A WEIGHTED TRAVEL TIME INDEX BASED ON DATA FROM E-HAILING TRIPS: AN APPLICATION FOR SÃO PAULO, BRAZIL

Renato S. Vieira
Eduardo A. Haddad

A Weighted Travel Time Index Based on Data From E-Hailing Trips: An Application for São Paulo, Brazil

Renato Schwambach Vieira and Eduardo Amaral Haddad

Abstract: In this paper, we combine data from Uber Movement and from a representative household travel survey to constructs a weighted travel time index for the Metropolitan Region of São Paulo. The index is calculated based on the average travel time of Uber trips taken between each pair of traffic zone and in each hour between January 1st, 2016 to December 31, 2018. The index is weighted based on the travel patterns reported in a representative household travel survey, thus the results reflect average congestion levels faced by individuals in the city. We show that the index has a strong correlation with traditional measures of congestion, however, it has a broader coverage of the road network. Finally, we run two analyses using the index: 1) we evaluate the trends of traffic congestion between 2016 and 2018, showing a significant decline in average time spent in traffic; 2) We analyze the effect of different events on traffic congestion in the city, including holidays, public transit strikes, road shutdowns, rain and Major sport events.

1. INTRODUCTION

Traffic congestion is a major component of transport systems efficiency. With higher levels of congestion, more time is spent in traffic and less time is available for productive activities and for leisure (Litman, 2009). Therefore, congestion represents a major source of costs and inefficiency for urban economies. Because of that, measuring congestion is a key element for monitoring and evaluating transport systems and for supporting policy decisions aimed to improve them. However, traditional methods for measuring congestion such as loop detectors are costly and ineffective, especially for tracking congestion with a highly granular temporospatial coverage (Kalinic & Krisp, 2019). The costs of these traditional methods are particularly relevant for developing country cities where traffic agencies face stricter financial constraints (Vaziri, Jafarabady, & Bindra, 2007).

However, with technological improvements and the ubiquitous spread of cellphones with geographical tracking and transportation applications, in-vehicle data is now broadly available, creating the opportunity for the development of alternative methods to measure characteristics of transport systems with detailed coverage both in terms of time and space and with costs that represent a fraction of those required to install and maintain traditional

monitoring structures. One example of this new type of data is the information compiled by the *Uber Movement* website,[1] which provides the average travel time of trips made by Uber vehicles on selected cities throughout the world by hour and by pair of neighborhood origin and destination.

In this paper, we explore data from the Uber Movement (UM) website and combine it with a representative household travel survey for the Metropolitan Region of São Paulo[2] (MRSP) to create a virtually cost-free Travel Time Index (TTI) that estimates trip delays due to congestion experienced by residents of São Paulo in every hour throughout the last three years and in virtually all neighborhoods of the city. We compare this index with traditional congestion measures calculated by the government of São Paulo, and we show that while there is a strong correlation between the measures, our index covers a much broader set of roads. Finally, we present two examples of analysis based on the index. First, we evaluate the aggregated evolution of congestion in the three years period that we analyzed. We show a significant decline in overall congestion in the period. Second, we estimate the impact of different events on congestion, identifying their direction and magnitude.

The remaining of this paper is divided as follows: Section 2 describes the UM data and the household travel survey used to construct the index. On section 3 we explain the index and present some of its descriptive characteristics. Section 4 shows the two analyses we carried using the index and Section 5 concludes.

2. DATA

2.1 Uber Movement

The UM project provides data about travel times of trips made by Uber vehicles in selected cities throughout the world. Cities are divided into neighborhoods according to

---

[1] https://movement.uber.com/

[2] With approximately 20.8 million residents, the Metropolitan Region of São Paulo is the largest urban agglomeration of the Southern Hemisphere

official traffic zones or administrative boundaries, and the dataset available in the website includes the average travel time of uber trips made between any two given neighborhoods at a given time interval.[3] That is, for a given neighbor of origin o, and a neighborhood of destination d, if there was a minimum number of Uber trips between those neighborhoods during a time period p, then the website includes the average travel time t_{odp} of these trips.

In the case of São Paulo, the metropolitan area is divided into the 517 traffic zones (TZs) defined by the 2017 Origin-Destination Household Travel Survey carried by the city's Subway Company (described in further details later in this Section), and average travel times are available from January 1st, 2016 to the most recent finished quarter. While different level of data aggregation can be extracted directly from the UM website, we requested from the project's maintainers all the hourly data, for all neighborhood pairs of MRSP in all dates between January 1st 2016 to December 31, 2018. This dataset includes the average travel time observed in 370 million combinations of origin-destination-date-hour. It covers almost all dates[4] and all hours from January 1st 2016 to December 31, 2018 (1,094 dates and 26,256 hours) from 98,063 different TZ pairs.

2.2   The 2017 São Paulo Household Travel Survey

Thee UM data is restricted to travel times. However, our goal is to create a congestion measure that accounts for the travel patterns of São Paulo residents. For example, in the early hours of the day, there is a larger flow of travelers going from residential neighborhoods to the CBD. Therefore, higher levels of congestion in those routes in the

---

[3] Further details about the methodology used to calculate the data is available at the project's technical note available at https://movement.uber.com/_static/76002ded222a46a02ae89f207e91e335.pdf. The release of Uber Movement data is very recent and its caveats and limitations have not yet been formally discussed by independent studies. Wu (2019), compared Uber Movement travel times with predictions from Google Maps API for the Greater Sydney Region, observing that Uber Movement travel times tended to be lower if compared to the Google API predictions. Possible explanations for this difference include differences in travel speed between Uber drivers and other travelers and a conservative time buffer in the case of Google Maps API. An example of academic work using Uber Movement data is Aryandoust, Vliet, & Patt (2019), who calculated parking density maps for 39 different cities throughout the world.

[4] Due to issues in the system of extraction, data from October 27-28, 2018 was not collected.

morning affect more individuals than an equivalent level of congestion in those same routes during the evening.

To account for these differences in flows, we use data from the 2017 Origin Destination Household Travel Survey (OD17) carried by the city's Subway Company. The survey interviewed 31,487 households in the MRSP between June, 2017 to October, 2018, and collected information about 157,992 trips made by 86,318 individuals in the business day immediately before the interview date (METRO, 2019). The survey was designed to be statistically representative for the population of the whole metropolitan region, so each observation included in the survey has an associated survey weight that can be used to estimate population totals.

