Pontifícia Universidade Católica
do Rio de Janeiro

Bruno Guberfain do Amaral

A visual analysis of bus GPS data in Rio

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–Graduação em Informática of the Departamento de Informática, PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Informática

Advisor: Prof. Hélio Côrtes Vieira Lopes

Rio de Janeiro
June 2015

Pontifícia Universidade Católica
do Rio de Janeiro

Bruno Guberfain do Amaral

A visual analysis of bus GPS data in Rio

Dissertation presented to the Programa de Pós–Graduação em Informática of the Departamento de Informática, PUC–Rio as partial fulfillment of the requirements for the degree of Mestre em Informática.
Approved by the following commission:

Prof. Hélio Côrtes Vieira Lopes
Advisor
Departamento de Informática — PUC–Rio

Prof. Marco Antonio Casanova
Departamento de Informática — PUC–Rio

Prof. José Eugênio Leal
Departamento de Engenharia Industrial — PUC–Rio

Prof. Hugo Varela Repolho
Departamento de Engenharia Industrial — PUC–Rio

Prof. José Eugênio Leal
Coordinator of the Centro Técnico Científico — PUC–Rio

Rio de Janeiro, June 12th, 2015

Bruno Guberfain do Amaral

Bachelor degree in Informatics and Information Technology at State University of Rio de Janeiro (UERJ) in 2006. Works at Petrobras since 2008 as System Analyst.

Bibliographic data

do Amaral, Bruno Guberfain

A visual analysis of bus GPS data in Rio / Bruno Guberfain do Amaral; advisor: Hélio Côrtes Vieira Lopes. — 2015.

45 f. : il. (color); 30 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2015.

Inclui bibliografia.

1. Informática – Teses. 2. Ciência de Dados. 3. Cidades Inteligentes. 4. Big Data. 5. Visualização. 6. Ônibus. 7. GPS. 8. Rio de Janeiro. I. Lopes, Hélio Côrtes Vieira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

To my wife Renata, my soon-to-be-born son Rafael, my family and professor Hélio. Those whose incentive made possible the conclusion of this work.

Acknowledgements

Prima facie, I am grateful to God for the good health and well-being that were necessary to complete this work.

Furthermore, I wish to express my sincere thanks to:

- My wife, *Renata*, that in a daily-basis helped me in the development of this work;
- My soon-to-be-born son, *Rafael*, that shortened the conclusion of this work in 6 months;
- My parents, *Helvio* and *Vera*, for understanding when I couldn't visit them for some weeks;
- My sister, *Renata*, and her husband, *Marcos*, for always being together with me;
- Professor *Helio Lopes*, for spending some long and entertained afternoons discussing this work and always being helpful for whatever I needed;
- Professors *Paulo Eustáquio* and *Alexandre Sztajnberg*, for being my tutors during graduation and providing recommendations, even 10 years after the conclusion of my bachelor degree;
- My bosses at Petrobras, *Bruno Abud* and *Ana Claudia Lima Pinheiro*, for providing me a good balance between studies and work, and the means for concluding this course;
- All my friends, a lot of them, for supporting me and understanding why I drank less beer in the previous months.

Abstract

do Amaral, Bruno Guberfain; Lopes, Hélio Côrtes Vieira (Advisor). **A visual analysis of bus GPS data in Rio**. Rio de Janeiro, 2015. 45p. MSc. Dissertation — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Smart cities is a current subject of interest for public administrators and researchers. Getting the cities smarter is one of the challenges for the near future, due to the growing demand for public services. In particular, public transportation is one of most visible aspects of a living city and, therefore, its implementation must be very efficient. The public transportation system of the City of Rio de Janeiro is historically deficient, mostly because it is based on an old bus system. To change it, the City Hall took some actions, such as the development of an open data project that shows, at about every minute, the GPS instant position of all buses in the city. Although it is not a new technology, it is the first initiative to be developed in Rio. This work presents simple tools for the visual exploration of this big dataset based on the historical information from this service, which reaches a total of more than one billion samples. With these tools one is able to discover trends, identify patterns, and locate abnormalities within the massive collection of the buses' GPS data.

Keywords

Data Science; Smart Cities; Big Data; Visualization; Buses; GPS; Rio de Janeiro.

Resumo

do Amaral, Bruno Guberfain; Lopes, Hélio Côrtes Vieira. **Uma análise visual dos dados de GPS dos ônibus no Rio**. Rio de Janeiro, 2015. 45p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O tema *Cidades Inteligentes* é um assunto de interesse para gestores públicos e pesquisadores. Desenvolver cidades mais inteligentes é um dos desafios para o futuro próximo devido à crescente demanda por serviços públicos. Em particular, o transporte público é um dos aspectos mais visíveis de uma cidade viva e, portanto, a sua implementação deve ser muito eficiente. O sistema de transporte público da cidade do Rio de Janeiro é historicamente deficiente, principalmente porque ele é baseado em um antigo sistema de ônibus. Para melhorá-lo, a Prefeitura tomou algumas ações, como o desenvolvimento de um projeto de dados aberto que mostra, a quase a cada minuto, a posição instantânea de todos os ônibus na cidade. Embora não seja uma nova tecnologia, esta é a primeira iniciativa a ser desenvolvida no Rio. Neste trabalho, são aprsentadas ferramentas simples para a exploração visual deste grande conjunto de dados com base em informações históricas, que chega a um total de mais de um bilhão de amostras. Com essas ferramentas, um usuário será capaz de verificar as tendências, identificar padrões e localizar anomalias nesta grande quantidade de dados de GPS dos ônibus.

Palavras–chave

Ciência de Dados; Cidades Inteligentes; Big Data; Visualização; Ônibus; GPS; Rio de Janeiro.

Contents

1 Introduction **12**

2 Related work **14**

3 Data preparation **16**
3.1 Acquisition 16
3.2 Route matching 17
3.3 Bus stops matching 19

4 Route-based graphical tools for visual data analysis **21**
4.1 Overview 21
4.2 Space × Speed graph 22
4.3 Space × Time graph 23
4.4 Bus concentration heatmap 25

5 Typical bus traffic **26**
5.1 Overview 26
5.2 Shared routes segments detection 27
5.3 Computation of maximum speed for each *shared section* 27
5.4 Map generation 28

6 Visual analysis **30**
6.1 Space × Speed 30
6.2 Space × Time 30
6.3 Bus concentration heatmap 31
6.4 Typical bus traffic map 32

7 Trends, Patterns, and Abnormalities **33**
7.1 Seasonal patterns 33
7.2 Detection of outdated route geometry 34
7.3 Mirror pattern of opposite routes 34
7.4 Stacking graphs 34
7.5 Map of a typical workday traffic 36
7.6 Analysis of BRS São Clemente 37

8 Conclusion and future works **41**

9 Bibliography **42**

List of Figures

3.1 GPS sampling interval time distribution of a bus line. 16
3.2 Route matching using different approaches. 18
3.3 Route matching using two samples. 18
3.4 GPS samples at the Copacabana neighbourhood. 19

4.1 Colors representing the sample time and hue degree. 21
4.2 Program functionalities. 23
4.3 An example of *Space* × *Speed* graph. 24
4.4 State diagram to detect the *departure-sample*. 24
4.5 An example of *Space* × *Time* graph. 24
4.6 Heatmap colors. 25
4.7 An example of *Bus concentration heatmap*. 25

5.1 An excerpt of *shared sections*. 27
5.2 Effect of different percentiles to calculate the maximum speed. 28
5.3 Typical bus traffic in and around Lagoa. 29

6.1 Another example of *Space* × *Speed* graph. 30
6.2 Routes for the same line number showing opposite pattern. 31
6.3 Another example of *Bus concentration heatmap*. 31
6.4 Bus traffic at Botafogo along a week. 32

7.1 A route with seasonal section 33
7.2 A line with seasonal routes. 34
7.3 Example of a route with outdated geometry. 34
7.4 Mirror pattern on routes of the same line number, but in opposite directions. 35
7.5 Graph stack of bus lines # 636 and # 778. 36
7.6 Graph stack of bus lines # 862 and # 338. 37
7.7 Typical traffic on Botafogo during workdays. 38
7.8 Geometry of BRS Botafogo. 38
7.9 Graph stack of BRS São Clemente at different months. 39
7.10 Heatmap of a street interdiction. 40
7.11 Relevant tweets during the events analysed. 40
7.12 Heatmap of an abnormal traffic. 40

List of Tables

5.1 *Traffic color scale* used to define the traffic on a *Typical bus traffic* map. The open ranges, e.g. $]60, 50[$, are treated as a linear gradient between the two colors. 26

List of Algorithms

3.1 Pseudocode for data gathering 17
3.2 Pseudocode bus stop matching 20

1 Introduction

The main purpose of a *Data Science* project is the extraction of knowledge from data [1]. To do so, it utilizes techniques and theories from many fields (such as mathematics, statistics, computer science, . . .) in order to answer important questions made by decision makers based on the available data. The *Data Science* discipline is very interested in methods that scale to *Big Data* [2], which is a general term for data sets that are so large or complex that traditional data processing applications are considered inadequate.

It is recognized that *visualization* is a fundamental tool to comprehend huge amounts of data [3], since it allows the perception of emergent properties that were not anticipated, enables problems with the data to become immediately apparent, facilitates understanding of both large-scale and small-scale features of the data and speed up hypothesis formation.

In a *Big Data* setting, *Smart cities*, i.e., a city that uses the huge amount of information it produces in a "smart" way [4], is a current subject of interest for public administrators and researchers. Getting the cities smarter is one of the challenges for the near future, due to the growing demand for public services. In particular, public transportation is one of most visible aspects of a living city and, therefore, its implementation must be very efficient.

The public transportation system of the City of *Rio de Janeiro* is historically deficient, mainly because it is based on an old bus system. To change this state, the City Hall took some actions, such as the development of an open data project that shows, at about every minute, the GPS instant position of all buses in the city. Although it is not a new technology, it is the first initiative to be developed in Rio.

In this dissertation, a large dataset containing the GPS positions (more than 1.2 billion samples by May, 2015) of all buses that operate in the City of Rio de Janeiro since mid-2014 was build from this service. For future studies, we organize this data and made it available at the URL's (`http://www.inf.puc-rio.br/~bguberfain` or `http://www.inf.puc-rio.br/~lopes`).

The preparation and processing of this dataset is presented together with some graphical tools that can visually summarize relevant information for decision makers and urban planners. Using this tools, one can discover trends, identify patterns and locate abnormalities within the massive collection of the buses' GPS data. After all, we found that the dataset contains useful information and some knowledge of Rio's bus system can be extracted from it.

This dissertation is organized as following: Chapter 2 overviews some of the works related to traffic data; Chapter 3 describes the process of collecting and processing the data gathered from the City Hall's server; Chapter 4 and 5 introduces graphical tools for data analysis – the former generates graphs focused on a given route while the latter for a giving region; The following Chapter, 6, discuss general analyzes from the generated graphics; Chapter 7 shows examples of trends, patterns and abnormalities found within the dataset; The final Chapter 8 displays an overview of the contributions and proposes future works based on the dataset.

