Using Bundling to Visualize Multivariate Urban Mobility Structure Patterns in the Sao Paulo Metropolitan Area

Tallys Gustavo Martins

University of Sao Paulo Institute of Mathematics and Statistics: Universidade de Sao Paulo Instituto de Matematica e Estatistica

Nelson Lago

University of Sao Paulo Institute of Mathematics and Statistics: Universidade de Sao Paulo Instituto de Matematica e Estatistica

Eduardo Felipe Zambom Santana

University of Sao Paulo Institute of Mathematics and Statistics: Universidade de Sao Paulo Instituto de Matematica e Estatistica

Alexandru Telea

Utrecht University Department of Information and Computing Sciences

Fabio Kon

University of Sao Paulo Institute of Mathematics and Statistics: Universidade de Sao Paulo Instituto de Matematica e Estatistica

Higor Amario de Souza (✉ hamario@ime.usp.br)

University of Sao Paulo Institute of Mathematics and Statistics <https://orcid.org/0000-0003-4233-5987>

Research

Keywords: Bundling, Urban Mobility, Data Visualization, Travel Surveys, Smart Cities, Open Data

Posted Date: November 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101282/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Using bundling to visualize multivariate urban mobility structure patterns in the São Paulo Metropolitan Area

Tallys G. Martins^{1*}, Nelson Lago¹, Eduardo F. Z. Santana¹, Alexandru Telea², Fabio Kon¹ and Higor A. de Souza¹

*Correspondence:

tallys@ime.usp.br

¹Department of Computer

Science, University of São Paulo,

São Paulo, Brazil

Full list of author information is available at the end of the article

Abstract

Internet-based technologies such as IoT, GPS-based systems, and cellular networks enable the collection of geolocated mobility data of millions of people in large metropolitan areas. In addition, large, public datasets are made available on the Internet by open government programs, providing ways for citizens, NGOs, scientists, and public managers to perform a multitude of data analysis with the goal of better understanding the city dynamics to provide means for evidence-based public policymaking. However, it is challenging to visualize huge amounts of data from mobility datasets. Plotting raw trajectories on a map often causes data occlusion, impairing the visual analysis. Displaying the multiple attributes that these trajectories come with is an even larger challenge. One approach to solve this problem is trail bundling, which groups motion trails that are spatially close in a simplified representation. In this paper, we augment a recent bundling technique to support multi-attribute trail datasets for the visual analysis of urban mobility. Our case study is based on the travel survey from the São Paulo Metropolitan Area, which is one of the most intense traffic areas in the world. The results show that bundling helps the identification and analysis of various mobility patterns for different data attributes, such as peak hours, social strata, and transportation modes.

Keywords: Bundling; Urban Mobility; Data Visualization; Travel Surveys; Smart Cities; Open Data

1 Introduction

Governments around the world are providing open datasets about city resources, such as citizen surveys, data collected with the help of Internet of Things devices, accountability and transparency reports, and so on. The Internet facilitates the gathering and the availability of a large variety of data sources, which allows citizens, researchers, companies, non-profit organizations, and public agents to perform analyses on their matters of interest regarding city-related issues. In the public administration realm, these data sources can be used to provide better services for citizens, to improve the management of urban infrastructure, or to reduce bureaucracy costs, supporting evidence-based policymaking [1, 2].

Urban mobility is a great concern for citizens and governments. It directly affects people's quality of life, causes prejudices to the environment, and has a high economical impact. For example, the traffic congestion in the São Paulo Metropolitan Area (SPMA) is estimated to affect 89% of work-related commuting trips^[1], causing monetary losses of seven billion Brazilian reais (~US\$1.8 billion) every year [3]. Thus, the development of more efficient transportation systems is a critical issue that cities should tackle.

Several data sources can be used for urban mobility analysis such as data captured with IoT devices such as traffic cameras, GPS tracking, bike-sharing systems, as well as censuses and trip surveys [4, 5]. In the SPMA, every ten years since 1967, the São Paulo Metropolitan Company (Metrô), which manages the subway system in the city of São Paulo, conducts a travel study called *Origin-Destination (OD) survey*. This survey is performed by interviewing citizens about their life and commuting activities on a typical working day, resulting in a comprehensive panorama of the mobility behavior of the population over the SPMA. The last OD survey (2017) shows that there are around 42 million trips over the 24 hours of a regular working day. Beyond a trip's origin (O) and destination (D) data itself, the survey also gathers a wide number of trip-related and social-economic aspects, or data attributes, such as transportation modes used, trip reasons, age, gender, and household income. Hence, the OD survey generates a large and multivariate dataset.

While the above-mentioned OD survey is very comprehensive and accurate, urban planners need proper tools to analyze this large amount of multivariate data.

^[1]We use the terms trip, travel, trail, or trajectory interchangeably during the paper

¹Spreadsheets and statistics programs help producing tables and charts with aggre-¹
²gated information showing, *e.g.*, the number of users of the public transportation²
³system over the years or the number of women *vs* men that commute for work. Vi-³
⁴sual techniques, such as density maps, can help to answer geolocated data-related⁴
⁵questions, such as finding regions that concentrate most of the mobility flow dur-⁵
⁶ing the day or the most common origin-destination pairs. However, considering the⁶
⁷huge amount of information (and attributes) that is produced every day in any city,⁷
⁸translating geolocated data into meaningful images is very challenging.⁸

⁹ A recent survey about traffic visualization methods indicates the common use of⁹
¹⁰*line-based* visualization techniques to study the structure of mobility data, which¹⁰
¹¹often includes pattern discovery and clustering tasks [6]. Still, drawing OD lines can¹¹
¹²only handle thousands of trajectories. Visualizing 42 million trajectories, as recorded¹²
¹³daily in the last OD survey, would only generate a fully cluttered image (as can be¹³
¹⁴seen in Figure 5). *Bundling techniques* improve upon line-based techniques and can¹⁴
¹⁵depict high volumes of trajectory data by essentially simplifying trajectories in the¹⁵
¹⁶image space, grouping spatially close and data-wise similar trajectories together [7,¹⁶
¹⁷8].¹⁷

¹⁸ In this paper, we use bundling to visualize the structure and patterns of urban¹⁸
¹⁹mobility in the São Paulo Metropolitan Area. For this, we adapt the recent CUBu¹⁹
²⁰framework [9] – Compute Unified Device Architecture (CUDA) bundling, which²⁰
²¹has proven to be useful in other scenarios using large movement datasets (flight²¹
²²traffic, eye tracking) [10, 11, 12]. CUBu generates different bundling styles, grouping²²
²³trips per distance, density, or direction. We adapt CUBu to use bundling to find²³
²⁴and visually explore mobility patterns created by different combinations of the²⁴
²⁵attributes present in the OD survey, such as traffic on peak hours, trips per social²⁵
²⁶strata, different transportation modes, and trips made for distinct reasons. We show²⁶
²⁷how the bundled visualizations help identify different structural patterns of urban²⁷
²⁸mobility in São Paulo, providing insights into the data characteristics that were²⁸
²⁹analyzed.²⁹

³⁰ This paper is an extended version of a paper accepted for publication in the³⁰
³¹IV Brazilian Workshop of Urban Computing (CoUrb 2020) [13]. It is organized³¹
³²as follows. Section 2 outlines related work with a focus on OD data visualization³²
³³and trajectory bundling. Section 3 presents the main characteristics of the SPMA³³