3. THE WEIGHTED TRAVEL TIME INDEX

A congestion measurement has the objective of allowing quantitative assessments for transportation planners and information to the general public and policy makers. Desirable characteristics of these measurements include: easiness of communication, applicability to different geographical scales, comparability to a certain standard, be measured in a continuous scale, be based on travel time data and be able to describe very congested conditions (Levinson & Lomax, 1996). One type of metric that satisfies all these conditions and that has been commonly used both by academic works and other types of publications[5] are the travel time indexes (TTI), that are calculated by comparing the travel time of a given trip relative to the expected duration of that same trip under a certain baseline, usually the free-flow conditions. Formally, the index is calculated by th formula:

$$i_{odp} = \frac{t_{odp}}{t_{od}^*} - 1 \qquad (1)$$

---

[5] Examples of TTIs used or discussed in academic papers include Li, Hong, & Zhang (2017), Bertini (2006), Mehran & Nakamura (2009), Sweet & Chen (2011), Bennecke et al. (2011), Karim, Daissaoui, & Boulmakoul (2017), (Chen, Tong, Lu, & Wang (2018). The TTI is also among the performance indicators used in Urban Mobility Report, a periodical that compares congestion in the 494 largest American urban areas (Schrank, Eisele, & Lomax, 2019), and is also the indicator used to construct the TomTom Traffic Index that compares congestion levels in 403 cities across 56 countries (TomTom, 2019).

Where the term i_{odp} represents the TTI of trips made between a point of origin o and a point of destination d during the period p. On the right side of the equation, the term t_{odp} represents the travel time observed for this route in period p, while t^*_{od} indicates the travel time of that same route in the free flow conditions. For example, suppose that a certain trip takes on average 32 minutes during the morning peak on business days, and that this same trip takes 20 minutes in free-flowing conditions. In this case, the TTI would be calculated as $32/20 – 1 = 0.6$. That is, the TTI indicates that during the morning peak, this route has an average level of congestion of 60%.

A key advantage of this metric is that it has a simple and intuitive interpretation. Another important characteristic is that the indicator can be aggregated both in time and space, allowing the analysis amd comparison of different regions and different periods. In addition, the index aggregation may be weighted to consider the different travel volumes between regions at different points in time. The aggregation of the TTI can be descried by the following formula:

$$I_{RP} = \sum_{odp} \left(\frac{t_{odp}}{t^*_{od}} \times \frac{v_{odp}}{V_{RP}} \right) \quad | \quad (o, d \in R) \quad \& \quad (p \in P) \tag{2}$$

Where I_{RP} is the aggregated TTI for region R over a time period P. Besides that, v_{odp} is the number of trips observed between the traffic zones o and d that make up region R, and all periods p that compose P. V_{RP} is the total number of trips between all traffic zone pairs within region R and all the periods that make up the period P. For example, if o and d are traffic zones, and p are hours of the day, P can be defined as a certain date, and R as a Metropolitan Region as a whole. The result of this aggregation can be directly interpreted as the average level of congestion among all trips made in region R throughout the whole period P.

3.1 Adjustments for the implementation in São Paulo using UM data and the OD17

From the elements of Equation 2, UM data includes the average travel time by hour for each pair of traffic zones in the SPMR (t_{odp}). So, to calculate the weighted TTI, we still need: 1) the free-flow time between each pair of TZ (t_{od}^*); 2) the number of trips made between each TZ pair at each period (v_{odp}) and the corresponding aggregation (V_{RP}). Next, we describe how each of these elements are calculated to build the index.

3.1.1 Free-flow travel time (t_{od}^*)

The free-flow travel time is the time that a trip would take if the roads were completely free of other vehicles and other factors causing slowness, thus being a theoretical measure. While simulation methods could be used to estimate this metric for the TZ pairs included in our study, however, such models would require detailed information about the road network such as speed limits and the distribution of origins and destination exact coordinates of trips, both of which are not easily available. Therefore, to overcome these limitations, we estimate free-flow travel time relying on the UM data and the assumption that there are periods of the day when observed travel times approximate free-flow conditions, most often during late night hours.

Based on these assumptions, we define free-flow travel time for each TZ pair as the second lowest average travel time value per hour during the most recent quarter of data from UM. The second lowest value is used instead of the first lowest to avoid selecting eventual outlier observations from pairs with fewer Uber trips.[6] The most recent quarter is used because as Uber expands its services, the density of data from each TZ pair also increases.

To illustrate this approach, Figure 1 shows the average travel time by hour in a selected TZ pair from the SPMR during the last quarter of 2018. As expected, average travel times by hour are lower during late night hours and increase during peak periods.

---

[6] The second lowest hourly average travel time corresponds to approximately the 8th percentile of the data per pair.

6

Figure 1: Example of Free Flow Travel Time Estimation

(Chácara Itaim to Aeroporto de Guarulhos – 4ᵗʰ Quarter of 2018)
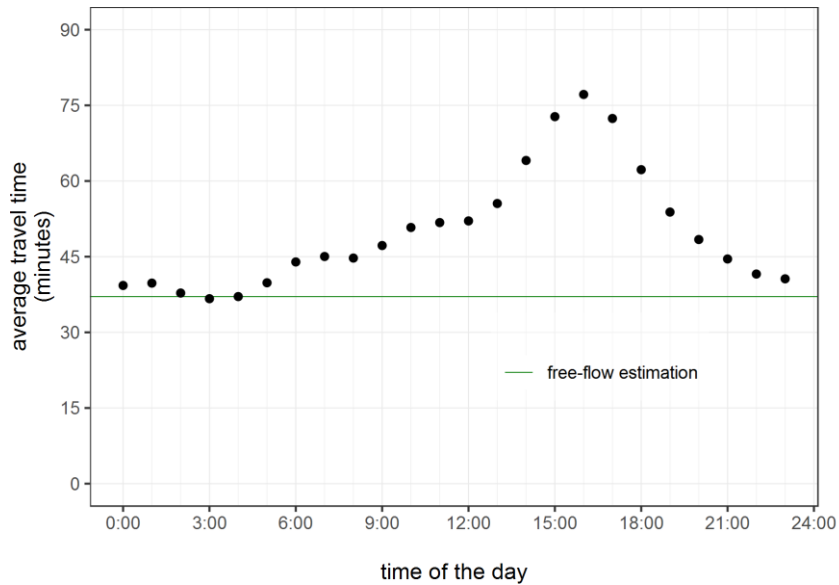
### 3.1.2   Number of trips by region pair (v_{odp})

In order to identify the travel flows between each zone of traffic and each period, we used the micro-data from the OD17 Survey. Although the survey is designed to be statistically representative of the displacement patterns in the SPMR, there are two important practical constraints that need to be considered in the use of its data to calculate the weighting of the TTI:

1) The number of TZ pairs observed on UM is not constant; there are days and times when travel time information between a given TZ pair does not exist on the platform because the number of trips made with the Uber application is not sufficient for the inclusion in the system.

