Graph Analytics and Optimization Methods for Insights from the Uber Movement Data

Arun V Sathanur
arun.sathanur@pnnl.gov
Pacific Northwest National Laboratory
Seattle, Washington

Vinay Amatya
vinay.amatya@pnnl.gov
Pacific Northwest National Laboratory
Richland, Washington

Arif Khan
arif.khan@pnnl.gov
Pacific Northwest National Laboratory
Richland, Washington

Robert Rallo
robert.rallo@pnnl.gov
Pacific Northwest National Laboratory
Richland, Washington

Kelsey Maass
kelsey.maass@pnnl.gov
Pacific Northwest National Laboratory
Richland, Washington

ABSTRACT

In this work, we leverage the Uber movement dataset for the Los Angeles (LA) area where partial TAZ to TAZ (Traffic Analysis Zone) trip time data is available, to predict travel time patterns on the full TAZ to TAZ network. We first create a TAZ-TAZ network based on nearest neighbors and propose a model that allows us to complete the $(O-D)$ (Origin-Destination) travel time matrix, using optimization methods such as non-negative least squares. We apply these algorithms to several communities in the TAZ-TAZ network and present insights in the form of completed $(O-D)$ matrices and associated temporal trends. We qualify the error performance and scalability of our flows. We conclude by pointing out the directions in our ongoing work to improve the quality and scale of travel time estimation.

CCS CONCEPTS

• **Applied computing → Engineering**; **Operations research**; **Transportation**.

KEYWORDS

transportation datasets, arterial travel time, optimization, graph analytics, traffic analysis zone

ACM Reference Format:

Arun V Sathanur, Vinay Amatya, Arif Khan, Robert Rallo, and Kelsey Maass. 2019. Graph Analytics and Optimization Methods for Insights from the Uber Movement Data. In *The 2nd ACM/EIGSCC Symposium On Smart Cities and Communities (SCC '19), September 10–12, 2019, Portland, OR, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3357492.3358625

1 INTRODUCTION

Data-driven mobility modeling and prediction are important aspects of modern urban planning. Instead of making predictions at the street-level, metropolitan areas are divided into a number of small geographical units called Traffic Analysis Zones (TAZ) and modeling and prediction tasks are done at the TAZ-level. Each of the TAZs are characterized by factors such as the total population, type of population, employment etc. Specifically, with respect to travel forecasting, the two major areas of research are demand modeling and travel time estimation [2]. Demand modeling involves generation of accurate statistics of number of trips from origin TAZs to various destination TAZs in the form of an $(O-D)$ demand matrix. Similarly travel time estimation also involves prediction of travel times for trips given a pair of origin and destination TAZs, also leading to an $(O-D)$ travel time matrix.

When we consider existing research on travel time estimation, modeling of interstate links has received disproportionate attention in the transportation research community primarily due to the available of large amounts of data from freeway sensors. While being equally important, the same is not true when it comes to arterial modeling where the coverage is limited due to cost issues with installation of large number of probe sensors and associated infrastructure. Under these circumstances, significant insights can be gained with datasets such as the Uber Movement data at a fraction of the cost. Uber datasets [23] provide anonymized, aggregated, and coarse-grained origin-destination $(O-D)$ travel times at the TAZ level for a multitude of metropolitan areas around the world. While these datasets can be coarse-grained, they do allow for coverage over large areas and are available for multiple metropolitan areas allowing for generalizability.

Our work is concerned with filling this gap in the arterial travel time estimation. We consider a hierarchical approach to arterial travel time estimation where in we first analyze the incomplete Uber data at the TAZ level and in the second phase, we consider the major road segments that connect the TAZs under consideration. The work presented in this paper describes our research on the first stage, namely the analytics at the TAZ level. For this purpose, we first create a TAZ-TAZ graph and leverage various graph analytical and optimization flows to find the travel times along the edges of this abstract TAZ-TAZ graph. This allows us to complete the travel time $(O-D)$ matrix at the TAZ level and examine the same for different times of the day and for different densely connected clusters of TAZs with varying data sparsity. We then examine the

SCC '19, September 10–12, 2019, Portland, OR, USA

ACM ISBN 978-1-4503-6978-7/19/09…$15.00
https://doi.org/10.1145/3357492.3358625

temporal correlations of travel times along contiguous TAZ-TAZ paths to serve as a high-level validation of the travel time estimates. We also examine the accuracy and scaling issues with our approach,

The paper is organized as follows. Section 2 describes the related work in the area of graph analytics and machine learning / optimization as related to road networks. Section 3 is related to the network preparation task under which we describe the Uber Movement data that is used in this work (Section 3.1), the construction of the LA area road network and the TAZ-TAZ network (Section 3.2), the comparison in basic statistics between the two networks (Section 3.3), and the communities in the TAZ-TAZ network (Section 3.4). In the subsequent section (Section 4), we describe the forward model (Section 4.1) and the optimization methodology (Section 4.2) utilized in this work. Section 5 presents our principal experimental results. We first present the results related to convergence and scaling in Section 5.1, followed by the results related to the travel time ($O - D$) matrix completion and temporal correlations in Section 5.2. We conclude the paper with a note on future work in Section 6.

2 RELATED WORK

There is a rich history of using graph analytical measures and machine-learning models in urban traffic modeling. These fall into two major categories. The first set of approaches use graph representations of road networks, employ different weighting schemes and use traditional graph analytical measures. The second set of approaches involve a graph-based fabric upon which traffic dynamics is imposed and variety of machine learning models are proposed and optimization methods are employed to estimate the model parameters. The detailed survey presented in [24] provides an overview on a number of these methods.

