Spatial-focused Multimodality Analysis in the City of Lisbon

Carlos Eduardo Rodrigues Lemonde de Macedo

Dissertation to obtain the Master of Science Degree in

Computer Science and Engineering

Supervisors: Prof. Rui Miguel Carrasqueiro Henriques
Dr. Elisabete Maria Mourinho Arsénio

Examination Committee

Chairperson: Prof. Alberto Manuel Rodrigues da Silva

Supervisor: Prof. Rui Miguel Carrasqueiro Henriques

Member of the Committee: Prof. Alexandra Sofia Martins de Carvalho

January 2021

Acknowledgements

The present work it certainly would not have been possible without the valuable contribution of a group of people to whom the author wishes to express his thanks: To my advisors, Professor Rui Henriques and Engª. Elisabete Arsénio, for their trust, for the criticisms and suggestions, for the availability and friendship. To Engº José Barateiro, for the interest in the work and the means and conditions provided. To my colleagues that also contributed to the project ILU, and helped me many times. To my father, for his support and for being an inspiration in the field of engineering. For the rest of my family and friends for their unconditional support. And last but not least, to my lovely girlfriend for believing in me.

Resumo

A Câmara Municipal de Lisboa está a desenvolver esforços para recolher dados de tráfego urbano e o seu contexto situacional para obter uma visão mais abrangente das mudanças em curso na mobilidade multimodal e apoiar decisões em conformidade. O presente trabalho contribui para o projeto pioneiro de pesquisa e inovação “Integrative Learning from Urban Data” (ILU), e descreve uma metodologia para identificar padrões de mobilidade multimodal através da análise de índices espaço-temporais de multimodalidade em transporte público de passageiros em relação ao contexto situacional disponível. A análise foi realizada através da aplicação de dois índices socioeconómicos (Coeficiente de Gini e Índice de Herfindahl) na cidade de Lisboa organizada por uma unidade geográfica sintética, as Zonas de Análise de Tráfego (TAZ). Os resultados demonstraram que o centro da cidade, rico em pólos de geração e atração de tráfego, beneficia de um uso mais multimodal do sistema de transporte público. Adicionalmente, foi construída uma ferramenta de software com o objetivo de auxiliar especialistas da área, a encontrarem inconsistências na rede de transporte público da cidade de Lisboa.

Palavras Chave

Multimodalidade; Mobilidade Sustentável; Análise de Dados; Análise de Dados Espaço-temporais; Transporte Público.

Abstract

The effects of the rise of car ownership in urban centers are known, such as heavy congestion, road accidents, fuel consumption and air and noise pollution, lowering people's quality of life, and loss of competitiveness of certain urban areas. These increasing concerns prompt modern cities to reevaluate their transportation system and promote a shift towards more efficient and sustainable modes. Multimodality, the use of different modes of transport in a single journey, can support the shift to a low carbon economy by taking advantage of the benefits of different transport types to ease pressure on Europe's congested roads, whilst also contributing to safer and cheaper transportation. In this context, the Lisbon's City Council is establishing efforts to collect urban traffic data and their situational context for gaining more comprehensive views of the ongoing multimodal mobility changes and support decisions accordingly.

The present work is anchored in the pioneer research and innovation project "Integrative Learning from Urban Data" (ILU), and describes a methodology to identify multimodal mobility patterns through the analysis of spatiotemporal indices of multimodality in passengers' public transport against the available situational context. The analysis was conducted by applying two social-economic indices (Gini Coefficient and Herfindahl index) in the city of Lisbon organized by a synthetic geographic unit, the Traffic Analysis Zones (TAZ). Results demonstrate that the center of the city, abundant in traffic generation and attraction poles, benefit from a more multimodal usage of the public transport system. Additionally, a software tool was built in order to aid specialists in the field, to find inconsistencies on the public transportation network of the city of Lisbon.

Keywords

Multimodality; Sustainable Mobility; Data Analysis; Spatio-temporal Data Analysis; Public Transportation.

Contents

1 Introduction **1**

1.1 Research Problem . . . 3

1.2 Contributions . . . 4

1.3 Organization of the Document . . . 5

2 Background **7**

2.1 Public Transport in Lisbon . . . 9

2.2 Multimodality . . . 10

2.3 Spatial-temporal Analysis . . . 11

3 Literature Review **15**

3.1 Multimodality Patterns . . . 17

3.2 Inequality Measurement . . . 18

3.3 Multimodality Traffic Performance Measurement . . . 24

4 Methodology **25**

4.1 Public Traffic Data Analysis . . . 27

4.2 Multimodality Index Data Analysis . . . 30

4.3 Incorporating Situational Context . . . 33

5 Results **35**

5.1 The Dataset . . . 37

5.2 Public Transport Data Analysis . . . 40

5.3 Multimodality Indices Analysis . . . 48

5.4 Software Tool . . . 51

6 Conclusion **53**

6.1 Concluding Remarks . . . 55

6.2 Community and Scientific Acceptance . . . 55

6.3 Future Work . . . 56

A Auxiliary Tables **61**

B Auxiliary Figures **69**

Acronyms

PGIL	Intelligent Management Platform of the City of Lisbon
ILU	Integrative Learning from Urban Data
INESC	Instituto de Engenharia de Sistemas e Computadores
IST	Instituto Superior Técnico
LNEC	Laboratório Nacional de Engenharia Civil
LMA	Lisbon Metropolitan Area
CP	Comboios de Portugal
OTLIS	Operadores de Transportes da Regiao de Lisboa
TAZ	Transport Analysis Zone
DTW	Dynamic Time Warping
DBA	DTW Barycenter Averaging
PCC	Pearson’s Correlation Coefficient
SRC	Spearman’s Rank Correlation
DXA	Detrended Cross-correlation Analysis
AFC	Automatic Fare Collection System
HH	Herfindahl-Hirschman (index)
NHH	Normalized Herfindahl-Hirschman (index)

List of Figures

2.1 Lisbon Metropolitan Area and its municipalities. 9
2.2 Example of DTW Barycenter Averaging (DBA). 12

3.1 Lorenz curve. 20
3.2 Two Lorenz curves with the same Gini coefficient. 21
3.3 Lorenz Curve and Robin Hood Index. 22

4.1 Zoning: geographical decomposition of the Lisbon city at different granularities. 29

5.1 Lisbon's stations location. a) Carris. b) Metro c) Gira. 37
5.2 Weekly mode share distribution of TAZ nº66. a) Week days. b) Weekends. 37
5.3 Sample of Carris smart card's validations data. 38
5.4 Sample of Metro smart card's validations data (entries). 38
5.5 Sample of Gira stations' data. 39
5.6 Major traffic generation poles: commercial (blue), schools and institutes (green), and health centres (red). 39
5.7 TAZs of the Municipality of Lisbon. 41
5.8 Parishes of the Municipality of Lisbon. 41
5.9 TAZs with three modes of transport (Metro, Carris, Gira). 42
5.10 Parishes with three modes of transport (Metro, Carris, Gira). 42
5.11 Cycling-bus-subway market quota (modal trip share) for major traffic analysis zones (TAZ) in Lisbon. 43
5.12 Weekly volume and variation of validations in TAZ nº66. 43
5.13 Daily volume and variation of validations in TAZ nº66. a) Week days. b) Weekends. 45
5.14 Barycenter averaging (DTW and Euclidean) for weekly volume of validations in TAZ nº66. 46
5.15 Barycenter averaging (DTW and Euclidean) for daily volume of validations in TAZ nº66 (logarithmic scale). a) Week days. b) Weekends. 47
5.16 Weekly Pearson correlation heatmap between modes of TAZ nº66. 48

5.17 Weekly Spearman correlation heatmap between modes of TAZ nº66. 48
5.18 Weekly DXA correlation heatmap between modes of TAZ nº66. 48
5.19 Gini index TAZ map (week days). a) 8h. b) 12h. c) 17h. d) 21h. 49
5.20 HH index TAZ map (week days). a) 8h. b) 12h. c) 17h. d) 21h. 50
5.21 Weekly Gini index lineplot. 50
5.22 Weekly HH index lineplot. 51
5.23 Home page of the ILU App. 51
5.24 Multimodal Patterns Analysis Menu (ILU App). 52
5.25 Multimodal Indices Analysis Menu (ILU App). 52

B.1 Daytime Carris Network Diagram. 70
B.2 Metropolitano de Lisboa Network Diagram. 71
B.3 CP/Metro Network Diagram. 72

List of Tables

A.1 TAZs of the Municipality of Lisbon 61
A.2 Parishes of the Municipality of Lisbon 66

1

Introduction

Contents

1.1 Research Problem 3
1.2 Contributions 4
1.3 Organization of the Document 5

1.1 Research Problem

In the last decades, road traffic and mobility needs have increased significantly in Europe, especially in urban and metropolitan areas, as a result of the ongoing economic growth and other socioeconomic changes, causing higher CO_2 emissions, congestion and air pollution, compromising quality of life of European citizens. To reach climate goals set by the Paris Agreement, the European Commission have already recognised the importance of multimodal passenger transport and is committed to increase the use of public transport and other active modes such as walking, cycling, and shared mobility options.[1] Multimodality, the use of different transport modes on the same journey, can offer more efficient transport solutions whilst contributing to a more sustainable and integrated transport system, by taking advantage of the benefits of the different modes, such as convenience, reliability, cost, speed and predictability.

Mobility in major European capitals is not yet sustainable, prompting those capitals to reevaluate their public transport system to meet environmental goals. Lisbon, capital of Portugal, has an estimated resident population of 506.654 inhabitants within its administrative area of 100 km^2, and comprises fourteen public transport operators. Still, there are half a million cars circulating everyday in Lisbon. It is estimated that 370.000 cars enter the capital each day, joining the 200.000 that already circulate within the city.[2] With a diversity of public transports available and a population density considered low compared to other European Capitals, it raises awareness about the intense traffic flow from individual transport (private cars).[3] However, two major problems still remain: the perceived duration of the journey by public transport that negatively influences the mode choice, giving preference to the car; and, the fact that public transport operators make isolated decisions in managing and planning their transport modes, without having in consideration the arrangement of the other modes that share the same space, making their use inefficient.

The Lisbon's City Council is making efforts in becoming sensorized by collecting heterogeneous urban data for a better understanding of the city mobility patterns. Big data are currently being consolidated in the Intelligent Management Platform of the City of Lisbon (PGIL) to meet various purposes.[4] Still, the potentialities of exploring the multiplicity of available urban data sources in an integrative manner for reaching sustainable mobility goals are still untapped.

---

[1] 2018 - Year of Multimodality: `https://ec.europa.eu/transport/themes/logistics-and-multimodal-transport/2018-year-multimodality_en`.

[2] pordata: `https://www.pordata.pt`.

[3] `https://www.worldometers.info/population/countries-in-europe-by-population/`.

[4] PGIL: `https://https://lisboainteligente.cm-lisboa.pt`.

1.2 Contributions

This work is anchored in the pioneer research and innovation project “Integrative Learning from Urban Data” (ILU), a project that joins the Lisbon City Council and two research institutes (INESC/IST and LNEC), bridging the ongoing research on urban mobility with recent advances from artificial intelligence.[5] The project ILU proposes to address the following challenges:

- The lack of an integrative analysis capable of combining different sources of urban data collected from sensors in the city and of validation tickets in different public transport modes;
- The absence of a situational context in the forecast and recommendation of circulation in the city.

This work aims at contributing to these lines of research by proposing a methodology for the comprehensive understanding of multimodal synergies in demanding urban areas of Lisbon, using the available data collected from different sources, and to relate that knowledge with relevant situational context. To achieve this, three major activities are pursued with this dissertation:

1. *Comprehensive descriptive analytics*:

   With the initiatives established by the Lisbon City Council towards sustainable mobility, enabling the access and consolidation of numerous sources of urban data, it was possible to easily prepare the data to get relevant knowledge in a straightforward manner. The description of statistical measurements to model the expected demand in public transportation at different times of the day and the week, in different zones. And, the computation of the expected behavior through the application of barycenter averaging.

2. *Multimodal pattern analysis*:

   Assessment of Lisbon’s multimodal patterns, using the mathematical properties of socioeconomic indices, and correlating these indices with the available situational context. This type of analysis will allow to have a global view on the performance of multimodality in several areas of the city of Lisbon.

3. *Software Tool*:

   Deployment of a software tool comprising the previously mentioned analytics. In the form of a web application, called *ILU App*, where specialists can visually interact with the data of the different types of transport modes in the city of Lisbon, applying different parameters (e.g. temporal, spatial, etc.) for analytic reporting to aid their research.

---

[5] ILU: `https://https://https://web.ist.utl.pt/rmch/ilu/`.

Additionally, the research developed in the scope of this dissertation was submitted for international review in the form of two articles. First the article “Boosting Multimodality Mobility Decisions using Big Data in the City of Lisbon: ongoing and future challenges”, accepted by the scientific committee of the 14th Conference on Transport Engineering (CIT 2020). And then the article “Exploring multimodal mobility patterns with big data in the city of Lisbon”, submitted to the scientific committee of the 48th European Transport Conference (ETC 2020).

1.3 Organization of the Document

The rest of the present work is organized as follows: Chapter 2 introduces concepts related to multimodality and spatial temporal analysis, also including a description of Lisbon’s public transport network; Chapter 3 presents insights of multimodality behavior related studies, and related work on inequality measurement, and on multimodality performance measurement; In Chapter 4 a methodology to the assessment of multimodality at a spatial level is described; and Chapter 5 presents results of applying our approach in order to evaluate multimodal patterns of the public transportation in Lisbon; finally, Chapter 6 concludes this work and proposes some potential expansions for future work.

2

Background

Contents

2.1 Public Transport in Lisbon 9
2.2 Multimodality 10
2.3 Spatial-temporal Analysis 11

This chapter describes the essential concepts and notions necessary for an absolute understanding of the remaining chapters of this work. The chapter starts with a description of the public transportation available in the Municipality of Lisbon and its integrated fare collection system.

2.1 Public Transport in Lisbon

The available traffic data comes from various heterogeneous sources collected for the Lisbon Metropolitan Area (LMA). The LMA is an administrative division in Portugal centered in the municipality of Lisbon and covering more 17 municipalities (Figure 2.1). Although the reported research is directed towards the municipality of Lisbon, its contribution and results can be extended and applied to other nearby municipalities to enable more comprehensive analysis of inter-municipal commuting mobility patterns.

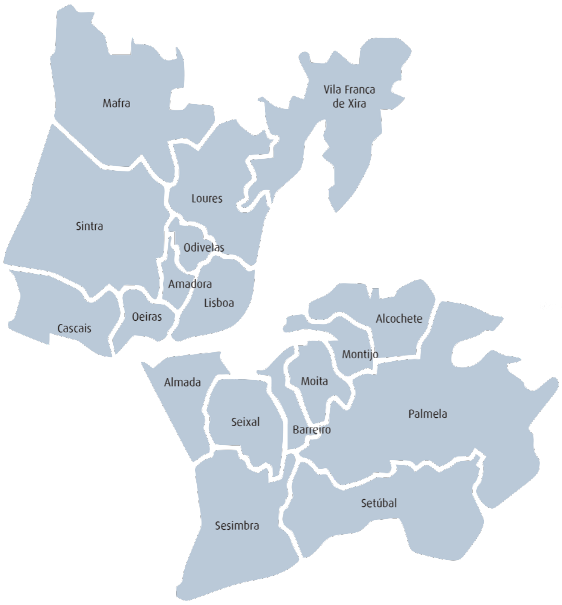


Figure 2.1: Lisbon Metropolitan Area and its municipalities.

Source: www.vimeca.pt

Like many European Capitals, Lisbon comprises a vast and diverse fleet of public transport, fulfilling any mobility needs of the population. Nearly all types of transport modes are available in the city:

- **Bus**: operated by a municipal company of public urban passenger transport surface called Carris, that also manages electrical trams, and urban lifts.[1] With a fleet of 706 buses, Carris covers all the area of Lisbon's Municipality (see Figure B.1).
- **Subway**: The Metropolitano de Lisboa (Metro) provides a public passenger transport service, in

---

[1] `https://www.carris.pt`.

subway mode.[2] It comprises four lines - Blue Line, Yellow Line, Green Line and Red Line - running on 44.5 kilometres of route and serving 56 stations (see Figure B.2).

- **Railway**: The rail transport service, in urban trains, of passengers in Lisbon is the responsibility of Comboios de Portugal (CP).[3] CP has a large network of transportation across Portugal, it includes four railway services that enable the travelling from the municipality of Lisbon to other Portuguese cities. Figure B.3 presents the modal interfaces between the CP service of Lisbon's urban rail transport and Metro.
- **Waterway**: Transtejo and Soflusa provide a public river transport service integrated in the global system of the Lisbon Metropolitan Area.[4] It allows crossing the Tagus river.
- **Cycling**: The city of Lisbon has a public bike sharing system known as Gira.[5] Bicycle stations are scattered across the city, where a person can pick up a bike using her/his smartphone, and deliver it to a station near her/his destination.

The providers of bus, subway, railway and inland waterway modes of transport are currently operating under an integrated fare collection system, enabled through the VIVA card initiative.[6] The VIVA card initiative, firstly established between the subway operator (Metro) and the major bus operator (Carris), was in 2017 extended to further encompass railway operator, Comboios de Portugal (CP), and in 2019 extended towards the remaining major carriers operating within (or interfacing with) the city of Lisbon. To this end, the early individual ticketing systems were consolidated into a unique ticketing system coordinated by OTLIS, the entity responsible to manage the information resources shared among carriers. The integrated fare collection offers the unprecedented possibility to trace the movements of each user throughout the modes of the public transportation system, providing an essential source of information to understand the true mobility dynamics in the city. In 2019, multimodal tariff plans were also released to create incentives towards a multimodal use of the public transportation system.

