Travel time estimation for public transportation systems

Renzo Massobrio[1], Diego Hernández[2], and Martín Hansz[2]

[1] Universidad de la República, Uruguay / Universidad de Cádiz, Spain
renzom@fing.edu.uy
[2] Universidad Católica del Uruguay, Uruguay
dhernandez@ucu.edu.uy, martinhansz@gmail.com

Abstract. Travel time is a widely used indicator of the accessibility offered by public transportation systems. This article proposes a comprehensive model for travel time estimation between different zones in a city. The proposed approach models the public transportation network as a graph and uses data regarding the roadway infrastructure, the public transportation lines, stops, and schedules; and accounts for restrictions on the maximum walkable distance and the maximum number of transfers allowed. The model is applied to the public transportation system in Montevideo, Uruguay, and validated using two alternative sources of travel time data. Due to the intense computational load, the travel time estimations are parallelized to benefit from multiple computing resources. Results indicate a strong correlation between the estimated travel times and the alternative data sources, thus validating the proposed approach. Finally, an interactive web application is presented, which allows displaying the estimated travel time matrix in an intuitive way.

Keywords: public transportation, travel time estimation, accessibility

1 Introduction

In urban scenarios, citizens are required to travel in order to engage in the social and economic activities of their city [1]. In this context, public transportation plays a major role, since it is the most efficient and socially-fair mean of transportation [3]. Understanding the accessibility of citizens to the public transportation service of a city is paramount in order to identify inequalities among the population and implement policies that aim at improving the quality of service offered to passengers. Several indicators may be considered to measure the accessibility offered by a transportation system, among them, travel time strikes as the most intuitive one, since it is tightly related to the perception of passengers of the quality of service of a transportation system [14].

Public transportation systems operate on predefined routes and depend on schedules that vary throughout the day. Additionally, travel speeds vary greatly due to traffic congestion, passenger demand, and road infrastructure. Thus, assuming constant speed of vehicles along routes usually results in significant travel time differences between the estimations of the model and the actual reality. Furthermore, trips in public transportation are usually comprised of several stages, including, walking to/from stops, waiting at stops for vehicles, actually traveling through the public transportation network, and may even include transfers between different lines. A comprehensive model for travel time estimation in public transportation networks needs to account for all these factors.

In this paper we present a model to estimate travel times in public transportation between different zones of a city. The proposed approach models the public transportation network as a graph and incorporates data regarding the roadway infrastructure, the public transportation lines, stops, and schedules; and allows imposing restrictions on the maximum walkable distance and the amount of transfers involved in a trip. Travel times are estimated using a shortest-path algorithm on the generated graph. The public transportation system in Montevideo, Uruguay, is used as a case study. Due to the intense computational burden, the shortest path calculations are parallelized to benefit from multiple computing resources. For the studied scenario, a travel time matrix is built and compared against results from a government web application and to the findings of a recent mobility survey. Results indicate that the proposed approach is suitable to accurately estimate travel times between different zones in the city, showing a strong correlation with the other sources of travel time data. Finally, an interactive web application is presented, which allows conveying the information of the estimated travel time matrix in an intuitive way.

The remainder of the paper is organized as follows. Section 2 reviews works in the related literature. Section 3 outlines the proposed model for travel time estimation. The application of the model to the case study is presented in Section 4. The web application to visualize the computed results is presented in Section 5 and the conclusions and future work are presented in Section 6.

2 Related works

Lei and Church (2010) presented a short review on measuring the accessibility in public transportation systems [8]. The survey showed that several authors are concentrated on the physical aspects of a system (e.g., distance to a bus stop) instead of focusing on the travel time between pairs of locations. Furthermore, previous works which do focus on travel times usually make assumptions which significantly impact the accuracy of their estimations, e.g., constant transfer and waiting times, average speed of vehicles, or not considering bus schedules at all. The authors propose an extended GIS data structure to account for the temporal dimension of public transportation systems which is applied to the public transportation system in Santa Barbara, California.

Salonen and Toivonen (2013) presented a comparison of different travel time measures [14]. The work covers both travel times using private vehicles and public transportation. Regarding the latter, three models are outlined and applied to a case study in the capital region of Finland: a simple model which does not include vehicle schedule information at all, an intermediate model which uses schedules only to estimate the average waiting time, and an advanced model which queries a government API with up-to-date schedules and uses its routing engine as a black-box to compute travel times. The proposed models identified travel time disparities accross modes (i.e., private vs. public transportation), with a lower effect in areas near the city center.

Previous works have addressed the public transportation system in Montevideo, Uruguay, which is used as a case study in this article. Massobrio (2018) presents an urban data analysis approach to understand mobility in the city using different sources of urban data [10]. Origin-destination matrices, which describe mobility, were built using ticket sales data. Other studies have measured the quality of service offered by the public transportation service in Montevideo, by analyzing punctuality based on GPS bus location data [11, 2].