2
Related work

Public transportation with GPS. [5] proposed an intelligent management system that has a data-driven approach for modelling, analysis and decision-making to better control the traffic and to plan the mobility in a city. The main kind of data that serve as input in such systems is trajectories based on GPS installed on vehicles [6], such as taxis [7] and buses [8]. This work deals with GPS data installed in all buses for public transportation in the City of Rio de Janeiro.

GPS data streaming applications. To estimate traffic states from GPS data streams is an important task to improve the efficiency of traffic systems. According to [5], traffic applications using GPS data streams can be divided into two main groups: centralized and distributed. The first group uses traffic data from multiple GPS devices simultaneously, while the second group of applications uses individual GPS data. Traffic state estimation [9], queue profile estimation [10], detection of traffic anomalies [11] are examples of applications of the centralized applications. Applications of the second group include: vehicle performance analysis [12], vehicle monitoring [13], and vehicle anomaly detection [14]. This paper could be classified in the first group of applications.

Map matching. According to [15], the vehicles trajectories sampled from GPS devices is uncertain and to this data become useful to applications, it should be related to the underlying road network by means of map matching algorithms. Many methods have been proposed to solve map matching problems. Their objective is to aligning a sequence of observed positions with the road network on a digital map [16]. This work considers a simplified version of this problem: the observed bus positions should be matched with a given set of routes.

Visual tools for movement and traffic analysis. [17] proposed the Aitvs, a Web based traffic visualization system, which provides visualization components to analyze and to monitor traffic conditions. [18] presented a real-time surveillance system with a rule-based behavior and event-recognition module for traffic videos. [19] developed HOLMES, which is a system for highway operation monitoring and evaluation. [20] proposed the TripVista

(Triple Perspective Visual Trajectory Analytics) that is an interactive visual analytics system for complex traffic trajectory data visual exploration and analysis. [21] introduced the VAIT (A Visual Analytics System for Metropolitan Transportation), which is a system to study large-scale transportation data, integrating visualization and data analytics methods. Recently, [22] proposed a system for the visual exploration of sparse traffic trajectory data that was recorded from the movement of vehicles only at predefined locations. [23] proposed the use of scientific visualization tools for the visual exploration of big spatio-temporal urban data, and apply it to the study of New York City taxi trips. Recently, very modern visualization tools have been presented to visualize and explore passenger mobility in a public transportation system [24]. A complete survey of visual analytics techniques for movements can be found in [25].

Bus GPS data in Rio. A specific study about the bus traffic in the City of Rio de Janeiro was done by [26], which proposes a predictive model for the bus travel time using the same GPS data provided by the City Hall in [27]. The authors of this paper presented in [28] how to acquire the bus GPS data from [27] and how to use this historical data to build traffic maps. To the best of our knowledge, no other work for the visual analysis of bus GPS data in Rio de Janeiro exists in the literature.

3
Data preparation

3.1
Acquisition

The data presented in this section, and used throughout the dissertation, has been continuously acquired from [27] since June, 2014. Until now, more than 1 billion samples were stored, where each sample contains a *timestamp*, the *bus identifier*, the *line number*, the *position* (as latitude and longitude) and the *speed*. Samples from the same bus are updated at about every 1min30sec. Figure 3.1 shows the GPS sampling interval time distribution from the dataset.

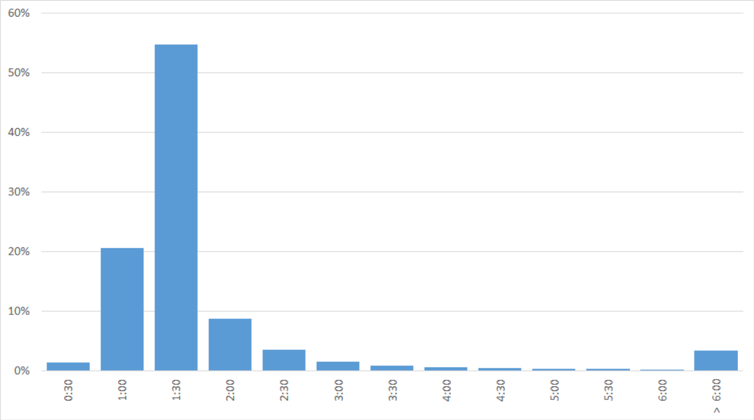


Figure 3.1: GPS sampling interval time distribution of a bus line.

The public data service from [27] offers only the instantaneous data, that is, no historical data is available. For this reason, it was implemented a service that queries this data periodically and stores its entries for future processing. This service was implemented using *Groovy* [29] language and it runs 24×7 in a machine at the Microsoft Azure Cloud. A pseudo-code of its logic can be seen at Algorithm 3.1. The historical data was organized and made available at the URL (`http://www.inf.puc-rio.br/~bguberfain` or `http://www.inf.puc-rio.br/~lopes`).

The log file generated by Algorithm 3.1 is rolled daily and contains the following information for each sample:

- Timestamp;
- Bus id (order number);
- Line number;
- Coordinates (latitude and longitude);
- Instant speed;
- Delay of timestamp, relative to [27] server's date.

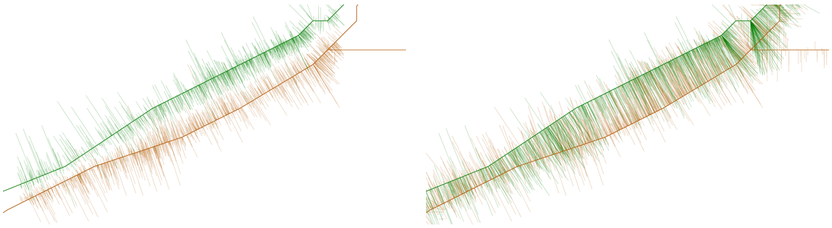
Algorithm 3.1 Pseudocode for data gathering

```
$currentState \leftarrow$ GetRealTimeData()
Wait(40s)
loop
    $newState \leftarrow$ GetRealTimeData()
    $differenceState \leftarrow newState - currentState$
    if Not-Empty($differenceState$) then
        Save(differenceState)
        Wait(40s)
        $currentState \leftarrow newState$
    end if
    Wait(4s)
end loop
```

3.2 Route matching

Every bus line has a set of expected routes. Matching the samples to one of these routes unveils useful information, such as the distance travelled or the time to pass through a street section (enabling the average velocity computation per street segment). The geometry of each route, defined by a list of consecutive coordinates, is provided by the City Hall on [27]. However, the data provided by the GPS service has no information about which route the bus is following at that moment. To find it, one can simply match the current bus position with the closest route, with results shown on the thin lines of Figure 3.2(a). Unfortunately, this can lead to errors when samples are placed between two closed routes.

A better approach is to find the most probable route by looking at two consecutive samples of a bus. These two consecutive positions determine a vector $\overrightarrow{V}$, see Figure 3.3. These two positions are also projected on each route R_i, respecting the route direction, to create the vectors $\overrightarrow{V_{R_i}}$. The chosen route is the one that has the smallest angle with $\overrightarrow{V}$. An example of the result of this method is show on Figure 3.2(b).



3.2(a): Route match by closest point

3.2(b): Route match by direction analysis

Figure 3.2: Route matching using different approaches.

This problem is a simplification of what is know as *map matching*, which is a subject of a specific research such as [16]. In this work there was no needed to use sophisticated algorithms since the samples should be matched with a small set of possible routes.

Sample distance tolerance Some factors can bring samples apart from the route:

- *Noise*: this is mostly generated by GPS reading errors from the devices that are attached to the buses, but it can be related to the presence of high buildings too. Figure 3.4 illustrates the latter, as samples along the beach are less uncertain (small spread) than those from inner streets, that mostly contains high buildings. In this figure, the colors indicate the direction of movement;
- *Outdated route geometry*: this is due to use of an old metadata of route geometry;
- *Wrong line number*: sometimes the server may return an incorrect line number, leading to the wrong geometry being used by the matching algorithm;
- *Bus going off-route*: buses can arbitrary go outside the route if it is interrupted or for any other reason;

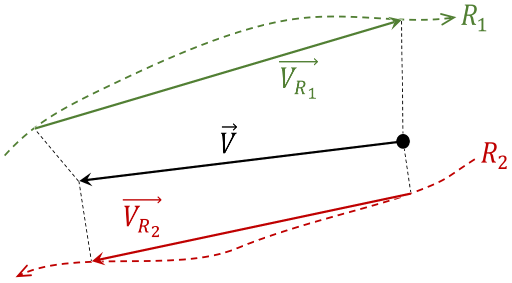


Figure 3.3: Route matching using two samples.

In order to prevent these factors from affecting other measures, all samples that are farther than $40m$ from the route are discarded.

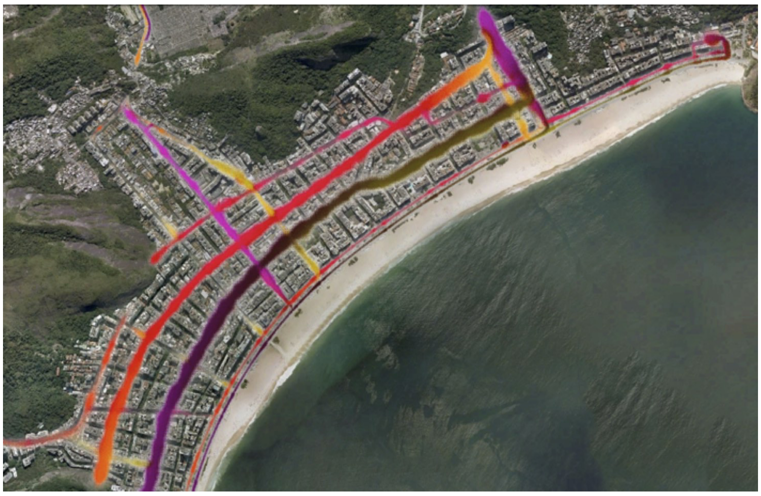


Figure 3.4: GPS samples at the Copacabana neighbourhood.

3.3 Bus stops matching

Annotating every route with its bus stops may be useful for analysing patterns specific these locations. A list of the bus stops for every bus line is available at [27]. It contains the coordinate (latitude and longitude), the line number and an incremental number associated to the sequence of the stop along the route. Thus, no information of which route the bus stop belongs is presented. For this reason, a method to find the best route was developed. Its implementation can seen on Algorithm 3.2.

