

# Visualizing the structure of urban mobility with bundling: A case study of the city of São Paulo

Tallys Gustavo Martins<sup>1</sup>, Nelson Lago<sup>1</sup>, Higor Amario de Souza<sup>1</sup>,  
Eduardo Felipe Zambom Santana<sup>1</sup>, Alexandru Telea<sup>2</sup>, Fabio Kon<sup>1</sup>

<sup>1</sup>Department of Computer Science – IME – University of São Paulo  
São Paulo – SP – Brazil

<sup>2</sup>Department of Information and Computing Sciences, Utrecht University, Netherlands

{tallys, lago, hamario, efzambom, kon}@ime.usp.br, a.c.telea@uu.nl

**Abstract.** *Visualization of urban mobility data can facilitate the analysis and the decision-making process by public managers. However, mobility datasets tend to be very large and pose several challenges to the use of visualization, such as algorithm scalability and data occlusion. One approach to solve this problem is trail bundling, which groups motion trails that are spatially close in a simplified representation. This paper presents the results of adapting and using a recent bundling technique on a big dataset of urban mobility in São Paulo. The results show that bundling allows the visualization of various mobility patterns in the city.*

## 1. Introduction

Urban mobility is a great concern for citizens and governments. It directly affects people's quality of life and has a high impact on the economy, both strong motivations for the development of more efficient transportation systems. For example, traffic congestion in the São Paulo metropolitan area is estimated to impact 89% of work-related commuting trips, causing monetary losses of seven billion reais (~US\$1.8 billion) every year [Vale 2018].

Urban mobility data can support decision-making activities aiming to tackle such issues. This data comes from several sources: GPS tracking, traffic cameras, censuses, and travel surveys. Among these, 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*. In it, they interview citizens to gather data 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 São Paulo metropolitan area. The OD Survey data contains the origin and destination points of trips made by a representative sample of the population, which results in a large amount of data. For instance, in the last OD Survey (2017), there were around 42 million trips accounted over 24 hours of a regular working day. The survey also covers a wide number of socio-economic aspects such as age, gender, family income, and transportation modes used.

While the above-mentioned OD Survey is very comprehensive and accurate, policymakers need proper tools to analyze this large amount of data. Spreadsheets and statistics programs help producing tables and charts that show aggregated insights showing *e.g.* the number of users of the public transportation system over the years or the number of men and women that commute for work. Density maps can help answering geolocated

data-related questions, such as finding regions that concentrate most of the mobility flow during the day or the most common origin-destination pairs.

However, considering the huge amount of information that is produced every day in any city, translating geolocated data into meaningful images can be very challenging. A recent survey about traffic visualization methods points to the common use of *line-based* visualization techniques to study the structure of mobility data, which often includes pattern discovery and clustering tasks [Chen et al. 2015]. Still, drawing OD lines can only handle thousands of trajectories. Visualizing 42 million trajectories, as recorded daily in the OD Survey, would only generate a fully cluttered image. (see Figure 4 discussed later in the paper). *Bundling techniques* improve upon line-based techniques and can depict high volumes of trajectory data by essentially simplifying trajectories in the image space, via grouping spatially close and data-wise similar trajectories together [Zhou et al. 2013, Lhuillier et al. 2017].

In this paper, we propose the use of a bundling model to visualize the structure and characteristics of urban mobility in the city of São Paulo as obtained by the OD Survey. We adapt a recent bundling framework known as CUBu [Van Der Zwan et al. 2016], which has proven to be useful in other scenarios using movement data (flight traffic, eye tracking). However, to our knowledge, neither CUBu nor other bundling techniques have been used so far to visualize such a high volume of trajectories as in our case. We show how the bundled visualizations help finding structural patterns in urban mobility and reveal new insights regarding São Paulo’s transportation systems and city infrastructure.

This paper is organized as follows. Section 2 outlines related work with a focus on OD data visualization and trajectory bundling. Section 3 details our data. Section 4 describes our visualization approach. Section 5 presents the obtained results and discusses the use of bundling in urban mobility visualization. Finally, Section 6 concludes the paper.

## 2. Background and Related Work

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

*Trajectories* (or for short, trails) are typically sets of spatial points  $\mathbf{t} = \{\mathbf{x}_i\}$  recorded at consecutive time moments  $t_i$  that describe the motion of an object over time. Origin-Destination (OD) data are particular cases of trails which contain only the first and last recorded point. Trails can have additional data attributes, recorded either at each sample point  $\mathbf{x}_i$ , *e.g.*, speed, or globally for the entire trail, *e.g.*, vehicle type. Trail data can be directly plotted atop of 2D maps, with data attributes encoded optionally into visual variables such as color, line thickness, and line style. Besides statically drawing entire trail-sets, these can be also shown by animating particles along their points  $\mathbf{x}_i$  [Hurter 2015, Klein et al. 2014].