2.2 Multimodality

Multimodality is commonly defined as the use of more than one transport mode to complete a trip within a certain time period. By contrast, **monomodality** generally refers to the exclusive use of one mode of transport [29]. Buehler and Hamre (2016) state that multimodality is a subfield of a larger body of research on intrapersonal variability of travel behaviour, which consists of four dimensions: temporal, spatial, purpose and modal. Where the “modal” dimension describes the variability in the use of means of transport over time, referring to the multimodality research [6]. Nobis (2007) emphasizes the fact that the general definition of multimodality must be observed along individual trips to ensure its separation from the monomodality concept [29]. This distinction relates to the chosen time period, the longer the time period is, the higher is the probability that a person uses more than one mode of transportation. For instance, Nobis (2007) uses in her study a loose definition of multimodality, where any person who uses more than one mode of transportation within one week is a multimodal transport user.

---

[2] `https://www.metrolisboa.pt`.
[3] `https://www.cp.pt`.
[4] `https://ttsl.pt`.
[5] `https://www.gira-bicicletasdelisboa.pt`.
[6] `https://www.portalviva.pt`.

The following terms are related to the spatial criteria applied to the multimodality analysis (see Section 4.1.2).

A **Traffic Analysis Zone** (TAZ) is a geographical unit used in transportation planning models to assess socio-economic data [27]. The use of TAZ delineation instead of other more commonly used zones, like municipalities or parishes (i.e. administrative zones) is justified by the reduced noise level of the data for traffic modelling, and at the same time, the decrease of geographical error of the trip end location.

The concept of **Traffic Generation and Attraction Poles** refers to commercial areas, employment centres such as business parks and enterprises, and collective equipment like hospitals, schools and stadiums, that generate or attract a significant volume of vehicle trips, either from contributors, visitors or providers [19].

2.3 Spatial-temporal Analysis

Spatial-temporal analysis techniques are required to assess urban mobility behaviours. Spatio-temporal data differs from relational data in that both spatial and temporal attributes are available in addition to the actual measurements/attributes. Contrary to traditional data mining that deals with distinct objects (or data instances) having well-defined features, in Spatio-temporal data mining, one can define objects and features in a variety of ways. This section describes mathematical concepts that were applied in this research work [3].

Time Series: A time series represents a collection of values obtained from sequential measurements over time [5, 13]. A *time series* x is an ordered sequence of t real-valued observations from a random variable,

$$x = (x_1, ..., x_t), x_i \in \mathbb{R}. \tag{2.1}$$

Multivariate Time Series: The previous definition refers to univariate time series. When multiple variables are monitored along the same time range, the gathered observations form a multivariate time series [5, 13],

$$\vec{x}_t = [x_{1,t}, .., x_{m,t}], x_{i,t} \in \mathbb{R}. \tag{2.2}$$

Dynamic Time Warping: Dynamic Time Warping (DTW) is based on the Levenshtein distance, it finds the optimal alignment (or coupling) between two numeric time series, and captures flexible similarities by aligning the coordinates inside both series [31]. The cost of the optimal alignment can be recursively computed by,

$$D(A_i, B_j) = \delta(a_i, b_i) + min \begin{Bmatrix} D(A_{i-1}, B_{j-1}) \\ D(A_i, B_{j-1}) \\ D(A_{i-1}, B_j) \end{Bmatrix}, \tag{2.3}$$

where A_i is the subsequence $\langle a_1, ..., a_i \rangle$, B_j is the subsequence $\langle b_1, ..., b_j \rangle$, and δ is a distance between elements of two series. The overall similarity is given by:

$$D(A_{|A|}, B_{|B|}) = D(A_X, B_X). \tag{2.4}$$

Average Time Series (Barycenter): Given a set of time series $D = \{x_1, .., x_n\}$ in a space E induced by Dynamic Time Warping, the average time series $\overline{x}$ is the time series that minimizes [30]:

$$argmin \sum_{i=1}^{n} DTW^2(\overline{x}, x_i). \tag{2.5}$$

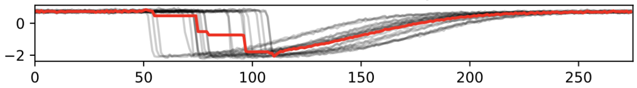


Figure 2.2: Example of DTW Barycenter Averaging (DBA).

Pearson's Correlation: For numeric attributes, it's possible to evaluate the correlation between two attributes, A and B, by computing the Pearson's Correlation Coefficient (PCC) [15]. It can also be applied to time series by pairing observations by time points and ignoring time dependencies between observations,

$$r_{A,B} = \frac{\sum_{i=1}^{n}(a_i - \overline{A})(b_i - \overline{B})}{n\sigma_A\sigma_B}, \tag{2.6}$$

where n is the number of tuples, a_i and b_i are the respective values of A and B in tuple i, $\overline{A}$ and $\overline{B}$ are the respective mean values of A and B, σ_A and σ_B are the respective standard deviations of A and B. And $-1 \leqslant r_{A,B} \geqslant +1$. If $r_{A,B}$ is greater than 0, then A and B are positively correlated, meaning that the values of A increase as the values of B increase. The higher the value, the stronger the correlation. Hence, a higher value may indicate that A (or B) may be removed as a redundancy. If the resulting value is equal to 0, then A and B are independent and there is no correlation between them. If the resulting value is less than 0, then A and B are negatively correlated, where the values of one attribute increase as the values of the other attribute decrease. This means that each attribute discourages the other.

Spearman's Correlation: Spearman's Rank Correlation (SRC) is a non-parametric (distribution-free) correlation measure which equals the Pearson correlation computed from the ranks of the observations, only if all ranks are distinct integers, usually designated as r_S [40],

$$r_S = 1 - \frac{6 \sum d_i^2}{n^3 - n}, \tag{2.7}$$

where $d_i = rank(X_i) - rank(Y_i)$. If each of the n measurements of one of the time series is denoted as X_i (i.e. $X_1, X_2, ..., X_n$), then $rank(X_i)$ may represent the rank of X_i, where each rank is an integer, from 1 through n, indicating relative magnitude. The measurements may be ranked from high to low (e.g. rank 1 indicates the fastest car, rank 2 the next fastest, and so on, with rank n the slowest) or from low to high (rank 1 denotes the slowest and rank n the fastest). Similarly, each of the n measurements of the second time series may be denoted as Y_i (i.e. $Y1$, $Y2$, ... , Yn), and $R(Y_i)$ would denote the rank of Y_i, where the sequence of ranking (either high to low or low to high) is the same as for $R(X_i)$. An $r_S = 0$ ("no correlation") indicates that the magnitudes of the ranks of one time series are independent of the magnitudes of the ranks of the second time series. A positive value of r_S ("positive correlation") indicates that the $rank(Xi)$ increases as $rank(Y_i)$ increases; a negative r_S ("negative correlation") indicates that the $rank(X_i)$ decreases as $rank(Y_i)$ increases. If the sequence of ranks were identical for the two time series, meaning there was a perfect positive correlation, and r_S = 1.0. A perfect negative correlation (where $r_S = -1.0$) would be one in which the magnitudes of the ranks for one variable vary inversely with the sizes of the ranks of the second.

Covariance: In probability theory and statistics, correlation and covariance are two similar measures for assessing how much two attributes change together [15]. Considering two numeric attributes A and B, and a set of n observations $(a_1, b_1), ..., (a_n, b_n)$. The mean values of A and B, also known as the expected values on A and B, respectively:

$$E(A) = \overline{A} = \frac{\sum_{i=1}^{n} a_i}{n}, \tag{2.8}$$

and

$$E(B) = \overline{B} = \frac{\sum_{i=1}^{n} b_i}{n}. \tag{2.9}$$

Hence, the covariance between A and B is defined as:

$$Cov(A, B) = E((A - \overline{A})(B - \overline{B})) = \frac{\sum_{i=1}^{n} (a_i - \overline{A})(b_i - \overline{B})}{n}. \tag{2.10}$$

Equation (2.6) can be reformulated using equation 2.10 as:

$$r_{A,B} = \frac{Cov(A,B)}{\sigma_A \sigma_B}. \tag{2.11}$$

It can also be shown that,

$$Cov(A,B) = E(A \cdot B) - \overline{AB}. \tag{2.12}$$

If A is larger than $\overline{A}$ (the expected value of A), then B is likely to be larger than $\overline{B}$ (the expected value of B). Therefore, the covariance between A and B is positive. Conversely, if one of the attributes tends to be above its expected value when the other attribute is below its expected value, then the covariance of A and B is negative.

Detrended Cross-Correlation Analysis: Podobnik and Stanley (2008) proposed a modification of the above covariance equation, called Detrended Cross-correlation Analysis (DXA), in order to quantify long-range cross-correlations when non-stationarities are present [32]. In this procedure, two long-range cross-correlated time series y_i and y'_i of equal length N, are divided into $N-n$ overlapping boxes, each containing $n+1$ values. Defining two integrated signals $R_k = \sum_{i=1}^{k} y_i$ and $R_k = \sum_{i=1}^{k} y'_i$, where $k = 1, .., N$, a local trend to be the ordinate of a linear least-squares fit, $\tilde{R}_{i,k}$ where $i \leq k \leq i+n$, and the detrended walk as the difference between the original walk and the local trend. The covariance of the residuals in each box is calculated by:

$$f^2_{DXA}(n,i) = \frac{1}{n-1} \sum_{k=i}^{i+n} (R_k - \tilde{R}_{k,i})(R'_k - \tilde{R}'_{k,i}). \tag{2.13}$$

Finally, the detrended covariance is given by:

$$F^2_{DXA}(n) = \sum_{i=1}^{N-n} f^2_{DXA}(n,i). \tag{2.14}$$

3

Literature Review

Contents

3.1 Multimodality Patterns 17

3.2 Inequality Measurement 18

3.3 Multimodality Traffic Performance Measurement 24

The present chapter summarizes research work related to the study of urban multimodality. The structure of this chapter follows a line of thought reflected in the next two chapters. It starts by referencing other works that analysed multimodal traffic patterns and behaviors. Followed by a complete definition of inequality measures present in the social and economic literature. And ends with mentions to other work that used those socio-economic indices to measure traffic multimodality.

3.1 Multimodality Patterns

Comparison of findings on multimodality analyses and initiatives is challenging, because of different geographic locations, data sources, timing, and definitions of multimodality. However, some relevant results are common among studies: the percentage of multimodal persons decreases with advancing age [7, 22, 29]; car availability is negatively correlated with multimodal behaviour, and positively correlated with monomodal driving [11, 22, 29]; and having a driver's license is negatively associated with multimodal users [21, 29].

Transfers affect the attractiveness of passenger transport [24]. Therefore, examining transfer patterns can be beneficial for public transport management. Jang (2010) illustrates that multimodal transfer data can be used to locate the critical transfer points that need improvement [20]. The dataset used in his research came from the automatic fare collection (AFC) system of Seoul in South Korea. Contrary to the fare collection method adopted in Lisbon, the AFC in Seoul is distance-based, where the fare is calculated on the basis of the total distance run by buses, subway trains, or both from boarding to alighting. Unlike the public transport operator Carris in Lisbon, the Seoul's buses are equipped with two smart card readers located at the doors, for boarding and alighting, so it is possible to obtain the whole itinerary of each individual trip from the departing location to the final destination, including intermediate transfers. In his research, Jang codes transfer patterns into series of alphanumeric letters. Each letter represents a public transport mode associated with it. For example, "BRB" represents a bus-subway-bus trip. The data collected from the AFC allows to identify stops or stations that have heavy transfer demands, pointing out the areas that need improvements to enable a seamless transfer between modes.

Heinen (2018) analysed the association between multimodality and the intention to change transport mode [16]. The research by Heinen aimed to demonstrate that multimodality can be a predictor of variability of mode choice over time, diverging from other common predictors of variability such as, gender, employment, car ownership and life events. Using five indicators of variability and cluster analysis, Heinen explored to what extent multimodality was associated with the intention to change the level of cycling, walking, car use and public transport use. Research findings revealed that the more multimodal individuals were, the more likely they intended to decrease their car use. However, the analysis by the mentioned author did not provide conclusive evidence that the level of multimodality is associated with the intention to change.

3.2 Inequality Measurement

Assessing multimodality can be seen as a particular instance of a more general issue, the measure of diversity and inequality [12]. This section presents the properties of well-established inequality measures from the branch of socio-economics, and how to compute and model inequality.

3.2.1 Properties of an Inequality Measure

An inequality measure is usually a function that describes a value to a specific distribution of income in a way that allows direct and objective comparisons across different distributions. An inequality measure should have certain properties and behave in a certain way given certain events. Cowell (2011) proposes five criteria to build a particular class of mathematical functions for use as inequality measures [8]:

1. *Weak Principle of Transfers*:

   An inequality measure satisfies the weak principle of transfers if the following statement is true. Considering any two individuals, one with income y, the other, a richer person, with income $y + \delta$ where δ is positive ($\delta > 0$). The richer person transfers a positive amount of income Δy to the poorer person, where Δy is less than 2δ. Inequality should then definitely decrease.

2. *Scale Independence*:

   This states that the measured inequality of the slices of the cake should not depend on the size of the cake. If everyone's income changes by the same proportion then it can be argued that there has been no essential alteration in the income distribution, and thus that the value of the inequality measure should remain the same.

3. *Principle of Population*:

   This means that the inequality of the cake distribution should not depend on the number of cake-receivers. If the measured inequality in a particular economy with n people in it, is then merged with the economy of another identical one, it returns a combined economy with a population of $2n$, and with the same proportion of the population receiving any given income. To satisfy the principle of population, the measured inequality must be the same for any replication of the economy.

4. *Decomposability*:

   This property indicates that there should be a coherent relationship between inequality in the whole of society and inequality in its constituent parts.

5. *Strong Principle of Transfers*:

   Consider a distance measure given by:

$$d = h(s_1) - h(s_2),$$

   where s_2 is greater than s_1, and $h(s)$ is a decreasing function defined as $h(s) = \frac{1-s^\beta}{\beta}$. Then consider a transfer from rich person 2 to poor person 1. We say that the inequality measure satisfies the principle of transfers in the strong sense if the amount of the reduction in inequality depends only on d, the distance, no matter which two individuals we choose.

Aaberge (1986) noted that a desirable property of an inequality measure is that it should equal zero when the underlying distribution function expresses perfect equality [1].

3.2.2 Representation of Inequality

Income distribution is a central topic in welfare economics. Deardorff (2014) defines welfare economics as the branch of economics concerning with social welfare, including especially various propositions relating competitive general equilibrium to the efficiency and desirability of an allocation [9]. In other words, is the study of how the allocation of resources and goods affects social welfare. This relates directly to the study of economic efficiency and income distribution. In order to assess income distribution in populations, economists apply various inequality measures and principles.

Before describing those measures, a clear definition of income distribution and inequality should be made. Cowell (2011) states that "inequality" suggests a departure from some idea of equality [8], in which case "equality" just represents the fact that two or more given quantities are the same size, and "inequality" merely relates to differences in these quantities. Depending on the problem in hand, inequality can be defined in terms of some personal attribute, such as consumption of a particular good, life expectancy, land ownership, etc; but usually it's expressed regarding income. In Deardorffs' Glossary of International Economics, "income" is the amount of money (nominal or real) received by a person, household, or other economic unit per unit time. May be (or not) in return for services provided or goods sold [9].

Towards establishing an inequality measurement, Cowell (2011) sets the essential ingredients of a "Principle of Inequality Measurement" [8]:

- The specification of an individual social unit such as a single person, the nuclear family or the extended family, depicted as "persons".
- Description of a particular attribute (or attributes) such as income, wealth, land-ownership or voting strength, depicted as "income".

- A method of representation or aggregation of the allocation of “income” among the “persons” in a given population.

The third ingredient refers to inequality measurement which can be summarized as a scalar numerical representation of the interpersonal differences in income within a given population.

The Lorenz curve is one of the simplest representations of inequality. Specialists use the Lorenz curve [25] to represent graphically the degree of inequality in the distribution of income in societies. Lorenz (1905) has pointed out the mathematical inaccuracy of certain commonly used methods, and has suggested a graphic solution. Individuals holding assets of varying size, are arranged in order, poorest to richest. The horizontal axis represents the percent of people and the vertical axis the percent of income those people receive. Equality of distribution would give a series of points in a straight line. The Lorenz curve is obtained by plotting the cumulative proportion of income against the cumulative proportion of population, represented in Figure 3.1.

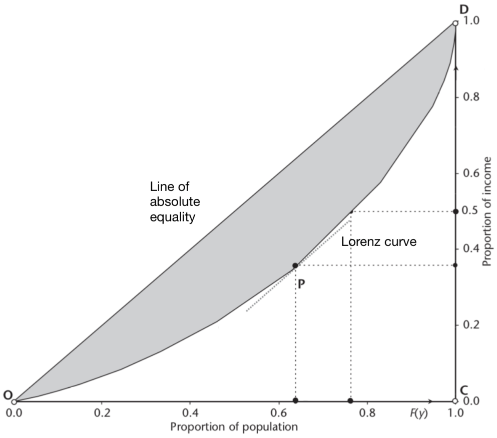


Figure 3.1: Lorenz curve.

Source: Economic Trends, November 1987

A common use of the Lorenz curve is to derive the Gini coefficient, expressed as the ratio of the shaded area in Figure 3.1 to the area OCD:

$$G = \frac{OPD}{OCD}. \tag{3.1}$$

The Gini coefficient was developed by the Italian statistician Corrado Gini [14] as a summary measure of income inequality in society. The Gini coefficient can be presented as a value between 0 and 1 or as a percentage. A coefficient of 0 reflects a perfectly equal society in which all income is equally shared; in this case the Lorenz curve would follow the line of equality. The more the Lorenz curve deviates from the line of equality, the higher will be the resulting value of the Gini coefficient. A coefficient of 1 (or 100%) represents a perfectly unequal society, where all income is earned by one individual in an infinite population. The equation (3.1) can be applied in practical terms as the mean of the difference between every possible pair of individuals, in a population of size n, divided by the mean size μ [35]:

$$Gini = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \mu}. \tag{3.2}$$

The Gini coefficient as with many inequality measures, it is a synthetic index. Therefore, it does not contain all the information in the Lorenz curve, and it has been pointed out that different Lorenz curves can have the same Gini coefficient [39] (see Figure 3.2).