2) The OD17 Survey is not statistical representative at the level of TZ pairs per hour. The Survey includes information about 157,992 trips made in a typical working day, however, there is a total of 6.4 million combinations of TZ pairs*hours (517 TZs x 517 TZs x 24 hours), therefore the Survey data is not sufficiently dense to inform about travel flows in the same disaggregation level as the travel times from UM.

Regarding restriction 1), the main issue is that, with the expansion of Uber's activities, a direct comparison of congestion levels between distinct periods is not necessarily valid. If the TTI is calculated without considering the differences in composition over time, the results of the analysis may be biased. For example, it is possible that the pairs observed in 2016 are mostly in the more central regions of the MRSP where congestion levels are naturally higher. On the other hand, it is possible that in the case of the observations from 2018, the proportion of suburban TZ pairs become larger. In this case, even if the congestion levels were unchanged in the MRSP, a TTI that does not consider the composition difference between the periods would indicate a congestion drop over time.

To work around this problem, the solution is to keep constant the TZ pairs included in the index construction. We selected the pairs observed in at least three quarters of the days in the analyzed period. With this criterion, 23,807 pairs were selected, and they represent 28.8% of motorized trips in the MRSP according to according to the OD17. Figure 2 shows that although the selected pairs correspond to only 8.9% of the total possible combinations between pairs of zones, they cover practically the entire urbanized area of the MRSP, with the exception of some of the most isolated districts.

**Figure 2: Traffic Zones Included in Selected Pairs of Zones for Analysis**
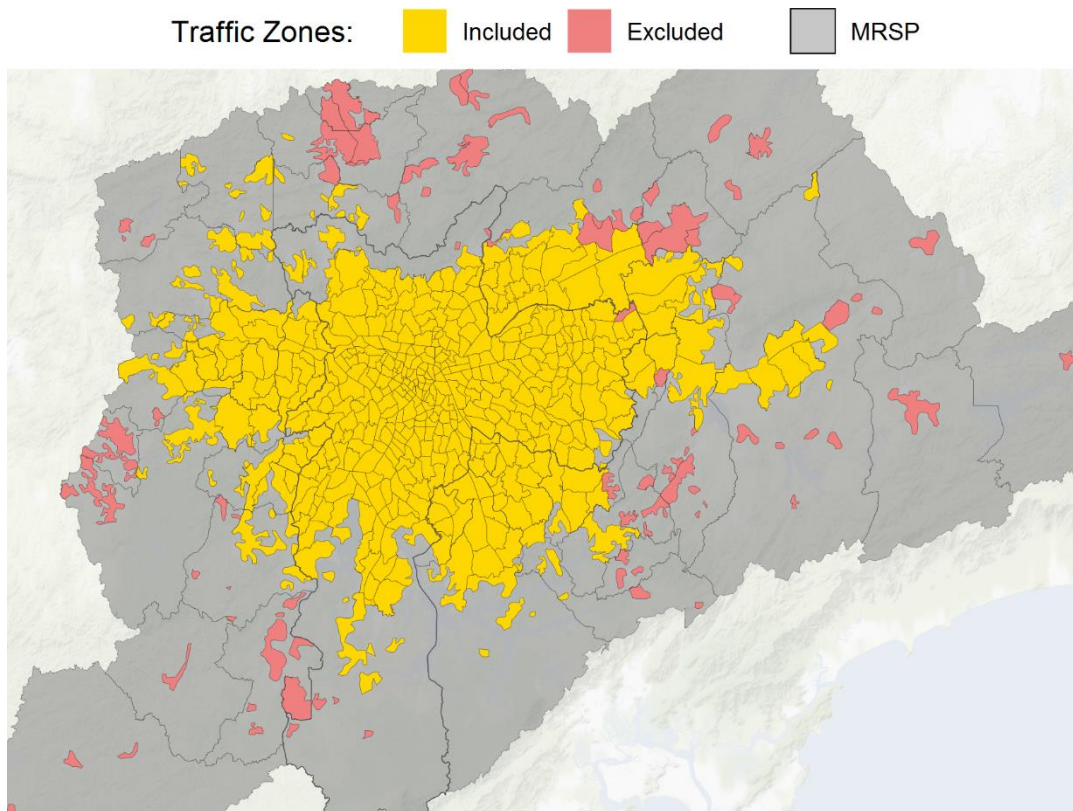
Notes: The map is restricted to the urban area of TZs.

Regarding restriction 2), that is, the non-statistical representativity of the OD17 Survey at the level of TZ pairs by hour, the workaround was to approximate the weighting calculation using a higher level of temporospatial aggregation. Specifically, the MRSP was divided into nine macro-regions and the weights were calculated based on the number of trips observed in the OD17 between pairs of macro-regions during periods of the day[7] rather than the number of trips between each pair of TZ by hour, thus ensuring the statistical representativity of the patterns observed between each pair used for the weighting. The macro-regions used in the study were defined according to Figure 3 below.

---

[7] Late-night (midnight to 7 a.m.), morning peak (7 a.m. to 10 a.m.), midday (10 a.m. to 6 p.m.), afternoon peak (4 p.m. to 7 p.m.) and night (7 p.m. to midnight). These are the periods defined on the global UM plataform.