Regarding accessibility, Hansz (2016) studied the disparity between transport provision and transport needs in Montevideo [5]. An *Index of Public Transport Provision* was defined to measure provision at a given area, which combines bus frequencies and number of bus stops. Results showed that transport provision is highly skewed towards the city center. More recently, Hernández (2018) studied the inequalities in access to jobs and education across different social classes due to the public transportation offer [6]. Travel times were used to measure accessibility, which were obtained through a government web application which is described next. Results showed an unequal distribution of mobility opportunities, especially for commuting citizens and students of upper level public education.

Regarding travel times, city authorities in Montevideo offer a webapp named *Cómo Ir*[1], which allows passengers to search for the best route that connects two points in the city via public transportation. The site allows users to select two points in a map and displays a list of routes that connect the origin to the destination. Additionally, users may indicate their desired time of departure or arrival to get the estimated travel time based on the line schedule. A mobile version is also available, which incorporates real-time bus location information from on-board GPS units. Computed routes only consider up to one bus transfer. This is a clear limitation of the service, specially when trying to find routes between two poorly-connected areas of the city. The model proposed in this article allows to configure the maximum number of transfers on routes, thus providing with more flexibility when computing travel times.

Another source of mobility data in Montevideo comes from the metropolitan household survey conducted in 2016 [12]. The survey aimed to update mobility information in the city, which dated back to 2009. The survey was designed in order to obtain a broad picture of mobility in the city, considering all modes of transportation and also taking into account the metropolitan area, which includes towns and villages outside of Montevideo. Face-to-face interviews were carried out during working days from August to October 2016 in 2230 households to 5946 individuals. Each surveyed individual was inquired about each trip done between 4.00 a.m. on the previous day of the interview to 4.00 a.m. of that day. Regarding travel times, each individual was asked to estimate the total time of

[1] comoir.montevideo.gub.uy

each surveyed trip. It is worth noting that travel times were recorded as reported by the interviewed subjects. Thus, personal biases may affect the estimations, with passenger over- or under-estimating their actual travel times.

According to the literature reviewed, few previous works have applied a systematic procedure to estimate travel times by public transportation in Montevideo, Uruguay. Additionally, the model proposed in this article combines several sources of information and models the public transport network as a graph, to compute travel times between different zones in the city. Moreover, it allows configuring the maximum walkable distance and the maximum number of transfers within the route, which represents a clear improvement over the reviewed works in the literature and the existing web solutions.

3 Proposed model for travel time estimation

This section describes the proposed model for travel time estimation in public transportation systems.

3.1 Overview of the model

Although the proposed model for travel time estimation is suitable to different means of public transportation, for the sake of clarity, we would focus on trips done by bus, which are the ones used in the case study presented in Section 4. When using public transportation, several alternatives routes might exist that connect a given origin to a given destination. Trips may include several stages, e.g., walking from the origin to the bus stop, waiting for the bus to arrive, traveling inside the bus, and finally walking from the alighting bus stop to the destination. Additionally, passengers may transfer between different bus lines, which involves walking between intermediate bus stops and waiting for more buses to arrive. Finally, passengers may opt to walk between the origin and the destination directly, specially for short trips where boarding a bus might be counterproductive. The proposed model accounts for all these alternatives for each origin-destination pair and selects the fastest route to estimate the travel time, assuming passengers take optimal decisions when planning their routes.

The model receives the following data as input:

- A set of zones for which pairwise travel times need to be estimated.
- A roadside network, indicating the streets in the city.
- A bus network, including bus lines and bus stops.
- A bus schedule, with the arrival time of each bus at each bus stop.

The travel time between zones is approximated by the travel time between the centroids of each zone. Thus, all trips starting/ending on a given zone are assumed to start/end at the centroid of that zone. To consider all potential routes that connect each centroid (i.e., each origin-destination pair) the model builds a graph, where a shortest path algorithm is later ran to compute the fastest travel time between each pair of points. More specifically, a directed weighted multigraph is built, i.e., edges have an assigned direction and weight, and nodes may be connected by multiple edges. Nodes represent centroids and bus stops. The weight of an edge indicates the travel time between the connected nodes. Two types of edges are considered, namely, *walking edges* and *bus edges*.

Weights on walking edges are computed using the roadside infrastructure. For this purpose, a routing engine is used, which is able to compute routed distances between a pair of points over the roadside topology given as input. In some cases it might be necessary to approximate nodes by the closest point that falls within the roadside network in order to ensure connectivity. The routing engine may assume constant walking speed, incorporate road characteristics, or even use historic information to estimate walking travel times with a higher accuracy.

The procedure to compute weights on bus edges is described next.