¹and the OD survey data. Section 4 describes our visualization approach. Section 5¹
²presents the results of our analysis and discusses the use of bundling in urban mobil-²
³ity visualization. In addition to the original explorations presented in [13], Section 5³
⁴shows the visual exploration of several other attribute combinations: density per⁴
⁵social strata (Sec. 5.5), mobility of young students per social strata (Sec. 5.6), di-⁵
⁶rections at peak hours (Sec. 5.7), density by transportation mode (Sec. 5.8), and⁶
⁷trip distance per trip reasons (Sec. 5.9). Section 6 concludes the paper. ⁷

⁸ **2 Background and related work** ⁸

⁹We next discuss related work on trajectory bundling with a focus on OD data. ⁹

¹⁰*Trajectories* (or for short, trails) are typically sets of spatial points $\mathbf{t} = \{\mathbf{x}_i\}$ ¹⁰
¹¹recorded at consecutive time moments t_i that describe the motion of an object¹¹
¹²over time. Origin-Destination (OD) data are particular cases of trails which contain¹²
¹³only the first and last recorded point. Trails can have additional data attributes $\mathbf{a}_j$,¹³
¹⁴recorded either at each sample point, *e.g.*, speed $\mathbf{a}_j(\mathbf{x}_i)$; or globally for the entire¹⁴
¹⁵trail, *e.g.*, vehicle type $\mathbf{a}_j$. As such, an OD dataset can be seen as a multivariate¹⁵
¹⁶dataset with trails being the observations and the measured attributes being the¹⁶
¹⁷dimensions or variables. Trail data can be directly plotted atop of 2D maps, with¹⁷
¹⁸selected data attributes encoded into visual variables such as color, line thickness,¹⁸
¹⁹and line style. Besides statically drawing entire trail-sets, these can be also shown¹⁹
²⁰by animating particles along their points $\mathbf{x}_i$ [14, 15]. ²⁰

²¹Visualizing trail-sets having thousands of trails or more, each with multiple at-²¹
²²tributes, is challenging. Simply drawing the trails yields significant clutter which²²
²³makes finding even the simplest spatial patterns very hard. Particle techniques also²³
²⁴do not scale well – with over roughly 50 thousand points, the resulting visual pattern²⁴
²⁵resembles Brownian motion. *Bundling* methods aim to solve this visual scalability²⁵
²⁶issue by grouping trail fragments which are spatially close and, optionally, have sim-²⁶
²⁷ilar data attributes. This creates more *overdraw* between (similar) trails, but also²⁷
²⁸generates visually empty space between trail groups that reduces overall *clutter*. ²⁸

²⁹Bundling is used in several other studies on movement data. Graser et al. [16]²⁹
³⁰created a bundled visualization to study the characteristics of bird migrations. Klein³⁰
³¹et al. [15] developed a dynamic visualization for aircraft flight data where the trail-³¹
³²set itself varies over time. Willems et al. [10] analyzed maritime vessel traffic using³²
³³

¹bundling techniques. Separately, bundling was used to simplify eye-tracking data to¹
²infer reading patterns [11, 12]. Lhuillier et al. [8] compiled a survey on the state of²
³the art on bundling techniques. This survey distinguishes between bundling graph³
⁴data (straight-line drawings of graph layouts) and trail data (such as our OD data),⁴
⁵highlighting the best methods for both cases. ⁵

⁶ *Image based* bundling (IBB) techniques are a particular class. They essentially⁶
⁷exploit mean shift clustering [17], well known in image processing. We outline this⁷
⁸process for the Kernel Density Estimation Edge Bundling (KDEEB) technique [18],⁸
⁹although all other IBB techniques we are aware of work similarly and share the same⁹
¹⁰properties (see also Fig. 1). Let $T = \{\mathbf{t}_i\}$ be the trail set to bundle, and let $S = \{\mathbf{p}_i\}$ ¹⁰
¹¹be a set of points that densely sample T . First, a density map $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is¹¹
¹²computed from S by kernel density estimation (KDE), *i.e.* by convolving the points¹²
¹³in S with an Epanechnikov kernel of radius k . Hence, ρ will be high in regions of high¹³
¹⁴density of the sampling points – and by implication, of trails – and low elsewhere.¹⁴
¹⁵Second, the points in S are advected upwards in the density gradient $\nabla\rho$ with a¹⁵
¹⁶small step. Terminal (O and D) points are kept fixed so that one can still identify¹⁶
¹⁷the origin and destination of the bundled trails next. Finally, a 1D Laplacian filter¹⁷
¹⁸is used to smooth the trails in T , thereby removing inherent numerical noise in¹⁸
¹⁹the advection. The process is repeated for N iterations, whereby k is continuously¹⁹
²⁰decreased so as to ‘slow down’ the advection and make trails converge around the²⁰
²¹local density maxima. The main advantages of IBB are simplicity of implementation,²¹
²²control over the simplification induced by the bundling (given by the parameter k),²²
²³and, notably, speed, as image-space operations map well to GPU functionality. ²³

²⁴ Zeng et al. [19] adapted the KDEEB technique into Road Aware Edge Bundling²⁴
²⁵(RAEB), which constraints the bundles along the road network on which the respec-²⁵
²⁶tive trails are recorded. RAEB was demonstrated on 166K taxi trajectories from²⁶
²⁷New York City. RAEB is the only work that uses bundling with urban mobility²⁷
²⁸data that we know of. However, RAEB requires detailed trails data (not OD data)²⁸
²⁹and the layout of the city road network. ²⁹

³⁰ CUBu, which stands for CUDA bundling, is an improved implementation of³⁰
³¹KDEEB [9]. It leverages CUDA^[2] to implement all IBB steps (trail sampling, KDE³¹
³²computation, advection, smoothing, and final rendering) on the GPU, thereby sur-³²

³³^[2]developer.nvidia.com/cuda-zone

¹passing KDEEB and older IBB methods [20, 21] in speed by one to two orders¹
²of magnitude. Additionally, CUBu supports several bundling styles, most notably²
³*directional* bundling [20], which separates spatially close trails running in opposite³
⁴directions in different bundles. Rendering-wise, CUBu supports a large palette of⁴
⁵options, such as pseudo-shading to emphasize high-density bundles [22], encoding⁵
⁶bundle importance into opacity, and color mapping trail attributes (*e.g.*, direction,⁶
⁷time, length).⁷

⁸ Overall, from all bundling methods surveyed in Lhuillier et al. [8], CUBu is the⁸
⁹fastest and offers the most options to visually encode data. While these options⁹
¹⁰are well described in Van Der Zwan et al. [9], they have however not been, to our¹⁰
¹¹knowledge, fine-tuned and actually *deployed* to support a real-world use-case of¹¹
¹²analyzing very large trail sets, let alone trails from mobility data. Indeed, CUBu¹²
¹³is explicitly proposed as a *framework* to realize many types of bundling. However,¹³
¹⁴it is up to the application domain expert to customize this framework to obtain¹⁴
¹⁵actual results. In the remainder of this paper, we show how we parameterize and¹⁵
¹⁶customize CUBu to analyze our subject urban mobility dataset, along the following¹⁶
¹⁷dimensions:¹⁷