Algorithm 3.2 Pseudocode bus stop matching

```
for all line : lines do
    routes ← ROUTES(line)
    stops ← STOPS(line)
    stopsOnRoute ← NEW-MAP(key : Route, value : NEW-LIST(Stop))

Create an initial stop for each route on the beginning of its geometry
    for all route : routes do
        firstStop ← FIRST-COORDINATE(route)
        stopsOnRoute[route] ≪ firstStop
    end for

The stops are read in order by its sequence field
    for all sequence : SEQUENCES-FROM(stops) do
        stopsForSequence ← STOPS-FOR-SEQUENCE(stops, sequence)

This creates a table of routes × stopsForSequence with the distance of every
stopsForSequence to the last stop of every routes
        distanceToLastStopOnRoute ← NEW-TABLE(rows : Route, cols : Stop)
        for all stop : stopsForSequence do
            for all route : routes do
                lastStop ← LAST-STOP(route)
                distance ← DISTANCE-ON-ROUTE-BETWEEN(lastStop, stop, route)
                distanceToLastStopOnRoute[route, stop] ← distance
            end for
        end for

For every stop on distanceToLastStopOnRoute we find the route whose
distance to last stop is minimized and store it on stopsOnRoute
        for all stop : distanceToLastStopOnRoute do
            choosenRoute ← SHORTEST-DISTANCE-ON(distanceToLastStopOnRoute[stop])
            stopsOnRoute[choosenRoute] ≪ stop
        end for
    end for

The last step is to persist the structure stopsOnRoute
    SAVE(stopsOnRoute)
end for
```

4
Route-based graphical tools for visual data analysis

4.1
Overview

The analysis of bus traffic along a route is useful to find patterns that are specific to that route. This chapter proposes three graphical tools for the visual analysis of bus trajectory data:

- *Space* × *Speed*: graph showing the average speed at each section of a route;
- *Space* × *Time*: graph showing the time the bus spent on the route since it departs from the origin;
- *Bus concentration*: a heatmap graph that focus on showing the density of buses along a route during the day.

All graphics share the same X axis that represents the traversed arc length along the route. This is useful to display each of these three graphical tools on top of the other. By doing so, the correlation between them is emphasized.

Colors The *Space* × *Speed* and *Space* × *Time* graphics share the same color scale, which represents the sampled time during the day. It was adopted the *hue degree* for this purpose, since it has the same cyclic pattern as the time along a day. The color scale values are presented on Figure 4.1.

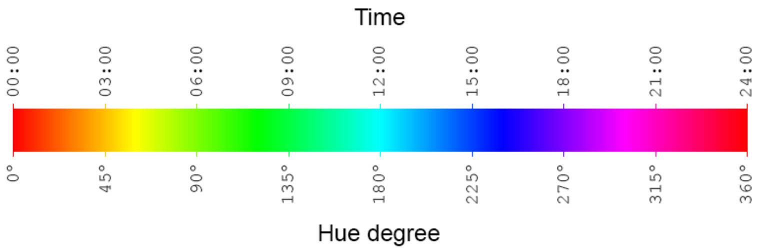


Figure 4.1: Colors representing the sample time and hue degree.

It is important to note that, although the color scales are the same on the *Space* × *Speed* and *Space* × *Time* graphics, they are used in a different manner on each of them. For the *Space* × *Speed* graphic, the color represents the sample's time and varies along the route whereas, for the *Space* × *Time* graphics, it represents the moment the bus left the origin, therefore, it has the same color along the route.

Transparency Due to the huge amount of data presented on each *Space* × *Speed* and *Space* × *Time* graphics, reaching more than 700.000 samples on some, a transparency should be used to draw the lines. For this work, we choose a value of 2.3 % or 6/255 for these graphics.

Interactivity The graphs were generated by a custom Java tool which provides some interactivity functionality, such as:

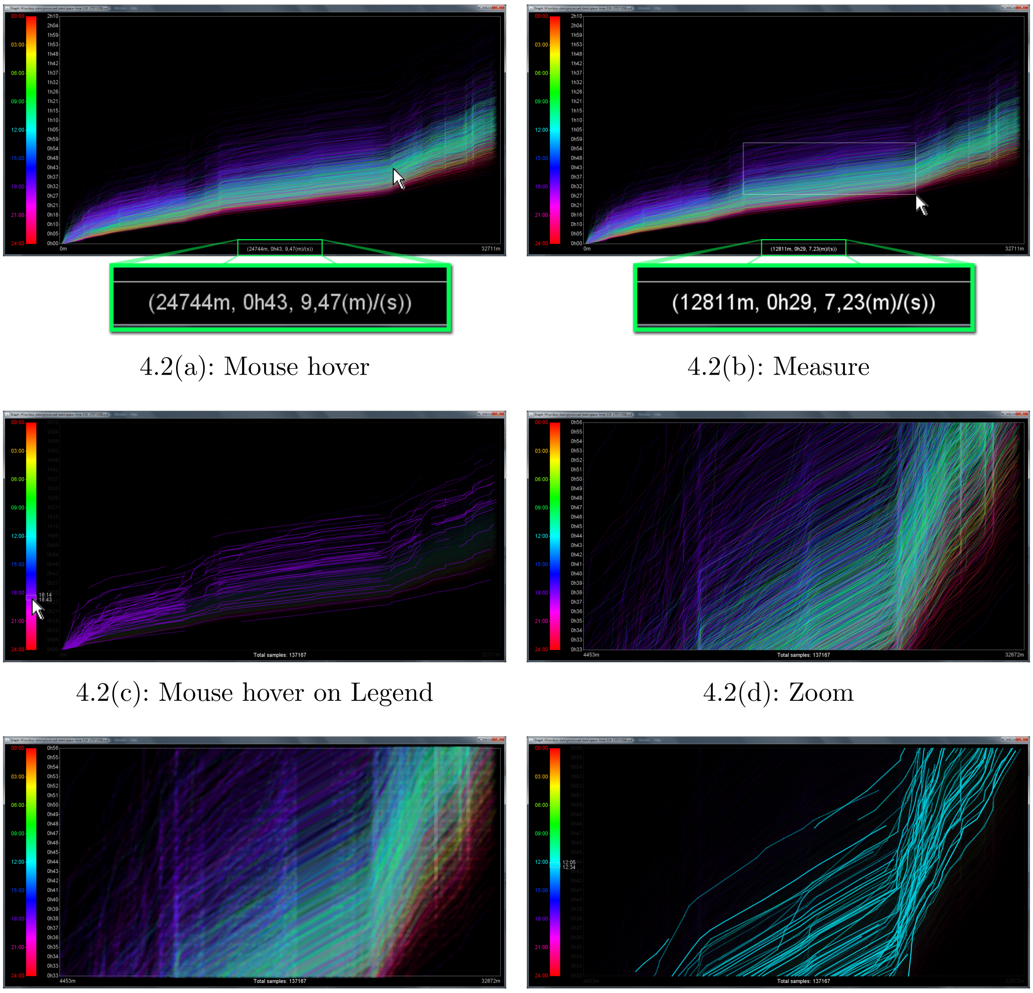
- *Mouse hover*: displays information on the bottom of screen according to cursor positiom (Figure 4.2(a));
- *Measure*: allows mouse drag to calculate distance on both axis and its relation (Figure 4.2(b));
- *Mouse hover on Legend*: highlights a range of legend's data (Figure 4.2(c));
- *Zoom*: zoom functionality by dragging the cursor on a area of the map (Figure 4.2(d));
- *Stroke configuration*: allows a dynamic definition of stroke width and transparency used to draw the data (Figure 4.2(e)).

4.2 Space × Speed graph

This graph presents the variance of speed along a specific route. This is useful to identify bottlenecks or time-related patterns. For this graph, one must first compute the distance between every two consecutive samples of the same bus along the geometry of the route. Then, the average speed the bus go through on this section is calculated by relating it to the time difference between these samples. A resulting view of this graph is shown on Figure 4.3, which was computed for one of the routes of bus line # 638.

For this graph, some outliers are removed to prevent them to interfere with the results:

- Speed above 100 km/h;
- Samples that are separated by more than 3.3km. This number corresponds to a sample travelling at maximum speed for 2 minutes.



4.2(a): Mouse hover

4.2(b): Measure

4.2(c): Mouse hover on Legend

4.2(d): Zoom

4.2(e): Stroke configuration

4.2(f): Combined features of zoom, highlight and stroke configuration

Figure 4.2: Program functionalities.

4.3 Space × Time graph

This graph shows the trip time along a route. It is useful to visualize patterns or clusters of trips during the day. In order to measure the duration of a trip it is necessary to find, for each bus, its departure time from the origin. For this, it was developed a procedure that finds a good sample representing the departure, called *departure-sample*. A state machine that runs for each vehicle and evaluates its position regarding the origin of the route ("inside" or "outside") is used for this task, as shown on Figure 4.4.

A sample will be considered the *departure-sample* if it changes the state machine from *At origin* to *In route*. Afterwards, every pair of samples that keeps the state machine on *In route* will be drown on the graph. When the final state is reached, the state machine is restarted.

This process has two other conditions that restart the process as well:

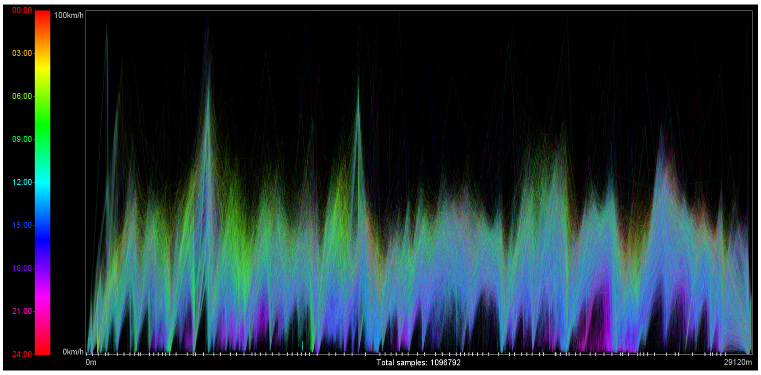


Figure 4.3: An example of *Space* × *Speed* graph.

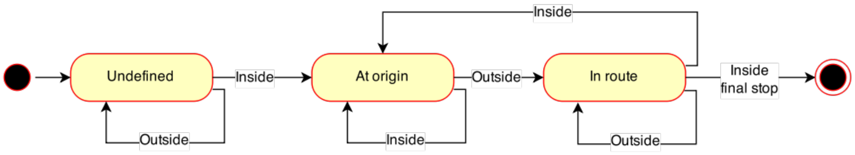


Figure 4.4: State diagram to detect the *departure-sample*.

– The bus has spend more than 4 hours on the route;

– The bus changed its route.

An example of this graph, drawn for the same route as Figure 4.3, can be viewed on Figure 4.5.

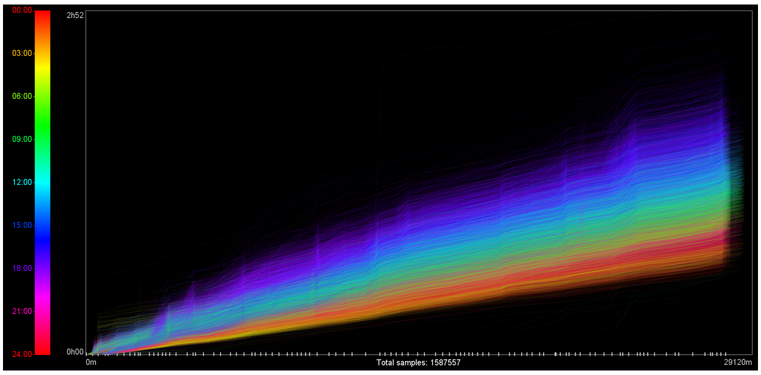


Figure 4.5: An example of *Space* × *Time* graph.