3.2 Computing average travel times by bus

Weights on bus edges correspond to the average travel time by bus between the connecting nodes, considering the waiting time and the travel time inside the bus. These travel times are defined using data from the bus network and timetable schedule. For each trip in the schedule, the effective travel time between each bus stop is computed, by processing the arrival time of each line at each bus stop. Then, all travel times of trips belonging to the same bus line are grouped to get an average travel time of the line for each pair of connected bus stops. Similarly, the average *headway* of each line on each bus stop is computed. Headway is defined as the time between the arrival of a vehicle to a given bus stop until the next vehicle of the same line arrives at the same stop. The average headway is estimated by dividing the number of buses of a given line that visit a certain stop by the length of the timeframe considered for the analysis. In summary, after processing the bus schedule we get, for each pair of bus stops, the average travel time and the average headway of each bus line that connects those stops.

In some cases, a bus line may connect two stops but involving a large detour, resulting in very large travel times. Thus, passengers will not consider such a line as an attractive option to connect both bus stops. Therefore, these lines should not be considered when computing average travel times between stops. For this purpose, lines with an average travel time larger than the time to connect the same pair of stops by the fastest line (considering waiting and effective travel time) are discarded. More formally, for a pair of stops s_1 and s_2, lines l_i which connect s_1 and s_2 are not considered when computing average travel times if another line l_k exists, which also connects s_1 with s_2, and the following inequality holds: $t_{l_i}^{\text{travel}} > h_{l_k} + t_{l_k}^{\text{travel}}$, where $t_{l_j}^{\text{travel}}$ is the average travel time of line l_j when connecting stops s_1 and s_2, and h_{l_j} is the average *headway* of line l_j at stop s_1.

After filtering out lines with large travel times, the average travel time and average waiting time is computed for each pair of stops. Average travel time between a pair of stops is computed by weighing the average travel time of each bus line that connects those stops according to their average headway, as indicated on Equation 1

$$E(T^{\text{travel}}) = \sum_{l_i} P(l_i) \cdot t_{l_i}^{\text{travel}} = \sum_{l_i} \frac{\overbrace{\frac{\min_{l_k} h_{l_k}}{h_{l_i}}}^{\text{number of buses of line } l_i}}{\underbrace{\sum_{l_j} \frac{\min_{l_k} h_{l_k}}{h_{l_j}}}_{\text{total number of buses}}} \cdot t_{l_i}^{\text{travel}} \tag{1}$$

Consider two lines l_1 and l_2 which connect stop s_1 with s_2. The average travel time when using line l_1 is 24 minutes ($t_{l_1}^{\text{travel}} = 24$), whereas the average travel time between the considered stops when using line l_2 is 22 minutes ($t_{l_2}^{\text{travel}} = 22$). Average headways of lines l_1 and l_2 at stop s_1 are $h_{l_1} = 5$ and $h_{l_2} = 10$ minutes, respectively. The average travel time between stops s_1 and s_2 is computed by replacing these values on Equation 1, as shown in Equation 2. It can be seen that, by weighing travel times according to headways (i.e., the inverse of the line frequency), the average travel time is larger than the mean travel time of both lines, since the slower line has a higher frequency than the faster one.

$$E(T^{\text{travel}}) = \Big(\frac{5/5}{5/5 + 5/10}\Big) \cdot 24 + \Big(\frac{5/10}{5/5 + 5/10}\Big) \cdot 22 = 23.3 \text{ min} \tag{2}$$

The procedure to compute the average waiting time at a given stop s_1, when waiting for a set of lines that connect s_1 with s_2, is described next. The proposed model considers the passage times of each line l_i by stop s_1 as a Poisson process of intensity $1/h_{l_i}$, independent from the passage time of the other lines. Therefore, the average waiting time can be computed as outlined in Equation 3.

$$\mathbb{P}(T^{\text{waiting}} > t) = \prod_{l_i} \overbrace{e^{-t/h_{l_i}}}^{\text{no bus of line } l_i \text{ in } (t_0, t_0+t)} = e^{-t/h}, \text{ where } \frac{1}{h} = \sum_{l_i} \frac{1}{h_{l_i}} \tag{3}$$

Since the waiting time T^{waiting} follows an exponential distribution with parameter $1/h$, the expected value is $\mathbb{E}(T^{\text{waiting}}) = h$. Following the prior example, with two lines l_1 and l_2 with headways $h_{l_1} = 5$ and $h_{l_2} = 10$ minutes, the average waiting time is: $\mathbb{E}(T^{\text{waiting}}) = \frac{1}{1/5 + 1/10} = 3.33$ minutes.

The average travel times between each pair of bus stops (considering waiting time and effective travel time) are used during the graph construction phase, which is described next.