Figure 3: Macro-regions Used for the TTI Weighting
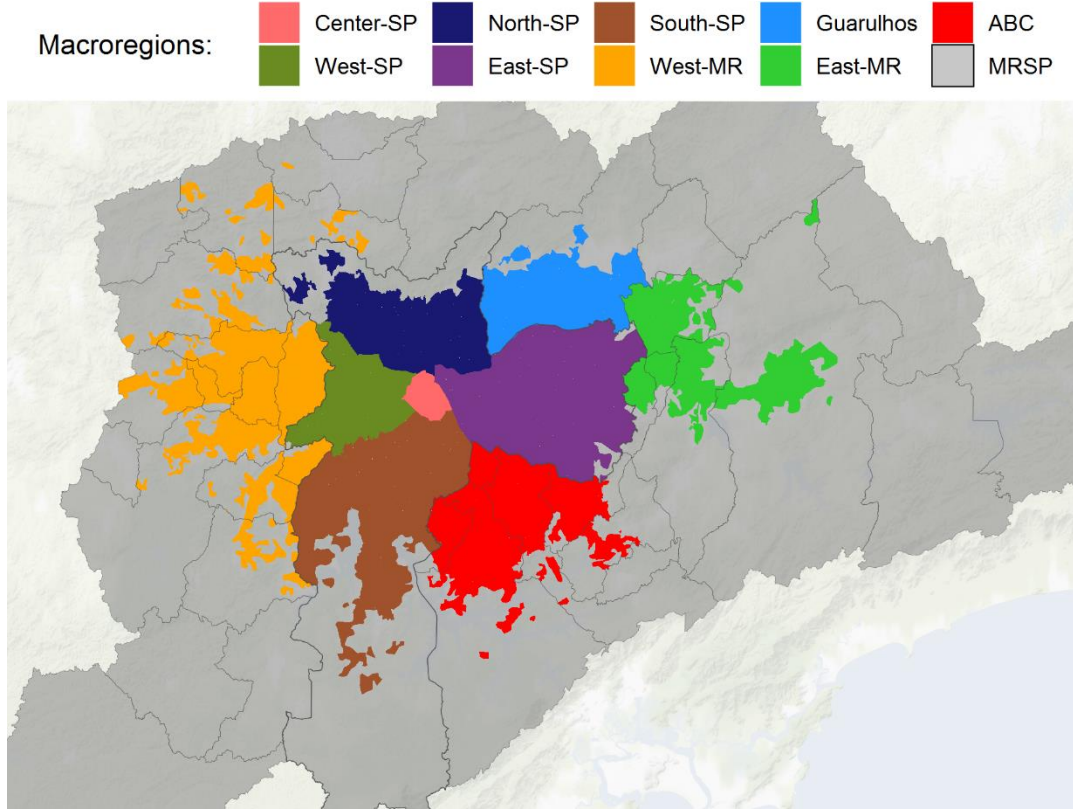
Notes: Each region was defined from the aggregation of TZs from the OD17. The regions within the city of São Paulo were defined according to the spatial division used by the São Paulo Traffic Authority (CET) to report their traditional congestion indicator at cetsp1.cetsp.com.br/monitransmapa/agora/. The regions composed of TZs from the other MRSP municipalities were defined by their geographic location. TZs without a sufficient number of observations on the UM platform were excluded.

Finally, to ensure that the proportion of trips between each macro-region is constant even if the number of TZ pairs change over time in in the UM data, an adjustment factor was added to the TTI formula in order to maintain constant the total weight of each macro-region pair.

$$w_{odp} = \frac{(v_{\bar{O}\bar{D}p})/N_{\bar{O}\bar{D}p}}{V_{RP}} \tag{3}$$

Where $v_{\bar{O}\bar{D}p}$ is the total number of trips observed in the OD17 Survey between the OD macro-region pair during period P, such as $o \in O$ and $d \in D$. In addition, $N_{\bar{O}\bar{D}p}$ is the total number of TZ pairs that make up the macro-region $\bar{O}\bar{D}$ and which have travel time

information in period p on the UM platform. This adjustment ensures that even if the number of TZ pairs fluctuate throughout time, the TTI results will always correspond to the static aggregated traffic volumes observed in the OD17. That is, if a pair of macro-regions contains 10% of the total travel flows in the OD17, then the sum of the weights of the TZ that make up that macro-region pair will always be equal to 10%, and this equivalence will be independent of the number of TZ pairs observed in the UM data in a given day.[8]

Therefore, the final adjusted calculation of the weighted TTI is described by the following formula:

$$I_{RP} = \sum_{odp} \left(\frac{t_{odp}}{t_{od}^*} \times w_{odp} \right) \qquad (4)$$

3.2   TTI Descriptive Results

Based on the formula described in the previous sub-section, the TTI was calculated for all dates from January 1st, 2016 to December 31, 2018. For each day, the index was also calculated for each period of the day and for each macro-region of the MRSP.

The average value for the total TTI was of 34.88%, indicating that on average, trips made in São Paulo are 34.88% longer than if made on free-flow conditions. Figure 4 plot the average TTI by different types of segmentation. In all plots, the dotted red line indicates the overall mean of 34.88% for the period and the whole MRSP. Panel A shows the average TTI by year (2016, 2017 and 2018), showing a decrease of approximately 6.1 percentage points between 2016 (39.6%) to 2017 (33.5%), and a new reduction of 2.1 points in 2018 (31.4%). Panel B shows the within year seasonality of the TTI mean by comparing the average levels of the index by month, showing that in the months of school holidays (January and July) congestion levels are below average (28.3% and 27.9%, respectively). Meanwhile, March is the month with the highest average TTI (40.3%).

---

[8] Attachment A illustrating this weighting adjustment system.

