (upper)	0.4295	0.429	0.43	
	(middle) × (upper)	-2.1598	-2.1603	-2.1593	
	(all) × (men lower)	-3.2646	-3.2655	-3.2637	
	(all) × (men-middle)	-0.6351	-0.636	-0.6343	
	(all) × (men-upper)	-2.562	-2.5628	-2.5611	
	(all) × (women lower)	-2.775	-2.7759	-2.7741	
	(all) × (women-middle)	-0.0381	-0.039	-0.0373	
	(all) × (women-upper)	-2.9615	-2.9624	-2.9606	
	(men lower) × (men-middle)	2.6294	2.6286	2.6303	
	(men lower) × (men-upper)	0.7026	0.7018	0.7035	
	(men lower) × (women lower)	0.4896	0.4887	0.4905	
	(men lower) × (women-middle)	3.2265	3.2256	3.2273	
	(men lower) × (women-upper)	0.3031	0.3022	0.304	
	(men-middle) × (men-upper)	-1.9268	-1.9277	-1.9259	
	(men-middle) × (women lower)	-2.1398	-2.1407	-2.139	
	(men-middle) × (women-middle)	0.597	0.5961	0.5979	
	(men-middle) × (women-upper)	-2.3264	-2.3272	-2.3255	
	(men-upper) × (women lower)	-0.213	-0.2139	-0.2121	
	(men-upper) × (women-middle)	2.5238	2.5229	2.5247	
	(men-upper) × (women-upper)	-0.3996	-0.4004	-0.3987	
	(women lower) × (women-middle)	2.7368	2.736	2.7377	
	(women lower) × (women-upper)	-0.1865	-0.1874	-0.1857	
	(women-middle) × (women-upper)	-2.9234	-2.9242	-2.9225	
work	(all) × (men)	-0.3503	-0.3509	-0.3497	***
	(all) × (women)	0.1219	0.1213	0.1226	
	(men) × (women)	0.4722	0.4716	0.4729	
	(all) × (lower)	86.5667	86.5655	86.5679	
	(all) × (middle)	89.7687	89.7674	89.7699	
	(all) × (upper)	87.5448	87.5436	87.546	
	(lower) × (middle)	3.202	3.2007	3.2032	
	(lower) × (upper)	0.9781	0.9768	0.9793	
	(middle) × (upper)	-2.2239	-2.2251	-2.2227	
	(all) × (men lower)	-4.2172	-4.2194	-4.2151	
	(all) × (men-middle)	-0.8862	-0.8884	-0.8841	
	(all) × (men-upper)	-2.8996	-2.9017	-2.8975	
	(all) × (women lower)	-3.8529	-3.855	-3.8508	
	(all) × (women-middle)	-0.1643	-0.1665	-0.1622	
	(all) × (women-upper)	-3.8952	-3.8974	-3.8931	
	(men lower) × (men-middle)	3.331	3.3289	3.3331	
	(men lower) × (men-upper)	1.3176	1.3155	1.3198	
	(men lower) × (women lower)	0.3643	0.3622	0.3665	
	(men lower) × (women-middle)	4.0529	4.0508	4.055	
	(men lower) × (women-upper)	0.322	0.3199	0.3241	
	(men-middle) × (men-upper)	-2.0134	-2.0155	-2.0112	
	(men-middle) × (women lower)	-2.9667	-2.9688	-2.9645	
	(men-middle) × (women-middle)	0.7219	0.7198	0.724	
	(men-middle) × (women-upper)	-3.009	-3.0111	-3.0069	
	(men-upper) × (women lower)	-0.9533	-0.9554	-0.9512	
	(men-upper) × (women-middle)	2.7353	2.7331	2.7374	
	(men-upper) × (women-upper)	-0.9956	-0.9978	-0.9935	
	(women lower) × (women-middle)	3.6886	3.6864	3.6907	
	(women lower) × (women-upper)	-0.0423	-0.0445	-0.0402	
	(women-middle) × (women-upper)	-3.7309	-3.733	-3.7288	

Table A.14: Multi-group means comparisons of the distributions of the mobility diversity between different set of travels in SAO 2017 using Tukey's HSD test. See the caption of Table A.8 for the description of each column, and the notation.