3.3 Graph construction

The graph built to compute travel times has nodes which represent centroids (i.e., origins and destinations) and bus stops; and edges with weights that represent the travel time to connect the nodes, either walking or by bus. The proposed model for travel time estimation allows imposing restrictions on both the maximum walkable distance and the maximum number of bus transfers. To fulfill the first restriction, walking edges with a weight (i.e., a travel time) larger than a given threshold are not included in the graph. The model even allows imposing different thresholds for walks from the origin to the first bus stop, from the last bus stop to the destination, between bus stops during bus transfers, and between origins and destinations when walking directly. The procedure to impose restrictions on the number of bus transfers is described next.

The restriction on the number of allowed transfers is achieved by using duplicate nodes in the graph. Nodes corresponding to centroids are represented by two different nodes, which account for the centroid acting as the origin or the destination of the computed route. Thus, a centroid C_i is represented in a graph by two nodes: C_i^{orig} when the centroid acts as an origin in a route and C_i^{dest} when acting as a destination. Nodes C_i^{orig} have no incident edges (i.e., $deg^-(C_i^{orig}) = 0$) and their outward edges are always walking edges, since every trip starts with a walk (either to the bus stop or directly to the destination). Conversely, nodes C_i^{dest} have no outward edges (i.e., $deg^+(C_i^{dest}) = 0$) and only have incident edges of type walking, since all trips end with a walk (from the last bus stop in the route or directly from the origin when no buses are boarded).

Similarly, bus stops are represented by $2 \times B_{max}$ nodes, where B_{max} is the maximum number of bus trips allowed. This representation allows limiting the number of buses boarded in a given route and ensures that walking and bus edges are alternated when computing the shortest path in the graph. Each bus stop s_i is represented by the following set of nodes $\{s_i^{B'_{max}}, s_i^{B_{max}-1}, s_i^{B_{max}-1'}, \dots s_i^1, s_i^{1'}, s_i^0\}$. Nodes s_i^j represent the stop when arriving by bus and leaving by foot. Conversely, nodes $s_i^{j'}$ model the stop within the route when the stop is reached by foot and left using a bus trip. Thus, s_i^j nodes have incident edges of bus type and outward edges of walking type, whereas nodes $s_i^{j'}$ only have incident edges of walking type and outward edges of bus type.

Nodes s_i^j model stop s_i when it belongs to a partial path that still has j bus trips available, according to the maximum number of bus trips allowed. Nodes $s_i^{B'_{max}}$ may only have in-edges of type walking which connects them to nodes C_k^{orig} (representing the walk from the origin centroid to the first bus stop in the route) and out-edges of type bus which connects them to nodes of type $s_k^{B_{max}-1}$. Analogously, nodes s_i^0 only have in-edges of type bus coming from nodes $s_k^{1'}$ and out-edges of type walking which connects them to nodes C_k^{dest} (representing walks from the last bus stop in the route to the final destination). Other intermediate nodes, s_k^m with $m \in \{B_{max} - 1 \dots 1\}$, have in-edges of type bus and out-edges of type walking. These walking out-edges may be directed to other bus stops (accounting for walks done for a bus transfer), to a final destination (C_k^{dest}), or even to the same bus stop, using edges of null weight, to account for transfers within the same physical stop (i.e., with no walk involved).

Fig. 3.3 outlines an example with two centroids C_1 and C_2 and four bus stops s_1, s_2, s_3, s_4, where the maximum number of bus trips allowed is set to three ($B_{max} = 3$). The example graph has 28 nodes (two for each centroid and six for each stop). Walking edges are represented with a dashed line and bus edges with a solid line. The weight of each edge is indicated along the line. For walking edges, the weight accounts for the time needed to walk between the two nodes connected by the edge. For bus edges, the weight corresponds to the average waiting time for a bus at the starting node plus the average travel time by bus between the nodes connected by the edge. For the sake of clarity, edges which are not part of any path between both centroids are not displayed.

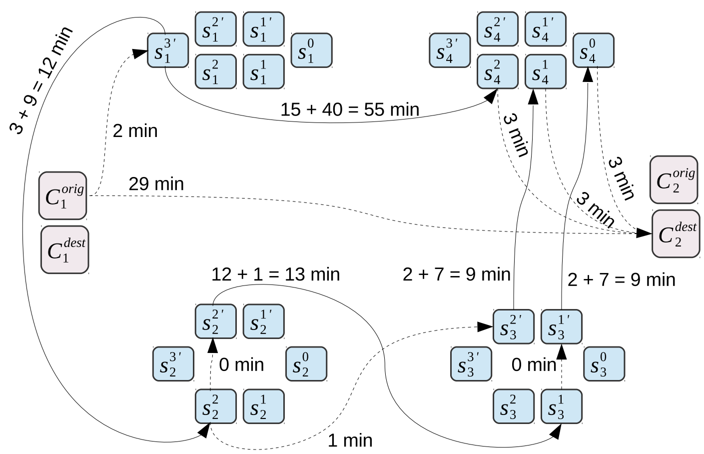

Fig. 1. Example graph and paths for two centroids and four bus stops

The travel time from centroid C_1 to C_2 is estimated by finding the shortest path (in terms of the weight of its edges) that connects the two centroids. Four paths exist which connect the two centroids in the example:

1. Walking directly from C_1 to C_2, without boarding a bus, with a total duration of 29 minutes.
2. Starting from C_1, walking to bus stop s_1, traveling by bus towards s_4, and finally walking to C_2. In total, this path involves one bus trip and has a total duration of 60 minutes.
3. Starting from C_1, walking to stop s_1 and taking a bus to stop s_2. Then, walking from s_2 to s_3 for a bus transfer, taking another bus from s_3 to s_4. Finally, walking from s_4 to C_2. Overall, this path involves two bus trips and a total duration of 27 minutes.
4. Starting from C_1, walking to stop s_1 and traveling by bus to stop s_2. At the same stop, take a new bus from s_2 to s_3. Then, making another transfer to take a new bus from s_3 to s_4. Finally, walking from s_4 to centroid C_2. Overall, this path involves three bus trips and a total duration of 39 minutes.

In this example, the shortest path is path number 3, with a total duration of 27 minutes. Thus, the estimated travel time from centroid C_1 to C_2 is set to 27 minutes.

4 Case study: public transportation system in Montevideo, Uruguay

This section presents the application of the proposed model to estimate travel times to the public transportation system of Montevideo, Uruguay.

4.1 Overview of the case study

Montevideo is one of the nineteen departments in Uruguay, where the capital city of the country is located. Situated in the southernmost part of the country, Montevideo extends to an area of only 530 km^2. In spite of accounting for only 0.3% of the total surface of Uruguay, Montevideo has an estimated population of 1 319 108, which represents nearly 40% of the total population of the country [7]. The surface of the city can be divided into 1063 zones named *census segments*, which are often used in surveys.

The public transportation system in Montevideo is comprised of 1528 buses operated by four private companies. The bus network consists of 145 bus lines and 4712 bus stops. However, each bus line usually has different variants, accounting for outward and return trips, as well as shorter versions of the same line. Thus, the total amount of bus lines when considering each variant individually ascends to 1383. The fare scheme of the system allows passengers to transfer between bus lines at any stop.

Fig. 2 shows a map of the city, its division into census segments, the centroids of each census segment and the bus lines of the public transportation system.

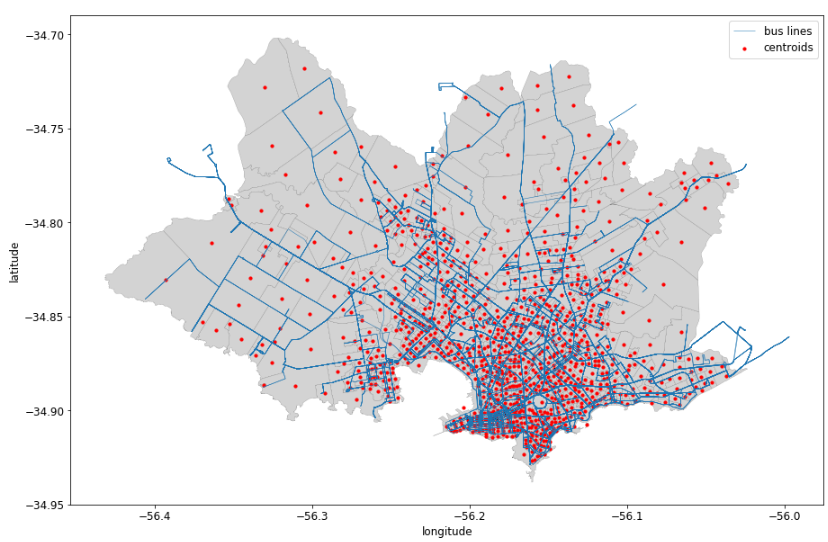

Fig. 2. Census segments, centroids and bus lines in Montevideo, Uruguay

4.2 Implementation details and parallel model

Several libraries were used to implement the proposed model for travel time estimation in public transportation systems.

The weights of walking edges were computed using Open Source Routing Machine (OSRM) [9]. OSRM is a high-performance routing engine implemented in C++, which combines sophisticated routing algorithms with roadmaps from OpenStreetMap (OSM) to efficiently compute routes within the road network. The *table* method was used, which allows computing pairwise travel times for a given list of geographic coordinates. The roadmap used for the network was provided by OSM[1]. OSRM is configured by *profiles*, which define average speeds related to the type of roads and impose driving rules (e.g., restrictions on turns). Weights of walking edges were computed using the *foot.lua* profile, which is provided by the library.