Visualizing trail-sets having thousands of trails or more is challenging. Simply drawing the trails yields significant clutter which makes finding even the simplest spatial patterns very hard. Particle techniques also do not scale well – with over roughly 50 thousand points, the resulting visual pattern resembles Brownian motion. *Bundling* methods aim to solve this visual scalability issue by grouping trail fragments which are spatially close and, optionally, have similar data attributes. Simply put, this creates more *overdraw* between (similar) trails, thus creating visually empty space that reduces *clutter*.

Bundling is used in several other studies on movement data. [Graser et al. 2017] created a bundled visualization to study the characteristics of bird migrations. [Klein et al. 2014] developed a dynamic visualization for aircraft flight data where the trail-set itself varies over time. [Willems et al. 2009] analyzed maritime vessel traffic using bundling techniques. Separately, bundling was used to simplify eye-tracking data to infer reading patterns [Hurter et al. 2014, Blascheck et al. 2017]. [Lhuillier et al. 2017] compiled a survey on the state of the art on bundling techniques. This survey distinguishes between bundling graph data (straight-line drawings of graph layouts) and trail data (such as our OD data), highlighting the best methods for both cases.

*Image based bundling* (IBB) techniques are a particular class. They essentially exploit the mean shift clustering [Comaniciu and Meer 2002] procedure, well known in image processing, as follows (we outline this process for the Kernel Density Estimation Edge Bundling (KDEEB) technique [Hurter et al. 2012], although all other IBB techniques we are aware of work similarly and share the same properties). Let  $T = \{t_i\}$  be the trail set to bundle, and let  $S = \{p_i\}$  be a set of points that densely sample  $T$ . First, a density map  $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}^+$  is computed from  $S$  by kernel density estimation (KDE), *i.e.* by convolving the points in  $S$  with an Epanechnikov kernel of radius  $k$ . Hence,  $\rho$  will be high in regions of high density of the sampling points – and by implication, of trails – and low elsewhere. Second, the points in  $S$  are advected upwards in the direction of the density gradient  $\nabla \rho$  with a small step. Terminal (O and D) points are however kept fixed so that one can still identify the origin and destination of the bundled trails next. Finally, a simple 1D Laplacian filter is used to smooth the trails in  $T$ , thereby removing inherent numerical noise in the advection. The process is repeated for  $N$  iterations, whereby  $k$  is continuously decreased so as to ‘slow down’ the advection and make trails converge around the local density maxima. The main advantages of IBB are simplicity of implementation, control over the simplification induced by the bundling (given by the parameter  $k$ ), and, notably, speed, as image-space operations map well to GPU functionality.

[Zeng et al. 2019] adapted the KDEEB technique into Road Aware Edge Bundling (RAEB), which constraints the bundles along the road network on which the respective trails are recorded. RAEB was demonstrated on 166K taxi trajectories from New York City. RAEB is the only work that uses bundling with urban mobility data that we know of. However, RAEB requires detailed trails data (not OD data) and the layout of the city road network.

CUBu, which stands for CUDA bundling, is an improved implementation of KDEEB [Van Der Zwan et al. 2016]. It leverages CUDA<sup>1</sup> to implement all IBB steps (trail sampling, KDE computation, advection, and smoothing) on the GPU, thereby surpassing KDEEB and older IBB methods [Peysakhovich et al. 2015, Lhuillier et al. 2017] in speed. Additionally, CUBu supports several bundling styles, most notably *directional* bundling [Peysakhovich et al. 2015], which separates spatially close trails running in opposite directions in different bundles. Rendering-wise, CUBu supports a large palette of options, such as pseudo-shading to emphasize high-density bundles [Telea and Ersoy 2010], encoding bundle importance into opacity, and color mapping trail attributes (direction, time, length).

---

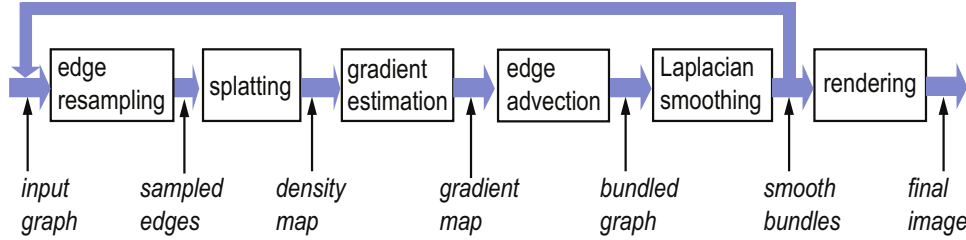
<sup>1</sup>[developer.nvidia.com/cuda-zone](https://developer.nvidia.com/cuda-zone)

Overall, from all bundling methods surveyed in [Lhuillier et al. 2017], CUBU is the fastest and offers the most options to visually encode data. While these options are well described in [Van Der Zwan et al. 2016], they have however not been, to our knowledge, fine-tuned and actually *deployed* to support a real-world use-case of analyzing very large trail sets, let alone trails from mobility data. Indeed, CUBu is explicitly proposed as a *framework* to realize many types of bundling. However, it is up to the application domain expert to customize this framework to obtain actual results.