Travels	Groups	Mean difference	95% Confidence interval Lower bound	Upper bound	Adjusted p-value
	(all) × (men)	0.0704	0.0702	0.0707	
	(all) × (women)	-0.1485	-0.1488	-0.1483	
	(men) × (women)	-0.219	-0.2192	-0.2187	
	(all) × (lower)	90.3673	90.3668	90.3678	
	(all) × (middle)	92.373	92.3725	92.3734	
	(all) × (upper)	88.2008	88.2004	88.2013	
	(lower) × (middle)	2.0057	2.0052	2.0061	
	(lower) × (upper)	-2.1665	-2.1669	-2.166	
	(middle) × (upper)	-4.1721	-4.1726	-4.1717	
	(all) × (men lower)	-2.3309	-2.3317	-2.3301	
	(all) × (men-middle)	-0.3343	-0.3351	-0.3335	
	(all) × (men-upper)	-4.4809	-4.4817	-4.4801	
	(all) × (women lower)	-2.6799	-2.6807	-2.6791	
	(all) × (women-middle)	-0.4041	-0.4049	-0.4033	
all	(all) × (women-upper)	-5.132	-5.1328	-5.1312	***
	(men lower) × (men-middle)	1.9966	1.9958	1.9974	
	(men lower) × (men-upper)	-2.15	-2.1508	-2.1492	
	(men lower) × (women lower)	-0.3489	-0.3497	-0.3481	
	(men lower) × (women-middle)	1.9268	1.926	1.9276	
	(men lower) × (women-upper)	-2.8011	-2.8019	-2.8003	
	(men-middle) × (men-upper)	-4.1466	-4.1474	-4.1458	
	(men-middle) × (women lower)	-2.3456	-2.3464	-2.3448	
	(men-middle) × (women-middle)	-0.0698	-0.0706	-0.069	
	(men-middle) × (women-upper)	-4.7977	-4.7985	-4.7969	
	(men-upper) × (women lower)	1.801	1.8002	1.8018	
	(men-upper) × (women-middle)	4.0768	4.076	4.0776	
	(men-upper) × (women-upper)	-0.6511	-0.6519	-0.6503	
	(women lower) × (women-middle)	2.2757	2.2749	2.2765	
	(women lower) × (women-upper)	-2.4522	-2.453	-2.4514	
	(women-middle) × (women-upper)	-4.7279	-4.7287	-4.7271	
	(all) × (men)	0.116	0.1154	0.1165	
	(all) × (women)	-0.4142	-0.4147	-0.4136	
	(men) × (women)	-0.5301	-0.5307	-0.5296	
	(all) × (lower)	89.0198	89.0187	89.0209	
	(all) × (middle)	91.8841	91.883	91.8851	
	(all) × (upper)	86.0713	86.0702	86.0724	
	(lower) × (middle)	2.8643	2.8632	2.8654	
	(lower) × (upper)	-2.9485	-2.9496	-2.9474	
	(middle) × (upper)	-5.8127	-5.8138	-5.8117	
	(all) × (men lower)	-3.2379	-3.2398	-3.2359	
	(all) × (men-middle)	-0.3976	-0.3995	-0.3956	
	(all) × (men-upper)	-6.1201	-6.1221	-6.1182	
	(all) × (women lower)	-4.6039	-4.6058	-4.6019	
	(all) × (women-middle)	-0.6298	-0.6318	-0.6279	
work	(all) × (women-upper)	-7.6548	-7.6567	-7.6528	***
	(men lower) × (men-middle)	2.8403	2.8384	2.8422	
	(men lower) × (men-upper)	-2.8822	-2.8842	-2.8803	
	(men lower) × (women lower)	-1.366	-1.368	-1.3641	
	(men lower) × (women-middle)	2.608	2.6061	2.61	
	(men lower) × (women-upper)	-4.4169	-4.4189	-4.415	
	(men-middle) × (men-upper)	-5.7225	-5.7245	-5.7206	
	(men-middle) × (women lower)	-4.2063	-4.2083	-4.2044	
	(men-middle) × (women-middle)	-0.2323	-0.2342	-0.2303	
	(men-middle) × (women-upper)	-7.2572	-7.2592	-7.2553	
	(men-upper) × (women lower)	1.5162	1.5143	1.5182	
	(men-upper) × (women-middle)	5.4903	5.4883	5.4922	
	(men-upper) × (women-upper)	-1.5347	-1.5366	-1.5327	
	(women lower) × (women-middle)	3.9741	3.9721	3.976	
	(women lower) × (women-upper)	-3.0509	-3.0528	-3.049	
	(women-middle) × (women-upper)	-7.025	-7.0269	-7.023	