Although these two types of graphs are commonly used in traffic studies [26], the use of a color scale indication the time of bus departure is new. This contribution facilitates the identification of several patterns in this massive data.

4.4 Bus concentration heatmap

In this graph, it is shown the concentration of buses along a route (X axis) during the day (Y axis). Hot colors are used for high concentration of buses in an area while cold ones used for sparse samples, as shown on Figure 4.6.

The standard scale used in this graph, if not quoted, assumes that 3 buses running on the same route is a sparse configuration. Therefore, a distance equivalent to $^1/_3$ of the route's length is used as a reference for low concentrations (colder color).

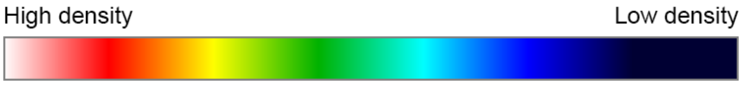


Figure 4.6: Heatmap colors.

To build this graph, the first step is to evaluate the distance between each pair of consecutive buses along the road, at a fixed rate. For every evaluation, these distances were converted to heat colors, based on a scale previously calculated for this route. In practice, the bus positions are evaluated as soon as new data is available from the service, which occurs at about every 1 minute. Samples older than 10 minutes are removed from the route.

The heatmap discretizes the time and the distance dimensions. Each dimension is divided using a fixed interval and, for every pair of values, the weighted average of all heat colors generated by the samples is computed. In this work a weighted average similar to the aliasing algorithm, mostly used in imaging processing, is used to avoid problems during this discretization. An example of this graph, generated for one of the routes of bus line # 802, is shown on Figure 4.7.

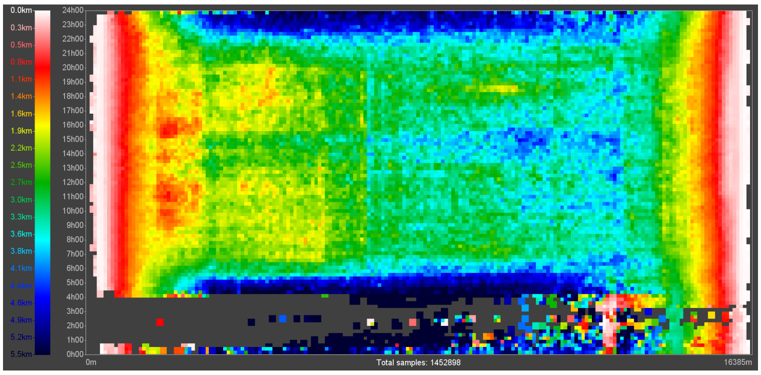


Figure 4.7: An example of *Bus concentration heatmap.*

This type of graph has not been presented in the literature, and it helps not only to identify patterns but also abnormalities.

5
Typical bus traffic

5.1
Overview

The *Typical bus traffic* is a map that emphasizes the traffic behaviour on a region along a giving period. It is useful to spot seasonal behaviours for a particular area.

In this chapter the data is computed for daily and weekly periods, each focusing on different aspects of the traffic. For example, one can use the computed daily data to observe the traffic along the hours of the day or the weekly data to spot the different behaviour at a specific time on weekends and workdays.

To compute the necessary information, the time dimension is discretized into *time slices* of fixed length, the routes are merged into *shared sections* and the average speed for each of these dimensions is computed. Afterwards a color scale is used to display the traffic information for a particular section at a giving *time slice*.

The color scale implemented in this work is based on a common scale used by popular tools, such as [30]. This scale is shown on column *Color* of Table 5.1.

The steps to build this graph are detailed on the following sections.

Label	Color	Relative speed (%)
Light	🟩	[100, 60]
Moderate	🟧	[50, 40]
Heavy	🟥	[30, 20]
Very Heavy	🟫	[10, 0]

Table 5.1: *Traffic color scale* used to define the traffic on a *Typical bus traffic* map. The open ranges, e.g. $]60, 50[$, are treated as a linear gradient between the two colors.

5.2
Shared routes segments detection

The first step to build this map is to find the common sections among all routes of all lines, called *shared sections*. These sections share the same spatial information (*from* and *to* coordinates), which are used to draw the map. Furthermore, once the sections are joint, the information from samples of multiple lines numbers can be computed together, generating more statistical information per segment.

In this dissertation, we accomplished this task by indexing each section of each route, using the *from* and *to* coordinate, then finding all common sections using this index. Afterwards, all sections from the same index are joint into a single *shared section*, resulting in a structure similar to Figure 5.1. In this figure, lines # 190 and # 472 (from the top) are joint with line # 523 (from the bottom) into a single *shared section* with all lines numbers (on the left).

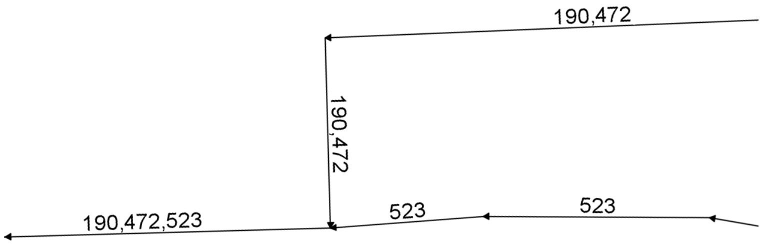


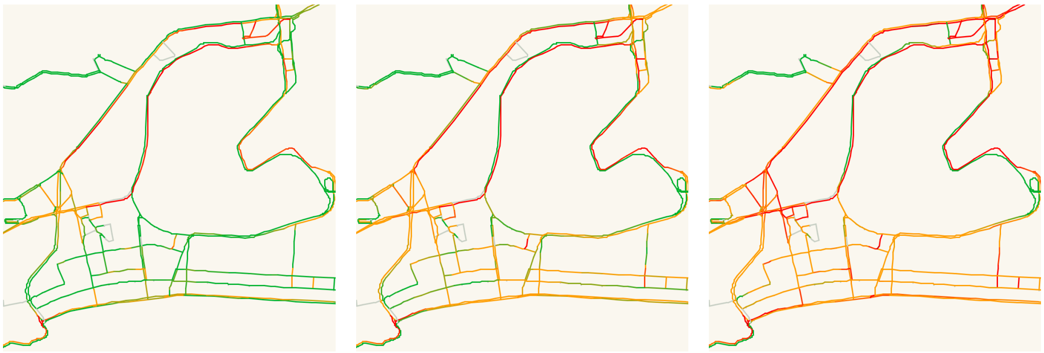
Figure 5.1: An excerpt of *shared sections*.

5.3
Computation of maximum speed for each shared section

The second step is to establish the maximum speed of each *shared section*, that is used to calculate the relative speed for a giving *time slice* on that segment. The computation of the maximum speed for each *shared section* is more robust than using an overall absolute value, due to specific behaviours of each segment (e.g., a speed of 20 km/h can be interpreted as *heavy* in a highway while it can be considered *light* on a tertiary street).

Due to noise samples, a simple approach, such as finding the fastest bus that passed through the *shared section*, would produce inaccurate results. Therefore, a statistical method must be used to prevent these errors. In this work, a high percentile of all samples speeds is used to establish the maximum speed for a *shared section*.

As the value for this percentile can interfere with the final color result, which is a subjective measure, a visual analysis was performed to establish this parameter, shown on Figure 5.2. For this work the 98 th percentile for maximum speed calculation was choose.



5.2(a): $P(90)$ 5.2(b): $P(95)$ 5.2(c): $P(98)$

Figure 5.2: Effect of different percentiles to calculate the maximum speed.

5.4 Map generation

In this section the map generation is discussed and, since it has some aspects linked to cartography, some concepts of this discipline are introduced.

5.4.1 Cartography background

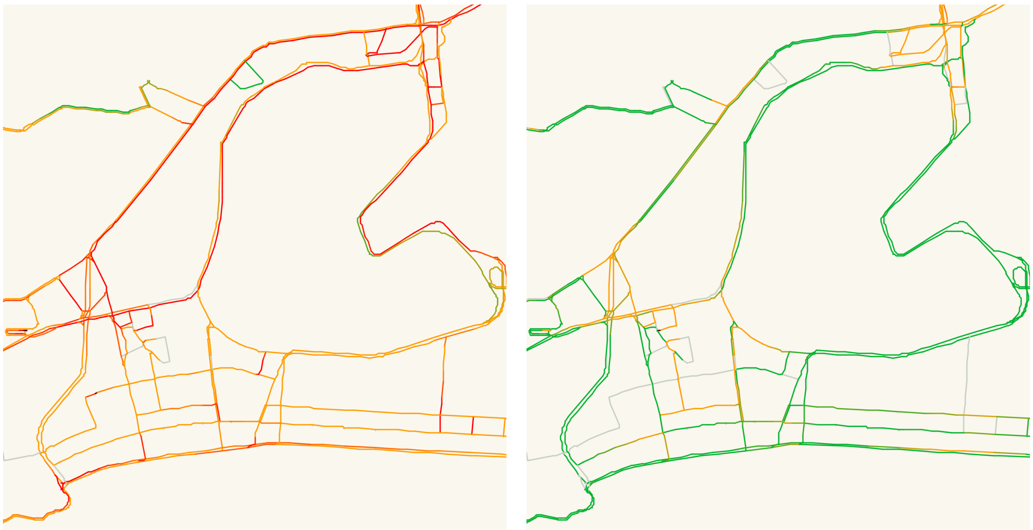
"A map projection is a systematic representation of all or part of the surface of a round body, especially the Earth, on a plane" [31]. The Mercator projection, introduced by Gerardus Mercator in 1569, "is one of the most widely known and has a long history of use for global-scale mapping" [32]. Its variation, named Web Mercator, "has now been readily adopted by Google Maps, Microsoft Bing Maps, Yahoo Maps, Esri's ArcGIS Online, OpenStreetMap, and The National Map of the US Geological Survey and therefore has become the *de facto* standard for online maps".

From a perceptual standpoint, Mercator and Web Mercator projections can be considered the same. Hence, both can be used to overlay maps together, even those from online sources. In this work the *Mercator* projection was used to build the maps, since it has easy-to-find and reliable implementations available.

5.4.2 Image generation

Intuitively, it is know that the traffic has a strong relation with time and, thus, the traffic maps images are generated for each *time slice*. Any region covered by the dataset can be defined as the area to be drown.

For this task, all *shared sections*, whose coordinates are projected to the image using the Mercator projection, are drown to the map using a color chosen from *Traffic color scale* (5.1). An example of a generated map for the same region, but at different *time slices*, can be viewed on Figure 5.3.