- ¹⁸ • *Density map*: We use the underlying density map ρ computed by the KDE¹⁸
¹⁹ process implemented in CUBu to show the local edge density in the produced¹⁹
²⁰ bundles. This way, one can easily separate visually high-density (important)²⁰
²¹ traffic flows from less important ones;²¹
- ²² • *Attribute filtering and coloring*: We use CUBu’s ability to color code trail²²
²³ directions or trail lengths to study mobility behavior. We also extended CUBu²³
²⁴ to filter specific attribute ranges, and select specific combinations of attributes²⁴
²⁵ to display. This allows the analyst to search for different types of patterns²⁵
²⁶ present in the data;²⁶
- ²⁷ • *City map*: We modified CUBu to blend its bundled result atop of a map of²⁷
²⁸ the SPMA where traffic is analyzed. This allows correlating the bundles with²⁸
²⁹ actual locations in the city.²⁹

³¹3 Mobility in São Paulo³¹

³²To provide a better understanding of the scenario of our analysis, this section de-³²
³³scribes the São Paulo Metropolitan Area and the last travel survey data performed³³

in that region. To avoid misunderstanding between the São Paulo Metropolitan Area (SPMA) and the city of São Paulo, we will refer to the city as capital or simply São Paulo. We will refer to the metropolitan area as SPMA or metro area.

3.1 São Paulo Metropolitan Area

The state of São Paulo is located in the Southeastern coast of Brazil. The São Paulo Metropolitan Area is the most populated region of South America. According to the Brazilian Institute of Geography and Statistics [23], the SPMA has 21.9 million citizens in 2020, which represents around 10% of the Brazilian population. The SPMA is composed of 39 cities in an area of 7946.84 km^2 . The city of São Paulo is the capital of the state of São Paulo and it is placed at the center of the metro area. Figure 2 shows the 39 cities that form the SPMA. The capital is the most populated city in Brazil, with 12.3 million inhabitants. In the SPMA, the other cities with more inhabitants are Guarulhos (1.4 million), São Bernardo do Campo (844 thousand), Santo André (721 thousand), and Osasco (699 thousand).

The capital and its nearest cities concentrate most of the job opportunities, public facilities and services, universities, museums, and entertainment options. Thus, there is huge daily commuting to the capital downtown and its surrounding neighborhoods. In São Paulo, neighborhoods close to the city center are highly valued, so living in these parts of the city is expensive. People with less financial conditions usually live in the peripheral areas of the capital or in other cities in the SPMA. Indeed, some cities in the SPMA are *dormitory towns* for those who cannot afford to buy or rent a house in the capital.

Regarding the transportation infrastructure, the capital has a subway system that serves the northern, southern, eastern, western, southwest, and southeast regions of the city. There are a few subway lines, most of them crossing the center of the

Figure 2: Municipalities of the SPMA.

capital, which limits access to the subway system to some regions of the city. The other cities do not have subway systems. There is a metropolitan railway system that serves several surrounding cities and also the capital. The railway system is integrated with the subway system.

Each city has its own bus system and there is an intercity bus system to link the neighboring cities. During the 20th Century until the late 1990s, most investments in transportation were guided by a car-centric approach in detriment of public transportation [24]. The aims were to enlarge roads and streets to attend the increasing demand for private cars. In the last two decades, the local governments have invested more in policies to incentivize public transportation, such as the implantation of bus corridors and the replacement of bus fleet, building new subway lines, and modernizing the railway system. Also, there is an increasing cycling infrastructure in the capital that is being expanded in the last decade. Although these investments have increased in recent years, the SPMA still suffers from traffic congestion, especially during peak hours [24, 3]. Thus, there is a need for a better understanding of the traffic behavior in the SPMA to propose new policies to improve mobility for its citizens.

3.2 Origin-Destination survey

The Origin-Destination (OD) survey is the primary source of mobility information in the SPMA. It is performed by the São Paulo Metropolitan Company (Metrô) every ten years since 1967. The last OD survey of 2017 (OD17) has information about 157 992 trips of people randomly sampled in the SPMA, which is a representative sample of the total population with an error margin of 6% and a confidence interval of 92% [25, p.18]. Trips occur for different reasons such as work, house, study, and leisure, and different trip modes such as by walk, car, subway, and train. Figure 3 shows the distribution of trips by transportation mode and by motivation. Figure 4 shows the distribution of trips by the hours of the day.

Figure 3: Trips by mode and by reason.

As Figure 3 shows, most trips in the city are made by pedestrians. Usually, pedestrian trips cover small distances (624m on average) and stay inside a single district of the city. However, there is also a significant number of trips by car and public transportation (bus, subway, and train), most of them longer (7.7 km on average), especially by train and subway (13.5 km on average). Looking at trips by hour (Figure 4), we see that traffic in the SPMA has three main peaks: in the morning (6 AM to 8 AM), at lunchtime (noon), and in the early evening (5 PM to 7 PM).

Figure 4: Trips by hour of the day.

The 157992 entries in the OD survey have several important attributes for our study, as follows (attribute type indicated between brackets): the most important ones are the *origin* and *destination* coordinates (quantitative), and the *expansion factor* of the trip (quantitative), which is the statistical extrapolation for the population size that each surveyed trip represents. That is, the expansion factor summed over all survey entries yields the 42 million trips on a typical working day. Other attributes are the *time* of departure (quantitative), the transportation *mode* (categorical, 17 values: train, subway, car driver, car passenger, bus from the São Paulo city and from other cities, intercity bus, monorail, chartered vehicle, school bus, regular taxi, non regular taxi, motorcycle driver, motorcycle passenger, pedestrians, bicycle, and others), and the *reason* for the trip (categorical: work, house, study, and leisure).

Figure 5 shows the OD survey trajectories plotted over the map of the SPMA, where each line represents an OD trail. The extreme clutter in this figure precludes the visualization of individual trajectories, traffic patterns, or connections between the regions of the map. From this image, we can only infer the existence of high traffic over the major part of the metropolitan area and a significant concentration of this traffic at São Paulo downtown. Moreover, this image does not show any of the available data attributes, except O and D. Hence, filtering and aggregation

techniques are essential to simplify the visualization and make it understandable¹
for the human eye.²

Figure 5: Naïve straight-line drawing of OD trails from the SPMA.

4 Analyzing the urban mobility structure of São Paulo

We used the OD17 data to explore how bundling can be used to visualize mobility¹⁸
patterns over the São Paulo Metropolitan Area (SPMA). The OD17 contains data¹⁹
that represents trips of a regular working day over the SPMA. We parameterized²⁰
and adapted CUBu to explore several properties of this data with different bundled²¹
visualizations. Figure 6 shows our entire pipeline. We detail each of its steps next.²²

Figure 6: Pipeline of the OD17 visual data exploration.