On Panel C, we show the dynamics of congestion within the week. On weekends, the mean TTI is well below average (15.6% on Sundays and 22.2% on Saturdays). On weekdays, there seems to be an increasing pattern, with Friday being the most congested day (46.2%). Panel D shows the dynamics of congestion within the different periods of the day. The afternoon peak, defined as the period between 4-7pm, has the highest average TTI (55.4%). Meanwhile, the night and late-night periods have lower average value, with mean TTIs of respectively 22.2% and 15.3%.
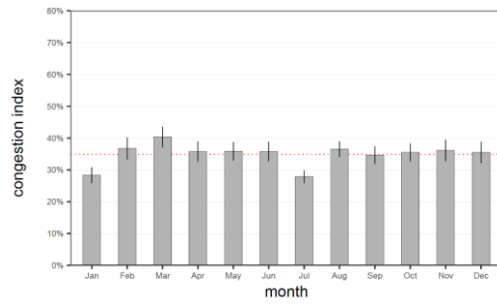
Finally, Panel E shows the average TTI within each of the MRSP macro-region. This comparison shows that the South, West and Central regions of São Paulo have the highest congestion averages (45.5%, 37.4% and 44.2%, respectively). On the other hand, the regions outside the capital (West of MR, Guarulhos, East of MR and ABC) present congestion levels below the overall average (21.9%, 26.4%, 16.3% and 24.8%).
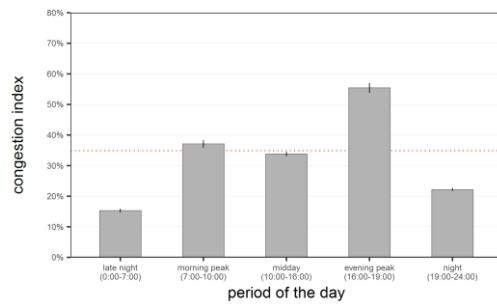
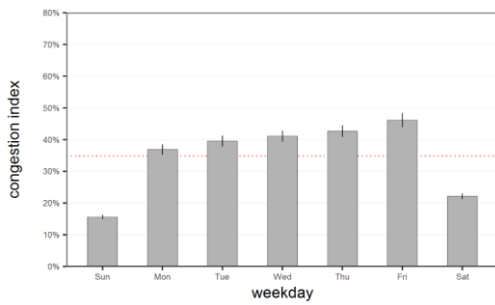**Figure 4: Average Congestion Index per Year**

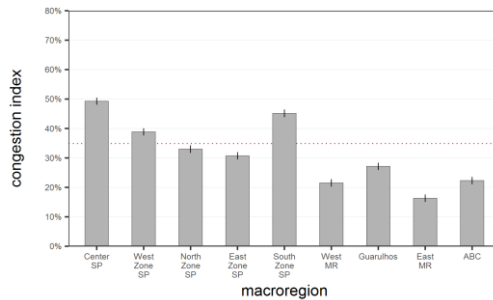*Panel A: Average TTI by Year*
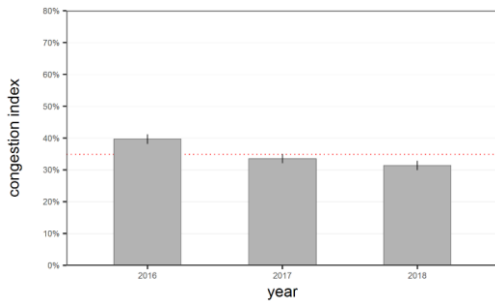
Panel B: Average TTI by Month

Panel C: Average TTI by Weekday

Panel D: Average TTI by Period of the Day

Panel E: Average TTI by Macro-Region of MRSP

Notes: Error bars indicates the 95% confidence interval for the estimated means. The red dotted line indicates the average TTI (34.88%) for the the whole period (2016-2018) and for the whole MRSP.

3.3 Total time spent due to congestion

Based on the TTI and the OD17 data, it is possible to estimate with a simple back-of-the-envelope calculation the average time lost due to congestion in a typical weekday in São Paulo. As shown in the previous section, the average TTI on weekdays is equal to 41.2%, and according to the OD17, in a typical business day, 3.62 million individuals travel by car in São Paulo spending a total of 5.47 million hours on these trips. Therefore, if we divide this total number of hours by one plus the average TTI, we have that those same

trips would spend only 3.87 million hours if they were always made under free-flow conditions. That is, congestion causes individuals who travel by car in São Paulo to spend, on average, 26.4 more minutes than if all trips were performed under free-flow conditions.

As pointed by Litman (2009), expecting all trips to be performed in free-flow conditions is not something reasonable, especially in dense urban environments. However, the objective of this type of calculation is not to set up a target of potential time savings. Instead, the most important value of this estimation is to translate into easily understandable metrics the potential outcomes of different scenarios. For example, we can calculate that the 6.1 percentage point reduction in the TTI between 2016 to 2017 corresponds to an average reduction of almost 3 minutes per day for all individuals who travel by private cars[9] in the MRSP.

3.4   Comparison with traditional congestion measures

To further validate the TTI calculated in this paper, we compare it with a traditional congestion metric that is currently calculated by CET[10], which counts the length of congestion on selected roads by hour. The CET measure is calculated based on the evaluation of technicians in official vehicles or positioned at the top of buildings. Only roads with the largest volume of vehicles are measured, corresponding to just over 800km of the 17,000km of the total roadway network in São Paulo[11] (CET, 2019).

Figure 5 shows the scatter plot of our TTI on the y-axis against the CET congestion measure on the x-axis. To make the values comparable, we restricted the TTI to the same 5 regions that are used in the CET measure, and we also restricted the analysis to business days. Therefore, each point in the plot represents the value for both measures by region

---

[9] An important limitation of this calculation is that it is restricted to trips made with private cars (47.4% of the trips made by motorized vehicles that circulate on the streets, according to Pesquisa OD 2007). Obviously other modes such as buses and bikes are also affected by congestion, but at different scales. Thus, the measure estimated here does not contemplate the total hours lost by all São Paulo citizens, but only by the portion that circulates by car. The estimation of total losses is undoubtedly an interesting exercise, but it would require additional data and analysis, thus being a suggestion for future studies.

[10] *Companhia de Engenharia e Tráfego*, the São Paulo Traffic Authority.

[11]   More   details   about   the   CET   measure   can   be   accesed   at http://cetsp1.cetsp.com.br/monitransmapa/agora/ajuda.htm

of São Paulo city and by day. The figure shows a clear and strong positive correlation between the 2 measure (0.873). However, it is worth noticing that the TTI values are less likely to be equal or very close to zero, mainly because the road network is not restricted to certain roads as in the case of the CET measure.