5.3(a): 18:30

5.3(b): 23:30

Figure 5.3: Typical bus traffic in and around Lagoa.

6
Visual analysis

6.1
Space × Speed

This graph can be used to identify bus stops along the route. In the same way as Figure 4.5, one can notice that on Figure 6.1, generated for bus line # 613, a stop accumulation occurs on certain areas as well. An additional information presented in these graphics are small marks along the X axis where the bus stops computed on 3.3 are plotted.

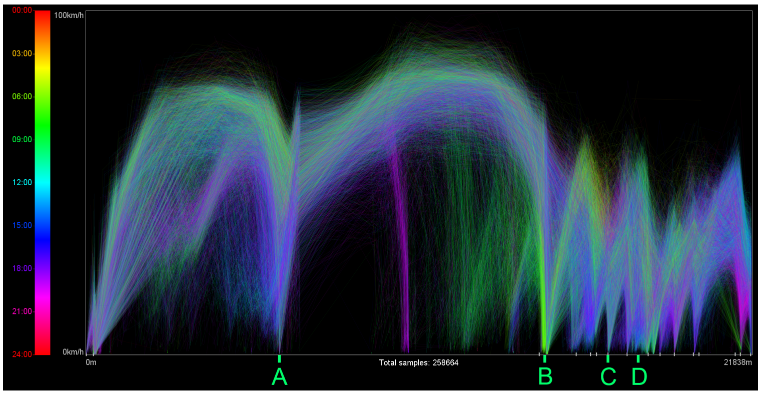


Figure 6.1: Another example of *Space* × *Speed* graph.

In this graph some bus stops can be identified, such as in B and C, and the locals where the traffic is heavier, such in points A and D. They differ as the latter has no bus stops defined on X axis. Furthermore, we can verify that between the origin and the point B the buses flow faster due to a highway (Linha Amarela), but with a toll plaza at point A.

6.2
Space × Time

This graphic presents the time spend to complete a specific route, grouped by time of the day, as well as a spread of this time, representing the uncertainty of the traffic.

On the graphics presented in Figure 6.2, it is possible to notice the direction of the route, related to the population movement. For the first route, which goes towards downtown, the traffic is heavier (higher on Y axis) on the morning while on the second route, which goes away from the city center, the traffic is heavier by the end of the afternoon.

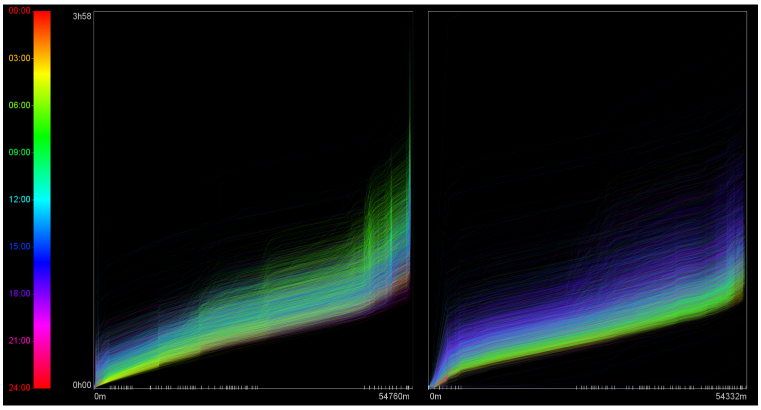


Figure 6.2: Routes for the same line number showing opposite pattern.

6.3 Bus concentration heatmap

This graph can be used to identify how the bus line works along the day, from the point of the bus concentration. Figure 4.7 shows that the Bus Line 802 does not offer service during the first hours of the day, representing by a high concentration of no-data (gray) or blue cells. On the other hand, a high concentration of red cells could probably indicate an inefficiency problem, when several buses pass to a bus stop in a very short interval. Figure 6.3, generated for bus line # 315, shows an example where heavier traffic imposes a high concentration of buses (horizontal pattern from hotter colors). Notice that in this graph some noise samples from bus terminals are presented on the early hours of the day.

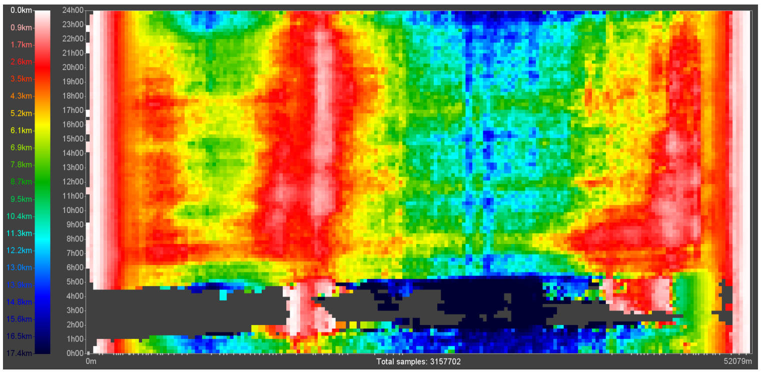
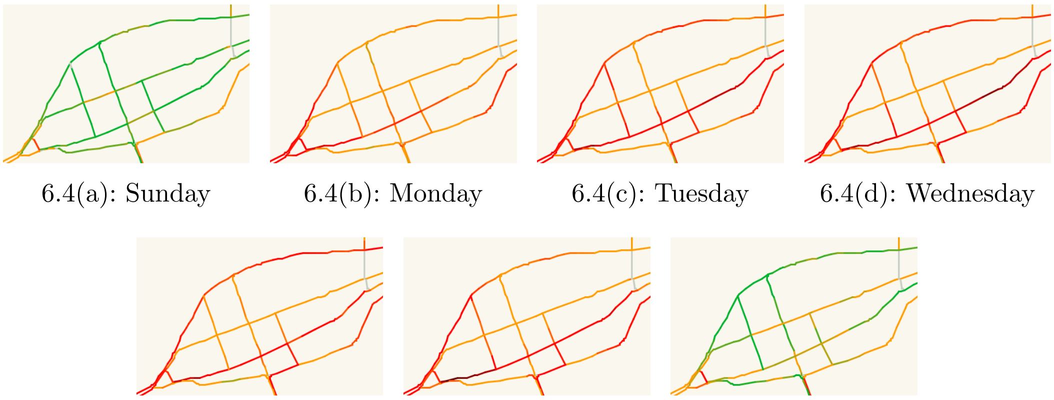


Figure 6.3: Another example of *Bus concentration heatmap*.

6.4 Typical bus traffic map

This graph allows an analysis of traffic for a region, rather than on a specific route. It displays an easy-to-understand map, similar to popular tools such as *Google Maps* [30], and can be generated for any period (daily, weekly, etc) and *time slice* (every 15 min, hourly, etc).

One possible analysis is to compare the traffic behaviour along a week at the same time of day, as shown on Figure 6.4, which displays the traffic for Botafogo neighbourhood between 18:00 h and 18:30 h. It is easy to spot the different behaviour between weekdays and weekends. Upon generating these images, the holidays were ignored.



6.4(a): Sunday 6.4(b): Monday 6.4(c): Tuesday 6.4(d): Wednesday

6.4(e): Thursday 6.4(f): Friday 6.4(g): Saturday

Figure 6.4: Bus traffic at Botafogo along a week.

7
Trends, Patterns, and Abnormalities

7.1
Seasonal patterns

Some routes can present a seasonal pattern along the day. These patterns can be easily spot on the *Space* × *Speed* graph when sections have a color cast.

7.1.1
Seasonal sections

In this case, a section of the route that presents a seasonal pattern is shown. In Figure 7.1 the interval between A and B has a color that mainly represents the morning hours, showing that this route has a different trajectory during this period. Furthermore, at point C there is an accumulation of samples with a high level of noise. This is explained by the presence of a tunnel (Santa Bárbara) in this segment.

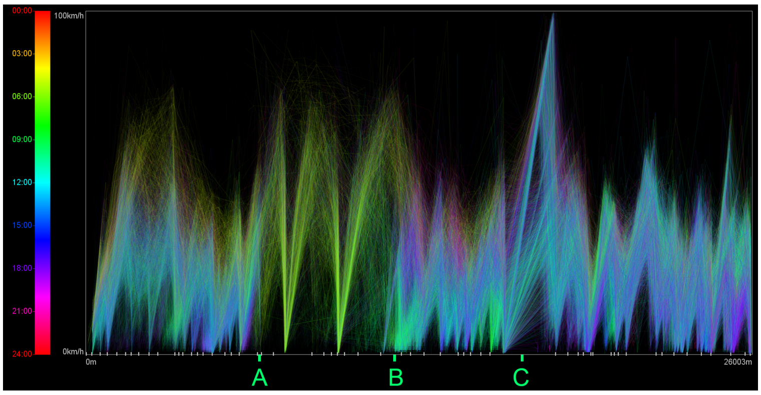
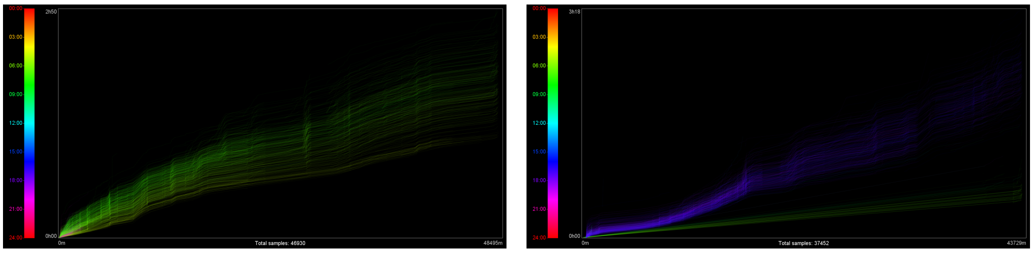


Figure 7.1: A route with seasonal section

7.1.2
Seasonal routes

The same analysis can be extend to the whole route. In this case, the whole graph presents the same color cast, as shown on Figure 7.2. For this line, the trip towards downtown runs only during the morning, while the return trip runs only by the end of the afternoon.



7.2(a): Route towards downtown

7.2(b): Route away from downtown

Figure 7.2: A line with seasonal routes.

7.2 Detection of outdated route geometry

Several routes changed since the last time the City Hall published the data. An example can be viewed on Figure 7.3, where a region of the route did not present samples.

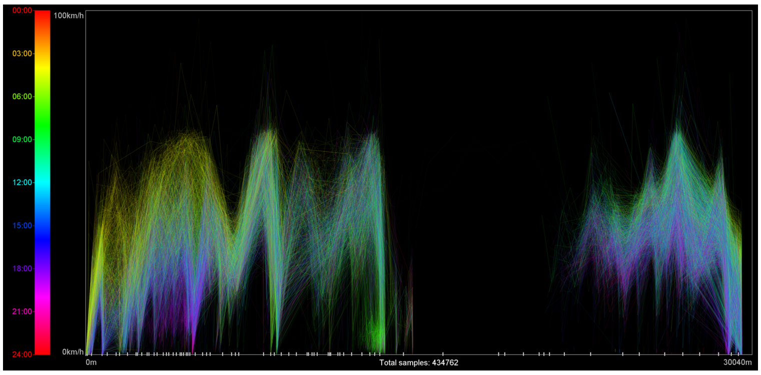


Figure 7.3: Example of a route with outdated geometry.