4.1 Data representation

The OD trajectory data we use as input for bundling is described by a table with³²
six columns: trail ID, transportation mode, departure time, arrival time, origin³³

¹coordinates, and destination coordinates. We extract this data from the OD17: the¹
²trail ID is the identifier obtained directly from the OD17 dataset. The origin and²
³destination coordinates are transformed to the latitude/longitude system (required³
⁴to geolocate these coordinates on a map). The transportation mode is stored as an⁴
⁵integer in the range 1 to 17. The expansion factor of each OD record (introduced⁵
⁶in Section 3.2) is used to replicate trails – for an OD17 trail with expansion factor⁶
⁷ E , we create E copies of that trail as input for bundling. This yields a complete⁷
⁸dataset of 42 million trails. For all such replicated trails, we keep the ID of the⁸
⁹underlying OD17 trail which generated them. This way, we can trace back which⁹
¹⁰bundled trails correspond to an OD17 record. To this base data, extra attributes¹⁰
¹¹from the OD17 survey can be added such as trip reason (see Sec. 3.2) and personal¹¹
¹²data (age, income). Overall, this yields a dataset with over ten attributes per trail.¹²
¹³Table 1 shows a sample of the generated trails. Here, trails corresponding to the¹³
¹⁴top two rows were created from the same OD17 trail with ID 50 and $E = 2$.¹⁴

Table 1: Format of data used as input for bundling.

ID	Mode	Depart	Arrive	Origin		Destination	
		time	time	Longitude	Latitude	Longitude	Latitude
50	1	6.45	7.10	−46.628 094	−23.551 692	−47.003 481	−23.393 563
50	1	6.45	7.10	−46.628 094	−23.551 692	−47.003 481	−23.393 563
51	4	8.30	9.05	−47.001 872	−23.398 469	−47.003 481	−23.393 563

4.2 Data preprocessing

²⁴Although CUBu is – according to its authors and also to our knowledge – the²⁴
²⁵fastest existing solution for trail bundling, it still cannot process 42 million trails at²⁵
²⁶interactive rates (which are required for interactive data exploration). To deal with²⁶
²⁷this problem, we reduced the OD17 dataset based on the expansion factor E .²⁷

²⁸First, we analyzed how representative were the sample records from the OD17²⁸
²⁹regarding the population of 42 million trips. We observed that all records with²⁹
³⁰ $E < 75$ represented less than 4% of the total population. Considering the error³⁰
³¹margin of the OD17 was below 6% (see Section 3.2), we could remove the records³¹
³²with $E < 75$ and still keep the representativeness of the dataset. Next, we divided³²
³³the E values of all remaining records by this threshold value of 75, resulting in³³

a lower number of duplicated edges that maintains the same data distribution.

Although 75 would be a reasonable threshold, as explained above, to be sure that we would not alter the data properties, we used in practice a less aggressive threshold of 55. Figure 7 shows the accumulated percentage of records that are removed by choosing different expansion factor values. This allowed us to reduce the number of trails from 42 million to 685 115, which resulted in less than 2.5% of information loss coupled with a significant reduction in the number of data points.

4.3 Filtering and slicing the dataset

Beyond the data reduction process explained above, we also applied two other simplification methods, *filtering* and *slicing*, to investigate specific characteristics of the dataset.

To see the relations between the different transportation modes and the traffic directions over the peak hours, we used visual filters that make undesired trajectories fully transparent in the visualization. This is particular helpful when we want to visualize relations between different data, like how the buses from SP and from other neighboring cities relates to each other, so it is suggestive to apply bundling in the entire data and then differentiate them somehow in the visualization, in this case we used different colors and filter options.

To perform the analyses related to trip reason, household income, and age of commuters, we sliced the whole dataset for each of these features separately and applied bundling to those subsets. This brings more details to the observed feature

¹because it removes the interference of other data. This is also more straightforward¹
²to execute, instead of implementing filters to all the features, albeit it is desired to²
³have both options in the analysis process. 3

⁴ 4 ⁵4.4 Parameter setting 5

⁶The literature is not clear on how to choose good bundling parameters [8]. This⁶
⁷was observed and explicitly studied in Zeng et al. [19], which also proposed ways to⁷
⁸compute good parameter settings. However, as they also mention, these settings are⁸
⁹valid for their method (RAEB) and would not generalize directly to other methods,⁹
¹⁰such as CUBu. Hence, we had to find good parameters for our study empirically.¹⁰
¹¹The obtained parameter set in this way was: image resolution $R = 512 \times 512$ pixels;¹¹
¹²trail sampling step: 10 pixels (in line with the recommendation in Van Der Zwan,¹²
¹³et al. [9] of using for this about 1% of the image diagonal dimension); kernel size¹³
¹⁴ $k = 18$ pixels; and number of bundling iterations $N = 15$. In our biggest dataset,¹⁴
¹⁵represented by all 685 115 trips, this configuration yielded around 3.2 million sample¹⁵
¹⁶points, which were bundled in around 52 milliseconds per iteration on a PC with¹⁶
¹⁷a 4GB NVidia GeForce 940MX graphics card GPU. We used this parameter set to¹⁷
¹⁸create most of the bundled visualizations shown next in Section 5, except for the¹⁸
¹⁹ones from Section 5.6 where we show the commuting patterns of different social¹⁹
²⁰strata classes. For these particular subsets, we lowered the sampling step to 5 pixels,²⁰
²¹and set alpha transparency of the trails to the fraction of 0.15. In practice, half²¹
²²of the sampling step parameter doubles the number of points in the visualization,²²
²³and also the changes in the transparency of trails reduces the amount of details²³
²⁴in the visualization. But we observed that these variations are helpful to highlight²⁴
²⁵the density spots that we seek in our analysis with no impact in the underlying²⁵
²⁶structure of the visualization. 26

²⁷ 27 ²⁸4.5 Visualization enhancements 28

²⁹To enrich our visualizations of São Paulo urban mobility, we made two important²⁸
³⁰additions to CUBu. First, we added the map of the metropolitan area as a back-²⁹
³¹ground image and plotted the subway and train lines atop of the bundled image (see³⁰
³²Sec. 5.2). This way, one can use the map to reason about where certain bundles are;³¹
³³and can use the rails to see how bundles correspond to specific train infrastructure.³²
³⁴Next, we implemented filters to select the subset of the 17 transportation modes to 33

¹display in the visualization, using different categorical colors for each of them. We¹
²used this feature to explore relations between the different modes and their impact²
³on urban mobility (see Sec. 5.4).³

⁴**5 Results**⁵

⁶This section presents the results of several analyses of the OD17. We started by⁶
⁷exploring the different visual encodings provided by CUBu (density, distance, di-⁷
⁸rection, and coloring) to visualize the distinct characteristics of the OD17 dataset.⁸
⁹These aspects are covered next in Sections 5.1, 5.2, 5.3, and 5.4. Based on these vi-⁹
¹⁰sual encodings, we next designed several visualizations for studying specific mobility¹⁰
¹¹patterns. These are detailed in Sections 5.5, 5.6, 5.7, 5.9, and 5.8.¹¹

¹²5.1 Adjusting density to visualize bundled trails¹²

¹³As explained in Section 2, standard bundling trades off clutter for overdraw. How-¹³
¹⁴ever, such a visualization does not tell us how many trails have been grouped in¹⁴
¹⁵a bundle. The typical solution for this, pioneered by Holten [26] is to draw semi-¹⁵
¹⁶transparent trails, each having a fixed transparency $\alpha < 1$. Thus, blending will show¹⁶
¹⁷high-density trails as more opaque and low-density ones as more transparent, re-¹⁷
¹⁸spectively. In addition to that (since transparency is not a strong quantitative visual¹⁸
¹⁹variable [27]), we also encode trail density, directly estimated by the KDE density¹⁹
²⁰ ρ , into color. Figure 8a shows a visualization obtained using color encoding for the²⁰
²¹entire OD17 dataset. We can see a few dense paths but the image still presents²¹
²²much clutter. This fact occurs because on consumer-grade GPUs, the transparency²²
²³ α is modeled, during blending, as an 8-bit integer value. Hence, only 255 different²³
²⁴transparency levels are possible, *i.e.*, only 255 different trail-density levels can be²⁴
²⁵displayed. Setting α too high would immediately saturate the transparency chan-²⁵
²⁶nel where higher trail densities occur – all densities above 255 are clamped to 255.²⁶
²⁷Setting α lower than $1/255$ would result in no image, since this would correspond²⁷
²⁸to zero opacity on the 8-bit representation.²⁸