The graph was modeled using the NetworkX library [4]. NetworkX is a Python library which provides efficient data structures to handle large graphs, along with a rich collection of graph algorithms. Due to the characteristics described in Section 3, the graph was modeled using the *MultiDiGraph* class. To compute the travel times between centroids the *shortest_path* method, which implements the famous Dijkstra algorithm, was used over the generated graph.

The pairwise shortest path computation is, by far, the most computationally intensive task of the travel time estimation procedure. Thus, a parallel implementation was developed to distribute the computations over several processing units. The parallel implementation follows a simple master-slave model, corresponding to a Single Instruction Multiple Data parallel approach. The master process assigns each slave one row of the travel time matrix for computation. This corresponds to assigning one centroid as origin and making the slave responsible for computing the shortest path from that origin to all other destinations. The master node is also in charge of balancing the computational load, by assigning new rows of the matrix to a slave as soon as they become available. In the end, the master node gathers data from all slaves and persist the final computed matrix to disk. To implement the parallel model the *joblib* library was used, which provides routines to easily parallelize Python code.

4.3 Model considerations

To apply the proposed model for travel time estimation to the public transportation system of Montevideo, Uruguay, the following data was used as input. Firstly, census segments were the zones considered as origins and destinations. Therefore, the pairwise travel time computation results in a 1063×1063 travel time matrix. Bus network (including bus lines and stops) as well as bus schedules were obtained from the government Open Data catalog[2] as of 2019-01-09. Only buses departing between 7 a.m. and 9 a.m. were considered during the

[1]Downloaded from download.geofabrik.de on 2019-01-14

[2]catalogodatos.gub.uy

experimental analysis, to account for the morning peak hour. The implemented model can easily be adapted to study different times of the day.

Regarding the parameters of the model, the constraints for maximum allowed walks and maximum number of transfers on a route were set as follows. Walks between centroids and bus stops, and between centroids directly, were limited to 30 minutes. Similarly, walks between bus stops (related to bus transfers) were limited to 20 minutes. The maximum number of boardings (B_{max}) was set to three, thus, at most two bus transfers are allowed on routes. These constraints were set based on the findings of the mobility survey reported in Section 2.

4.4 Experimental analysis

This section reports the computation of travel time matrices for Montevideo, Uruguay, and their validation against other sources of data.

Computing infrastructure Shortest path calculations on the modeled graph were executed over the cloud infrastructure at *ClusterUY*, the national center of supercomputing in Uruguay [13]. Executions were performed using a server comprised of 40 Intel Xeon Gold 6138 (2.00GHz) cores and 128 GB of RAM.

Computed results The graph built for the case study of Montevideo, Uruguay, is comprised of 30 398 nodes (1063 centroids × 2 + 4712 stops × 6) and 3 252 334 edges. The computation of the travel time matrix demanded 48 hours and 33 seconds, using the HPC infrastructure previously described. Travel time matrices are freely available at www.fing.edu.uy/~renzom/tt_bus.

Comparison against other data sources The estimated travel time matrix is validated against the two alternative sources described in Section 2: the mobility survey and the Cómo Ir webapp. Although each data source has its own shortcomings, it would be expectable to find correlation among the travel times estimated with each different data source.

The validation procedure studied a random sample of 85 trips, corresponding to morning commutes using public transportation, according to the mobility survey. The trips were assumed to start and end at the centroid of the census segment, since the actual starting/ending point within the segment is unknown. Thus, for each trip in the sample, three travel time estimations were obtained: i) *model*, which corresponds to the estimated travel time between the centroids of the origin and destination census segment using the procedure described in Section 3; ii) *webapp*, which corresponds to the travel time returned from the Cómo Ir webapp between the closest address to the centroids of the census segment of origin and destination; iii) *survey*, which corresponds to the time declared by the passenger in the household mobility survey (in this case the origin and destination may be different from the centroids of the census segments). Three trips returned no route when using the Cómo Ir webapp, thus, the comparisons against that source uses 82 samples in total.

Firstly, the Pearson coefficients (i.e., linear correlation) between the different travel time estimations were computed. The Pearson coefficient between the proposed model and the Cómo Ir webapp is 0.90, and between the proposed model and the mobility survey is 0.83, both with a 99% of statistical significance. These results indicate a strong correlation that suggests that the variance on the estimated travel times when evaluating different origin-destination pairs is consistent with the other data sources used for validation. Moreover, this conclusion is reinforced by the fact that the lowest Pearson correlation is the one between the two sources of validation, with a value of 0.73.

Then, the absolute differences (in minutes) between the travel time estimation using the proposed model and the two sources of validation were computed. For this purpose, we subtract the estimated travel time (TT) using the proposed model to the travel times corresponding to the Cómo Ir webapp and the mobility survey. For each source of validation two results are presented: one that considers the direction in the difference and one which accounts for the absolute difference between the travel times without considering the sign. Overall, the later indicator is more relevant to determine the precision of the proposed model. Results are presented in Table 1.