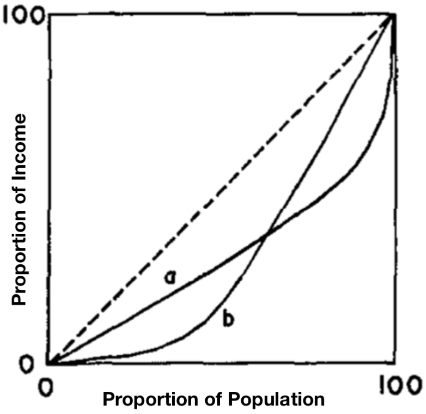


Figure 3.2: Two Lorenz curves with the same Gini coefficient.

The inability of the Gini coefficient brought other inequality measures into the field of social welfare. One of these measures is the Atkinson index, which is used to evaluate the effectiveness and fairness of social distribution [35]. It ranges from 0 to 1, where 0 means the highest equality in distribution, while 1 means it is most unequal in distribution. Let T_i be the income in the ith income range, f_i the proportion of the population in ith group, and $\overline{T}$ the mean household income. The Atkinson equation is defined as follows [2]:

$$Atk = 1 - \left[\sum_{i=1}^{n} \left(\frac{T_i}{\overline{T}} \right)^{1-\epsilon} f_i(T_i) \right]^{\frac{1}{1-\epsilon}}. \tag{3.3}$$

The Atkinson index evaluates the distributional effect of household income with the ϵ parameter, which represents the weight attached by society to inequality in distribution, and by consequence measures the degree of inequality aversion among various groups. The ϵ parameter gives the opportunity to define how sensitively the Atkinson index reacts to income inequalities. When ϵ increases, less distribution is transferred to high income groups and more distribution is transferred to low income groups. The ϵ parameter defines the sensitivity of Atkinson index to income inequalities. For $\epsilon = 0$ (no aversion to inequality), the society ignores inequality in distribution, whereas when ϵ tends to ∞ (infinite aversion to inequality) the society only considers distributing all money to low income people.

Similar to the Gini coefficient, an additional measure based on the Lorenz curve is the Hoover index, also known as the Robin Hood index. It is used to measure the deviation from the preferred equal distribution. It can be graphically represented as the maximum vertical distance between the Lorenz curve and the 45-degree line that represents perfect equality of incomes [33] (see Figure 3.3). The value of the index approximates the share of total income that has to be transferred from households above the mean to those below the mean to achieve equality in the distribution of incomes, and is defined as [18]:

$$RH = \frac{\sum_i |y_i - \overline{y}|}{2\sum_i y_i}, \tag{3.4}$$

where y_i denotes the income of the ith individual and $\overline{y}$ the mean income. Rogerson & Plane (2013) apply the Hoover index for the measurement of the degree of population concentration in a region of the earth, such as a country, that is disaggregated into a set of subregions (e.g. states, counties, or census tracts). In their study, the index ranges from a low of 0 to a high that approaches 100, with larger values indicating greater degrees of concentration. The value of the index can be interpreted as the percentage of the total population that would need to be redistributed across subregions to achieve equal population densities in all subregions [33].

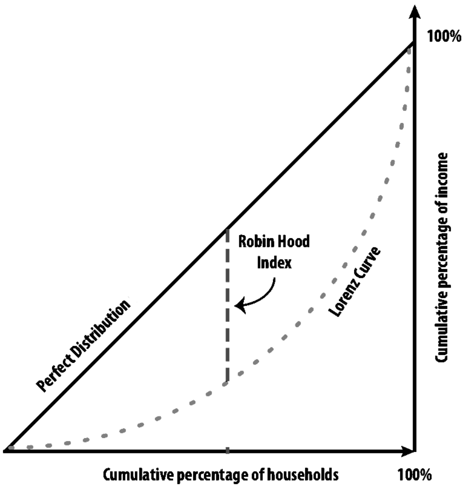

100%
Robin Hood
Index
Lorenz Curve
Perfect Distribution
Cumulative percentage of income
Cumulative percentage of households
100%

Figure 3.3: Lorenz Curve and Robin Hood Index.

The measure of inequality and diversity is also studied in widely different research scopes, and can also be derived from the field of information theory. The study of information theory is concerned with representing data in a compact fashion (i.e. data compression or source coding), as well as with transmitting and storing it in a way that is robust to errors (i.e. as error correction or channel coding) [28]. The task of quantifying information is related to measuring how much surprise there is in an event. Rare events (with low probability) are more surprising and therefore have more information than those events that are common (with high probability). The Shannon entropy of a distribution is the expected amount of information in an event drawn from that distribution and is formally defined as [36]:

$$H(X) = -\sum_{i} p(x_i) \log p(x_i), \tag{3.5}$$

where X is a random variable with possible outcomes $x_i ..., x_n$ which occur with probabilities $p(x_1), .., p(x_n)$. Theil (1967) argued that the entropy concept provides a useful device for inequality measurement [38]. The Theil entropy index addresses a question frequently encountered in the analysis of income inequality; to what extent the inequality in the total population can be attributed to income differences between major population subgroups (e.g. age, sex, race, occupation...) [37]. Let $y = (y_1, ..., y_n)$ be the income distribution vector for a population of n individuals, and $\overline{y}$ the mean income, the Theil index can be written as:

$$T(y) = \frac{1}{n} \sum_{i} \frac{y_i}{\overline{y}} \log \frac{y_i}{\overline{y}}. \tag{3.6}$$

The values for this measure vary between 0 in the case of complete equality, and infinity (or 1 if normalized) indicating larger inequality in the distribution.

Other interesting measure, slightly different from the previous mentioned inequality measures, is the Herfindahl-Hirschman (HH) index. The HH index is a measure of the size of firms in relation to the industry and an indicator of the amount of competition among them [17]. It can range from $1/n$ to 1 moving from a perfect competitive environment (all firms operate with equal market share) to a single monopolistic producer [4]. A commonly accepted measure for market concentration, the general form of the HH index is expressed as the sum of the squares of the market shares $s_i (i = 1, 2, ..., n)$ of all entities in the industry:

$$HH = \sum_{i=1}^{n} (s_i)^2, \tag{3.7}$$

if the following constraint holds:

$$\sum_{i=1}^{n} s_i = 1. \tag{3.8}$$

A modified version of HH is called the Normalized Herfindahl–Hirshman Index (NHH). Unlike the HH index, the NHH ranges between values from 0 to 1 and is calculated as follows [21]:

$$NHH = \frac{HH - \frac{1}{n}}{1 - \frac{1}{n}}. \tag{3.9}$$

3.3 Multimodality Traffic Performance Measurement

Multimodality is generally measured by considering the fraction of users that use a given number of travel modes. For example, Nobis (2007) showed that car and public transportation users tend to be between 10 and 25 years old, with the largest group consisting of people aged 18–25, in Germany [29]. While Buehler and Hamre (2016) indicate that 87% of all trips in the United States are made by car and 90% of Americans use automobiles in their commuting trips for work purposes [6]. Most of these works don't have in consideration the intensity of use of each mode.

Diana and Pirra (2016) targeted the problem of measuring multimodality at the individual level, by finding a multimodality index that comprises both descriptive statistics on the number of travel means, and the intensity of use of each mode [12]. They analysed some of the socio-economic measures, presented in the previous section, as multimodal indices, and in order to easily assess the different indices, they rewrote them as a function of a common set of parameters, reasoning at the end that there is not an index that outperforms the others, still, some measures give best results for different cases.

4

Methodology

Contents

4.1 Public Traffic Data Analysis 27
4.2 Multimodality Index Data Analysis 30
4.3 Incorporating Situational Context 33

Analysing multimodal mobility patterns is a complex task due to the heterogeneity of urban data sources, which requires a specific approach combining steps from descriptive analytics. The following sections describe those methodological steps from the processing of traffic data, to the incorporation with situational context.

4.1 Public Traffic Data Analysis

4.1.1 Data Preprocessing

The first step of the methodology comprises the collection, characterization, preprocessing and uniformization of the available traffic data from the different sources.

In the collection phase, the traffic sources are chosen and their corresponding data is collected. That choice can be based on the following factors:

- **Relevance**: the usage by the population, quantified by the number of validations per day, week or another temporal granularity (Section 4.1.2).
- **Accessibility**: the mode of transport must be available to a large number of the population. Its usage must not be compromised by any type constraint, in other words, the access to its stations/stops must be straightforward and scattered across the chosen spatial granularity (Section 4.1.2).
- **Privacy Policy**: even if working with modes of transport managed by public entities, that does not mean that the usage and disclosure of the respective traffic data can be easily available. Enforcement of privacy policies by those entities can be a hindrance for a complete and adequate analysis, by not sharing the requested urban data or just sharing a portion of the relevant information (i.e. incomplete data).
- **Quality of the Data**: the public transport entities can have open privacy policies, however, if the technology used to collect the data is not reliable, the obtained information can be useless and not adequate for analysis. Another issue is if those entities only collect insufficient traffic data, like for example, collecting only the number of validations and not registering a timestamp associated to the validations. Low-quality data will lead to low-quality mining results, affecting accuracy, completeness, consistency, timeliness, believability, and interpretability [15].

The step of collecting relevant traffic data is followed by its profiling stage. This phase is about getting familiar with the data, obtain knowledge about the data before preprocessing it. Data delivered from different urban sources can come in different file formats and data structures, but generally, data is presented as tabular data. Studying the various attribute types is a first way to approach it, these includes characterizing its nominal attributes, binary attributes, ordinal attributes, and numeric attributes. Basic statistic descriptions can then be applied, such as measures of central tendency – mean (average value), median (middle value), and mode (most common value) – which give us an idea of the "middle" or the center of distribution. Plotting the measures of central tendency shows us if the data are symmetric or skewed [15].

After characterising the data, the next step is the preprocessing. Knowing basic statistics regarding each attribute, in the profiling phase, makes it easier to fill in missing values, smooth noisy values, and spot outliers during data preprocessing. Quantile plots, histograms, and scatter plots are graphic displays of basic statistical descriptions that can be useful during data preprocessing and can provide insight into areas for mining. Data preprocessing is composed by the following tasks [15]:

1. **Data Cleaning**: applied to remove noise and correct inconsistencies in the data. This task is responsible for filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. If the data is "dirty", the analysis will not present trustworthy results.

2. **Data Integration**: merging of data from different data sources into the target consolidated repository, and the identification of shared dimensions among the diversity of data sources.

3. **Data Reduction**: obtains a reduced representation of the data set that is much smaller in volume, but still produces the same (or almost the same) analytical results. Data reduction comprises dimensionality reduction, where data encoding schemes are applied to obtain a "compressed" representation of the original data; and numerosity reduction, which consists of replacing the data by alternative, smaller representations using parametric models (e.g. regression models) or nonparametric models (e.g., histograms, clusters, sampling, or data aggregation).

4. **Data Transformation**: Strategies for data transformation includes smoothing, which means removing noise from data; aggregation; normalization, where attributes are scaled within a smaller range; and discretization where the raw values of a numeric attribute are replaced by interval labels (e.g., 0–10, 11–20, etc.) or conceptual labels (e.g., youth, adult, senior).

4.1.2 Selecting Spatial and Temporal Criteria

Multimodal pattern analysis can be conducted at different spatial and temporal granularities. In terms of spatial specifications, two major possibilities can be considered. One of them is to manually specify the target geographical region of interest using a polygon or a circular marking facility. The other, is to select predefined regions. We propose the following zoning maps for the Lisbon Metropolitan Area, that are represented in Figure 4.1:

- Traffic Analysis Zones (TAZ): geographical unit used in transportation planning models to assess socio-economic indicators.
- Administrative zones: coarsest geographical unit for the city, it can range from municipalities to parishes, depending on the geographical organization of the target city.
- Sections: finest geographical unit, comprising small districts and neighborhoods.

Under the selected spatial granularity, traffic events, such as card validations and trajectories, as well as the accompanying situational context data, are then linked to one or more Lisbon's zones in accordance with their spatial extent.

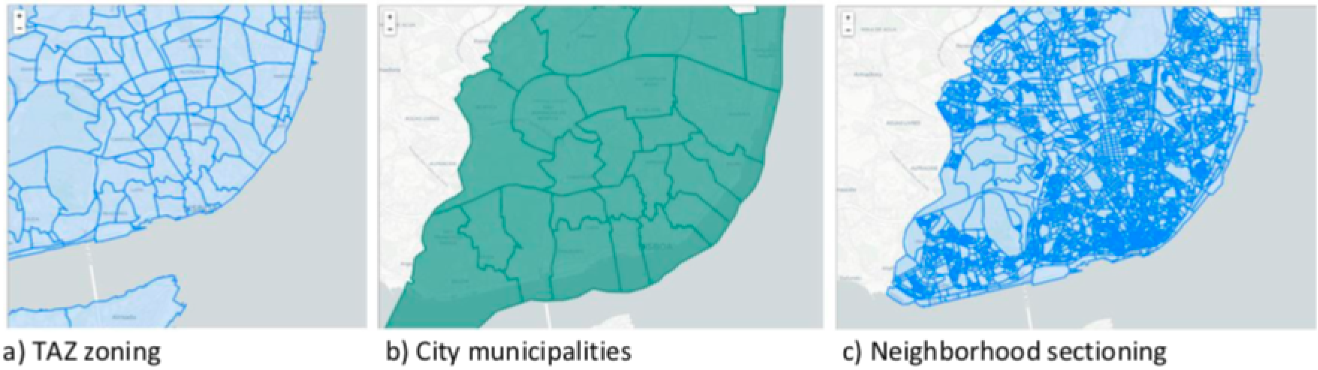

a) TAZ zoning
b) City municipalities
c) Neighborhood sectioning

Figure 4.1: Zoning: geographical decomposition of the Lisbon city at different granularities.

Source: ILU web application

Two major types of temporal constraints can be placed. First, calendrical constraints – such as day of the week (e.g. Mondays), weekdays, holidays or on/off-academic period calendars – can be placed to segment the available traffic data. Multimodal patterns can be represented per calendar or, alternatively, correction factors can be learned from calendrical annotations in order to guide the target tasks. Second, time intervals (e.g. on/off-peak hour intervals) or a fixed time granularity (e.g. 15-minute) can be optionally specified to guide traffic data descriptors. For instance, passenger volume series in public transport can be resampled from card validations. In the absence of a minimum time granularity, the data analysis can be conducted at the raw event level or under multiple time aggregations.

4.1.3 Consolidating Traffic Data

Once these constraints are fixed, multi-dimensional querying can be automatically derived to produce the consolidated data. In addition, data mappings are generally further applied to transform the retrieved spatiotemporal data structures into georeferenced multivariate time series structures [26]. These time structures can be aggregated at different granularities and DTW averaging can be applied for a more consistent analysis. Inspired by the work of Santos (2020), who studied the correlation between the demand for bicycles in Gira stations and the weather [34], correlation between time series of different modes of transport can aid into understanding multimodal synergies, by using linear correlation coefficients (e.g. Pearson's, Spearman's or Kendall's) and also detrended cross-correlation analysis for correlating time series under non-stationarity.

4.2 Multimodality Index Data Analysis

For detecting vulnerabilities associated with multimodal transportation, two major options are made available. First, the inference of multimodal origin-destination matrices. The origin-destination matrices, currently provided for Carris and Metro operators, are inferred from shared card identifiers along the public passenger transport operators. Entries in these matrices are marked with statistics, including number of cross-carrier and cross-mode commutes necessary to accomplish a complete origin-destination trip, that support the analysis of commuting susceptibilities. Second, the user can select one of the introduced indices of multimodality and compute them for different regions and time periods.

4.2.1 Challenges

To the best of my knowledge, this type of research has not yet been comprehensively attempted in the literature, where spatial multimodality is measured as a possible tool for mobility management. The choice of a suitable index for our problem is therefore a challenging task.

The first issue concerns the heterogeneity of transport modes. The different types of transport modes are land transport, airway and waterway transport, and each of these types are subdivided in further categories of transports. In the scope of this research, we are focused on land public transportation, which encompasses bus, subway, railway and cycling. Vehicles (motorized or not) necessary for transport according to the chosen mode, are intrinsically unique, with different travel speeds, used for different kinds of trips and with different frequencies.

The second issue is a consequence of the previous one, and deals with the measurement unit. Different results can be produced from considering different units, such as travel times, trip distances or number of trips. So, the best measurement unit to use, could depend on the context of the problem being solved.

The third difficulty, is the need to balance the variability of the set of transport modes and the variability in the intensities of usage.

4.2.2 Desired Properties

The properties of the inequality measures described in Section 3.2, must be adapted to the context of transportation before describing the multimodal indices. M. Diana & M. Pirra (2016) reformulate them as follows [12]:

1. *Weak Principle of Transfers*:

   Consider two travel modes whose intensities of use are I and $I\delta$, where $\delta > 0$. If the intensity of the most used mode decreases and that of the least used increases by the same quantity $I < 2\delta$ then the multimodality index should increase.

2. *Scale Independence*:

   If the frequency of use of each mode changes by the same proportion, the multimodality index should remain the same.

3. *Principle of Population*:

   The multimodality index should remain the same for any replication of the modes with their corresponding intensities of use. The choice set of the modes represents our population and 'replicating a mode' can be seen in our context simply as an increase in the population size due to the consideration of an additional number of modes with the same intensities of use of those already in the choice set.

4. *Decomposability*:

   Multimodality rankings of alternative distributions of intensities of use in the whole set of travel modes, should match the multimodality rankings of the corresponding distributions of intensities within any of the subgroups in which the whole set of travel modes can be composed.

5. *Strong Principle of Transfers*:

   Considering the following distance measure

   $$d = h(\frac{I_1}{I_{total}}) - h(\frac{I_2}{I_{total}}),$$

   for modes 1 and 2, with $I1 < I2$, where I_{total} is the sum of all intensities and h is a decreasing function defined as

   $$h(I) = \frac{(1 - I\beta)}{\beta},$$

   with β a parameter. If the intensity of the most used mode I_2 decreases and the one of the least used I_1 increases, the variation of the index depends only on the variation of d. The ratios I_i/I_{total}

are the ‘intensity shares’ of mode i; the larger the share, the more predominant is the use of that mode compared to others. The function h is introduced to decrease the distance, and therefore the effect on the index, when the modal transfer is taking place between two modes that are progressively more predominant, even if the difference in their relative intensity shares is constant.