Figure 5: Dispersion Graph – Travel Time Index and Kilometers of Congestion (CET) -

Business Days 2016-2018
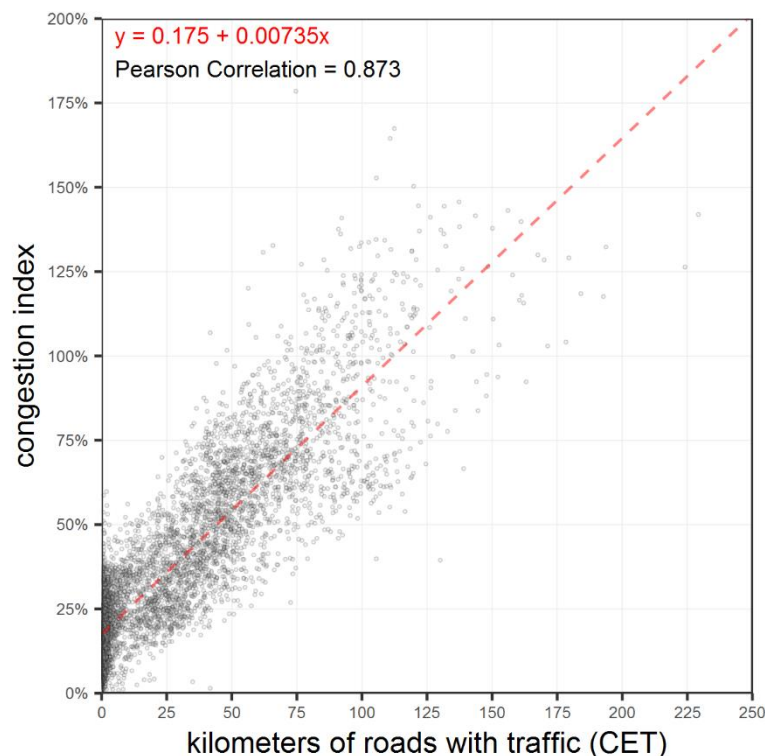
Notes: Each point represents the value of both variables for the 2016-2018 business days. The dashed red line is the prediction of the bivariate model that estimates the congestion index based on the kilometers of slowness plus an intercept.

Next, we present on Figure 6 the separated plots for each region of the city. The results of these analyses show that most series have a high degree of correlation (above 0.7). The only exception is the North Zone, where the correlation coefficient is equal to 0.499, which can only be considered a moderate positive association between the series. One possible explanation for this difference is the more restricted spatial coverage of the CET indicator in the North Zone.

Figure 6: Dispersion Graph - Congestion Index and Kilometers of Slowdown by Region of the City of São Paulo
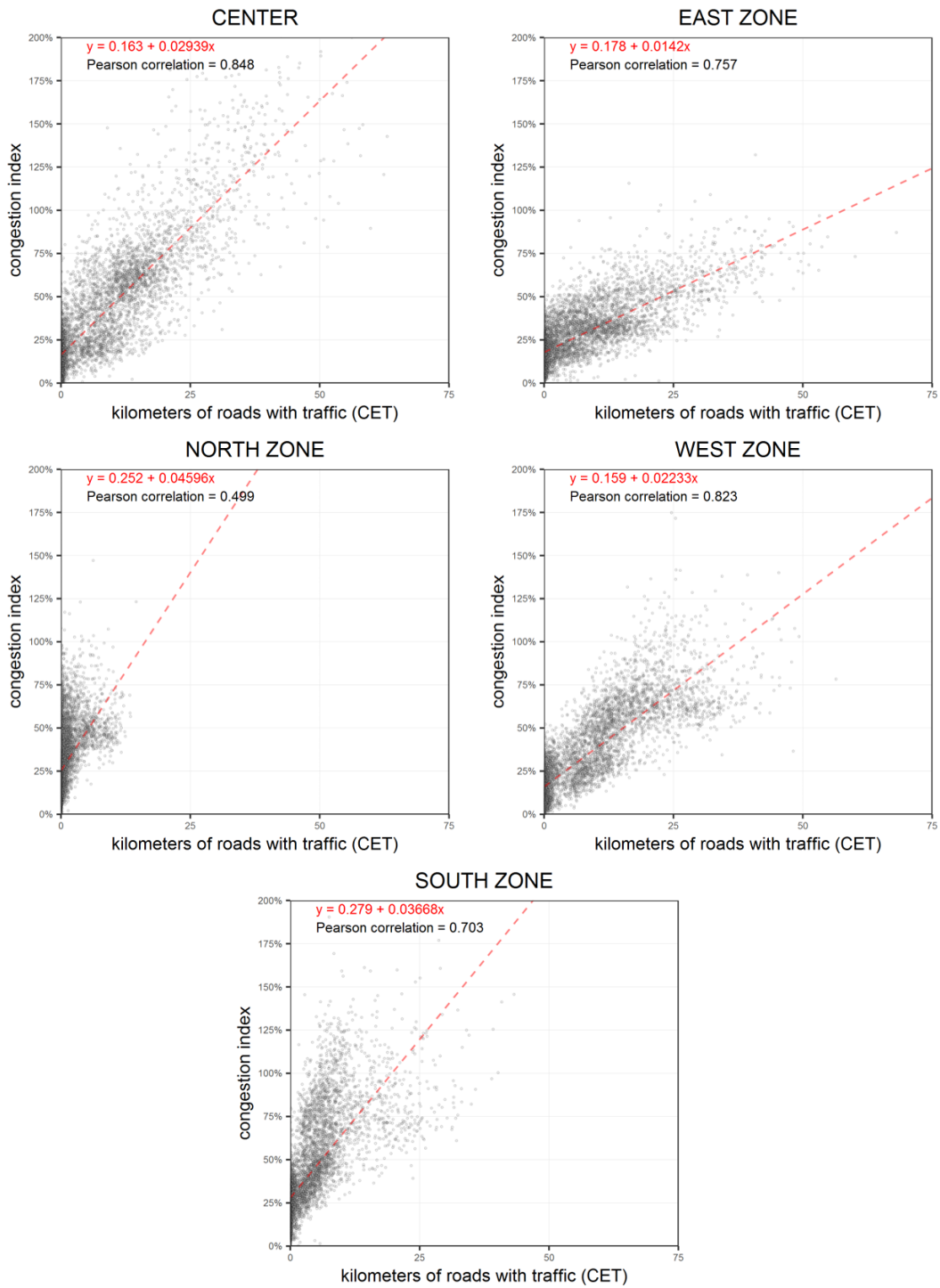
Notes: Each point represents the value of both variables for the period of a day observed between 2016-2018. The dashed red line is the prediction of the bivariate model that estimates the congestion index based on the kilometers of slowness plus an intercept.