In the remainder of this paper, we show how we parameterize (and customize) CUBu to this end and for our urban mobility data, along the following dimensions:

- *Density map*: We use the underlying density map  $\rho$  computed by the KDE process implemented in CUBu to show the local edge density in the produced bundles. This way, one can easily separate visually high-density (important) traffic flows from less important ones;
- *Attribute filtering and coloring*: We use CUBu’s ability to color code trail directions or trail lengths to study mobility behavior. We also extended CUBu to filter specific attribute ranges, allowing one to focus on specific traffic types;
- *City map*: We modified CUBu to blend its bundled result atop of a map of the city where traffic is analyzed. This allows correlating the bundles with actual locations in the city.

Figure 1 shows the bundling pipeline steps of the KDEEB algorithm, which receives an input graph and produces a bitmap image visualization of it.



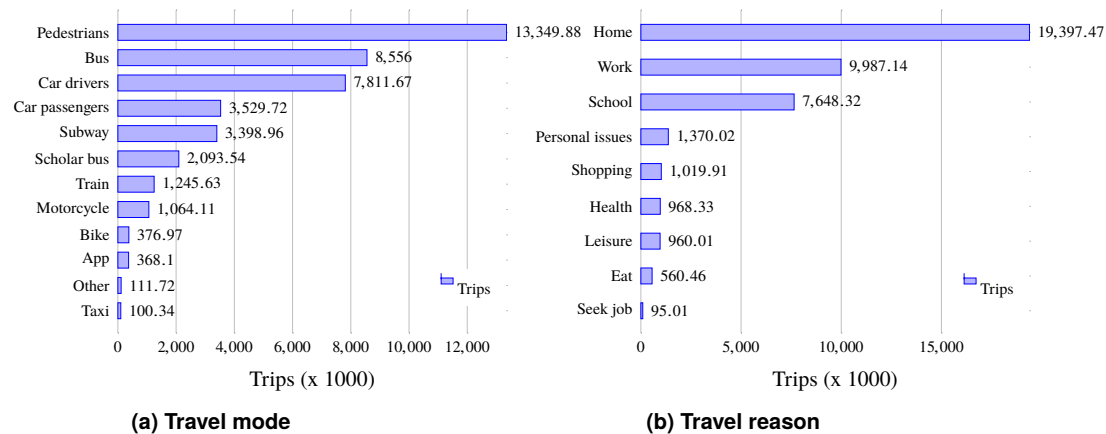
**Figure 1. Pipeline of the KDEEB algorithm (from [Hurter et al. 2012]).**

### 3. Mobility in São Paulo

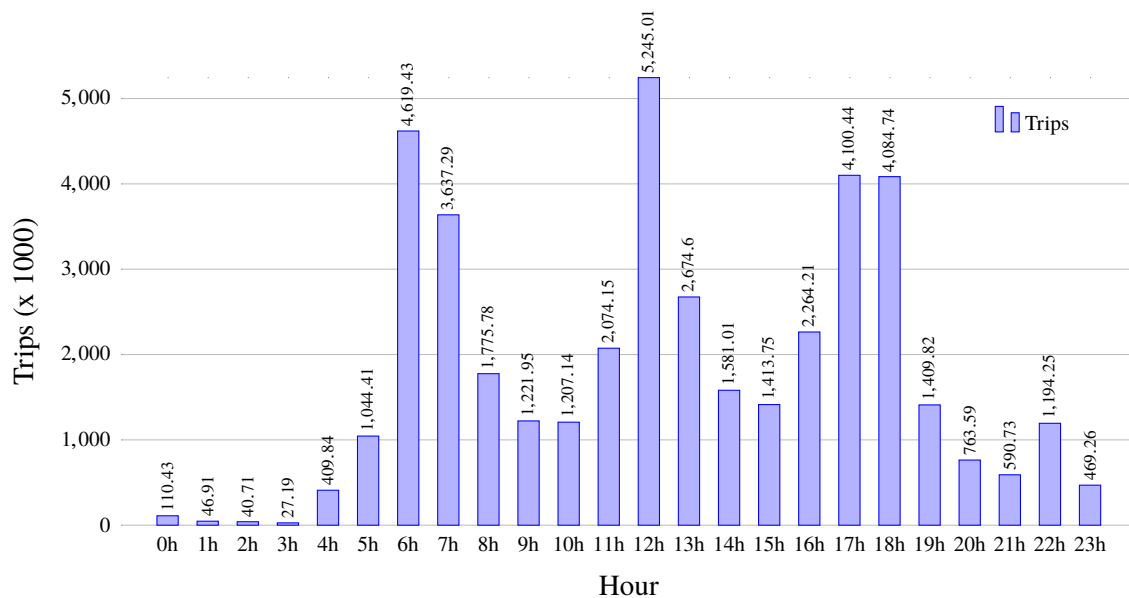
The Origin-Destination (OD) Survey is the primary source of mobility information in the São Paulo metropolitan area. It is performed by the São Paulo Metropolitan Company (Metrô) every ten years since 1967.