7.3 Mirror pattern of opposite routes

Routes with opposite directions of the same line have a similar pattern, but in a mirrored way. Figure 7.4 shows, on the bottom, the way back of the route presented on the top, but with a mirrored X axis (form the end to the beginning). Note how both graphs show a similar pattern. Although they seem to be very similar, they differ subtly on the way the buses accelerate. We can observe that accelerating is slower than stopping, probably due the traffic ahead.

7.4 Stacking graphs

This sections presents a short analysis that shows the advantages of stacking the graphs, as this reinforces patterns between them.

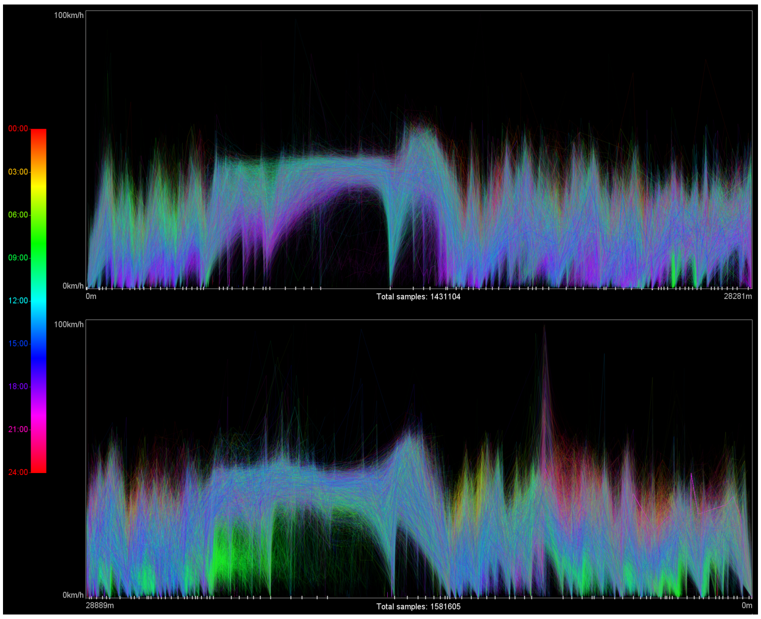


Figure 7.4: Mirror pattern on routes of the same line number, but in opposite directions.

High offer bus lines

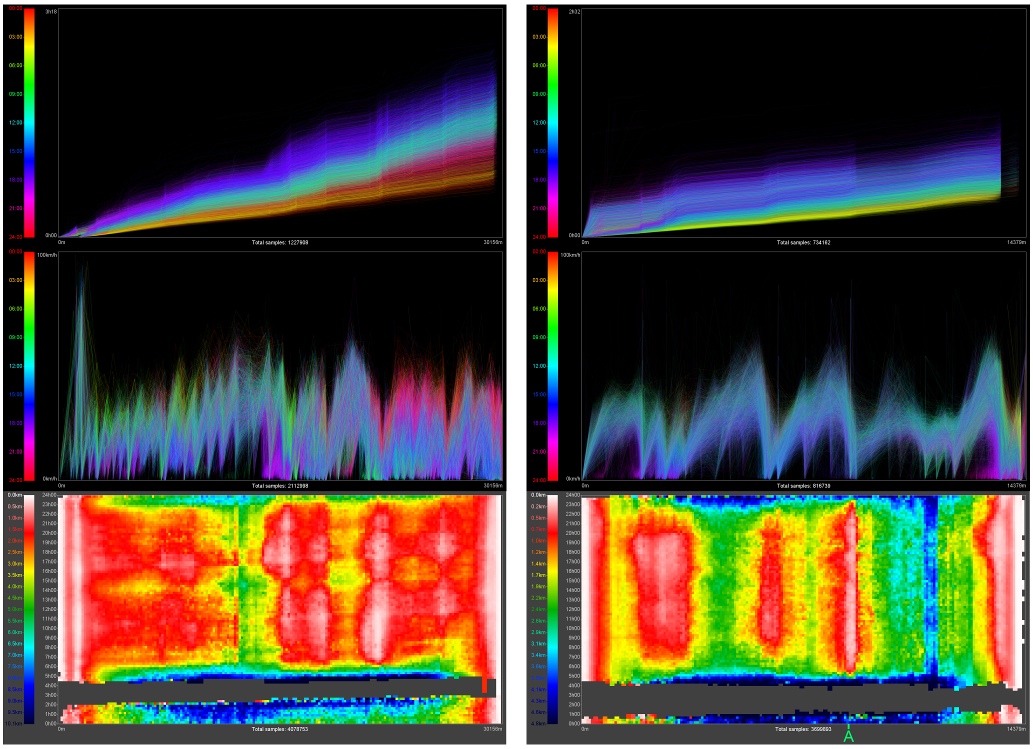
A high offer of buses can be spotted by a stack of graphs presenting an overall low-speed, with lots of stops, and a high concentration on the number of buses serving at the same time, as seen in Figure 7.5(a).

Localized bus offer

For most routes the bus offer is concentrated on both extremes. But for some, as shown in Figure 7.5(b), an extra offer along the way (marked on the bottom graph as A) can be seen. A high concentration of buses and a relative slower speed distinguish this pattern. The *Space* × *Time* graph shows an interlaced pattern in this region, showing that the buses changed its expected behaviour.

Typical traffic patterns

Traffic patterns can be easily seen when there is a high concentration of buses between specific times and a color concentration on lower speeds at the speed graph, as seen in Figure 7.6(a).



7.5(a): Bus Line # 636 with a high offer pattern.

7.5(b): Bus Line # 778 with a localized offer highlighted by *A*.

Figure 7.5: Graph stack of bus lines # 636 and # 778.

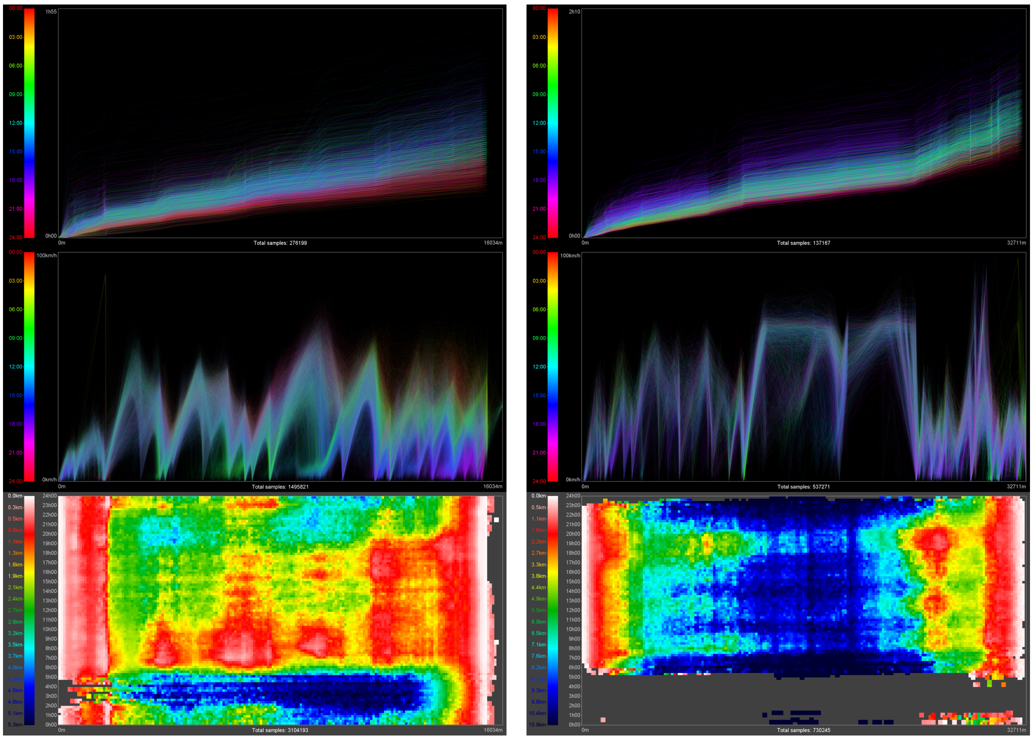
Highways patterns

Routes that go through a highway show a pattern of high speeds, without stops, and a low concentration of buses, as shown on Figure 7.6(b).

7.5 Map of a typical workday traffic

The analysis of the traffic along workdays is important do find time-related bottlenecks and to help workers to plan theirs daily routine. The latter is an advantage of a map representation of traffic, since it has widely used applications.

To create relevant maps for this task, the *Typical bus traffic map* should be generated in a daily basis, including only workdays, and its desired *time-slices* used to generate images of the traffic. Figure 7.7 shows an example of these maps, generated for every 2 hours from 6:00 until 20:00.



7.6(a): Bus Line # 862, showing a typical traffic pattern.

7.6(b): Bus Line # 338, showing a highway pattern.

Figure 7.6: Graph stack of bus lines # 862 and # 338.

7.6 Analysis of BRS São Clemente

The BRS São Clemente started on August 9th 2015 at Botafogo with an length of about 3640m, as shown on Figure 7.8.

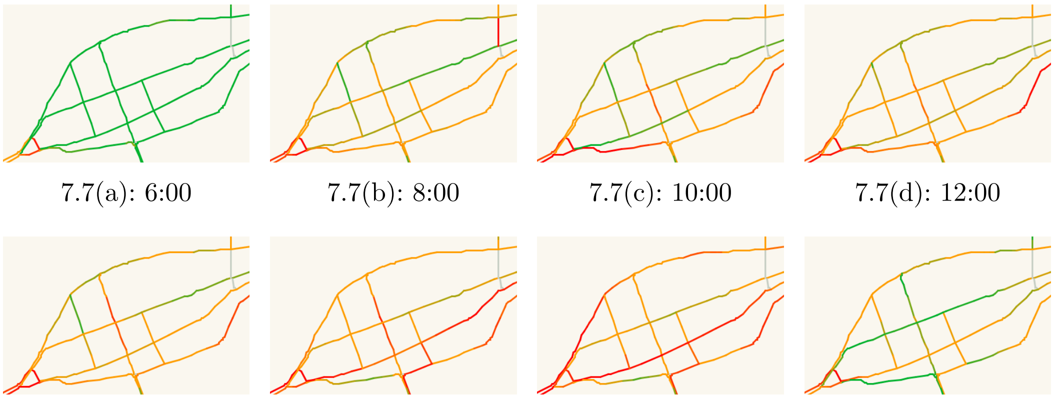
Many bus lines passes through it and, therefore, a considerable amount of information can be extracted from its geometry. In the following sections this geometry is analysed to detect some patterns and abnormalities.