The high correlation between our TTI and the traditional CET indicator serves as a validation for the measure developed in this study, encouraging its use for analyses that extrapolate the space and time limitations of the CET congestion indicator. While the series are similar, it is important to highlight a key difference between the indicators: the CET measurement is based on the level of traffic flow slowness observed in the main avenues of São Paulo, meanwhile, the TTI is calculated from actual travel times of real trips, therefore it already accounts for drivers routing optimization and adjustments to real traffic conditions. Thus, in addition to having a simpler and more direct interpretation, the TTI results can be directly translated into travel costs.

In the next section, we explore and present additional and more detailed examples of analyses that can be performed using the TTI results that were calculated in this Section.

4. ANALYSING THE IMPACT OF DIFFERENT EVENTS OF TRAFFIC CONGESTION

There are several factors that can affect the level of congestion in a city like São Paulo, including climatic events, strikes, political demonstrations, festivities, historical trends and other atypical events. Given the granularity of the time series of the congestion index, it is possible to estimate and compare the impact of different events in the city traffic through a multivariate analysis. In the period of our analysis, by way of illustration, we identified the following factors that could potentially affect São Paulo traffic:

- Rainy days;
- A national truck drivers' strike between May 21 to May 31, 2018;
- The 2018 FIFA World Cup (especially when the Brazilian national team was playing);
- School holidays;
- The closing of part of Marginal Pinheiros due to the collapse of a bridge (Nov. 15, 2018).

To evaluate the impact of each of these events, we estimate the following multivariate regression model:

17

$$C_t = \alpha_t + \delta_t + \beta R_t + \sigma S_t + \gamma W_t + \gamma W B_t + \eta H_t + \theta E_t + \zeta S H_t + \omega_t t_t + \varepsilon_t \qquad (5)$$

Where:

C_t: the congestion index C in the MRSP during each date t;

α_t: a vector of year specific intercepts;

δ_t: a vector of fixed effects associated with each day of the week (Mon, Tue, ...);

R_t: a dummy variable indicating days with more than 0.1mm of accumulated rain in the RMSP;

S_t: a dummy indicating the dates of the truck drivers' strike (May 21-31, 2018);

W_t: the 2018 FIFA World Cup dates (06/14/2013-07/15/2018)

WB_t: The Brazilian National team games during the World Cup;
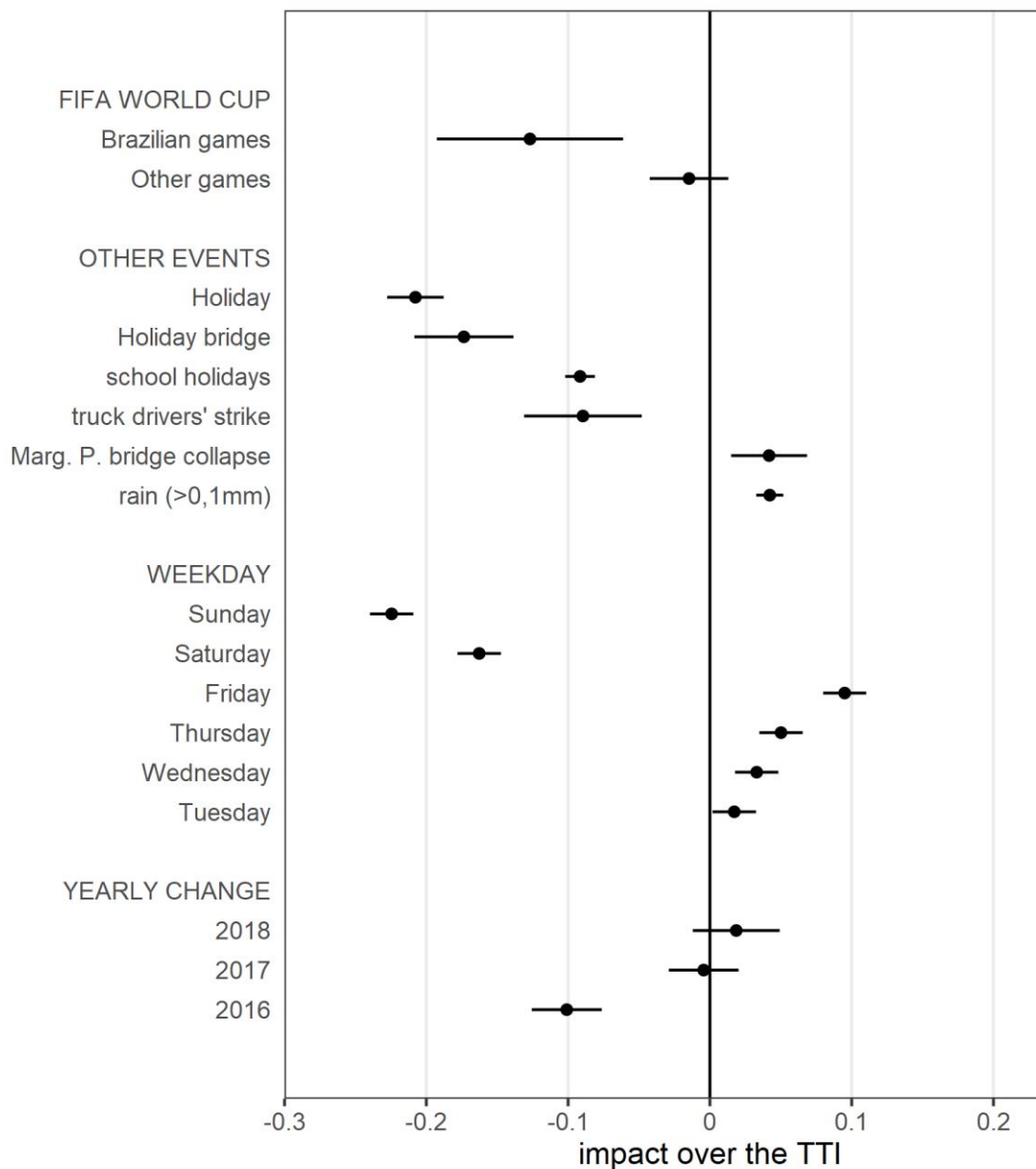
H_t: holidays;

E_t: holiday bridges;[12]

SH_t: school holidays (winter and summer);

ω_t: year specific linear time trends.