The 2017 OD Survey has information about 157 992 trips of people randomly sampled in the São Paulo metropolitan area, which is a representative sample of the total population (estimated to be 21.5 million in 2018) with an error margin of 6% and a confidence interval of 92% [Metrô SP 2018, p.18]. The trips occur for different reasons such as work, house, study, and leisure, and different travel modes such as by walk, car, subway, and train. Figure 2 shows the distribution of trips by transportation mode and by motivation. Figure 3 shows the distribution along the hours of the day.

As Figure 2 shows, most trips in the city are made by pedestrians. Usually, pedestrian trips cover small distances (624 m on average) and stay inside a single district of the



**Figure 2. Trips by mode and by reason.**



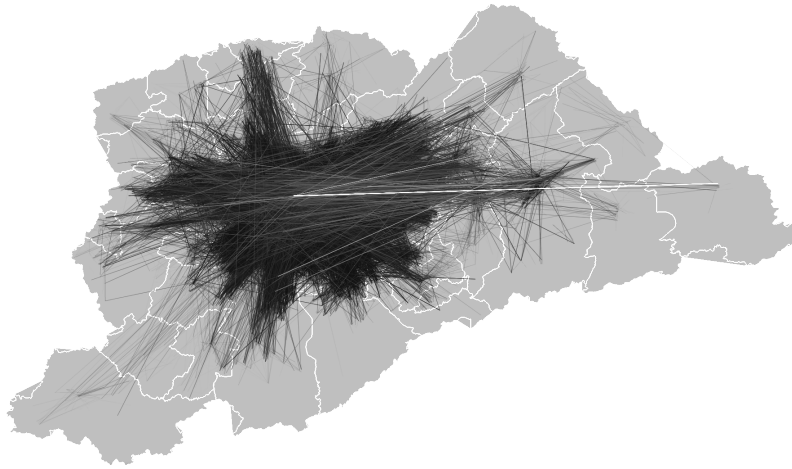
**Figure 3. Trips by hour of the day.**

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 3), we see that traffic in São Paulo has three very well-defined peaks: one in the morning (6 AM to 8 AM), another at lunchtime (noon), and the last one in the early evening (5 PM to 7 PM).

The 157 992 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 important 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 4 shows the OD Survey trajectories plotted over the map of the São Paulo metropolitan area, 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. In scenarios with large amounts of data like this one, filtering and aggregation techniques are essential to simplify the visualization and make it understandable for the human eye.



**Figure 4. Naïve straight-line drawing of OD trails form the Metropolitan area of São Paulo.**

## **4. Proposed visualization approach**

To analyze the structure of urban mobility, we focused on the visualization of the traffic of one complete day over the São Paulo metropolitan area using mobility data from the 2017 OD Survey. We parameterized and adapted CUBu to create different visualizations and explore several of the properties of this data. Our study includes the investigation of regions that concentrate most of the traffic and how this traffic relates to the city infrastructure. Moreover, we use the transport type information present in the OD survey to show the interaction between the bus systems from the São Paulo city and from neighboring cities. We next describe the data transformations we applied and CUBu customizations we performed to support our goals.

### **4.1. Data Representation**

The OD trajectory data we use as input for bundling is described by a table with six columns: trail ID, transportation mode, origin coordinates, and destination coordinates. We extract this data from the OD Survey: The trail ID is directly copied from the OD Survey data and serves as an identifier. The origin and destination coordinates are transformed to the latitude/longitude system (required to geolocate these coordinates on a map). The transportation mode is stored as an integer in the range 1 to 17. The expansion factor of each OD record (introduced in Sec. 3) is used to replicate trails – for

an OD Survey trail with expansion factor  $E$ , we create  $E$  copies of that trail as input for bundling. This yields a complete dataset of 42 million trails. For all such replicated trails, we keep the ID of the underlying OD Survey trail which generated them. This way, we can trace back which bundled trails correspond to an OD Survey record. Table 1 shows a sample of the generated trails. Here, trails corresponding to the top two rows were created from the same OD Survey trail with ID 50 and  $E = 2$ .

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

ID	Mode	Long. Origin	Lat. Origin	Long. Dest	Lat. Dest
50	1	-46.628 093 770	-23.551 691 866	-47.003 481 044	-23.393 563 283
50	1	-46.628 093 770	-23.551 691 866	-47.003 481 044	-23.393 563 283
51	4	-47.001 872 312	-23.398 468 606	-47.003 481 044	-23.393 563 283

## 4.2. Data Preprocessing

Although CUBu is – according to its authors and also to our knowledge – the fastest existing solution for trail bundling, it still cannot process 42 million trails at interactive rates (which are required for interactive data exploration). To alleviate this problem, we reduce the dataset from the OD Survey based on the expansion factor  $E$ .