²⁹We address this problem by actually mapping the density ρ in both transparency²⁹
³⁰and color. Since ρ is computed with floating-point accuracy on the GPU during³⁰
³¹KDE, no clamping or rounding-off issues occur. This density-based transparency³¹
³²modulation is useful to highlight even more the high-density areas and reduce gen-³²
³³eral clutter – an alternative to the use of higher kernel sizes that would create³³

stronger bundles but also cause more edge distortion. Figure 8b shows the resulting visualization on the same data as in Figure 8a. Now high-density bundles appear more salient – there is more dynamic range in this image. The image suggests that the traffic network of the metropolitan area can be divided into a few core branches that are strongly connected to the central area, where the city of São Paulo is placed. Indeed, this makes sense because this is the most populous part of the metropolitan area. Moreover, most transportation systems cross the capital downtown area, including subway and bus lines, and the main expressways.

5.2 Overlapping subway and train infrastructure vs bundled trails

The public transportation system is the most used by the citizens of the SPMA.

The impact of the rail mesh over the commuting of people is clear when we plot the rail lines over the bundled trajectories, as regions with the highest flows correspond to the paths of the rail lines (see Figure 9, where the rail lines are the black lines).

This is an expected result since according to the OD survey, about 44% of daily trips by public transportation involve the subway or trains. More interestingly, one may wonder whether the rail system was accurately planned to supply the demand, as the bundled visualization suggests, or whether the availability of this transportation option influenced the existence of such dense flows. While we may not know how to answer this question, traffic managers might use this kind of visualization to devise policies for public transportation. The high correlation between the paths and the rail lines also indicate good parameter settings for the bundled visualization in the metropolitan scale.

Note that this type of correlation (of bundles with rails) is not the same as RAEB [19]. In RAEB, the bundling was *explicitly* made to follow roads. In our case, roads are *superimposed* by a bundling that uses only the OD data. One may argue that RAEB, in this sense, produces more ‘correct’ bundles since these are constrained to follow the roads. However, upon a closer look, we can see that RAEB *cannot* have all bundles precisely follow the roads their trails went along – doing so would basically block any bundling. Moreover, RAEB requires registering OD trails with an accurate road network to function, and is significantly more complex to implement and more expensive to run than our CUBu-based solution. Hence, we argue that, by using relatively small kernels k , thus limiting the distortion of the original trails, our CUBu-based solution is an overall better alternative to RAEB.

5.3 Mapping length and direction attributes to bundled trails

To explore urban mobility from different perspectives, we need means to visualize its multiple data attributes. However, this variety of attributes requires distinct visualization strategies. Two important trail attributes for the study of mobility patterns are trail *length* and trail *direction*. Figures 10 and 11 show the visualization of the entire OD17 dataset using length and direction, respectively.

Figure 10 displays color-coded trail lengths with density-modulated transparency¹ as explained in Section 5.1 using the same rainbow colormap as in Figure 9. We² see a single roughly horizontal red curve in this image, but with maximal opacity.³ This implies there are many long trips, all which perfectly map to this trajectory⁴ between the same origin and destination (if they did not, we would see a fanning-out⁵ bundle rather than a precise curve). This is an interesting finding that, we argue,⁶ could not be easily found using non-visual methods. Besides this outlier, other⁷ trails, in general, run over regular distances. Long-distance edges can indicate lack⁸ of services or resources that do not satisfy local regions, forcing people to commute⁹ long distances to access them. The OD17 contains more information that may help¹⁰ to investigate the reason behind these long trips.¹¹

Figure 11 shows the same data as in Figure 10, but with color encoding trail²⁷ directions. Also, we used here the directional bundling mode of CUBu which sepa-²⁸ rates different-direction, close-location, trails into two roughly parallel bundles. We²⁹ can clearly see the existence of parallel trajectories over the bundles, which is not³⁰ surprising because the OD survey records the typical two-way commuting of people³¹ going from and then coming back to their origins. However, this symmetry would³² possibly not be seen if we analyzed a short time period of the day.³³

5.4 Coloring transportation modes: local vs intercity buses

The OD17 dataset contains 17 transportation modes. While it would be ideal to be able to see the 17 categories all at once in our bundled visualization, that would not be easy to do, since it would require the simultaneous encoding of 17 different categorical attributes. Instead, we use transparency to hide trails according to a user-set selector that filters them by transportation mode. Figure 12 shows how we can use these filters to visualize the integration between buses from the city of São Paulo (local buses) and intercity buses. Each transportation mode has a distinct color – olive for local buses and blue for intercity buses.

Figure 12 also highlights that these different transportation systems appear to complement each other. The city of São Paulo has a very active commerce and industry, and many people from neighboring cities work there. Thus, the availability of public transportation and its integration is very important for these people. This kind of filtering along with bundling helps to better understand correlations between data attributes – in this case, transportation modes.

5.5 Density per social strata

We used our bundle visualization to study how citizens with different economical conditions commute in the SPMA. The Brazilian Economical Classification Criterion (BECC) [28] is the official social-economic index used in the Brazilian Demo-

graphic Census, which is performed by the Brazilian Institute of Geography and Statistics. It measures the purchasing power of the Brazilian society. The BECC is divided into six levels or strata (Table 2). This index is used in the OD17 survey to complement the mobility data. Table 2 also shows the average monthly income in the local currency (Brazilian reals) and in US dollars, and the number of trips in the SPMA for each BECC level considering the whole population and only citizens with age between 6 and 18 years that commute for study (see Section 5.6).

To compare the mobility patterns of different BECC social strata, we bundled the trails in each stratum separately, as shown in Figures 13 to 19. We see significant differences in the mobility patterns between the highest and lowest income levels as shown in Figure 13. The *A* level (on the left-hand side) has a high density in the center of SPMA, which includes the capital downtown surroundings. The highest density is located in the west, southwest, and northeast neighborhoods near downtown. There are density flows between the capital and the cities of Barueri and Cotia, which have high-income residential areas. There are other high dense flows linking the capital to the cities of São Bernardo do Campo and Santo André. Comparing *A* to the *D-E* level (on the right-hand side of Figure 13), we see that *D-E* has the highest dense flows in the capital eastern region. In the *D-E* level

map, we can see the absence of high-dense flows in regions that are nearest to the capital downtown; in contrast, these are present in the *A* level map. We can see more details of *A* and *D-E* strata in Figures 14 and 19

Figure 13: Comparing density of trails between social strata *A* (left-hand side) and *D-E* (right-hand side).

Figure 14: Density of trails of social stratum *A*.