4.2.3 Candidate Multimodality Indices

4.2.3.A Herfindahl–Hirschman index

The Herfindahl–Hirschman Index is a typical measure of market concentration and is used to determine market competitiveness, as described in Section 3.2. This measure can be adapted to the context of urban mobility, where the value of the index is closer to zero when a lot of different travel means are used and no means is very intensively used, while the value increases when the use of a smaller number of modes tends to dominate [12]. The index can be defined as follows:

$$HH = \frac{1}{n}\left[\frac{n\sum_{i=1}^{n}(f_i - \overline{f})^2}{(\sum_{i=1}^{n} f_i)^2} + 1\right]. \tag{4.1}$$

In order to distinguish between the set of available modes and the set of effectively used modes, M. Diana & M. Pirra (2016) proposed a variant of Equation (**??**)eqn:HH* which takes into account only the m elements different from zero (the effectively used modes) [12]:

$$HH_m = \frac{1}{m}\left[\frac{n\sum_{i=1}^{n}(f_i - \overline{f})^2}{(\sum_{i=1}^{n} f_i)^2} + 1\right]. \tag{4.2}$$

4.2.3.B Gini Coefficient

The Gini coefficient or Gini index, is a summary statistic of the Lorenz curve and is usually used as a measure of inequality in a population. It considers the differences among values of a frequency distribution. M. Diana & M. Pirra (2016) translated the usual formulation of the index, presented in Section 3.2, in the context of multimodality where f_i is the intensity of use of ith mode and n the total number of modes, formulated as [12]:

$$Gini = \frac{2}{n}\frac{\sum_{i=1}^{n} i \cdot f_i}{\sum_{i=1}^{n} f_i} - \frac{n+1}{n}. \tag{4.3}$$

The Gini coefficient ranges from a minimum value of zero to a maximum of one. The former one corresponds to an equal usage of all modes, while the latter refers to an infinite population of modes in which all of them except one are not used (monomodality).

4.2.3.C Theil Index

The Theil index defined in Section 3.2, can also be adapted to the context of multimodality with the lowest value of the index indicating the usage of all means with the same frequency, and the highest one, instead, is referred to a situation of only one mode used among all the n possible choices. The reformulation of the index in terms of intensities of use of all modes is expressed as [12]:

$$T = \frac{1}{n} \sum_{i} \frac{f_i}{\overline{f}} \log \frac{f_i}{\overline{f}}. \tag{4.4}$$

Diana and Mokhtarian (2008) reinterpreted the concept of information theory by considering an hypothetical mode choice experiment, where the uncertainty of the outcome is proportional to past multimodality behaviours of the traveller, thus defining a multimodality index given by [10]:

$$OM\_PI = \sum_{i=1}^{n} \left[\frac{f_i}{\sum_{j=1}^{n} f_j} log_n \left(\frac{\sum_{j=1}^{n} f_j}{f_i} \right) \right]. \tag{4.5}$$

They also suggested a variant of *OM_PI* that is sensitive to the mean mobility level of individuals. With M as the absolute maximum reported frequency of utilisation of any mode, and nM as the potential maximum total frequency across all considered modes, hence defining a mobility-level-sensitive multimodality Index, given by:

$$OM\_MI = \sum_{i=1}^{n} \left[\frac{f_i}{nM} \left[1 + ln \left(\frac{M}{f_i} \right) \right] \right]. \tag{4.6}$$

4.2.3.D Atkinson Index

The Atkinson index described in Section 3.2, for n different modes of intensities f_i and mean $\overline{f}$ can be redefined as [12]:

$$Atk = 1 - \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{f_i}{\overline{f}} \right)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}. \tag{4.7}$$

4.3 Incorporating Situational Context

The analysis of multimodality indices can be complemented with the presence of situational context. The major constituent elements of such context are the traffic generation poles. The concept of traffic generation and attraction poles generally refers to commercial areas, employment centres such as business parks and enterprises, and collective equipment like hospitals, schools and stadiums, that generate or attract a significant volume of vehicle trips, either from contributors, visitors or providers. We currently maintain a complete localization of traffic generation poles for the city of Lisbon, as well as major city events (such as large concerts, congresses and soccer matches). Figure 5.6 provides a map of the city with some poles with impact on the city traffic. The combined analysis of the traffic generation/attraction poles maps with the computed multimodality indices, as well as station-route maps, providing a comprehensive and dynamic way of modelling the spatiotemporal distribution of traffic along the city. Additionally, the surveyed indices can be revised to further measure how the volume of passengers generated and attracted by nearby poles are being currently satisfied by the co-located modes of public transport.

5

Results

Contents

5.1 The Dataset 37

5.2 Public Transport Data Analysis 40

5.3 Multimodality Indices Analysis 48

5.4 Software Tool 51

5.1 The Dataset

Three public transport modes were chosen for this study - Carris, Metro and Gira - according to the choice factors presented in Section 4.1. Carris and Metro are the two most used public transport modes and their stations can be found almost in all the Municipality of Lisbon. Gira, the biking sharing system, however, can only be found in the center axis of the city and in the neighborhood of *Parque das Nações* (Figure 5.1); and the validations during the week are fewer than the other two modes (Figure 5.2), not fulfilling all the factors in Section 4.1. The *Relevance* and *Accessibility* constraints may not be satisfied, but, compared to the other modes not included in this work, Gira traffic data was easily available and its attributes were relevant for this analysis.

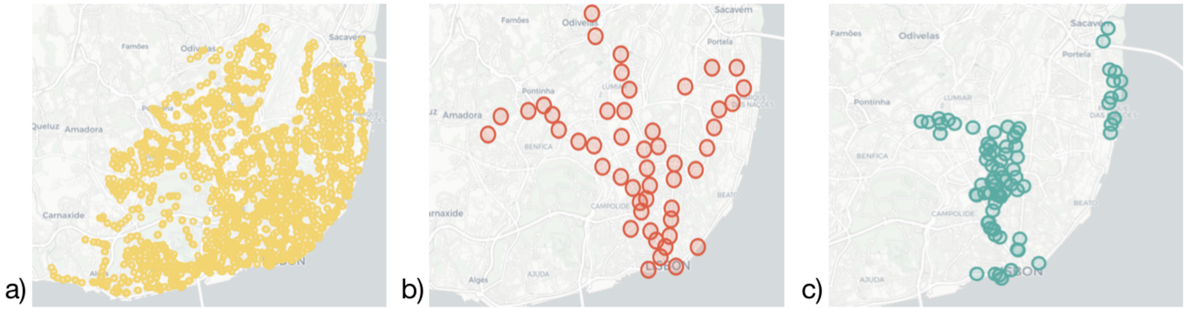

a)
Famões
Odivelas
Sacavém
Queluz
Amadora
Carnaxide
b)
Famões
Odivelas
Sacavém
Portela
Pontinha
Amadora
Benfica
Campolide
Algés
Ajuda
Lisbon
c)
Famões
Odivelas
Sacavém
Portela
Pontinha
Lumiar
Benfica
Campolide
Beato
Ajuda
Lisbon

Figure 5.1: Lisbon's stations location. a) Carris. b) Metro c) Gira.

As mentioned in Section 2.1, smart card technology (the VIVA card) was used to gather public transport traffic data. For Carris, the smart card data only monitors entries, estimators of existing validations can be used to infer the exits (not in the scope of this research). For the remaining modes of transport, Metro and Gira, we have access to both passengers' entry and exit records. The entities responsible for the modes Carris and Metro, allowed the use of their data from the month of October 2018, whereas for Gira, the records range from 13 December 2018 to 31 December 2018.

The data retrieved from Carris buses' smart card readers, is modeled as tabular data with 9865446

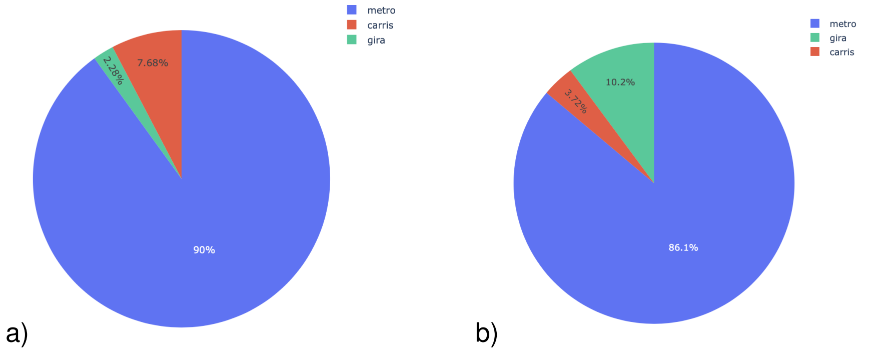
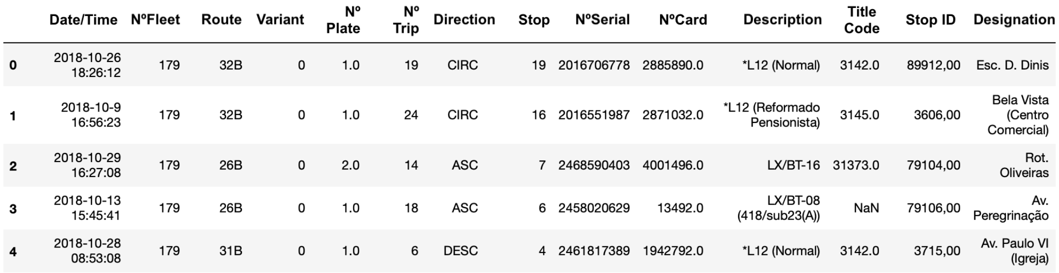


Figure 5.2: Weekly mode share distribution of TAZ nº66. a) Week days. b) Weekends.

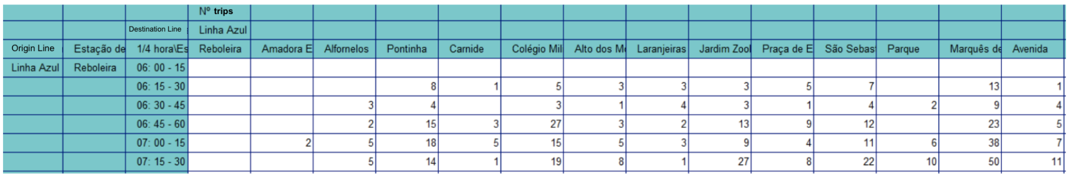
rows. Each row represents a smart card validation, which means that a passenger as entered the bus. Since the passengers can pay the bus ticket either with the smart card or with cash by buying a ticket at a Carris partner or directly to the bus driver, there is a portion of the validations that is not take into account. Yet, that small number of validations is not significant enough to influence the analysis, so that fact can be ignored. Figure 5.3 displays a sample (five rows) of the validations data gathered by the smart card readers. It has 14 attributes, 4 attributes are nominal (*Route*, *Direction*, *Description*, *Designation*), 9 are numeric (*NºFleet*, *Variant*, *NºPlate*, *NºTrip*, *Stop*, *NºSerial*, *NºCard*, *TitleCode*, *StopID*), and 1 is ordinal (*Date/Time*). But only *Date/Time* and *StopId* are relevant for this research. The attribute *Data/Hora* represents the date and the hour of a bus entry at the station identified by the attribute *Date/Time*, neither one of this columns contains missing or inconsistent values.



	Date/Time	**NºFleet**	**Route**	**Variant**	**Nº Plate**	**Nº Trip**	**Direction**	**Stop**	**NºSerial**	**NºCard**	**Description**	**Title Code**	**Stop ID**	**Designation**
0	2018-10-26 18:26:12	179	32B	0	1.0	19	CIRC	19	2016706778	2885890.0	*L12 (Normal)	3142.0	89912,00	Esc. D. Dinis
1	2018-10-9 16:56:23	179	32B	0	1.0	24	CIRC	16	2016551987	2871032.0	*L12 (Reformado Pensionista)	3145.0	3606,00	Bela Vista (Centro Comercial)
2	2018-10-29 16:27:08	179	26B	0	2.0	14	ASC	7	2468590403	4001496.0	LX/BT-16	31373.0	79104,00	Rot. Oliveiras
3	2018-10-13 15:45:41	179	26B	0	1.0	18	ASC	6	2458020629	13492.0	LX/BT-08 (418/sub23(A))	NaN	79106,00	Av. Peregrinação
4	2018-10-28 08:53:08	179	31B	0	1.0	6	DESC	4	2461817389	1942792.0	*L12 (Normal)	3142.0	3715,00	Av. Paulo VI (Igreja)

Figure 5.3: Sample of Carris smart card's validations data.

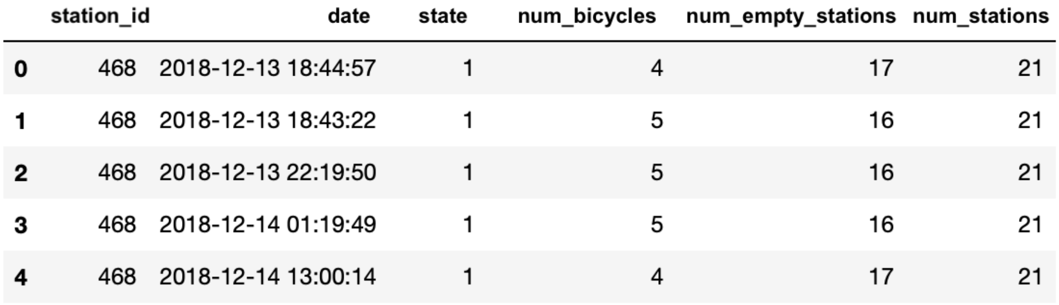
Metro data presents its data in the form of an origin-destination matrix (Figure 5.4). Contrary to Carris, Metro subway is equipped with two smart card readers, one at the boarding and the other at the alighting, thus, the data retrieved is organized in two matrix (entries and exits). The rows and columns are labeled with the Metro stations, and each cell has a numerical value describing the number of validations (entries or exits) from those stations, in a time range of 15 minutes, and that is the reason why Some cells contain missing values which will be filled with 0's.



			Nº trips													
		Destination Line	Linha Azul													
Origin Line	Estação de	1/4 hora\Es	Reboleira	Amadora E	Alfornelos	Pontinha	Carnide	Colégio Mil	Alto dos M	Laranjeiras	Jardim Zool	Praça de E	São Sebas	Parque	Marquês de	Avenida
Linha Azul	Reboleira	06: 00 - 15														
		06: 15 - 30				8	1	5	3	3	3	5	7		13	1
		06: 30 - 45			3	4		3	1	4	3	1	4	2	9	4
		06: 45 - 60			2	15	3	27	3	2	13	9	12		23	5
		07: 00 - 15		2	5	18	5	15	5	3	9	4	11	6	38	7
		07: 15 - 30			5	14	1	19	8	1	27	8	22	10	50	11

Figure 5.4: Sample of Metro smart card's validations data (entries).

For the Gira bike sharing system (for more information about this system see Section 2.1), the collected data is modeled by tabular data with a total of 88747 rows, where its columns have 6 attributes, 4 are numeric (*station_id*, *num_bicycles*, *num_empty_stations*, *num_stations*), 1 boolean (*state*), and 1 is ordinal (*date*) (Figure 5.5). The bike stations update the number of available bikes (*num_bicycles*) and the number of empty bike slots (*num_empty_stations*), in their records, every time a bike has been picked-up or dropped-off.



	station_id	date	state	num_bicycles	num_empty_stations	num_stations
0	468	2018-12-13 18:44:57	1	4	17	21
1	468	2018-12-13 18:43:22	1	5	16	21
2	468	2018-12-13 22:19:50	1	5	16	21
3	468	2018-12-14 01:19:49	1	5	16	21
4	468	2018-12-14 13:00:14	1	4	17	21

Figure 5.5: Sample of Gira stations' data.

Additionally to the information about Carris, Metro and Gira, other types of urban data are available. We have access to the localization of the stations of every mode, in geographic coordinates (Figure 5.1), but also the localization of traffic generation poles for the city of Lisbon (i.e. commercial areas, enterprises, hospitals, schools and stadiums), as well as major city events such as large concerts, congresses and soccer matches.

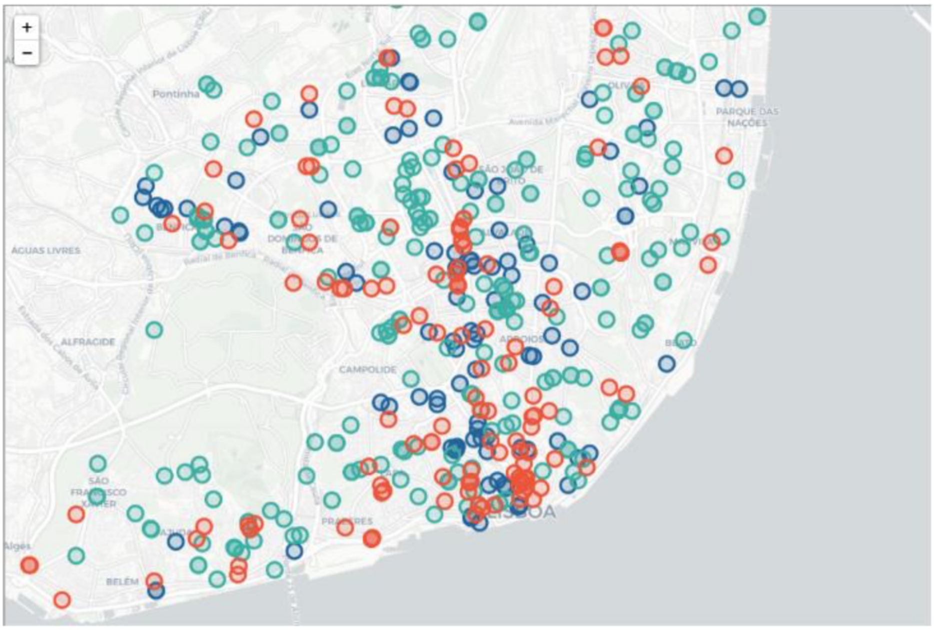


Figure 5.6: Major traffic generation poles: commercial (blue), schools and institutes (green), and health centres (red).