The results of this model estimation are presented in Figure 7 below:

---

[12] Holiday bridges occur on Mondays when Tuesdays are holidays, and Fridays when Thursdays are holidays.

Figure 14: Impacts of Different Events on the RMSP Daily Congestion Index

Notes: The points represent the estimated coefficients for each variable from Equation (5), and the horizontal bars indicate the 95% confidence interval of these estimates. Impacts can be interpreted as absolute variations of the index, that is, a coefficient of 0.1 indicates a TTI that is 10 percentage points higher for that specific variable.

During the 2018 World Cup, when the Brazilian National was not playing, the average congestion was not much different than usual, averaging 1.4 percentage points below expected (a nonsignificant result at 5% significance). However, on the days of Brazil's games, the index was on average 11.2 points below usual.

Regular holidays are associated with a congestion reduction of -20.7 percentage points and holiday bridges have a similar association (-17.3 points). During school holidays, congestion is 9.1 points lower. During the truck drivers' strike, the circulation of vehicles was greatly reduced due to the lack of fuel at filling stations. So not surprisingly, the congestion index during the strike was 8.9 points lower than usual. In the days following the closing of Marginal Pinheiros due to a bridge collapse, there was an increase of approximately 4.1 points in the TTI. Finally, on days with average rainfall above 0.1 mm per hour, congestion is on average 4.2 points higher than on non-rainy days.

As already noted in Section 2, the congestion patterns are heterogeneous throughout the days of the week. The multivariate regression estimated here confirms the same patterns. The reference group are Mondays. Therefore, the results indicate that the other days of the week present an average value of increasing congestion, with the highest value observed on Fridays (9.5 points higher than Mondays). On weekends, the TTI is well below the reference group, respectively -16.2 on Saturdays and -22.4 on Sundays. Finally, the year specific slopes show that while there was a significant decrease of 10 points in 2016, the TTI remained mostly stationary during 2017 and 2018.

5. CONCLUSIONS

As in most large metropolises around the world, traffic congestion is one of the greatest problems faced by the residents of São Paulo, generating economic and welfare losses for residents and for visitors. A first step for addressing the problem of traffic congestion is to measure, monitor and understand the phenomenon. However, one difficulty faced by technicians and researchers, particularly in developing world cities, is the lack of large-scale quantitative measures with high temporospatial granularity.

The TTI built in this study based on data from an e-hailing company aims to provide a new tool for congestion analysis that is virtually free and could potentially be extended to other cities and to other types of analysis. The indicator constructed here differs from traditional measures used in São Paulo by being based on actual travel time from real trips rather than road-based metrics. Because of that, the indicator directly reflects travel time costs while accounting for route adaptation and optimization. Additionally, the index

created here suggests a framework for integrating UM data with household travel surveys in order to create a weighting scheme that makes the index results to reflect the travel patterns and average delays experienced by drivers.

Still, the framework proposed in this paper is open to technical improvements and adjustments, such as refinement of free-flow metrics and travel weighing. In any case, the analyses presented here represent an initial outline of the possible applications of the measure. Thus, other studies using and improving the indicator, as well as exploring the Uber Movement, are highly recommended. It is hoped that such tools and studies will be used to objectively inform the urban mobility debate and assist public policy planners in their decisions.

6. BIBLIOGRAPHY

Aryandoust, A., Vliet, O. v., & Patt, A. (2019). City-scale car traffic and parking density maps from Uber Movement travel time data. *Scientific Data*, 1-18.

Bennecke, A., Friedrich, B., Friedrich, M., & Lohmiller, J. (2011). Time-dependent Service Quality of Network Sections. *Procedia Social and Behavioral Sciences*, 364-373.

Bertini, R. (2006). You are the traffic jam: an examination of congestion measures. *Transportation Research Board*.

CET. (2019, 11 4). *cetsp1*. Retrieved from monitransmapa: http://cetsp1.cetsp.com.br/monitransmapa/agora/ajuda.htm

Chen, P., Tong, R., Lu, G., & Wang, Y. (2018). Exploring Travel Time Distribution and Variability Patterns Using Probe Vehicle Data: Case Study in Beijing. *Journal of Advanced Transportation*, 1-13.

Haddad, E., & Vieira, R. (2015). Mobilidade, Acessibilidade e Produtividade: Nota Sobre a Valoração Econômica do Tempo de Viagem na RMSP. *Revista Economia Contemporânea*, 343-365.

Kalinic, M., & Krisp, J. M. (2019). Floating Car Data and Fuzzy Logic for classifying congestion indexes in the city of Shanghai. *Proceedings of the International Cartographic Association*.

Karim, L., Daissaoui, A., & Boulmakoul, A. (2017). Robust routing based on urban traffic congestion patterns. *Procedia Computer Science*, 698-703.

Levinson, H., & Lomax, T. (1996). Developing a Travel Time Congestion Index. *Transportation Research Records*, 1-10.

Li, Z., Hong, Y., & Zhang, Z. (2017). Do Ride-sharing Services Affect Traffic Congestion? An Empirical Study of Uber Entry. *SSRN*.

Litman, T. (2009). *Transportation Cost and Benefit Analysis.* Victoria: VTPI.

Ma, P., & Sun, X. (2015). Leveraging for big data regression. *Wiley Interdisciplinary Reviews: Computational Statistics, 7*, 70-76. doi:10.1002/wics.1324

Mehran, B., & Nakamura, H. (2009). Considering travel time reliability and safety for evaluation of congestion relief schemes on expressway segments. *IATSS Research*, 55-70.

METRO. (2019). *A Mobilidade Urbana da Região Metropolitana de São Paulo em Detalhes.* São Paulo: METRO.

Schrank, D., Eisele, B., & Lomax, T. (2019). *2019 Urban Mobility Report.* The Texas A&M Transportation Institute.

Sweet, M., & Chen, M. (2011). Does regional travel time unreliability influence mode choice? *Transportation*, 625-642.

TomTom. (2019, 11 04). *tomtom.com*. Retrieved from traffic-index: https://www.tomtom.com/en_gb/traffic-index/

Vaziri, M., Jafarabady, R., & Bindra, S. P. (2007). Modeling Highway Congestion Index for a Developing Country: The Iran Experience. *Scientia Iranica*, 1-10.

Wu, H. (2019). Comparing Google Maps and Uber Movement Travel Time Data. *Transport Findings*, 1-5.