First, we analyzed how representative were the sample records from the OD Survey regarding the population of 42 million trips. We observed that all records with  $E < 75$  represented less than 4% of the total population. Considering the error margin of the OD Survey was below 6% (see Section 3), we could remove the records with  $E < 75$  and still keep the statistical relevance of the dataset intact. Secondly, we divided the  $E$  values of all remaining records by this threshold value of 75, resulting in 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. 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. Parameter Setting

The literature is not clear on how to choose good bundling parameters [Lhuillier et al. 2017]. This was observed and explicitly studied in [Zeng et al. 2019], which also proposed ways to compute good parameter settings. However, as they also mention, these settings are valid for their method (RAEB) and would not generalize directly to other methods, such as CUBu. Hence, we had to find good parameters in our case empirically. The obtained parameter set in this way was: Image resolution  $R = 512 \times 512$  pixels; trail sampling step: 10 pixels (in line with the recommendation in [Van Der Zwan et al. 2016] of using for this about 1% of the image diagonal dimension); kernel size  $k = 18$  pixels; and number of bundling iterations  $N = 15$ . This configuration yielded roughly 3.2 million sample points, which were bundled in roughly 52 milliseconds per iteration on a PC with a 4GB NVidia GeForce 940MX graphics card GPU. We used this parameter set to create all bundled visualizations shown in Section 5.

#### 4.4. Visualization enhancements

To enrich our visualizations of São Paulo urban mobility, we made two important additions to CUBu. First, we added the map of the metropolitan area as a background image and plotted the train rail lines atop of the bundled image. This way, one can use the map to reason about where certain bundles are; and can use the rails to see how bundles correspond to specific train infrastructure. Secondly, we added filters to select the subset of the 17 transportation modes to display in the visualization, using different categorical colors for each of them. We used this feature to explore relations between the different modes and their impact on urban mobility.

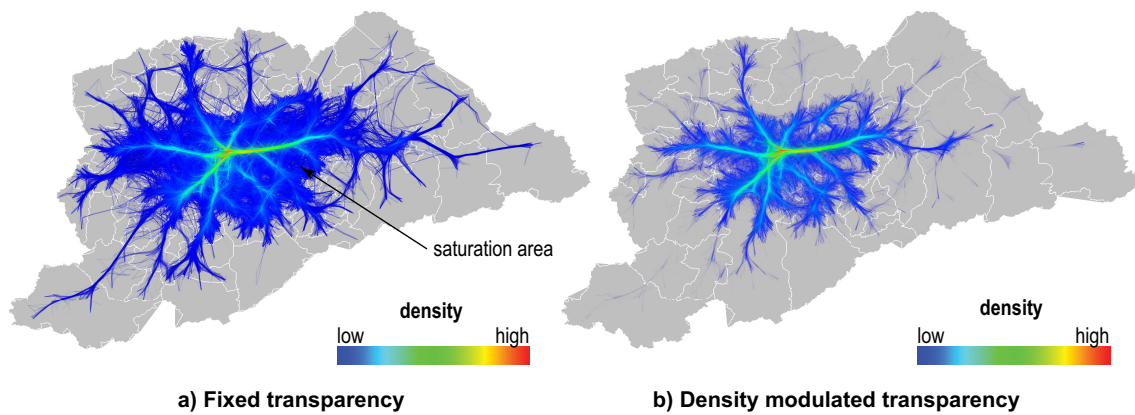
### 5. Results

We next present several results obtained with our proposed visualizations and explain, for each of them, how they can be useful in the study of mobility patterns.

#### 5.1. Understanding Traffic Density

As explained in Sec. 2, standard bundling trades off clutter for overdraw. However, such a visualization does not tell us how many trails have been grouped in a bundle. The typical solution for this, pioneered by [Holten 2006] is to draw semi-transparent trails, each having a fixed transparency  $\alpha < 1$ . Thus, blending will show high-density trails as more opaque and low-density ones as more transparent, respectively. In addition to that (since transparency is not a strong quantitative visual variable [Slocum et al. 2009]), we also encode trail density, directly estimated by the KDE density  $\rho$ , into color.

Figure 5a shows a visualization obtained this way. We can see a few dense paths but the image still presents much clutter. The explanation of this is simple: On consumer-grade GPUs, the transparency  $\alpha$  is modeled, during blending, as an 8-bit integer value. Hence, only 255 different transparency levels are possible, *i.e.*, only 255 different trail-density levels can be displayed. Setting  $\alpha$  too high would immediately saturate the transparency channel where higher trail densities occur – all densities above 255 are clamped to 255. Setting  $\alpha$  lower than  $1/255$  would result in no image, since this would correspond to zero opacity on the 8-bit representation.