5.2 Public Transport Data Analysis

This section models the traffic demand and explores the spatiotemporal content of the available traffic data, taking into consideration user-specific commutes in interface areas. The task of analysing public transport data is divided in three parts. First, the depiction of public transportation demand over time, through the analysis of time series. Second, we will see more accurately how temporal data from public transports behaves over time, using DTW time series averaging. And finally to see the correlation between the three modes of transport with three different correlation coefficients.

5.2.1 Spatial Granularity

The assessment of temporal data requires beforehand the specification of a spatial granularity. Among all the possibilities of spatial criteria defined in Section 4.1.2, we opted by choosing the Traffic Analysis Zones (TAZ) and the parishes of Lisbon. The TAZ are not administrative divisions, these form of spatial modelling is derived from trip generation densities processed by delineation algorithms that use the peaks of densities as the centre of a zone [27]. Figure 5.7 illustrates all the 103 TAZs of the Municipality of Lisbon, where their designation are indexed to their respective number on Table A.1. Only eleven TAZs can be considered for analysis since these are the only TAZs that enclose stations from all three chosen modes of transport (Figure 5.9). The parishes of Lisbon (called *freguesias* in Portuguese), contrary to the TAZs, have a coarser granularity, and is a territorial unit delineated by a Municipality or a State entity. The parishes of Lisbon are displayed in Figure 5.8 and their names are indexed in Table A.2. Like the TAZs, not all the parishes contain all three types of public transport stations, only ten (Figure 5.10). Multimodal pattern analysis using parishes as spatial granularity was not described in this dissertation (only TAZs), however, it is possible to use parishes as the geographic unit in the ILU software tool.

The location of the TAZs and parishes that contain stations of the three modes, Carris, Metro and Gira, reveals that the intermodal interfaces are located in the central axis of the city, and although there are other modes of transport besides those we are analyzing here, this demonstrates the urban planning priority given to the city's commercial zone and the high-density residential zone, and the lack of options in terms of public transport for the population in the medium and low-density resident zone. But these facts are already predictable considering the market share of each mode (Figure 5.11), which is highly irregular.

5.2.2 Temporal Series

Among all the Traffic Analysis Zones, the TAZ of *Avenidas Novas (Avenidas Novas — Este)* (TAZ nº66), was the chosen geographical zone for analysis. There's no particular reason that we chose that TAZ.

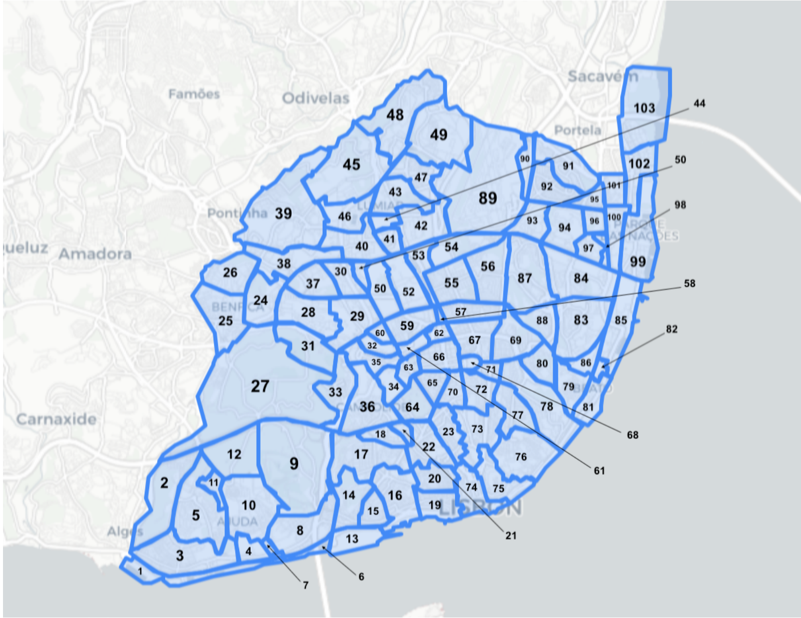

Sacavém
Famões
Odivelas
Portela
Pontinha
Queluz
Amadora
Benfica
Carnaxide
Algés
Ajuda
Lumiar
Parque das Nações
Lisbon
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103

Figure 5.7: TAZs of the Municipality of Lisbon.

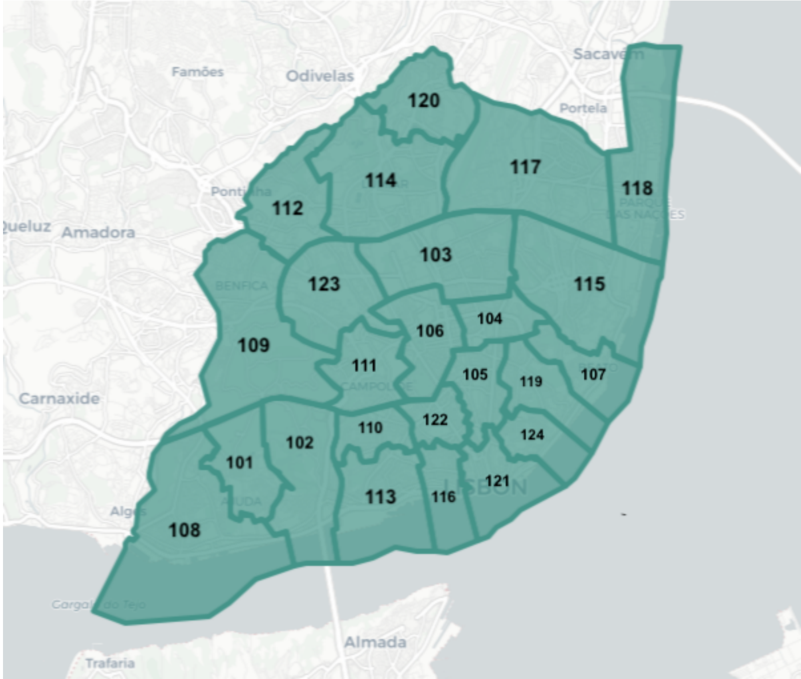

Sacavém
Famões
Odivelas
Portela
Pontinha
Queluz
Amadora
Benfica
Carnaxide
Algés
Ajuda
Campolide
Lisbon
Almada
Trafaria
120
114
117
118
112
103
123
115
109
106
104
111
105
119
107
122
110
124
101
102
113
116
121
108

Figure 5.8: Parishes of the Municipality of Lisbon.

However, its a zone enclosing stations from all three public modes of transport under study; and secondly, its a zone adjacent to *Saldanha*, which is an influential multimodal interface encompassing multiple modes of transport and characterized by the presence of business and cultural traffic generation poles.

Since we are studying the behavior of urban multimodal demand, only the check-ins are considered

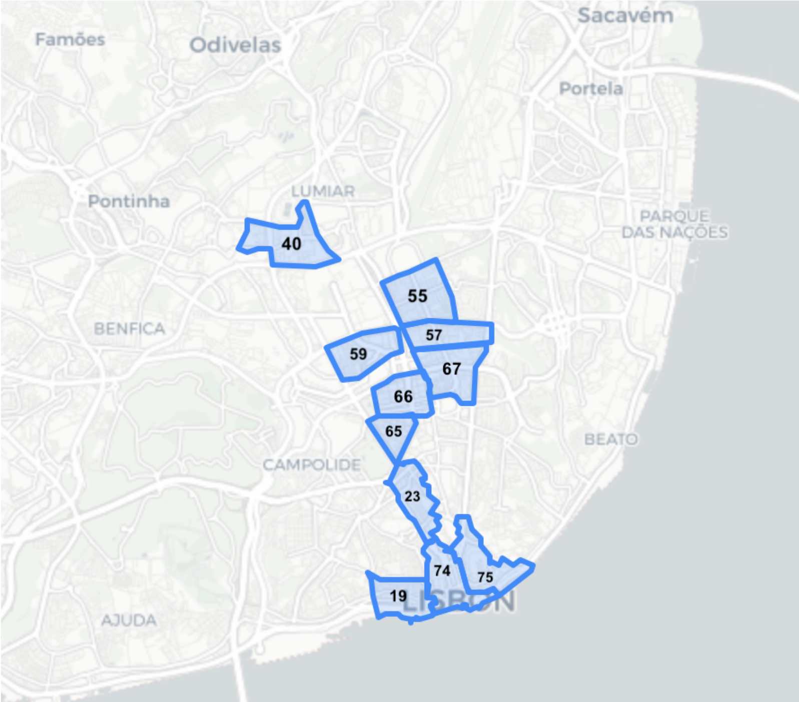

Famões
Odivelas
Sacavém
Portela
Pontinha
LUMIAR
PARQUE DAS NAÇÕES
BENFICA
CAMPOLIDE
BEATO
AJUDA
LISBON
40
55
57
59
67
66
65
23
74
75
19

Figure 5.9: TAZs with three modes of transport (Metro, Carris, Gira).

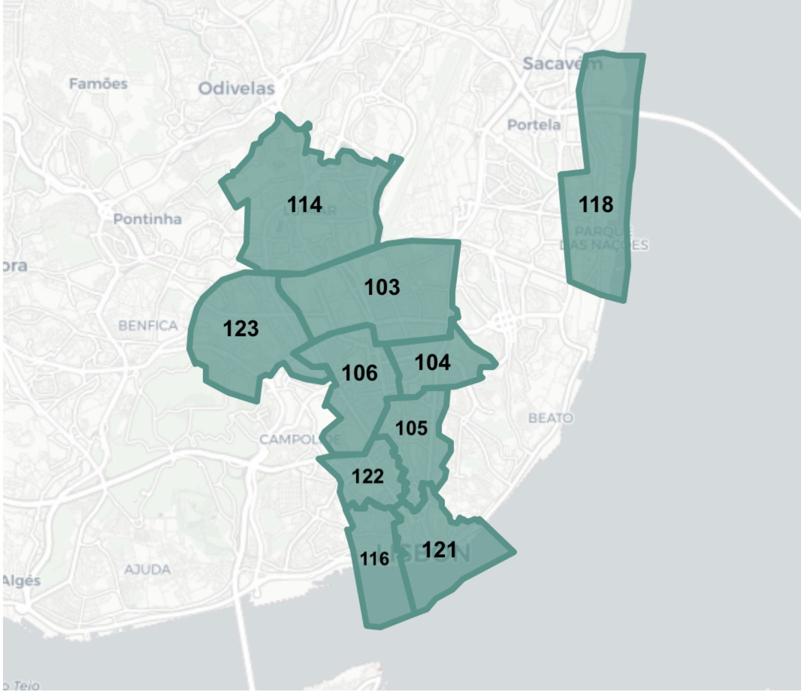

Famões
Odivelas
Sacavém
Portela
Pontinha
BENFICA
CAMPOLIDE
BEATO
AJUDA
Algés
114
103
123
106
104
105
122
116
121
118

Figure 5.10: Parishes with three modes of transport (Metro, Carris, Gira).

(in the case of Metro and Gira, cause Carris only validates entries). Figure 5.12 shows the volume and variation of validations in TAZ nº66 during a week. In terms of volume, Metro has roughly ten times

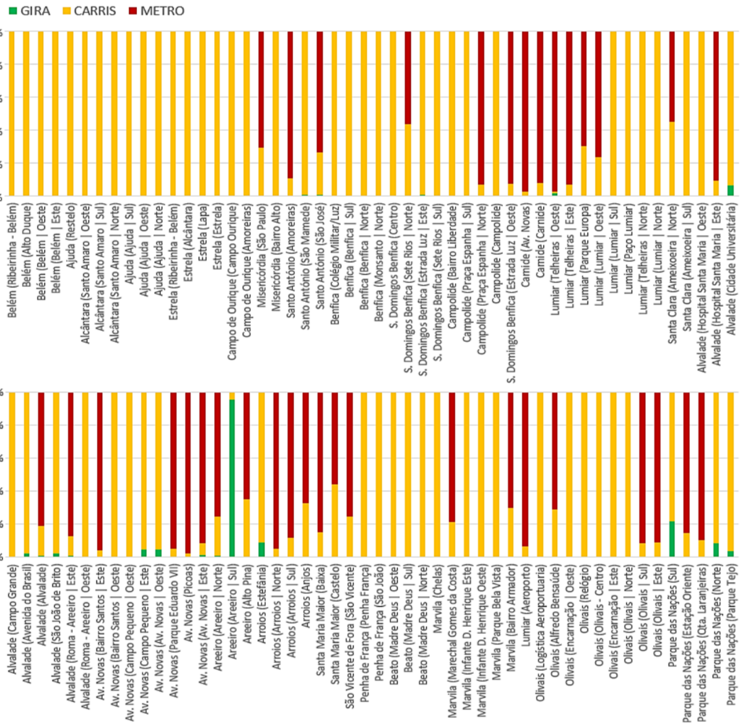


Figure 5.11: Cycling-bus-subway market quota (modal trip share) for major traffic analysis zones (TAZ) in Lisbon.

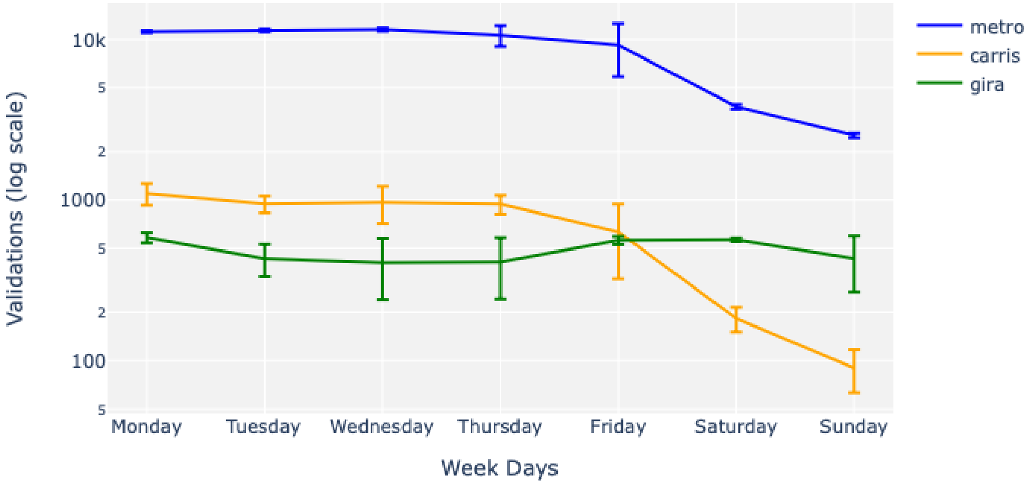
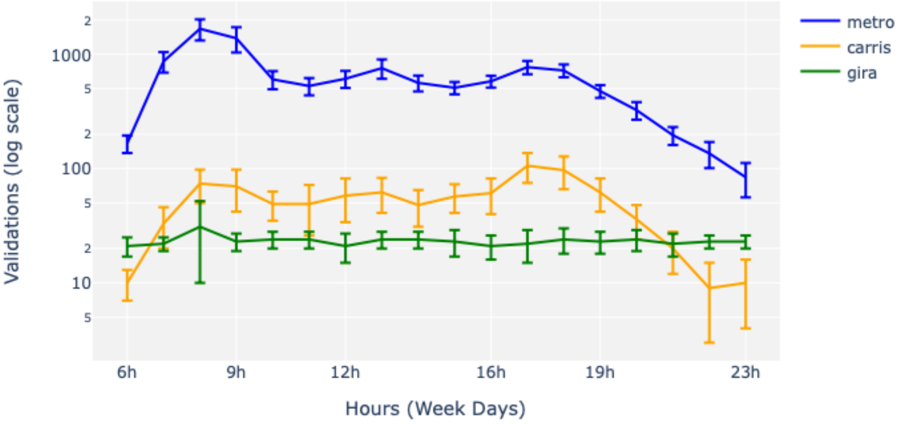


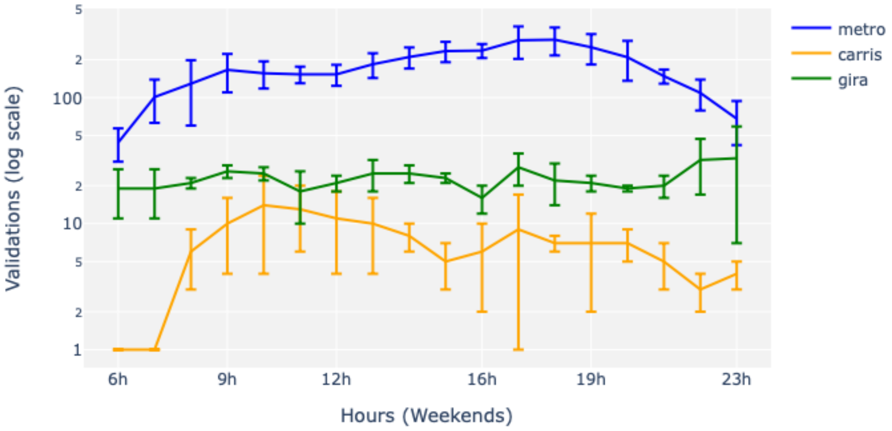
Figure 5.12: Weekly volume and variation of validations in TAZ nº66.

more validations than the other two modes. To better respond to the skewness towards large values, we applied a logarithmic scale for the Y axis. Metro and Carris have a similar behavior during the week, however the dispersion of values for Metro is minimal, where for Carris, the dispersion of values is slightly more substantial. The observations during the week for these two modes are approximately constant during the week, and then at Friday it starts to decrease till the end of the weekend. Gira demand behavior is relatively the opposite: its decreasing during the week days, and it increases during the weekend. The standard deviation for Gira during the weekend is low, where during the week days, the dispersion of values is significant. This difference between Gira and the other two modes, can be explained by the fact that bicycles are mostly used as leisure and sport transport, and not so much as transport for commuting (home-work), like Carris and Metro, which can also explain the high dispersion of values during the week. Another cause could be he fact that the bicycle is more accessible to the younger population, as it's a human-powered mode of transport, while the Carris and the Metro are used by people of all ages.