Table 1. Travel time differences against the alternative data sources.

	$\delta = TT_{\text{webapp}} - TT_{\text{model}}$		$\delta = TT_{\text{survey}} - TT_{\text{model}}$	
	δ	$\lvert\delta\rvert$	δ	$\lvert\delta\rvert$
mean	−1.61	7.74	0.60	11.34
percentile 10	−13.63	0.80	−21.95	1.09
percentile 20	−8.16	2.19	−9.84	2.62
percentile 30	−6.58	3.56	−6.10	3.89
percentile 40	−5.70	5.07	−1.55	6.28
percentile 50	−3.52	6.10	0.66	7.71
percentile 60	−1.08	7.46	3.34	10.81
percentile 70	0.67	8.60	6.68	13.98
percentile 80	3.83	12.13	12.07	19.85
percentile 90	10.25	16.45	19.62	25.70

Results show that the proposed model computes similar travel times to those computed by the Cómo Ir webapp. On average, the travel time difference between the proposed model and the Cómo Ir webapp is less than eight minutes. For half of the trips the difference in travel times was nearly six minutes. When comparing against the results of the mobility survey, differences are slightly larger, with an average absolute difference of over eleven minutes, a median difference below eight minutes, and 20% of cases with significant differences of over 19 minutes.

The validation procedure shows a strong consistency between the travel time estimations of the proposed model and the two alternative sources of data available. The travel time differences are low when considering the length of the trips involved due to commuting. It is worth noting that the sources used for comparison are not the ground truth for the travel times and, consequently, it is not possible to identify the source of error that produces the slight differences in travel times. For instance, the Cómo Ir webapp has certain limitations (described in Section 2) which prevented from computing travel times for three of the trips in the studied sample. Similarly, the results from the mobility survey correspond to the travel time estimated by the users, which might include a significant bias due to the perception of time of each surveyed passenger. Overall, the high consistency between the proposed model and the validation sources—indicated by the strong Pearson Correlation—as well as the low absolute difference in minutes between the travel times in the studied sample, suggest that the proposed model is a valid alternative to compute travel time matrices.

5 Travel time web application

Raw travel time data estimated with the proposed model is useful for city administrators and researchers working on transport accessibility. However, it would be desirable to offer a global picture of the computed results in order to communicate the main findings to the general public. For this purpose, an interactive web application was developed to show the computed travel time matrix in an intuitive and friendly manner. The travel time visualization tool allows users to select an area in the map and creates a heatmap indicating the number of minutes needed to travel from the selected area to all other areas in the map using public transportation. The tool was developed in Python using Pandas for data processing, Geopandas to display the map of the city, and Bokeh to provide interactivity to the visualization. The web application is freely available at www.fing.edu.uy/~renzom/tt_bus. Figure 3 shows the user interface of the developed tool and its main features are described next.

Users can select an area (i.e., a census segment) by clicking on the map. Then, the selected area is shown in a different color for the user to confirm the selection. Once confirmed, the application updates the color of all the areas in the map according to the amount of minutes necessary to travel from the selected area. A color bar is presented on the right to quantify the information visually displayed. Additionally, the application offers a hover tool, which displays information when the mouse cursor is over a given area. The displayed information includes the area identifier (id of the census segment) as well as the exact number of minutes to reach that destination. Finally, at every step of the visualization the user is able to export the displayed map as an image using the save button in the bottom right panel of the map.

6 Conclusions and future work

In this paper we presented a model to estimate travel times in public transportation between different zones of a city. The proposed model uses data regarding the roadway infrastructure, the public transportation lines, stops, and schedules;