When we compare all maps from the *A* to the *D-E* level (Figures 14 to 19), we see that the densest flows (red) tend to displace from the capital downtown to the eastern region of the city. The concentration of high-density flows is increasingly spreading from the center to the peripheral regions of the SPMA. Even the less dense flows are increasing and spreading over the SPMA. However, the *D-E* map shows that those flows diminish considerably for these social strata. This may indicate that

low-income citizens have less access to the urban mobility system. As a consequence, these people would have less access to the social, educational, health, and cultural services of the SPMA, as those facilities are concentrated in the center regions of the cities. It is worthy to note that those central regions also have more job opportunities. Looking at the $D-E$ map, we can see a “hole” in the capital west downtown. This region (Pinheiros district) concentrates a large number of jobs

related to information technology and financial services, which requires workers with high and medium education levels. Thus, the map shows that low-income citizens are not going to that region, which reflects the inequality of opportunities that these citizens face.

Table 2: Trips grouped per BECC income level, social stratum, and traveler age.

BECC level	Monthly income (Brazilian reais (R\$))	Monthly income (US dollars (~US\$))	Trips	Trips of 6 to 18 years old students
A	23,345	4,245	3,062,892	184,772
B1	10,386	1,888	3,854,040	260,652
B2	5,363	975	12,856,182	963,242
C1	2,965	539	11,277,159	976,745
C2	1,691	307	7,852,806	721,218
D-E	708	128	2,233,801	219,612

5.6 Mobility of young students from different social strata

To explore even more the mobility patterns showed bundling visualizations, we compared the trips of students from different social strata. We filtered citizens with age between 6 and 18 years whose commuting reason is study. We split them into two groups, the high- to moderate-income, which includes the BECC levels *A*, *B1*, *B2*, and *C1*; and the low-income, which includes levels *C2* and *D-E*. Figures 20 and 21 show the density maps for both groups.

The density map of the high- to moderate-income students (Figure 20) shows a large number of dense flows spread across the central region of SPMA. This part of SPMA concentrates most private schools, universities, and complementary colleges. In addition, high density flows are not as long as flows from other maps with all

Figure 20: Density of trails of young students from high-income households.

Figure 21: Density of trails of young students from low-income households.

the data (e.g., Figure 8). This indicates that trips to study are shorter than trips to work.

The density map of low-income students (Figure 21) shows that their mobility is very limited compared to the higher-income students. There are a few dense flows, most of them out of the capital downtown. The high density flows of low-income students are more present in the peripherals of the capital and also in the

¹neighboring cities. There is a concentration of both groups in the southwest region,¹
²where are the neighborhoods of the Campo Limpo district.²

³ It is worthy to note that the public schools in the SPMA are spread across the³
⁴central and peripheral parts of the cities. The students are enrolled in these schools⁴
⁵according to the proximity of their residences. Thus, they do not have to travel⁵
⁶long distances to reach their schools. Also, public schools have lower educational⁶
⁷performance than private schools in São Paulo. Thus, citizens with better financial⁷
⁸conditions use to put their children in private schools.⁸

⁹ The scarcity of flows from low-income students may indicate that they do not⁹
¹⁰have equal opportunities to study, being enrolled in schools in their neighborhood.¹⁰
¹¹They also do not use to go to the central region of the city and, thus, have less¹¹
¹²access to universities and complementary colleges. This inequality of opportunities¹²
¹³will probably impact these students' jobs and economical conditions.¹³

¹⁴ We also see that there are many more trails for the high- to moderate-income stu-¹⁴
¹⁵dents (Figure 20) than for low-income students (Figure 21). The high-to moderate-¹⁵
¹⁶income students also travel large distances to study, which indicates that they can¹⁶
¹⁷choose more flexibly where to study. This fact is corroborated by urban mobility¹⁷
¹⁸studies that indicate that people with better financial conditions have more mobility¹⁸
¹⁹than those with poorest conditions [29, 30].¹⁹

²⁰225.7 Directions at peak hours²⁰

²¹As discussed earlier in Sec. 3.2, Figure 4 shows the distribution of trips by hour of the²¹
²²day, with two main rush-hour peaks (6-9 AM and 5-8 PM). However, this aggregated²²
²³table does not give us insights in how the rush-hour patterns may differ. To see this,²³
²⁴we selected the two rush-hour time intervals and visualized them separately, using²⁴
²⁵directional bundling and color-coding.²⁵

²⁶ Comparing the peak hours, we can see that morning flows going to the SPMA²⁶
²⁷center (Figure 22, cyan bundle) are overall denser and longer than the flows coming²⁷
²⁸from the SPMA center during the afternoon/evening peak (Figure 23, red). This²⁸
²⁹suggests that in the morning people are in a hurry to reach their work, while they²⁹
³⁰are less in a hurry to go back home (or to other destinations like schools or the³⁰
³¹gym) in the afternoon/evening.³¹

Figure 22: Directions of trips between 6 to 9 AM

Figure 23: Directions of trips between 5 to 8 PM

In Figure 22, we can see that flows in the morning peak going to the capital downtown (cyan bundle coming from the east) are denser than opposite flows (red bundle going to the east). Although flows leaving the capital downtown in the morning are thinner than their opposite ones, they also concentrate a large number of trips, especially to the east and southwest. In Figure 23, the opposite flows seem more equally distributed.

5.8 Density by transportation mode

We next split the OD17 data by transportation mode to compare the flow patterns for four different transportation modes: pedestrians, bicycles, cars, and subway. Figures 24 to 27 show the respective visualizations.

Pedestrian trails (Figure 24) form several low-density ‘islands’ spread across the SPMA, with the densest one (red in figure) being in the capital downtown. Most

trails are quite short, which is expected (pedestrians). However, we see a few longer bundles between the capital downtown and the south and north regions of the city. Dense flows are also present in the neighboring cities of Diadema, Taboão da Serra, Osasco, Guarulhos, Poá, and Mogi das Cruzes. Upon examination, we found these dense flows to match the cities' downtown and commercial areas. This information

¹could be useful to find places that could deserve the attention of local governments¹
²to provide improvements for pedestrians.²

³
⁴ As most of the pedestrian trips are short, the bundling technique forms a few⁴
⁵flows over the SPMA. Using bundling for those short trips result in low-density⁵
⁶trails, which is less useful compared to long trips. Thus, in these cases it may not⁶
⁷be necessary to use bundling. In the upper left area of Figure 24, we can see the OD⁷
⁸trails without using bundling, which are near identical to the main bundled area. ⁸

⁹
¹⁰ Bicycle trips (Figure 25) exhibit similar patterns to pedestrian ones. They are¹⁰
¹¹shorter than three kilometers on average. In this figure, we see some thin flows in¹¹
¹²the capital downtown area. There are also some more salient flows in the capital¹²
¹³northeast and in the neighboring cities of Suzano and Guarulhos. However, compar-¹³
¹⁴ing Figure 25 with all other transportation means, we immediately see that bicycle¹⁴
¹⁵trips are by far the least numerous, and exhibit a far sparser pattern, with few star-¹⁵
¹⁶shaped ‘hubs’ where many trails meet. This suggests that the cycling infrastructure¹⁶
¹⁷is quite limited, and fragmented. Figure 25 also shows trails without using bundling¹⁷
¹⁸in the upper left corner.¹⁸