It's also interesting to analyse the urban multimodal demand at a finer granularity of time. Figure 5.12 presents the volume and variation in TAZ nº66 for every hour during week days and weekends. At week days, Metro and Carris have approximately the same trend, with two peaks during the day, one at around 8am and the other ate the end of the day (around 5pm). Those volume peaks correspond to commuting travels, resulting in the morning and evening rush hours. The variability is practically the same throughout a week day for these two modes of transport. Again, Gira behaves differently from Carris and Metro, with a regular trend through the hours of a week day. A high standard deviation is present at around 8am for Gira, which can also be explained by the morning commute, where people not always use their car to go to work and opt for a more healthy alternative like the bicycle. At weekends, Gira volume is always higher than Carris at all hours, which was already perceived in Figure 5.12. The standard deviation is highly irregular for Carris, but more steady for the subway mode. This irregularities can be explained by the fact that during weekends there is not a pattern as in commuting travels (most people don't work in weekends), and the public transport modes are greatly affected by cultural events (which usually take place during weekends), and since the stations of Carris are very accessible and scatter in almost every street of Lisbon, it makes the preferred mode to attend this type of events.



a)



b)

Figure 5.13: Daily volume and variation of validations in TAZ nº66. a) Week days. b) Weekends.

5.2.3 Barycenter Averaging

The data used in the analysis carried out in Section 5.2.2 was processed using the mean of validations per station, causing to be sensitive to noisy data [31]. A more accurate procedure is applied in this section, where euclidean averaging and DTW barycenter averaging (DBA) is applied to the sequences of each mode. These sequences or time series are built from each week of the available date range. For example, Carris data is available for the month of October (2018), so the DBA is computed from four sequences (i.e. a cluster) corresponding to the validations of the weeks of that month. Figure 5.14 shows the weekly volume of validations through barycenter averaging in TAZ nº66, where there are no significant differences in trend of the series, compared to the previous analysed plot in Figure 5.12. However it's interesting to notice that in the days where there is a high variability of the validations, the euclidean sequence is affected by it and deviates from the DBA sequence. Thus, it enables to visualize that the difference between the demand in the last day of the week (Friday) and the first day of the weekend (Saturday) is much more distinct with this type of plots.

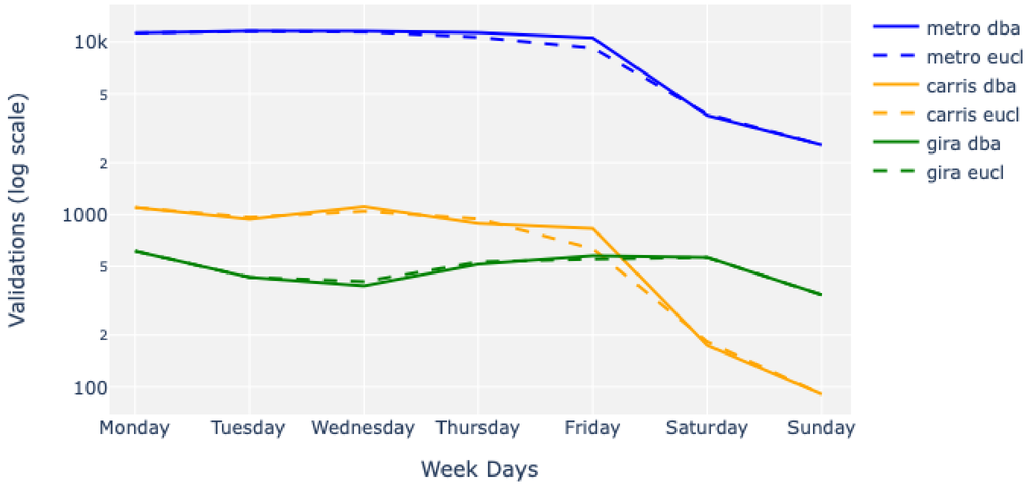
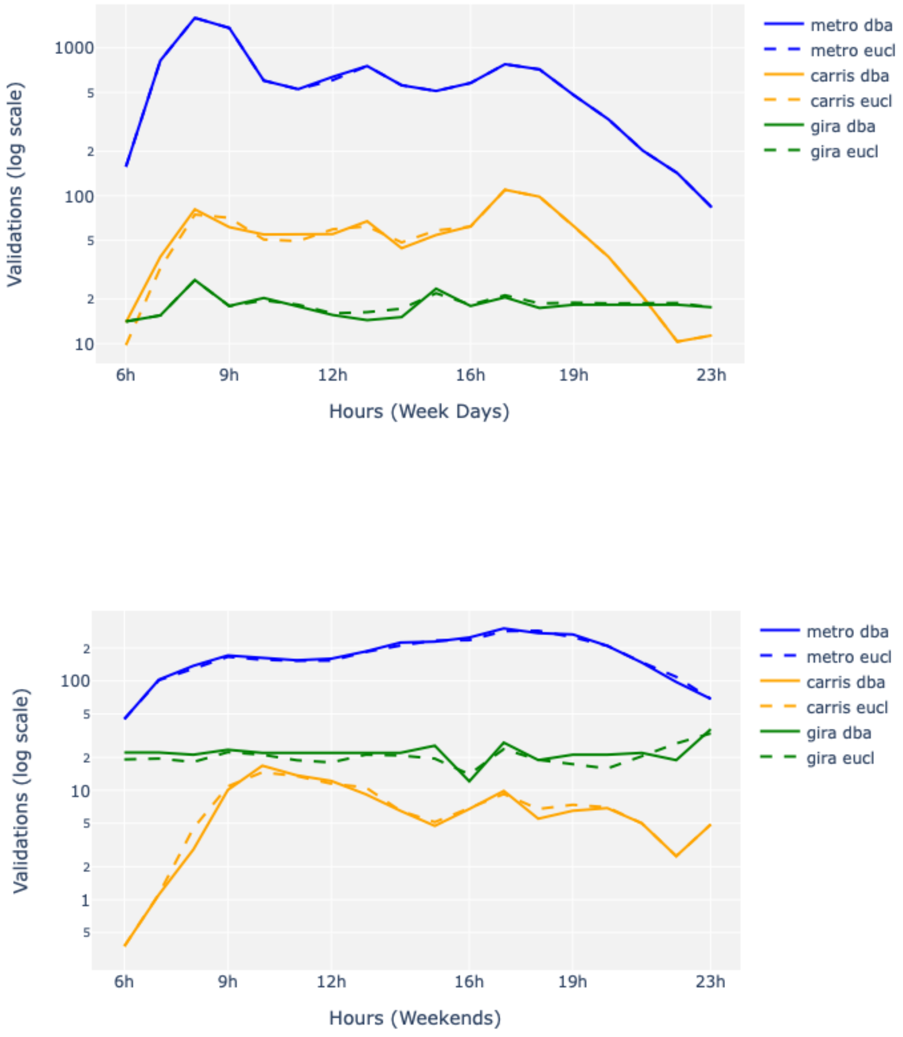


Figure 5.14: Barycenter averaging (DTW and Euclidean) for weekly volume of validations in TAZ nº66.

The daily volume of validations through barycenter averaging is described in Figure 5.15, and now the Gira sequence doesn't seem to behave constantly, as in Figure 5.13. For week days, we now can see that there is a peak at 8am, and another two peaks at around 3pm and 5pm. The 8am and 5pm peaks correspond to commute trips, also present in Metro and Carris. The 3pm Gira peak is not perceived in the other two modes, and may be caused by other currently unknown factors, probably some cultural event, or due to the fact that the Gira data analysed here is from December, so holiday season could influence those results. There is no major difference between the weekend hours series using DBA (Figure 5.15 b) and the aggregation series from Figure 5.13 b). What is interesting to notice here, is the difference between the euclidean barycenter and the DBA of the Gira sequence. At approximately 8pm and 10pm, the euclidean and the DBA are symmetric (when one is positive, the other is negative). Once again, this is explained by the dispersion of the values and the sensibility of the euclidean measure. This justifies the use of DBA in this type of analysis.



a)

b)

Figure 5.15: Barycenter averaging (DTW and Euclidean) for daily volume of validations in TAZ nº66 (logarithmic scale). a) Week days. b) Weekends.

5.2.4 Correlations

The final task of the public transport data analysis is to correlate the data of the available transport modes. Figures 5.16, 5.17 and 5.18 show respectively, the Pearson's, Spearman's and DXA correlation coefficients between the different modes weekly.[1] There is not a major difference between the used coefficients, the correlation varies in the same way between modes. Already perceived with the previous analysed line plots, but now we can confirm: Carris entries and Metro exits validations are highly correlated, whereas those volumes are negatively correlated with those from Gira.

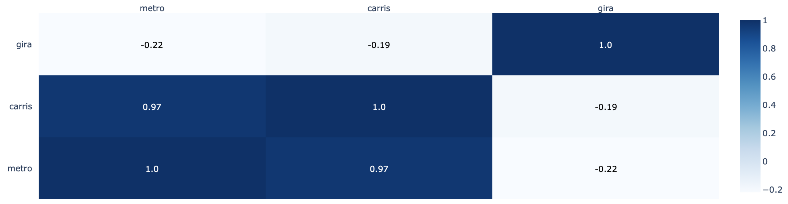


Figure 5.16: Weekly Pearson correlation heatmap between modes of TAZ nº66.

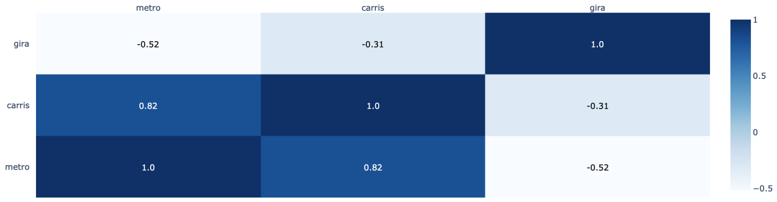


Figure 5.17: Weekly Spearman correlation heatmap between modes of TAZ nº66.

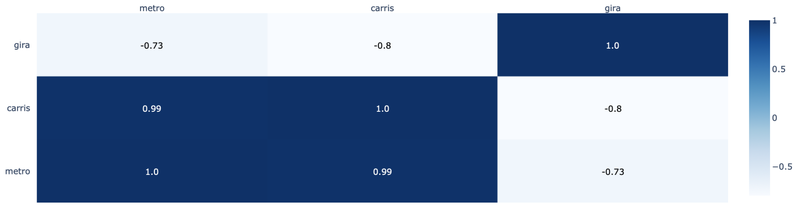


Figure 5.18: Weekly DXA correlation heatmap between modes of TAZ nº66.

5.3 Multimodality Indices Analysis

The previous section analysed multimodality in one zone. This section will now assess multimodality among all available geographical units (TAZs), in order to have a global vision of multimodality at the spatial level of the city of Lisbon. Among all candidate multimodality indices presented in Section 4.2.3, the Gini coefficient and the Herfindahl–Hirschman index were the measures selected for this analysis, because they are simple to implement and to adapt to the context of transportation and they differ in their properties but range between the same values. Figure 5.19 displays four TAZ maps of Lisbon coloured with the values of the Gini index (green is 0 corresponding to multimodality and 1 is red corresponding to monomodality) at different hours. There is not a major change in the index values among these hours. The TAZs with a higher degree of multimodality correspond mostly to the TAZs containing all the three modes (bus, subway and cycling) (see Figure 5.9) and encompass a large number of traffic generation poles (see Figure 5.6). Still there are TAZs containing all selected modes, but they have a medium index value (around 0.5); this is due to the fact that the Gini index is sensible to the intensity of usage of the modes (scale dependence, see Section 4.2.2). The results for the Herfindahl–Hirschman index (HH) are similar to the ones of the Gini index, with the exception that the HH index is highly sensible to the number of used modes (principle of population, see Section 4.2.2), that justifies the color red over almost all TAZs, since most of the TAZs out of the center only have two or one mode of transport (only including bus, subway and cycling).

---

[1] Results were computed for daily correlations, but they were very similar to the weekly correlations, so they were not included.

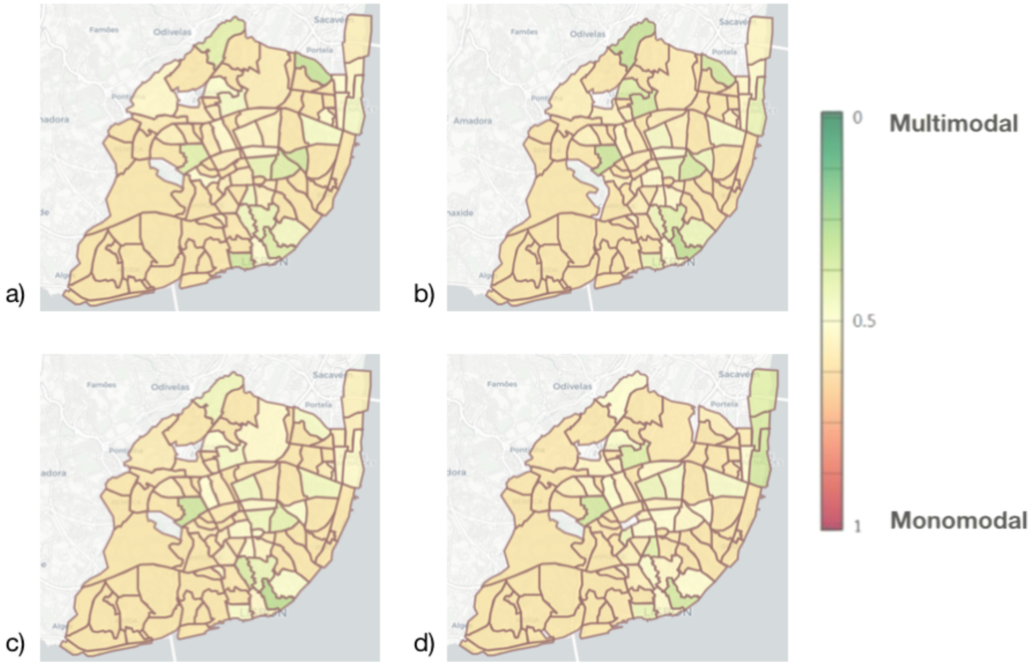


Figure 5.19: Gini index TAZ map (week days). a) 8h. b) 12h. c) 17h. d) 21h.

The variation of both indices was plotted throughout the week (Figures 5.21 and 5.22). And for both indices, the variation is almost identical. Until the middle of the week (Wednesday) the indices values rise, then start to decrease until Saturday, and at Sunday they slightly increase. The standard deviation was calculated for both plots, however, the variability was too large to be included in the plots, which

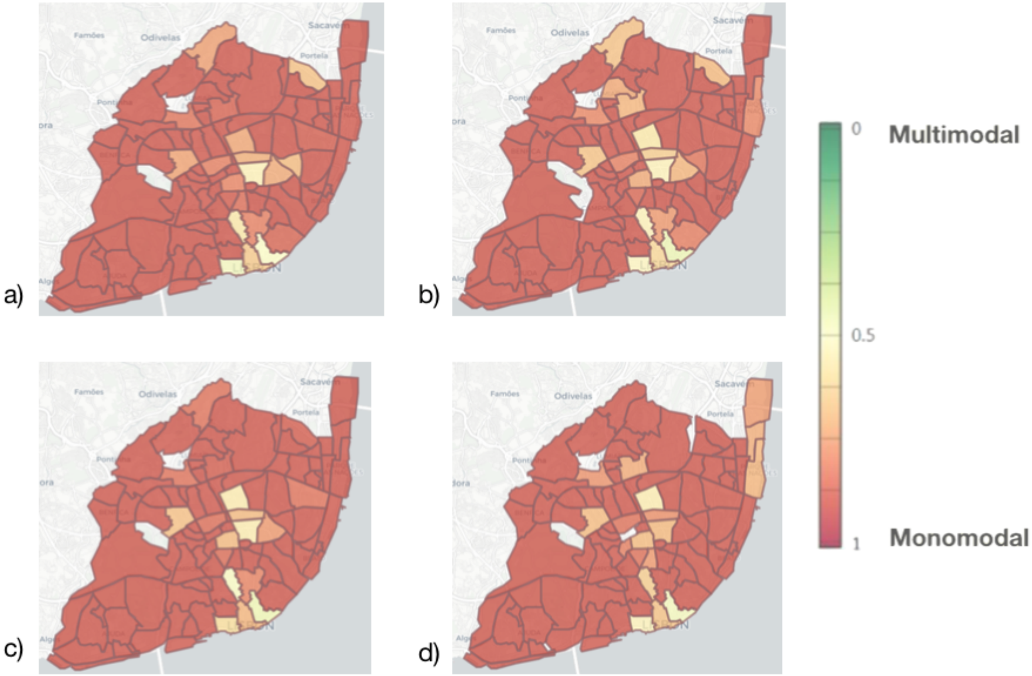

a)
b)
c)
d)
0
0.5
1
Multimodal
Monomodal
Famões
Odivelas
Sacavém
Portela

Figure 5.20: HH index TAZ map (week days). a) 8h. b) 12h. c) 17h. d) 21h.

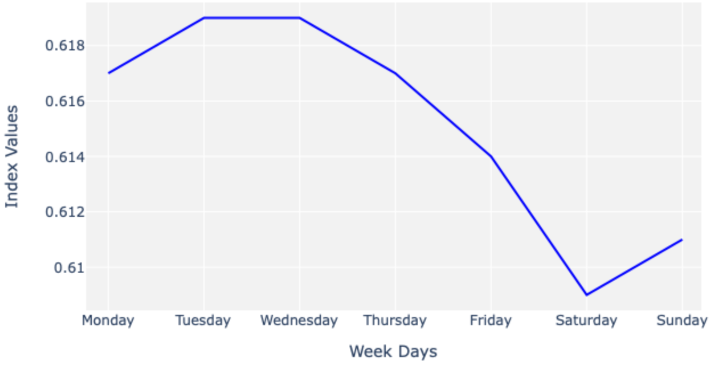


Figure 5.21: Weekly Gini index lineplot.

suggest that the weekly variation of the indices may not be significant enough to be considered in the analysis. This means that if the Y axis had coarser value ticks, both plots would look like a straight line.