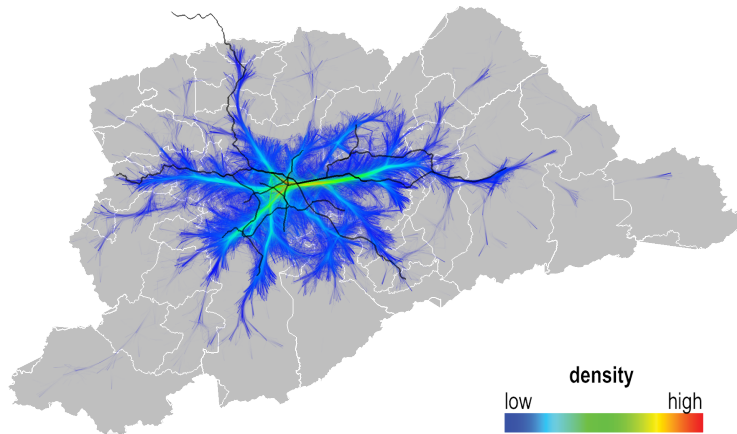
**Figure 5. Bundled trajectories colored by density in (a) fixed and (b) modulated transparency modes.**



We address this problem by actually mapping the density  $\rho$  in both transparency and color. Since  $\rho$  is computed with floating-point accuracy on the GPU during KDE, no clamping or rounding-off issues occur. This density-based transparency modulation is useful to highlight even more the high-density areas and reduce general clutter – an alternative to the use of higher kernel sizes that would create stronger bundles but also cause more edge distortion. Figure 5b shows the resulting visualization on the same data as in Figure 5a. 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. This makes indeed sense because this is the most populous part of the metropolitan area.

## 5.2. Subway and Train Infrastructure

The public transportation system is the main transportation mode for most people in São Paulo. The impact of the rails mesh over the motion 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 6). This is an expected result, since according to the OD Survey, about 44% of 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.



**Figure 6. Bundled trails colored by density and rail lines.**

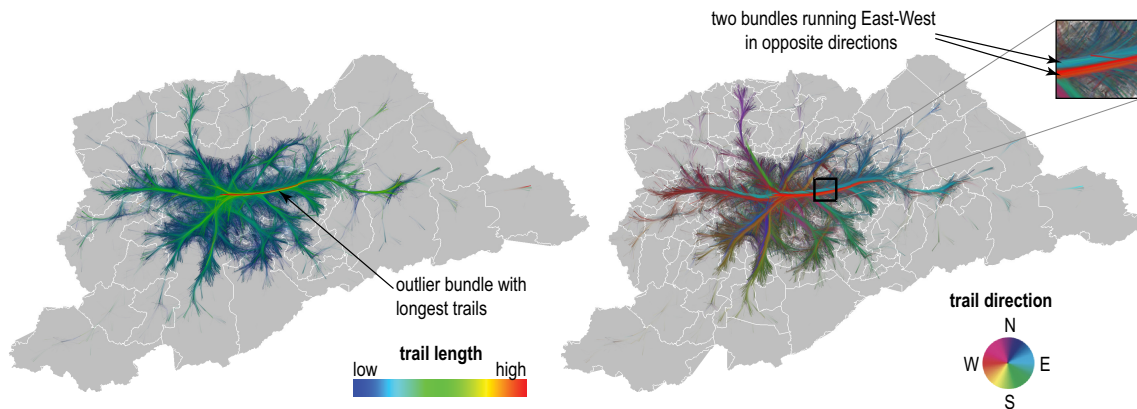
Note that this type of correlation (of bundles with rails) is not the same as RAEB [Zeng et al. 2019]. In RAEB, the bundling was *explicitly* made to follow roads. In our case, roads are *superimposed* by a bundling that uses only the OD data. One may argue that RAEB, in this sense, produces more ‘correct’ 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 Attributes to Bundled Trajectories

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*.

Figure 7a displays color-coded trail lengths with density-modulated transparency as explained in Sec. 5.1, using the same rainbow colormap as in Fig. 6. We see a single roughly horizontal red curve in this image, but with maximal opacity. This implies there are many long trips, all which perfectly map to this trajectory between the same origin and destination (if they did not, we would see a fanning-out bundle rather than a precise curve). This is an interesting finding that, we argue, could not be easily found using non-visual methods. Besides this outlier, other trails, in general, run over regular distances. Long-distance edges can indicate lack of services or resources that do not satisfy local regions, forcing people to travel long distances to access them. The OD Survey contains more information that may help to investigate the reason behind these long trips.