7.6.1 Traffic during school vacations

Along BRS São Clemente there are at least 2 schools, named Santo Inácio (SI) and Escola Alemã Corcovado (EAC). During January both are on vocation and one of its side-effects is the traffic along the BRS having a better flow.

Figure 7.9 shows a comparison of the months January and April. The *Bus Distance Heatmap* displays an increase of bus concentration at around 7:00 on Figure 7.9(b), just before the marks representing both schools (on the bottom of both images). This pattern repeats at around 13:00 too, but at a lower intensity.

The *Space* × *Time* graph displays an increase of trip-time during the late



7.7(a): 6:00

7.7(b): 8:00

7.7(c): 10:00

7.7(d): 12:00

7.7(e): 14:00

7.7(f): 16:00

7.7(g): 18:00

7.7(h): 20:00

Figure 7.7: Typical traffic on Botafogo during workdays.

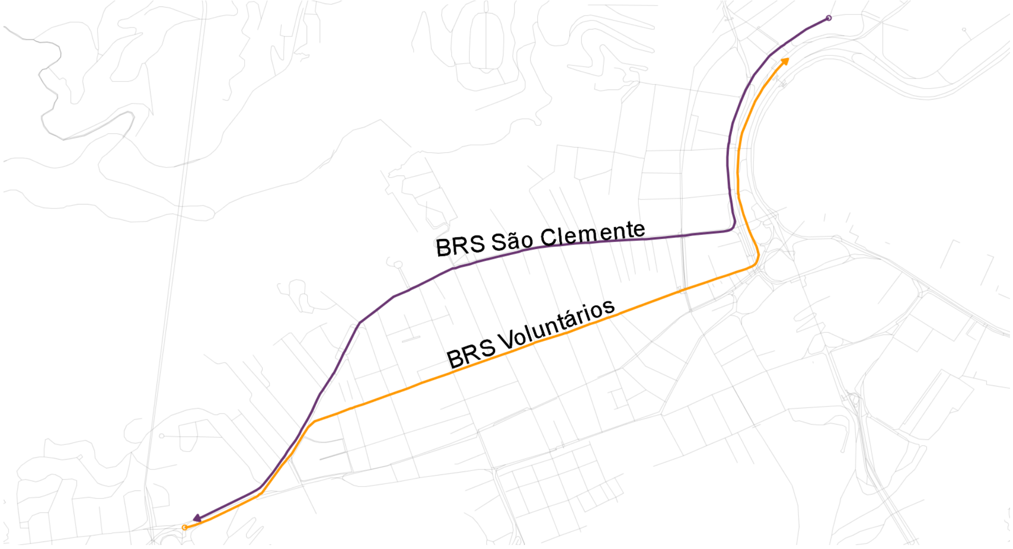


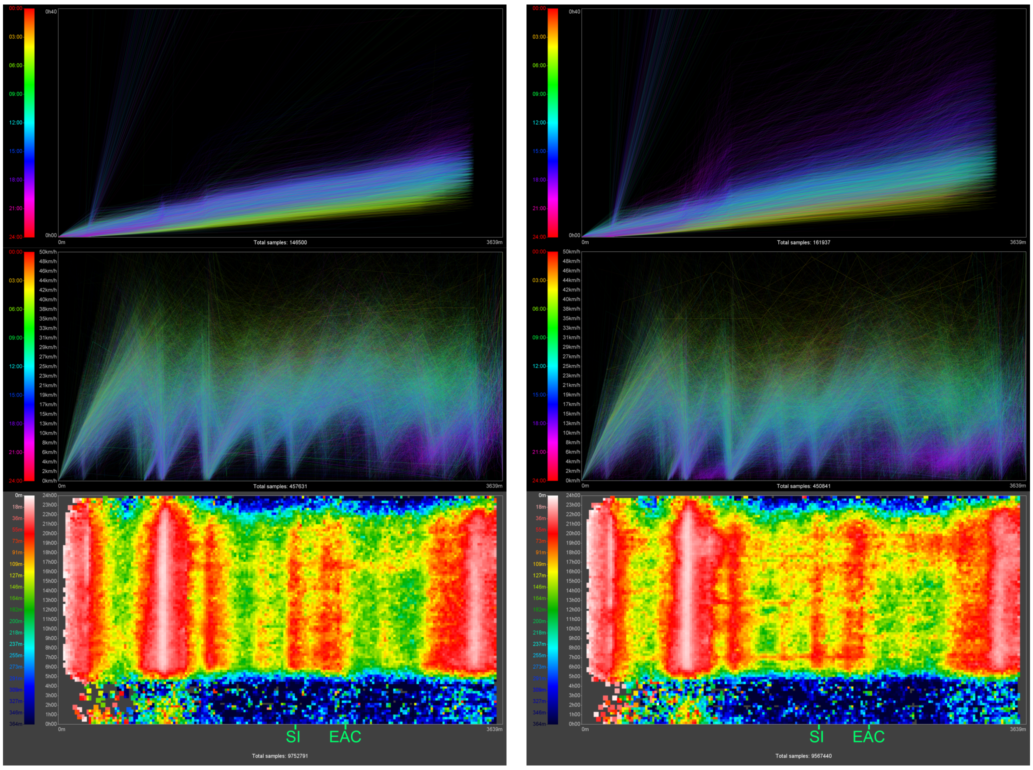
Figure 7.8: Geometry of BRS Botafogo.

afternoon, showing that the traffic is affected even outside school time. This may be due to the increase of people on vacations with their children.

7.6.2 Spotting daily problems on traffic using heatmap

Using a daily-generated *Distance heatmap* one can spot patterns specific to each day. The following sections presents some patterns found using these graphs and some information gathered from social networks to validate these analysis.

Note that a daily *Distance heatmap* can be done only for routes with a high number of samples, otherwise a noise (or empty) heatmap would be generated.



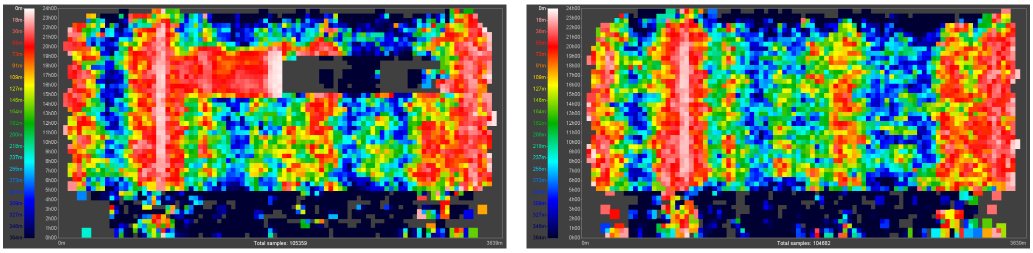
7.9(a): January

7.9(b): April

Figure 7.9: Graph stack of BRS São Clemente at different months.

Street interdiction due to an event Figure 7.10(a) displays a pattern of an event that occurred on 1st March 2015 that interdicted part of the analysed route. This interdiction caused the buses to accumulate at some point and no samples were presented afterwards. Figure 7.11(a) shows tweets (in Portuguese) from some relevant accounts (namely *@oglobo_VaiRio*, *@LeiSecaRJ* and *@WazeTrafficRIo*) within 1st Match containing the words "sao clemente", which reinforces the pattern found.

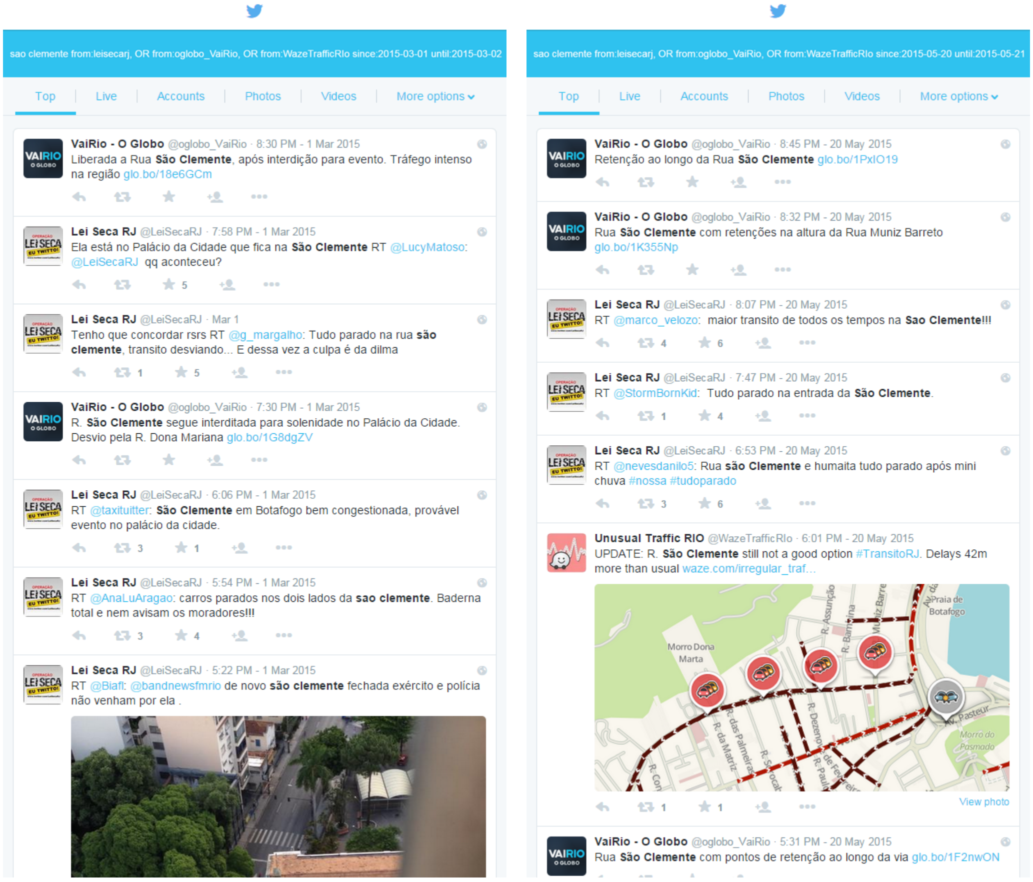
Abnormal traffic Figure 7.12(a) displays a pattern of an abnormal traffic that occurred on 20th May 2015. The graph shows a high and unusual concentration of buses from 18:00 to 21:00, representing by a large red-to-white strip. Figure 7.11(b) shows relevant tweets (as stated before) within 20th May that reinforces the pattern found.