¹⁹
²⁰ The car trips (Figure 26) show a pattern similar to the one displaying the entire¹⁹
²¹dataset, *i.e.*, all transportation modes (see *e.g.* Figure 8). For a start, this tells that²⁰
²²cars are *the* dominant form of transportation in the SPMA, accounting for the main²¹
²³traffic patterns. The highest-density flows occur in the capital downtown. There are²²
²⁴several high-density flows linking the downtown area to the other regions of the²³
²⁵capital, and also coming and going from the cities of Guarulhos, Barueri, Cotia,²⁴
²⁶São Bernardo do Campo, Santo André, Mauá, and Mogi das Cruzes. Compared²⁵
²⁷to all other transportation modes, cars show a far more ‘spread out’ pattern that²⁶
²⁸covers very large areas, indicating that cars are the prevalent transportation mode²⁷
²⁹in most parts of the SPMA.²⁸

³⁰
³¹ Finally, subway trips (Figure 27) show a strong star-shaped pattern, with very³⁰
³²high density bundles that connect the capital with the neighboring cities, due to³¹
³³the integration of the subway system with the train system. Compared to all other³²
³³transportation modes, subways show a clearer, simpler, trip pattern structure.³³

15.9 Different trip reasons

We next aim to study whether trips done for different reasons exhibit distinct trip patterns. For this, we create bundled visualizations from the OD17 data with trips grouped by work, health, education, and shopping. Figures 28 to 31 show the results.

Work-related trips (Figure 28) are overall longer than the other trip reasons, and also cover a larger area (see the central agglomeration in the figure). Interestingly, the longest trips, between the east side and the city center (red bundle), are similar in pattern to the longest trips for health and education. Trips for health reasons are sparser than work-related ones, and also show a more star-like pattern, with long bundles connecting to the central area. This may indicate that peripheral regions are not well served by health services. Trips for studying reasons (Figure 30) have the largest distances between the northeast and the western regions of the SPMA. Their pattern is somewhere in-between the work and health trips. Interestingly, education trips show several ‘loops’ in the center of the SPMA. Finally, shopping trips (Figure 31) show the least dense, and overall also shortest, patterns, apart from a few outliers like the red (important) bundle connecting the center to the northeast. This tells that, unlike health, education, and work, shopping facilities (which are actually provided by private companies) are better distributed over the SPMA. This outlines that bundled visualizations are useful not only when they show the *presence* of certain data, *e.g.* trails linking far-apart regions; the *absence* of patterns is also insightful, as in the case of the lack of long shopping trips.

6 Conclusions and future work

In this work, we explored the usage of trail bundling for creating visualizations of various aspects of the urban mobility data from the São Paulo Metropolitan Area. Our analyses on the characteristics of the 2017 OD survey shows that bundling can be used to identify and compare different mobility patterns implied by different subsets of the data and subsets of the available attributes. By suitably combining filtering (to reduce the amount of data and/or attributes to be explored) with bundling (to simplify the created visualizations and reduce visual clutter) and with the available visual channels (opacity, color, direction), we highlight different patterns in the OD17 dataset which would not have been easily obtainable by classical data mining and data analysis tools.

The bundled layout for the trajectory dataset highlights its centrality structure over the represented area. Moreover, this structure matches the subway and rail lines infrastructure of São Paulo. Albeit this was not a surprise, the correlation suggests that our parameters were well tuned for the visualization in the metropolitan scale. Our methodology to reduce the dataset complexity from 42 million trips to less than a million, and also our customization of a general-purpose bundling framework

²⁸(CUBu) to bundle specific subsets of data and/or attributes were key points that²⁸
²⁹made this analysis possible.²⁹

³⁰As future work, we intend to explore improvements in the usage and usability³⁰
³¹of bundled visualizations. From a visualization perspective, improvements in the³¹
³²map to display region divisions on top of the trajectories, as proposed in Klein et³²
³³al. [15], can help to better identify the connections between the regions. A different³³

¹approach to convey the density information of bundles could be to scale edge lines¹
²thickness proportionally to the expansion factor of each record instead of using²
³colors, similarly to Lhuillier et al. [21]. This design is worth exploring, as it would³
⁴eliminate the need to replicate edges and would therefore significantly reduce the⁴
⁵size of the dataset. These are important steps to enable the use of bundling for⁵
⁶real-time analysis over the Internet. 6

⁷ From an application perspective, there are many other possibilities for urban mo-⁷
⁸bility analysis using data from the OD survey itself and along with other datasets,⁸
⁹such as those from private mobility companies, IoT devices, and bike-sharing sys-⁹
¹⁰tems. Last but not least, we also intend to perform a user study to assess the¹⁰
¹¹usefulness of bundled visualizations with feedback from actual traffic managers. 11

12

13Abbreviations 13

BECC: Brazilian Economical Classification Criterion; CUBu: CUDA bundling; CUDA: Compute Unified Device
¹⁴Architecture; IBB: Image based bundling; KDEEB: Kernel Density Estimation Edge Bundling; OD17: 2017 14
¹⁵Origin-Destination survey; RAEB: Road Aware Edge Bundling; SPMA: São Paulo Metropolitan Area. 15

16Acknowledgments 16

We thank CNPq, CAPES, and FAPESP for the financial support granted to the work.

17Authors' contributions 17

¹⁸TGM has conducted the research with the guidance of the other authors. HAS has led the writing of this paper. NL,¹⁸
¹⁹EFZS, AT, FK, and HAS have contributed with ideas, discussion, and effective revision of all sections of the paper. 19
²⁰All authors read and approved the final manuscript. 20

Funding

²¹This research is part of the INCT of the Future Internet for Smart Cities funded by CNPq proc. 465446/2014-0, 21
²²Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, FAPESP 22
²³proc. 14/50937-1, and FAPESP proc. 15/24485-9. 23

24Availability of data and materials

²⁴The datasets used and/or analyzed during this study are available from the corresponding author on request. The 24
²⁵CUBu framework is available from AT on reasonable request. The complete 2017 OD survey dataset is publicly 25
²⁶available at <http://www.metro.sp.gov.br/pesquisa-od>. 26

27Competing interests

²⁷The authors declare that they have no competing interests. 27

28Author details 28

²⁹¹Department of Computer Science, University of São Paulo, São Paulo, Brazil. ²Department of Information and 29
³⁰Computing Sciences, Utrecht University, Utrecht, Netherlands. 30

31References 31

- ³¹1. Ahlgren, B., Hidell, M., Ngai, E.C.-H.: Internet of things for smart cities: Interoperability and open data. *IEEE* 31
³²*Internet Computing* **20**(6), 52–56 (2016) 32
- ³²2. Municipality of Seattle, U.: City of Seattle Open Data Portal 32
- ³³3. Vale, R.C.C.: The welfare costs of traffic congestion in São Paulo Metropolitan Area. Master's thesis, School of 33
³³Economics, Business Administration and Accounting at Ribeirão Preto — University of São Paulo (2018)