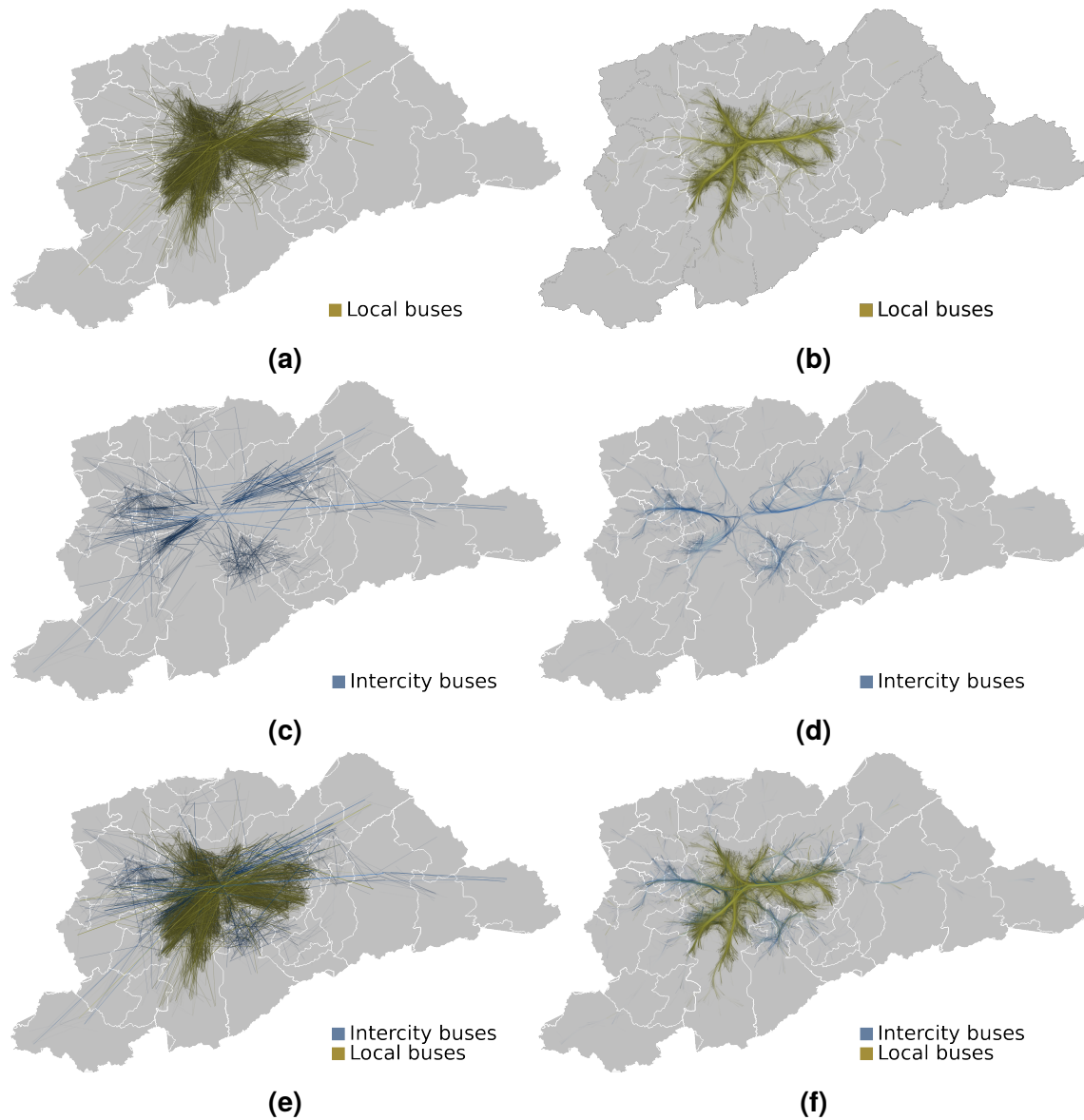
**Figure 7. Mapping (a) trail length and (b) trail direction to colors.**

Figure 7b shows the same data as in Figure 7a, but with color encoding trail directions. Also, we used here the directional bundling mode of CUBu which separates different-direction, close-location, trails into two roughly parallel bundles. We can clearly see the existence of parallel trajectories over the bundles, which is not surprising because the OD Survey records the typical two-way commuting of people going from and then coming back to their origins. However, this symmetry would possibly not be seen if we analyzed a short time period of the day.

### 5.4. Transport Modes in São Paulo: Local vs Intercity Buses

We also considered the 17 transportation modes recorded in the OD Survey. 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 8 shows how we can use these filters to show the integration between buses from the city of São Paulo (local buses) and from neighboring cities (intercity buses). Unbundled and bundled trajectories are presented on the left and right sides respectively. Each transportation mode has a distinct color – yellow for local buses and blue for intercity buses, respectively. Figure 8f 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.



**Figure 8. Edges filtered by transport modes: local buses and intercity buses. Original trails in the left (a, c, e) and bundled trails in the right (b, d, f).**

## 6. Conclusions and Future Work

In this work, we proposed several visualizations of urban mobility data from the São Paulo metropolitan area using trail bundling. We presented a bundled layout for the trajectory dataset that highlights its structure centrality over the represented area. We also showed that this structure highly 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 adaptations to a general-purpose bundling framework to highlight aspects of interest for our study, were key points to make the study of such flows possible.