7.10(a): Sunday, 1st March 2015

7.10(b): A typical Sunday

Figure 7.10: Heatmap of a street interdiction.



sao clemente from:leisecarj, OR from:oglobo_VaiRio, OR from:WazeTrafficRIo since:2015-03-01 until:2015-03-02

VaiRio - O Globo @oglobo_VaiRio · 8:30 PM - 1 Mar 2015
Liberada a Rua **São Clemente**, após interdição para evento. Tráfego intenso na região glo.bo/18e6GCm

Lei Seca RJ @LeiSecaRJ · 7:58 PM - 1 Mar 2015
Ela está no Palácio da Cidade que fica na **São Clemente** RT @LucyMatoso: @LeiSecaRJ qq aconteceu?
5

Lei Seca RJ @LeiSecaRJ · Mar 1
Tenho que concordar rsrs RT @g_margalho: Tudo parado na rua **são clemente**, transito desviando... E dessa vez a culpa é da dilma
1 5

VaiRio - O Globo @oglobo_VaiRio · 7:30 PM - 1 Mar 2015
R. **São Clemente** segue interditada para solenidade no Palácio da Cidade. Desvio pela R. Dona Mariana glo.bo/1G8dgZV

Lei Seca RJ @LeiSecaRJ · 6:06 PM - 1 Mar 2015
RT @taxituitter: **São Clemente** em Botafogo bem congestionada, provável evento no palácio da cidade.
3 1

Lei Seca RJ @LeiSecaRJ · 5:54 PM - 1 Mar 2015
RT @AnaLuAragao: carros parados nos dois lados da **sao clemente**. Baderna total e nem avisam os moradores!!!
3 4

Lei Seca RJ @LeiSecaRJ · 5:22 PM - 1 Mar 2015
RT @Biafl: @bandnewsfmrio de novo **são clemente** fechada exército e polícia não venham por ela .

7.11(a): Street block on Sunday, 1st March 2015

sao clemente from:leisecarj, OR from:oglobo_VaiRio, OR from:WazeTrafficRIo since:2015-05-20 until:2015-05-21

VaiRio - O Globo @oglobo_VaiRio · 8:45 PM - 20 May 2015
Retenção ao longo da Rua **São Clemente** glo.bo/1PxIO19

VaiRio - O Globo @oglobo_VaiRio · 8:32 PM - 20 May 2015
Rua **São Clemente** com retenções na altura da Rua Muniz Barreto glo.bo/1K355Np

Lei Seca RJ @LeiSecaRJ · 8:07 PM - 20 May 2015
RT @marco_velozo: maior transito de todos os tempos na **Sao Clemente**!!!
4 6

Lei Seca RJ @LeiSecaRJ · 7:47 PM - 20 May 2015
RT @StormBornKid: Tudo parado na entrada da **São Clemente**.
1 4

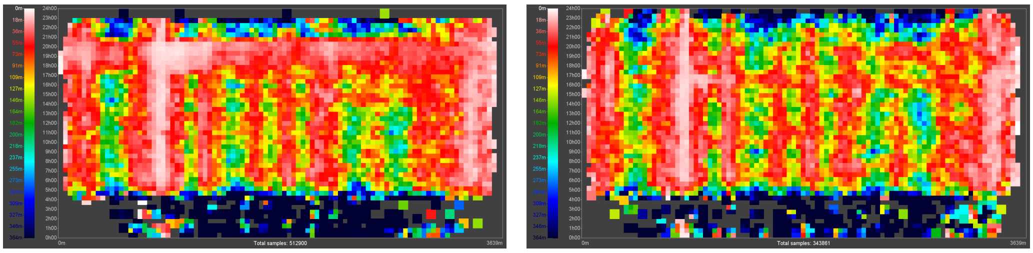
Lei Seca RJ @LeiSecaRJ · 6:53 PM - 20 May 2015
RT @nevesdanilo5: Rua **são Clemente** e humaita tudo parado após mini chuva #nossa #tudoparado
3 6

Unusual Traffic RIO @WazeTrafficRIo · 6:01 PM - 20 May 2015
UPDATE: R. **São Clemente** still not a good option #TransitoRJ. Delays 42m more than usual waze.com/irregular_traf...
1 1

VaiRio - O Globo @oglobo_VaiRio · 5:31 PM - 20 May 2015
Rua **São Clemente** com pontos de retenção ao longo da via glo.bo/1F2nwON

7.11(b): Unusual traffic on Wed., 20th May 2015

Figure 7.11: Relevant tweets during the events analysed.



7.12(a): Wed., 20th May 2015

7.12(b): A typical Wednesday

Figure 7.12: Heatmap of an abnormal traffic.

8 Conclusion and future works

In this dissertation, a system for the visual analysis of bus data was developed. It presents easy-to-understand graphical tools for a massive historical GPS data of a public bus transportation system. The proposed graphical tools are useful for identifying patterns and abnormalities and can help decision makers to better plan the city transportation services.

This work is only an initial step, there are several research avenues to investigate the bus GPS dataset used in the dissertation, including:

- Computation of bus routes based on the historical data and a map template (such as Open Street Maps). The data used in this work can be considered outdated (such as in 7.2) or inadequate in some cases. The routes derived from the dataset can be more reliable, including segmentation for time of the day or day of week;
- Computation of the bus stops along the routes and its demand, based on the time spent by the buses;
- Computation of bus frequency and average time of arrival at each bus stop;
- Once the dataset completes more than an year of data, one can compute annual changes in buses or traffic behaviour;
- The dataset can be used to develop and evaluate models for on-line prediction of estimated time of arrival;
- A real time system can be developed to spot traffic related problems on the city. An integration to social networks can improve this system even more;
- Crossing the information of maximum allowed speed on each street with the maximum speed observed on the dataset can spot hot zones of bus speeding, thus providing information for planing bus control.

9
Bibliography

1. ZUMEL, N.; MOUNT, J. **Practical Data Science with R**. [S.l.]: Manning Publications, 2014.

2. MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data: A revolution that will transform how we live, work, and think**. [S.l.]: Houghton Mifflin Harcourt, 2013.

3. WARE, C. **Information visualization: perception for design**. [S.l.]: Elsevier, 2012.

4. KHAN, Z. et al. Towards cloud based big data analytics for smart future cities. **Journal of Cloud Computing: Advances, Systems and Applications**, SpringerOpen, 2015.

5. ZHANG, J.-D.; XU, J.; LIAO, S. S. Aggregating and sampling methods for processing gps data streams for traffic state estimation. **Intelligent Transportation Systems, IEEE Transactions on**, IEEE, v. 14, n. 4, p. 1629–1641, 2013.

6. SHI, W.; KONG, Q.-J.; LIU, Y. A gps/gis integrated system for urban traffic flow analysis. In: IEEE. **Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on**. [S.l.], 2008. p. 844–849.

7. ZHU, B.; XU, X. Urban principal traffic flow analysis based on taxi trajectories mining. In: **Advances in Swarm and Computational Intelligence**. [S.l.]: Springer, 2015. p. 172–181.

8. SUNIL, N. G. et al. Dynamic bus timetable using gps. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, v. 3, n. 3, 2014.

9. GEISLER, S. et al. An evaluation framework for traffic information systems based on data streams. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 23, p. 29–55, 2012.

10. RAMEZANI, M.; GEROLIMINIS, N. Queue profile estimation in congested urban networks with probe data. **Computer-Aided Civil and Infrastructure Engineering**, Wiley Online Library, v. 30, n. 6, p. 414–432, 2015.

11. KUANG, W.; AN, S.; JIANG, H. Detecting traffic anomalies in urban areas using taxi gps data. **Mathematical Problems in Engineering**, Hindawi Publishing Corporation, v. 501, p. 809582, 2015.

12. KARGUPTA, H.; SARKAR, K.; GILLIGAN, M. Minefleet®: an overview of a widely adopted distributed vehicle performance data mining system. In: ACM. **Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2010. p. 37–46.

13. JOSE, D.; PRASAD, S.; SRIDHAR, V. Intelligent vehicle monitoring using global positioning system and cloud computing. **Procedia Computer Science**, Elsevier, v. 50, p. 440–446, 2015.

14. CHEN, C. et al. Real-time detection of anomalous taxi trajectories from gps traces. In: **Mobile and Ubiquitous Systems: Computing, Networking, and Services**. [S.l.]: Springer, 2012. p. 63–74.

15. BRAKATSOULAS, S. et al. On map-matching vehicle tracking data. In: VLDB ENDOWMENT. **Proceedings of the 31st international conference on Very large data bases**. [S.l.], 2005. p. 853–864.

16. LOU, Y. et al. Map-matching for low-sampling-rate gps trajectories. In: **Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**. New York, NY, USA: ACM, 2009. (GIS '09), p. 352–361. ISBN 978-1-60558-649-6. Disponível em: <http://doi.acm.org/10.1145/1653771.1653820>.

17. LU, C.-T.; BOEDIHARDJO, A. P.; ZHENG, J. Aitvs: Advanced interactive traffic visualization system. In: IEEE. **Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on**. [S.l.], 2006. p. 167–167.

18. KUMAR, P. et al. Framework for real-time behavior interpretation from traffic video. **Intelligent Transportation Systems, IEEE Transactions on**, IEEE, v. 6, n. 1, p. 43–53, 2005.

19. LU, C.-T. et al. Homes: highway operation monitoring and evaluation system. In: ACM. **Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems**. [S.l.], 2008. p. 85.

20. GUO, H. et al. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In: IEEE. **Pacific Visualization Symposium (PacificVis), 2011 IEEE**. [S.l.], 2011. p. 163–170.

21. LIU, S. et al. Vait: A visual analytics system for metropolitan transportation. **Intelligent Transportation Systems, IEEE Transactions on**, IEEE, v. 14, n. 4, p. 1586–1596, 2013.

22. WANG, Z. et al. Visual exploration of sparse traffic trajectory data. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, 2014.

23. FERREIRA, N. et al. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 19, n. 12, p. 2149–2158, 2013.

24. ZENG, W. et al. Visualizing mobility of public transportation system. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 20, n. 12, p. 1833–1842, 2014.

25. CHEN, W.; GUO, F.; WANG, F.-Y. A survey of traffic data visualization. **Intelligent Transportation Systems, IEEE Transactions on**, IEEE, 2015.

26. KORMAKSSON, M. et al. Bus Travel Time Predictions Using Additive Models. **arXiv preprint arXiv:1411.7973**, 2014. Disponível em: <http://arxiv.org/abs/1411.7973>.

27. Dados Rio. **GPS dos ônibus - Dados Rio - Prefeitura da Cidade do Rio de Janeiro**. 2014. Disponível em: <http://data.rio.rj.gov.br/dataset/gps-de-onibus>.

28. AMARAL, B. G. do; CASANOVA, M. A.; LOPES, H. Building traffic maps from bus gps data in rio. 2015.

29. Groovy. **Groovy Language**. 2007. Disponível em: <http://www.groovy-lang.org/>.

30. Google Maps. **Google Maps**. 2015. Disponível em: <https://maps.google.com/>.

31. SNYDER, J. P. **Map projections–A working manual**. [S.l.]: USGPO, 1987.

32. BATTERSBY, S. E. et al. Implications of web mercator and its use in online mapping. **Cartographica: The International Journal for Geographic Information and Geovisualization**, UT Press, v. 49, n. 2, p. 85–101, 2014.