- 1 4. Magrini, M., Moroni, D., Palazzese, G., Pieri, G., Leone, G., Salvetti, O.: Computer vision on embedded sensors¹
2 for traffic flow monitoring. In: Proceedings of the 2015 IEEE 18th International Conference on Intelligent
2 Transportation Systems, pp. 161–166 (2015)
- 3 5. Bongiorno, C., Santucci, D., Kon, F., Santi, P., Ratti, C.: Comparing bicycling and pedestrian mobility: Patterns³
4 of non-motorized human mobility in greater boston. *Journal of Transport Geography* **80**, 102501 (2019)
- 5 6. Chen, W., Guo, F., Wang, F.Y.: A Survey of Traffic Data Visualization. *IEEE Transactions on Intelligent*
5 *Transportation Systems* **16**(6), 2970–2984 (2015)
- 6 7. Zhou, H., Xu, P., Yuan, X., Qu, H.: Edge bundling in information visualization. *Tsinghua Science and*
6 *Technology* **18**(2), 145–156 (2013)
- 7 8. Lhuillier, A., Hurter, C., Telea, A.: State of the Art in Edge and Trail Bundling Techniques. *Computer Graphics*
7 *Forum* **36**(3), 619–645 (2017)
- 8 9. Van Der Zwan, M., Codreanu, V., Telea, A.: CUBu: Universal Real-Time Bundling for Large Graphs. *IEEE*
8 *Transactions on Visualization and Computer Graphics* **22**(12), 2550–2563 (2016)
- 9 10. Willems, N., Scheepens, R., Van De Wetering, H., Van Wijk, J.J.: Visualization of vessel traffic. *Situation*
10 *Awareness with Systems of Systems* (3), 73–87 (2009)
- 11 11. Hurter, C., Ersoy, O., Fabrikant, S.I., Klein, T.R., Telea, A.C.: Bundled Visualization of Dynamic Graph and
11 Trail Data. *IEEE Transactions on Visualization and Computer Graphics* **20**(8), 1141–1157 (2014)
- 12 12. Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T.: Visualization of eye tracking data:12
A taxonomy and survey. *Computer Graphics Forum* **36**(8), 260–284 (2017)
- 13 13. Martins, T.G., Lago, N., de Souza, H.A., Santana, E.F.Z., Telea, A., Kon, F.: Visualizing the structure of urban¹³
14 mobility with bundling: A case study of the city of São Paulo. In: Proceedings of the 4th Workshop of Urban
14 Computing. CoUrb 2020 (2020)
- 15 14. Hurter, C.: Image-Based Visualization: Interactive Multidimensional Data Exploration. Morgan Claypool, San
15 Rafael, California (2015)
- 16 15. Klein, T., van der Zwan, M., Telea, A.: Dynamic multiscale visualization of flight data. *Computer Vision*
16 *Theory and Applications (VISAPP), 2014 International Conference on* **1**, 104–114 (2014)
- 17 16. Graser, A., Schmidt, J., Roth, F., Brändle, N.: Untangling origin-destination flows in geographic information
17 systems. *Information Visualization* **18**(1), 153–172 (2017)
- 18 17. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on*
18 *Pattern Analysis and Machine Intelligence* **24**(5), 603–619 (2002)
- 19 18. Hurter, C., Ersoy, O., Telea, A.: Graph bundling by kernel density estimation. *Comput. Graph. Forum* **31**(3pt1),
19 865–874 (2012)
- 20 19. Zeng, W., Shen, Q., Jiang, Y., Telea, A.: Route-aware edge bundling for visualizing origin-destination trails in
20 urban traffic. *Computer Graphics Forum* **38**(3), 581–593 (2019)
- 21 20. Peysakhovich, V., Hurter, C., Telea, A.: Attribute-driven edge bundling for general graphs with applications in
21 trail analysis. In: Proceedings of the 2015 IEEE Pacific Visualization Symposium. PacificVis (2015)
- 22 21. Lhuillier, A., Hurter, C., Telea, A.: FFTEB: Edge bundling of huge graphs by the fast fourier transform. In:
22 Proceedings of the 2017 IEEE Pacific Visualization Symposium (PacificVis), pp. 190–199 (2017)
- 23 22. Telea, A., Ersoy, O.: Image-based edge bundles: Simplified visualization of large graphs. *Computer Graphics*
23 *Forum* **29**(3) (2010)
- 24 23. Brazilian Institute of Geography and Statistics (IBGE): Estimates of the resident population for Brazilian
24 municipalities and federative units with a reference date of July 1, 2020.
25 ftp://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2020/estimativa_dou_2020.pdf. In
26 Portuguese (2020)
- 27 24. Rolnik, R., Klintonowicz, D.: (Im)Mobility in the city of São Paulo. *Estudos Avançados (Advanced Studies)* **25**,
27 89–108 (2011)
- 28 25. São Paulo Metropolitan Company (Metrô SP): The Urban Mobility of the São Paulo Metropolitan Area in
28 details. Secretary of Metropolitan Transport. In Portuguese (2018)
- 29 26. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE*
29 *Transactions on Visualization and Computer Graphics* **12**(5), 741–748 (2006)
- 30 27. Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H.: Thematic Cartography and Geovisualization.
30

1	Pearson Prentice Hall, Upper Saddle River, NJ (2009)	1
2	28. of Research Companies (ABEP), B.A.: Critério de Classificação Econômica Brasil.	2
	http://www.academia.edu/download/48356646/CCE.pdf . In Portuguese (2008)	
3	29. Carruthers, R., Dick, M., Saurkar, A.: Affordability of public transport in developing countries. Transport	3
4	Papers TP-3 , 1–27 (2005)	4
	30. Lucas, K., Mattioli, G., Verlinghier, E., Guzman, A.: Transport poverty and its adverse social consequences. In:	
5	Proceedings of the Institution of Civil Engineers – Transport, vol. 169, pp. 353–365 (2016)	5
6		6
7		7
8		8
9		9
10		10
11		11
12		12
13		13
14		14
15		15
16		16
17		17
18		18
19		19
20		20
21		21
22		22
23		23
24		24
25		25
26		26
27		27
28		28
29		29
30		30
31		31
32		32
33		33

Figures

Figure 1

Pipeline of the KDEEB algorithm.

Figure 2

Municipalities of the SPMA. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

(a) Trip mode

(b) Trip reason

Figure 3

Trips by mode and by reason.

Figure 4

Trips by hour of the day.

Figure 5

Naïve straight-line drawing of OD trails from the SPMA. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 6

Pipeline of the OD17 visual data exploration.

Figure 7

Accumulated percentage of trip records lost by expansion factor

Figure 8

Bundled trajectories colored by density in (a) and (b) modulated transparency modes. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 9

Bundled trails colored by density and rail lines. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 10

Mapping trail length to colors. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 11

Mapping trail direction to colors. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 12

Edges by transportation modes: bundled trails of local and intercity buses. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 13

Comparing density of trails between social strata A (left-hand side) and D-E (right-hand side). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 14

Density of trails of social stratum A. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 15

Density of trails of social stratum B 1. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 16

Density of trails of social stratum B 2. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 17

Density of trails of social stratum C1. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 18

Density of trails of social stratum C2. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 19

Density of trails of social strata D-E . Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 20

Density of trails of young students from high-income households. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 21

Density of trails of young students from low-income households. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 22

Directions of trips between 6 to 9 AM Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 23

Directions of trips between 5 to 8 PM Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 24

Density of pedestrian trips Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 25

Density of bicycle trips Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 26

Density of car trips Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 27

Density of subway trips Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 28

Distance of trips for work reasons. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 29

Distance of trips for health-related reasons. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 30

Distance of trips for education reasons. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 31

Distance of trips for shopping reasons. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.