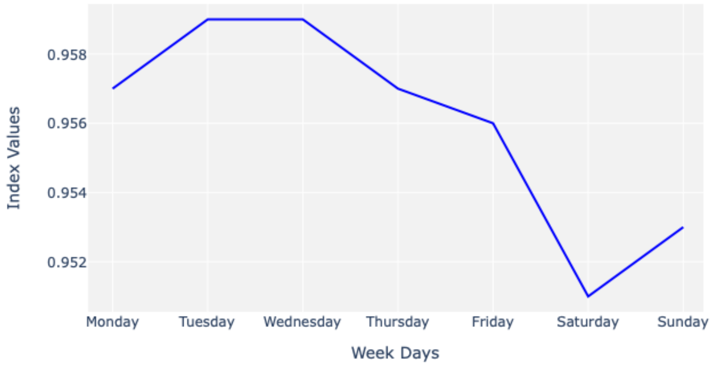


Figure 5.22: Weekly HH index lineplot.

5.4 Software Tool

The ILU project contributions are integrated in the ILU App web platform. The application was developed in Python with Dash, which is a framework for building machine learning and data science web applications, powered by Plotly[2]. Figure 5.23 presents the home page of the application. Each colourful option is associated with the respective task of the project ILU.

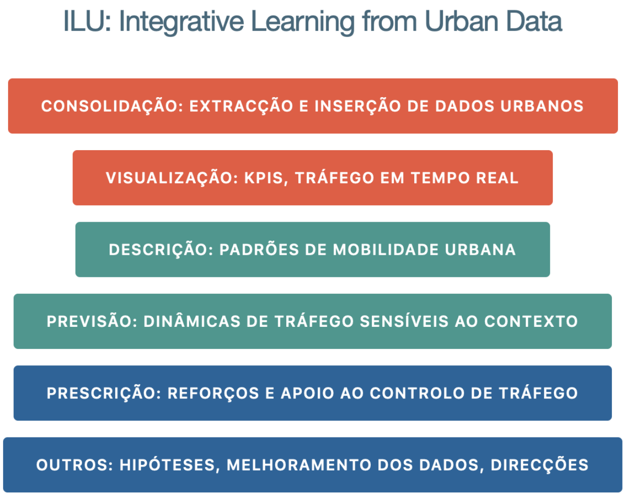
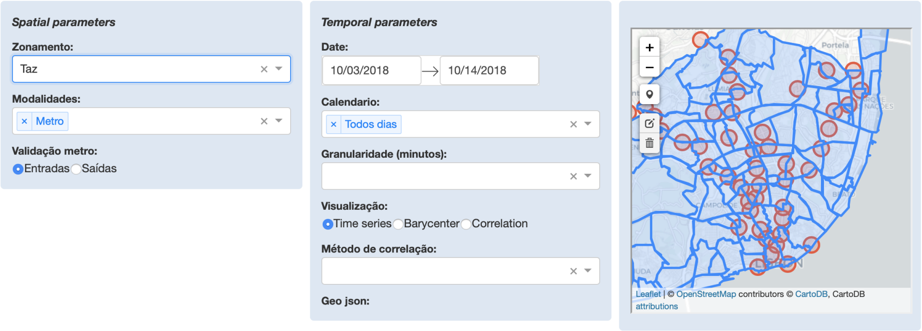

ILU: Integrative Learning from Urban Data
CONSOLIDAÇÃO: EXTRACÇÃO E INSERÇÃO DE DADOS URBANOS
VISUALIZAÇÃO: KPIS, TRÁFEGO EM TEMPO REAL
DESCRIÇÃO: PADRÕES DE MOBILIDADE URBANA
PREVISÃO: DINÂMICAS DE TRÁFEGO SENSÍVEIS AO CONTEXTO
PRESCRIÇÃO: REFORÇOS E APOIO AO CONTROLO DE TRÁFEGO
OUTROS: HIPÓTESES, MELHORAMENTO DOS DADOS, DIRECÇÕES

Figure 5.23: Home page of the ILU App.

This work contributes to the task *Descrição: Padrões de Mobilidade Urbana*. It contains two pages, one for the multimodal pattern analysis and other for the assessment of the multimodal indices. Figure

---

[2] Plotly: `https://plotly.com/dash/`.



Spatial parameters

Zonamento: Taz

Modalidades: Metro

Validação metro: ● Entradas ○ Saídas

Temporal parameters

Date: 10/03/2018 → 10/14/2018

Calendario: Todos dias

Granularidade (minutos):

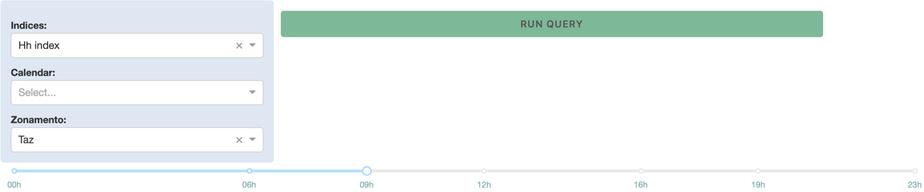
Visualização: ● Time series ○ Barycenter ○ Correlation

Método de correlação:

Geo json:

+
−
Portela
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions

Figure 5.24: Multimodal Patterns Analysis Menu (ILU App).



Indices: Hh index

Calendar: Select...

Zonamento: Taz

Figure 5.25: Multimodal Indices Analysis Menu (ILU App).

5.24 displays the interface of the Multimodal Pattern Analysis page, where the user can specify spatial parameters, such as the type of zone, the modes of transport, and in case the user chooses Metro, he can specify if he wants the validations from the exits or the entries. Next, the user can choose some temporal parameters, including the date range, the type of days (e.g. weekends), the granularity of the results (in minutes), the manner the results will be presented (time series, barycenter averaging or correlation heatmaps) and in case the user chooses to visualize correlation, he can choose which measure to be computed. After all the parameters are set, the user can select from the mini map the zone to be processed. In the Multimodal Indices page interface, presented in Figure 5.25, the user can choose the multimodal index, the type of days (e.g. week days) and the zoning method. The user can also select from the slider in the bottom a specific hour, and the page will present a map of Lisbon similar to the ones in Figure 5.19.

6

Conclusion

Contents

6.1 Concluding Remarks 55
6.2 Community and Scientific Acceptance 55
6.3 Future Work 56

6.1 Concluding Remarks

This Thesis offered a structured view of the multimodal synergies from heterogeneous sources of urban data. To the best of our knowledge, the literature only focused on the measurement of multimodality at the individual level, in other words, it studied to what degree people (or public transports passengers) were multimodal. So this thesis aimed at studying multimodality at the spatial level, in order to aid specialists in the field of transportation engineering and urban planning. The proposed methodology was applied to the city of Lisbon using three public modes of transport: Carris (bus), Metro (subway) and Gira (cycling). The results demonstrated that the cycling mode behaves differently than the bus and the subway mode, and the population prefers more environment friendly and healthy transport modes during the weekends (bicycles) than more traditional modes like the bus. The multimodality across the city was measured using two inequality measures: the Gini and the Herfindahl–Hirschman indices. Those measurements demonstrated that the commercial center of the city, rich in traffic generation and attraction poles, was more multimodal than the outer residential zones.

The inherent sensitivity of the properties of the used indices, is greatly influenced by the number of modes and their intensity, which could be attenuated if the other public transport operators shared their traffic data, providing the possibility to analyse more transport modes and to obtain a more extensive and complete view of the multimodal patterns of the city. The data received from Carris, Metro and Gira was also incomplete, relying only on information from the months of October and December 2018.

6.2 Community and Scientific Acceptance

The research pursued for this dissertation is being subject to international reviews through the submission of work to peer-review international Conferences in the field (the submission of articles is linked to Journal publications). The article proposal - “Boosting Multimodality Mobility Decisions using Big Data in the City of Lisbon: ongoing and future challenges”, was accepted by the scientific committee of the 14th Conference on Transport Engineering (CIT 2020) R-Evolution in Transport (this conference will be held in 2021 due to the current pandemic caused by COVID-19).[1] Additionally, the article – “Exploring multimodal mobility patterns with big data in the city of Lisbon”, was submitted to the scientific committee of the 48th European Transport Conference (ETC 2020) held online last September [23].[2] This latter article was presented by the former author at the ETC 2020 *Young Researchers and Practioners Forum* at the session *Mobility* and it received a very positive feedback from the international scientific community. It was also submitted to the European Transport Research Review Journal linked to the ETC 2020.

---

[1] `https://www3.ubu.es/cit2021/en/`

[2] `https://aetransport.org`

6.3 Future Work

This research presented an innovative method to evaluate traffic patterns, through the analysis of spatial multimodality data. The Traffic Analysis Zones were chosen as the geographical unit under study but other spatial granularities could equally be suggested. With more data, other types of temporal granularities could also be added to the research such as months and years. And, the use of the other mentioned inequality measures could provide more insights on the degree of spatial multimodality in the city. The multimodal data can also be correlated with other types of situational context, such as weather or public events (e.g. concerts), in order to evaluate the factors that influence variations in multimodality on the city. And finally, this study was applied to the city of Lisbon but would be equally interesting to explore multimodality behaviors in other cities.

Bibliography

[1] AABERGE, R. On the problem of measuring inequality. Tech. rep., Discussion Papers, 1986.

[2] ATKINSON, A. B., ET AL. On the measurement of inequality. *Journal of economic theory 2*, 3 (1970), 244–263.

[3] ATLURI, G., KARPATNE, A., AND KUMAR, V. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR) 51*, 4 (2018), 1–41.

[4] BREZINA, I., PEKÁR, J., ČIČKOVÁ, Z., AND REIFF, M. Herfindahl–hirschman index level of concentration values modification and analysis of their change. *Central European journal of operations research 24*, 1 (2016), 49–72.

[5] BROCKWELL, P. J., DAVIS, R. A., AND CALDER, M. V. *Introduction to time series and forecasting*, vol. 2. Springer, 2002.

[6] BUEHLER, R., AND HAMRE, A. An examination of recent trends in multimodal travel behavior among american motorists. *International journal of sustainable transportation 10*, 4 (2016), 354–364.

[7] CHLOND, B. Making people independent from the car–multimodality as a strategic concept to reduce co 2-emissions. In *Cars and Carbon*. Springer, 2012, pp. 269–293.

[8] COWELL, F. *Measuring inequality*. Oxford University Press, 2011.

[9] DEARDORFF, A. V. *Terms of trade: glossary of international economics*. World Scientific, 2014.

[10] DIANA, M., AND MOKHTARIAN, P. L. Travelers' segmentation based on multimodality behaviors and attitudes.

[11] DIANA, M., AND MOKHTARIAN, P. L. Desire to change one's multimodality and its relationship to the use of different transport means. *Transportation research part F: traffic psychology and behaviour 12*, 2 (2009), 107–119.

[12] DIANA, M., AND PIRRA, M. A comparative assessment of synthetic indices to measure multimodality behaviours. *Transportmetrica A: Transport Science 12*, 9 (2016), 771–793.

[13] ESLING, P., AND AGON, C. Time-series data mining. *ACM Computing Surveys (CSUR) 45*, 1 (2012), 1–34.

[14] GINI, C. Variabilità e mutabilità (variability and mutability). *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955) ed. Bologna* (1912).

[15] HAN, J., PEI, J., AND KAMBER, M. *Data mining: concepts and techniques*. Elsevier, 2011.

[16] HEINEN, E. Are multimodals more likely to change their travel behaviour? a cross-sectional analysis to explore the theoretical link between multimodality and the intention to change mode choice. *Transportation research part F: traffic psychology and behaviour 56* (2018), 200–214.

[17] HERFINDAHL, O. C. *Concentration in the steel industry*. PhD thesis, Columbia University New York, 1950.

[18] HOOVER, E. M. Interstate redistribution of population, 1850-1940. *The Journal of Economic History 1*, 2 (1941), 199–205.

[19] IMTT. *Guia para a elaboração de planos de mobilidade de empresas e pólos*.

[20] JANG, W. Travel time and transfer analysis using transit smart card data. *Transportation research record 2144*, 1 (2010), 142–149.

[21] KHURSHID, S., SINGH, R., AND SINGH, G. Levels and trends of competition among mutual funds in india. *Research Journal of Business Management 3*, 2 (2009), 47–67.

[22] KUHNIMHOF, T., CHLOND, B., AND VON DER RUHREN, S. Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. *Transportation research record 1985*, 1 (2006), 40–48.

[23] LEMONDE C, ARSÉNIO E, H. R. Exploring multimodal mobility patterns with big data in the city of lisbon. *European Transport Research Review* (2020).

[24] LIU, R., PENDYALA, R. M., AND POLZIN, S. Assessment of intermodal transfer penalties using stated preference data. *Transportation research record 1607*, 1 (1997), 74–80.

[25] LORENZ, M. O. Methods of measuring the concentration of wealth. *Publications of the American statistical association 9*, 70 (1905), 209–219.

[26] Mamoulis, N., Cao, H., Kollios, G., Hadjieleftheriou, M., Tao, Y., and Cheung, D. W. Mining, indexing, and querying historical spatiotemporal data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), pp. 236–245.

[27] Martínez, L. M., Viegas, J. M., and Silva, E. A. A traffic analysis zone definition: a new methodology and algorithm. *Transportation 36*, 5 (2009), 581–599.

[28] Murphy, K. P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[29] Nobis, C. Multimodality: facets and causes of sustainable mobility behavior. *Transportation Research Record 2010*, 1 (2007), 35–44.

[30] Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y., and Keogh, E. Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm. *Knowledge and Information Systems 47*, 1 (2016), 1–26.

[31] Petitjean, F., Ketterlin, A., and Gançarski, P. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition 44*, 3 (2011), 678–693.

[32] Podobnik, B., and Stanley, H. E. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters 100*, 8 (2008), 084102.

[33] Rogerson, P. A., and Plane, D. A. The hoover index of population concentration and the demographic components of change: An article in memory of andy isserman. *International Regional Science Review 36*, 1 (2013), 97–114.

[34] Santos, A. Study of the correlation between the weather and the demand for bicycles in gira stations.

[35] Sen, A. *On Economic Inequality*. Oxford University Press, 1973.

[36] Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal 27*, 3 (1948), 379–423.

[37] Shorrocks, A. F. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society* (1980), 613–625.

[38] Theil, H. Economic and information theory. *Co., Amsterdam* (1967).

[39] Weiner, J., and Solbrig, O. T. The meaning and measurement of size hierarchies in plant populations. *Oecologia 61*, 3 (1984), 334–336.

[40] Zar, J. H. Spearman rank correlation. *Encyclopedia of Biostatistics 7* (2005).

A

Auxiliary Tables

Table A.1: TAZs of the Municipality of Lisbon

Number	Description
1	Belém (Ribeirinha - Belém)
2	Belém (Alto Duque)
3	Belém (Belém — Oeste)
4	Belém (Belém — Este)
5	Ajuda (Restelo)
6	Alcântara (Ribeirinha - Belém)
7	Alcântara (Santo Amaro — Oeste)

Continued on next page

Table A.1 – continued from previous page

Number	Description
8	Alcântara (Santo Amaro — Sul)
9	Alcântara (Santo Amaro — Norte)
10	Ajuda (Ajuda — Sul)
11	Ajuda (Ajuda — Oeste)
12	Ajuda (Ajuda — Norte)
13	Estrela (Ribeirinha - Belém)
14	Estrela (Alcântara)
15	Estrela (Lapa)
16	Estrela (Estrela)
17	Campo de Ourique (Campo Ourique)
18	Campo de Ourique (Amoreiras)
19	Misericórdia (São Paulo)
20	Misericórdia (Bairro Alto)
21	Santo António (Amoreiras)
22	Santo António (São Mamede)
23	Santo António (São José)
24	Benfica (Colégio Militar/Luz)
25	Benfica (Benfica — Sul)
26	Benfica (Benfica — Norte)
27	Benfica (Monsanto — Norte)
28	São Domingos de Benfica (São Domingos de Benfica)
29	São Domingos de Benfica (Sete Rios — Norte)
30	São Domingos de Benfica (Estrada Luz — Este)

Continued on next page

Table A.1 – continued from previous page

Number	Description
31	São Domingos de Benfica (Monsanto — Norte)
32	São Domingos de Benfica (Sete Rios — Sul)
33	Campolide (Bairro Liberdade)
34	Campolide (Praça Espanha — Sul)
35	Campolide (Praça Espanha — Norte)
36	Campolide (Campolide)
37	São Domingos de Benfica (Estrada Luz — Oeste)
38	Carnide (Avenidas Novas)
39	Carnide (Carnide)
40	Lumiar (Telheiras — Oeste)
41	Lumiar (Telheiras — Este)
42	Lumiar (Parque Europa)
43	Lumiar (Lumiar — Oeste)
44	Lumiar (Lumiar — Sul)
45	Lumiar (Paço Lumiar)
46	Lumiar (Telheiras — Norte)
47	Lumiar (Lumiar — Norte)
48	Santa Clara (Ameixoeira — Norte)
49	Santa Clara (Ameixoeira — Sul)
50	Alvalade (Hospital Santa Maria — Oeste)
51	Alvalade (Hospital Santa Maria — Este)
52	Alvalade (Cidade Universitária)
53	Alvalade (Campo Grande)

Continued on next page

Table A.1 – continued from previous page

Number	Description
54	Alvalade (Avenida do Brasil)
55	Alvalade (Alvalade)
56	Alvalade (São João de Brito)
57	Alvalade (Roma - Areeiro — Este)
58	Alvalade (Roma - Areeiro — Oeste)
59	Avenidas Novas (Bairro Santos — Este)
60	Avenidas Novas (Bairro Santos — Oeste)
61	Avenidas Novas (Campo Pequeno — Oeste)
62	Avenidas Novas (Campo Pequeno — Este)
63	Avenidas Novas (Avenidas Novas — Oeste)
64	Avenidas Novas (Parque Eduardo VII)
65	Avenidas Novas (Picoas)
66	Avenidas Novas (Avenidas Novas — Este)
67	Areeiro (Areeiro — Norte)
68	Areeiro (Areeiro — Sul)
69	Areeiro (Alto Pina)
70	Arroios (Estefânia)
71	Arroios (Arroios — Norte)
72	Arroios (Arroios — Sul)
73	Arroios (Anjos)
74	Santa Maria Maior (Baixa)
75	Santa Maria Maior (Castelo)
76	São Vicente de Fora (São Vicente)