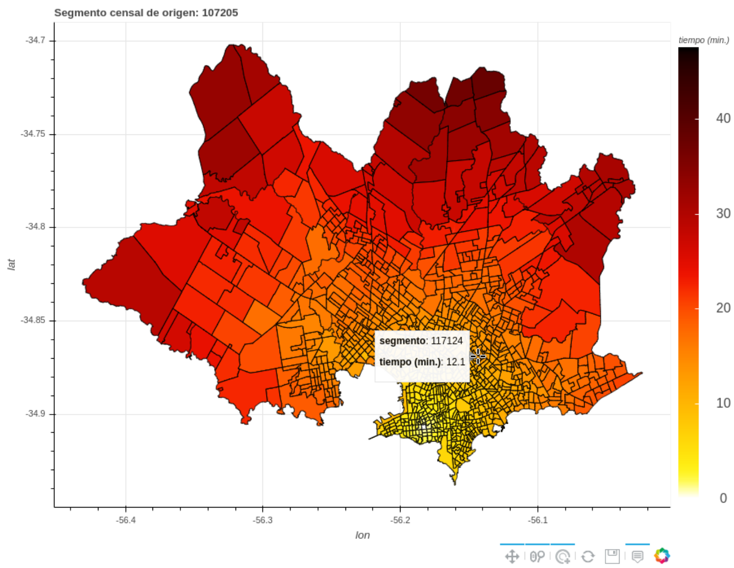

Fig. 3. Web application to display estimated travel times

and accounts for restrictions on the maximum walkable distance and the amount of transfers. To estimate travel times, a shortest-path algorithm is run in a parallel fashion over a graph that models the different travel options between all zones in the city. The proposed model was applied to the public transportation system in Montevideo, Uruguay, and validated against the results from a government web application and the findings of a household mobility survey. Results showed that the proposed approach has a strong correlation with the other data sources, thus proving the validity of the model. Finally, an interactive web application was presented, which allows displaying travel times in an intuitive way.

The main lines of future work are related to further refine the travel time estimations. For this purpose, it would be interesting to apply the model using GPS timestamps instead of relying on fixed bus schedules. This would account for traffic jams and peak hours, which fixed bus schedules may not model correctly. Then, the estimated travel matrix should be applied to assess the public transportation system, particularly in regards to its accessibility in different zones in the city. This would allow identifying potential inequalities in the potential mobility of passengers living in different area of the city.

Acknowledgments

The authors would like to thank ANII (Uruguay) for the support provided under grant FSDA_1_2017_1_143161 (Project: "El transporte público urbano y la accesibilidad a las oportunidades laborales"). The work of R. Massobrio was partly funded by ANII and PEDECIBA (Uruguay) and Fundación Carolina (Spain).

References

1. Cardozo, O.D., Rey, C.E.: La vulnerabilidad en la movilidad urbana: aportes teóricos y metodológicos. In: Foschiatti, A. (ed.) Aportes conceptuales y empíricos de la vulnerabilidad global, pp. 398–423. Editorial Universitaria de la Universidad Nacional del Nordeste (2007)
2. Denis, J., Massobrio, R., Nesmachnow, S., Cristóbal, A., Tchernykh, A., Meneses, E.: Parallel computing for processing data from Intelligent Transportation Systems. In: 10^{th} Int. Supercomputing Conf. in Mexico. pp. 1–15 (2019)
3. Grava, S.: Urban Transportation Systems. McGraw-Hill Professional (2000)
4. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) Proc. of the 7th Python in Science Conf. pp. 11 – 15. Pasadena, CA USA (2008)
5. Hansz, M.: Analysis of the spatial disparity in transport social needs and public transport provision in Montevideo. Master's thesis, University of Leeds (2016)
6. Hernandez, D.: Uneven mobilities, uneven opportunities: Social distribution of public transport accessibility to jobs and education in montevideo. Journal of Transport Geography 67, 119 – 125 (2018)
7. Instituto Nacional de Estadística, Uruguay: Resultados del censo de población 2011: población, crecimiento y estructura por sexo y edad (2012), retrieved from: http://www.ine.gub.uy/c/document_library/get_file?uuid=12d80f63-afe4-4b2c-bf5b-bff6666c0c80&groupId=10181 (Last accessed: 2018-08-30)
8. Lei, T.L., Church, R.L.: Mapping transit-based access: integrating GIS, routes and schedules. Int. Journal of Geographical Information Science 24(2), 283–304 (2010)
9. Luxen, D., Vetter, C.: Real-time routing with openstreetmap data. In: Proc. of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 513–516. GIS '11, ACM, New York, NY, USA (2011)
10. Massobrio, R.: Urban mobility data analysis in Montevideo, Uruguay. M.Sc. thesis, Universidad de la República (2018)
11. Massobrio, R., Pías, A., Vázquez, N., Nesmachnow, S.: Map-Reduce for Processing GPS Data from Public Transport in Montevideo, Uruguay. In: Simposio Argentino de Grandes Datos, 45 Jornadas Argentinas de Informática. pp. 41–54 (2016)
12. Mauttone, A., Hernández, D.: Encuesta de movilidad del área metropolitana de Montevideo. Principales resultados e indicadores. Tech. rep., CAF, Intendencia de Montevideo, Intendencia de Canelones, Intendencia de San José, Ministerio de Transporte y Obras Públicas, Universidad de la República, PNUD Uruguay (2017), retrieved from: http://scioteca.caf.com/handle/123456789/1078 (Last accessed: 2018-08-30)
13. Nesmachnow, S.: Computación científica de alto desempeño en la Facultad de Ingeniería, Universidad de la República. Revista de la Asociación de Ingenieros del Uruguay 61(1), 12–15 (2010)
14. Salonen, M., Toivonen, T.: Modelling travel time in urban networks: comparable measures for private car and public transport. Journal of Transport Geography 31, 143 – 153 (2013)