Our study is, to our knowledge, the first application of general-purpose trail bundling techniques to study large trail datasets, and the first to focus on urban mobility in particular. However, there are many additional possibilities for urban mobility exploration using data from the OD Survey. From a visualization perspective, improvements in the map to display region divisions on top of the trajectories can help to better identify the connections between the regions, as in [Klein et al. 2014]. A different approach to convey the density information of bundles could be to scale edge lines thickness proportionally to the expansion factor of each record instead of using colors, similarly to [Lhuillier et al. 2017]. This design is worth exploring, as it would eliminate the need to replicate edges and would therefore significantly reduce the size of the dataset. From an application perspective, there are many reasons to explore other data attributes, like gender, trip reason, family income, and the correlations between them. It would be also important to assess the usefulness of the visualizations with feedback from actual traffic managers to fill possible gaps between our mobility visualization research and practical solutions to real problems.

## Acknowledgments

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

## References

- Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., and Ertl, T. (2017). Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum*, 36(8):260–284.
- Chen, W., Guo, F., and Wang, F. Y. (2015). A Survey of Traffic Data Visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984. doi:10.1109/TITS.2015.2436897.
- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619.
- Graser, A., Schmidt, J., Roth, F., and Brändle, N. (2017). Untangling origin-destination flows in geographic information systems. *Information Visualization*, 18(1):153–172. doi:10.1177/1473871617738122.

- Holten, D. (2006). Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE TVCG*, 12(5):741–748.
- Hurter, C. (2015). *Image-Based Visualization: Interactive Multidimensional Data Exploration*. Morgan Claypool.
- Hurter, C., Ersoy, O., Fabrikant, S. I., Klein, T. R., and Telea, A. C. (2014). Bundled Visualization of Dynamic Graph and Trail Data. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1141–1157. doi:10.1109/TVCG.2013.246.
- Hurter, C., Ersoy, O., and Telea, A. (2012). Graph bundling by kernel density estimation. *Comput. Graph. Forum*, 31(3pt1):865–874. doi:10.1111/j.1467-8659.2012.03079.x.
- Klein, T., van der Zwan, M., and Telea, A. (2014). Dynamic multiscale visualization of flight data. *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, 1:104–114. URL: [ieeexplore.ieee.org/xpl/abstractMetrics.jsp?arnumber=7294794](http://ieeexplore.ieee.org/xpl/abstractMetrics.jsp?arnumber=7294794).
- Lhuillier, A., Hurter, C., and Telea, A. (2017). FFTEB: Edge bundling of huge graphs by the fast fourier transform. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 190–199. doi:10.1109/PACIFICVIS.2017.8031594.
- Lhuillier, A., Hurter, C., and Telea, A. (2017). State of the Art in Edge and Trail Bundling Techniques. *Comput. Graph. Forum*, 36(3):619–645. doi:10.1111/cgf.13213.
- Metrô SP (2018). *A mobilidade urbana da região metropolitana de São Paulo em detalhes*. Secretaria Estadual dos Transportes Metropolitanos e Companhia do Metropolitano de São Paulo, São Paulo. URL: [metro.sp.gov.br/pesquisa-od/arquivos/Ebook%20Pesquisa%20OD%202017\\_final\\_240719\\_versao\\_4.pdf](http://metro.sp.gov.br/pesquisa-od/arquivos/Ebook%20Pesquisa%20OD%202017_final_240719_versao_4.pdf).
- Peysakhovich, V., Hurter, C., and Telea, A. (2015). Attribute-driven edge bundling for general graphs with applications in trail analysis. In *Proc. IEEE PacificVis*.
- Slocum, T. A., McMaster, R. B., Kessler, F. C., and Howard, H. H. (2009). *Thematic Cartography and Geovisualization*. Pearson Prentice Hall.
- Telea, A. and Ersoy, O. (2010). Image-based edge bundles: Simplified visualization of large graphs. *Computer Graphics Forum*, 29(3).
- Vale, R. C. C. (2018). The welfare costs of traffic congestion in São Paulo Metropolitan Area. Master’s thesis, Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto — USP. doi:10.11606/D.96.2018.tde-01082018-091126.
- Van Der Zwan, M., Codreanu, V., and Telea, A. (2016). CUBu: Universal Real-Time Bundling for Large Graphs. *IEEE Transactions on Visualization and Computer Graphics*, 22(12):2550–2563. doi:10.1109/TVCG.2016.2515611.
- Willems, N., Scheepens, R., Van De Wetering, H., and Van Wijk, J. J. (2009). Visualization of vessel traffic. *Situation Awareness with Systems of Systems*, 9781461462(3):73–87. doi:10.1007/978-1-4614-6230-9\_5.

- Zeng, W., Shen, Q., Jiang, Y., and Telea, A. (2019). Route-aware edge bundling for visualizing origin-destination trails in urban traffic. *Computer Graphics Forum*, 38(3):581–593. doi:10.1111/cgf.13712.
- Zhou, H., Xu, P., Yuan, X., and Qu, H. (2013). Edge bundling in information visualization. *Tsinghua Science and Technology*, 18(2):145–156. doi:10.1109/TST.2013.6509098.