Continued on next page

Table A.1 – continued from previous page

Number	Description
77	Penha de França (Penha França)
78	Penha de França (São João)
79	Beato (Madre Deus — Oeste)
80	Beato (Picheleira)
81	Beato (Madre Deus — Sul)
82	Beato (Madre Deus — Norte)
83	Marvila (Chelas)
84	Marvila (Marechal Gomes da Costa)
85	Marvila (Infante Dom Henrique - Porto — Este)
86	Marvila (Infante Dom Henrique - Porto — Oeste)
87	Marvila (Parque Bela Vista)
88	Marvila (Bairro Armador)
89	Lumiar (Aeroporto)
90	Olivais (Logística Aeroportuaria)
91	Olivais (Alfredo Bensaúde)
92	Olivais (Encarnação — Oeste)
93	Olivais (Relógio)
94	Olivais (Olivais - Centro)
95	Olivais (Encarnação — Este)
96	Olivais (Olivais — Norte)
97	Olivais (Olivais — Sul)
98	Olivais (Olivais — Este)
99	Parque das Nações (Parque Nações — Sul)

Continued on next page

Table A.1 – continued from previous page

Number	Description
100	Parque das Nações (Estação Oriente)
101	Parque das Nações (Quinta Laranjeiras)
102	Parque das Nações (Parque Nações — Norte)
103	Parque das Nações (Parque Tejo)

Table A.2: Parishes of the Municipality of Lisbon

Number	Description
101	Ajuda
102	Alcantara
103	Alvalade
104	Areeiro
105	Arroios
106	Avenidas Novas
107	Beato
108	Belem
109	Benfica
110	Campo de Ourique
111	Campolide
112	Carnide
113	Estrela
114	Lumiar
115	Marvila

Continued on next page

Table A.2 – continued from previous page

Number	Description
116	Misericordia
117	Olivais
118	Parque das Nacoes
119	Penha de Franca
120	Santa Clara
121	Santa Maria Maior
122	Santo Antonio
123	Sao Domingos de Benfica
124	Sao Vicente

B

Auxiliary Figures

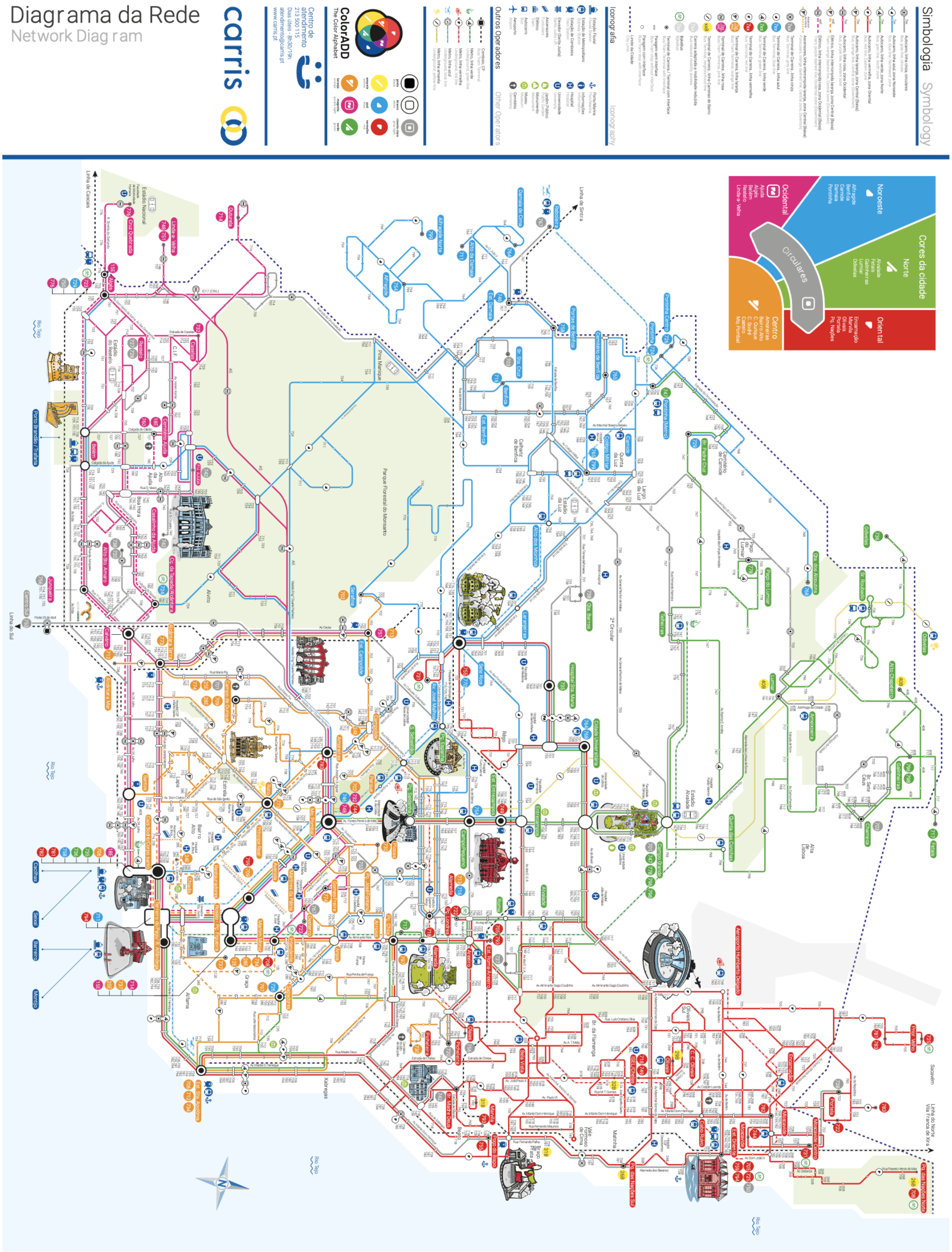


Diagrama da Rede
Network Diagram
carris
Simbologia Symbology
Iconografia Iconography
Outros Operadores Other Operators
ColorADD The Color Alphabet
Centro de atendimento
Cores da cidade
Noroeste
Norte
Oriental
Ocidental
Centro
Circulares
Parque Florestal do Monsanto
Rio Tejo

Figure B.1: Daytime Carris Network Diagram.

Source: www.carris.pt

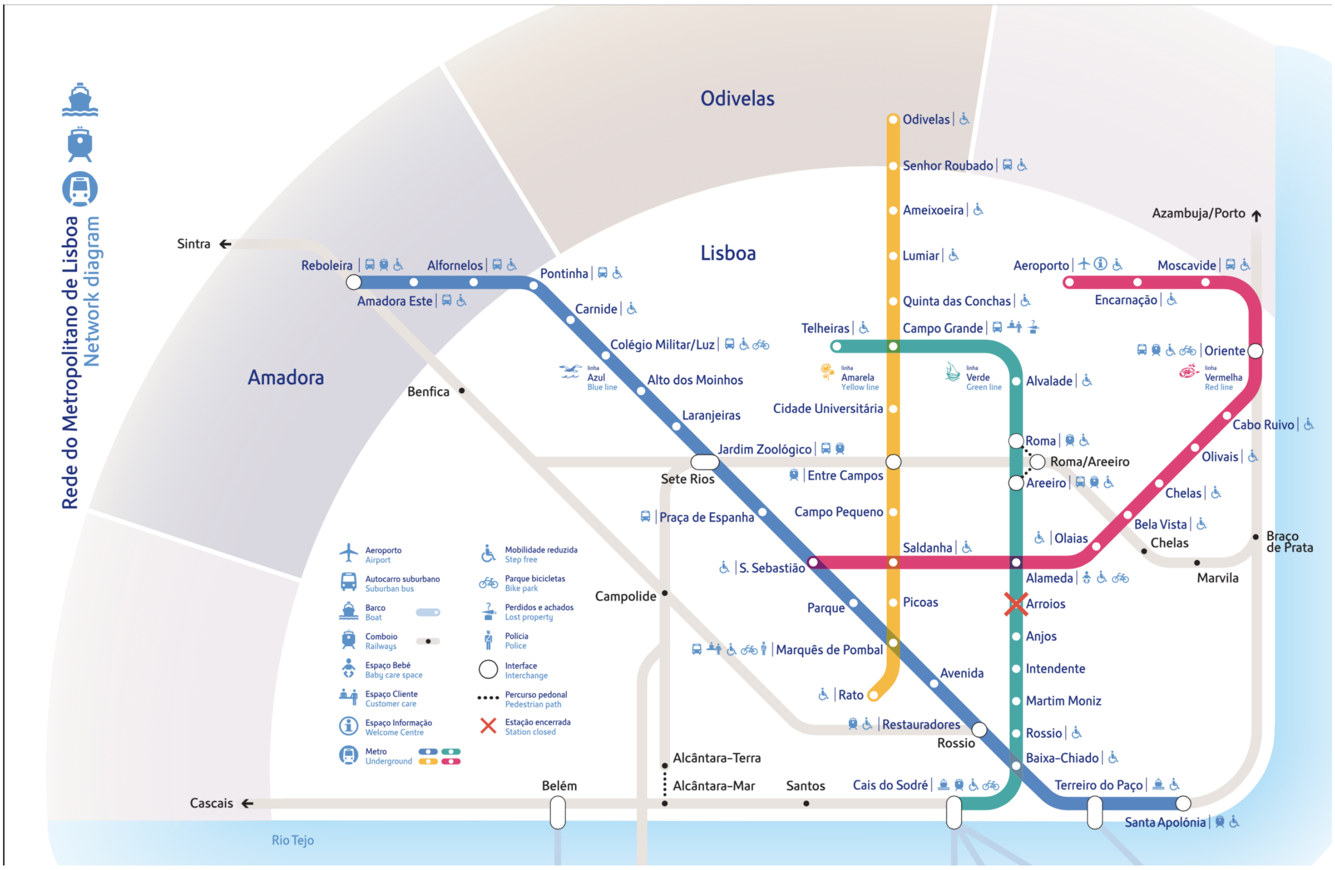

Rede do Metropolitano de Lisboa
Network diagram
Odivelas
Lisboa
Amadora
Sintra
Reboleira
Amadora Este
Alfornelos
Pontinha
Carnide
Colégio Militar/Luz
Alto dos Moinhos
Laranjeiras
Jardim Zoológico
Sete Rios
Praça de Espanha
S. Sebastião
Parque
Marquês de Pombal
Avenida
Restauradores
Rossio
Benfica
Campolide
Odivelas
Senhor Roubado
Ameixoeira
Lumiar
Quinta das Conchas
Campo Grande
Cidade Universitária
Entre Campos
Campo Pequeno
Saldanha
Picoas
Rato
Telheiras
Alvalade
Roma
Roma/Areeiro
Areeiro
Olaias
Alameda
Arroios
Anjos
Intendente
Martim Moniz
Rossio
Baixa-Chiado
Cais do Sodré
Terreiro do Paço
Santa Apolónia
Aeroporto
Encarnação
Moscavide
Oriente
Cabo Ruivo
Olivais
Chelas
Bela Vista
Chelas
Marvila
Braço de Prata
Azambuja/Porto
linha Azul Blue line
linha Amarela Yellow line
linha Verde Green line
linha Vermelha Red line
Alcântara-Terra
Alcântara-Mar
Santos
Belém
Cascais
Rio Tejo
Aeroporto Airport
Autocarro suburbano Suburban bus
Barco Boat
Comboio Railways
Espaço Bebé Baby care space
Espaço Cliente Customer care
Espaço Informação Welcome Centre
Metro Underground
Mobilidade reduzida Step free
Parque bicicletas Bike park
Perdidos e achados Lost property
Polícia Police
Interface Interchange
Percurso pedonal Pedestrian path
Estação encerrada Station closed

Figure B.2: Metropolitano de Lisboa Network Diagram.

Source: www.metrolisboa.pt

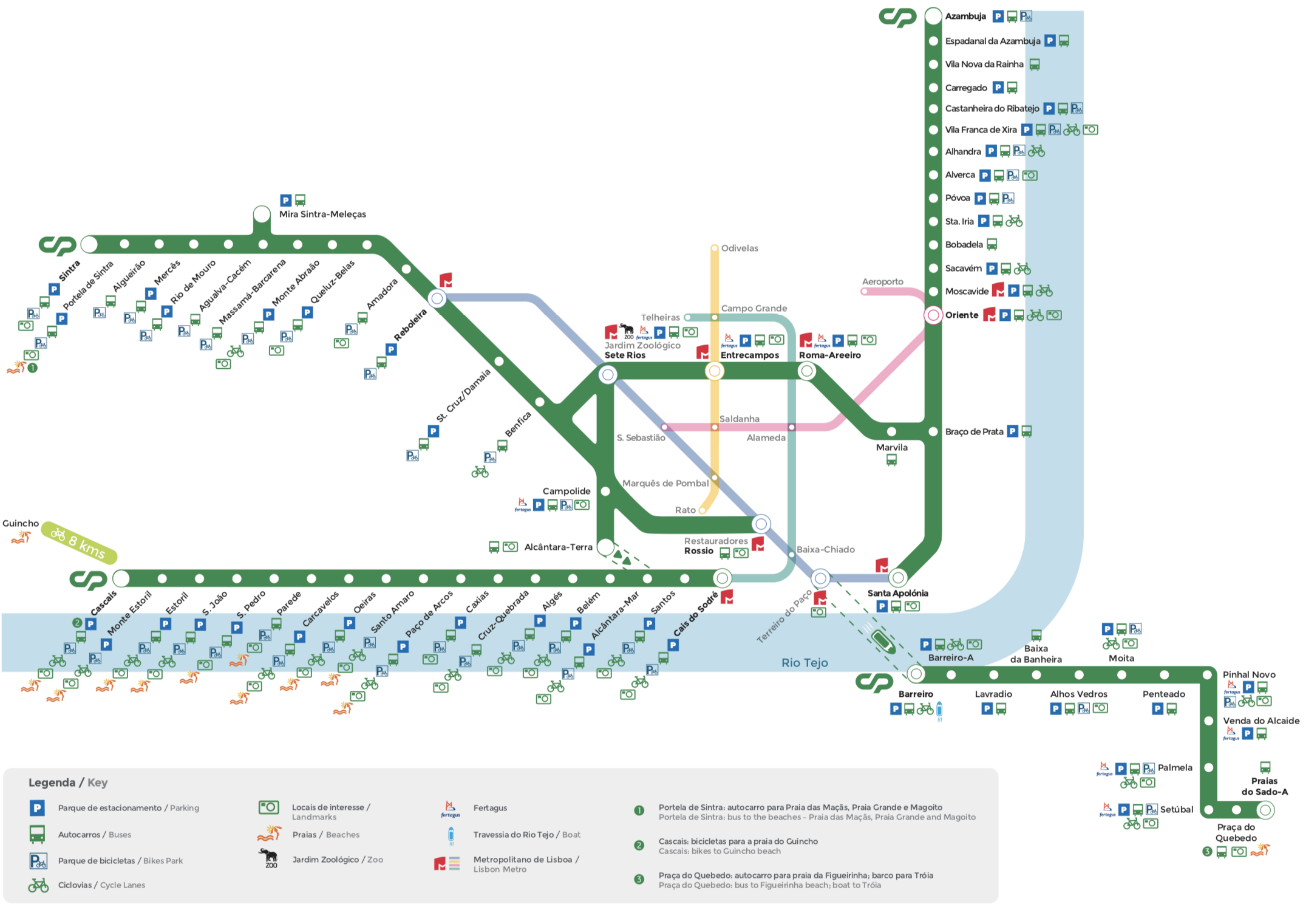

Azambuja
Espadanal da Azambuja
Vila Nova da Rainha
Carregado
Castanheira do Ribatejo
Vila Franca de Xira
Alhandra
Alverca
Póvoa
Sta. Iria
Bobadela
Sacavém
Moscavide
Oriente
Aeroporto
Braço de Prata
Marvila
Santa Apolónia
Mira Sintra-Meleças
Sintra
Portela de Sintra
Algueirão
Mercês
Rio de Mouro
Agualva-Cacém
Massamá-Barcarena
Monte Abraão
Queluz-Belas
Amadora
Reboleira
St. Cruz/Damaia
Benfica
Campolide
Alcântara-Terra
Odivelas
Telheiras
Campo Grande
ZOO
Jardim Zoológico
Sete Rios
Entrecampos
Roma-Areeiro
Saldanha
S. Sebastião
Alameda
Marquês de Pombal
Rato
Restauradores
Rossio
Baixa-Chiado
Terreiro do Paço
Guincho
8 kms
Cascais
Monte Estoril
Estoril
S. João
S. Pedro
Parede
Carcavelos
Oeiras
Santo Amaro
Paço de Arcos
Caxias
Cruz-Quebrada
Algés
Belém
Alcântara-Mar
Santos
Cais do Sodré
Rio Tejo
Barreiro-A
Baixa da Banheira
Moita
Barreiro
Lavradio
Alhos Vedros
Penteado
Pinhal Novo
Venda do Alcaide
Palmela
Praias do Sado-A
Setúbal
Praça do Quebedo
Legenda / Key
Parque de estacionamento / Parking
Autocarros / Buses
Parque de bicicletas / Bikes Park
Ciclovias / Cycle Lanes
Locais de interesse / Landmarks
Praias / Beaches
Jardim Zoológico / Zoo
Fertagus
Travessia do Rio Tejo / Boat
Metropolitano de Lisboa / Lisbon Metro
① Portela de Sintra: autocarro para Praia das Maçãs, Praia Grande e Magoito
Portela de Sintra: bus to the beaches - Praia das Maçãs, Praia Grande and Magoito
② Cascais: bicicletas para a praia do Guincho
Cascais: bikes to Guincho beach
③ Praça do Quebedo: autocarro para praia da Figueirinha; barco para Tróia
Praça do Quebedo: bus to Figueirinha beach; boat to Tróia

Figure B.3: CP/Metro Network Diagram.

Source: www.cp.pt