In a recent work, the authors in [22] use graph analytics on networks with three different weighting schemes to perform statistical characterization of the Beijing road network. Several research works consider betweenness centrality to be an important metric when applied to road networks since it is argued to be a direct predictor of important links in urban transport. Given a graph, betweenness centrality [10] of an edge is the fraction of how many shortest paths pass through that edge with respect to the total number of shortest paths. It has been shown that betweenness centrality correlates highly with the traffic flow count on the road network [9, 14, 19]. In real world, not all paths in a road network are equally travelled and travel pattern changes during different times of the day based on work and residential areas. The authors in [20] made use of these observations and defined an augmented betweenness centrality where shortest paths are weighted according to the traffic demand model based on census / traffic analysis zones (TAZ). The authors showed that the augmented betweenness centrality correlates better with the traffic flows than the other centrality measures. The authors in [8] employ graph analytics in the form of novel centrality measures to derive insights into the traffic flow patters in Singapore. The work presented in [17] use graph-based models to understand the urban road traffic patterns while utilizing heterogeneous datasets. In another interesting approach, the authors in [11] utilize a grid-based architecture and cellular automata for modeling arterial traffic while being computationally efficient.

The authors in [15] employ a deep learning approach, specifically a diffusion convolutional recurrent neural network to predict the freeway traffic counts in the short term on LA and the Bay area data. The work presented in [7] also uses a deep learning approach in the form of a combination of convolutional neural networks and recurrent neural networks with long short term memory units. This architecture is utilized for short-term traffic count forecasting at 349 locations in the Beijing road network. In a recent report [21], and publication [26], the authors address the issue of arterial travel times from probe data that uses Bluetooth and GPS sensors as well as propose novel methods for validation. The authors in [12], use a Coupled Hidden Markov Model (CHMM) to model the evolution of the traffic states and leverages the Expectation Maximization (EM) algorithm for model parameter estimation. They apply these methods to a sparse taxi-fleet data for the San Francisco Bay area road network. The authors in another publication [13], employ a dynamic bayesian network framework to learn the arterial dynamics using the same taxi-fleet dataset. Our approach is broadly based on the model and optimization approach developed in a recent work [3]. In this work the authors use shortest-path routing in a road network and a convex-relaxation based formulation to solve for the travel time.

3 NETWORK PREPARATION

In this section we discuss the details of the Uber movement data for the LA city area, LA metropolitan area road network and the equivalent TAZ-TAZ network and present some analytics on the LA city area road network and the TAZ-TAZ network derived from the Uber movement data for the LA city area.

3.1 The Uber Movement Data

Uber has released a trove of aggregated and anonymized data on trip time and average speed statistics for a large number of cities around the world where Uber operates [23]. Since gathering transportation data is an expensive and cumbersome process, leveraging the Uber data is expected to provide researchers and city planners conduct quick but fairly detailed analyses of the various aspects of vehicle mobility in urban settings. In this work, we mainly focus on the trip times data published by Uber. This data is available for a number of metropolitan areas and it provides statistics for trip times between two TAZs or census tracts along with the hour of the day and day of the week data. Further. in terms of the statistic used, our focus for this work is the arithmetic mean [3]. The formulation can handle geometric mean with minimal changes.

3.2 The LA city area road network and the TAZ-TAZ network

Before we delve into the coarse TAZ-TAZ network which is the main focus of this work, we provide some details on the graph representation of the full physical road network for the LA city area. The source for this network data is Open Street Maps [18], accessed via the python-based OSMNX package [6].

Next we construct the TAZ-TAZ network for the LA city area. Each TAZ is defined by set of polygons and each polygon is represented by a list of points in the lat-long coordinate system. The entire data is available as part of the Uber movement data in the

GeoJSON format. We then compute the centroids of these TAZs and then compute the all-pairs euclidean distance. For each TAZ, we then choose the nearest k TAZ centers to connect to, in order to prepare the network. There are other methods to construct the network. We could look for common polygon edges to connect the TAZs or use a radius value around a TAZ center to connect all TAZs within that radius. The first method is computationally expensive and can have some issues with very small errors in the co-ordinate values which can cause the algorithm to not identify many edges. The second method can sometimes not add any edges to large TAZs for a small radius value or too many edges to small TAZs for a larger radius value, Hence our method of utilizing k nearest neighbors addresses both the issues. We use $k = 5$ in this work.

3.3 Comparison of the LA road network and the TAZ-TAZ network

Before we move to the travel time estimation tasks, we briefly compare the basic network statistics of the of the two different graphs. Table 1 lists the comparison in terms of the basic network stats and finally Figure 1 shows the degree distributions in the form of the complementary cumulative distribution functions (CCDF). It's clear from the comparison that while the degree-related properties show similarities, the clustering coefficient shows a marked difference, The TAZ-TAZ network which is at a higher-level of abstraction shows a very good community structure as we see in Section 3.4 and this is responsible for the high clustering coefficient. The clustering coefficient of the road network is within the bounds of the clustering coefficient values derived from a large-scale network analysis of 27,000 US urban road networks [5] with more than 19,000 US cities and towns.

Network	Road network	TAZ network
Nodes	50296	2205
Links	138038	6464
Average degree	5.5	5.9
Min. Degree	2	5
Max Degree	12	10
Clust. Coeff.	0.036	0.42

Table 1: Comparison of the basic graph statistics between the LA road network and the constructed coarse TAZ-TAZ network.

3.4 Communities in the TAZ-TAZ network

Next we employ community detection on the LA area TAZ-TAZ network to identify closely connected TAZ groups. While this does reveal patterns in the TAZ network, our main motivation is to use community detection and the resultant clusters to expedite the solution to the optimization problem. This becomes apparent during the discussions on scalability presented in Section 5.1. Communities or clusters in a graph are groups of nodes such that they are densely connected within themselves with sparse interconnections between

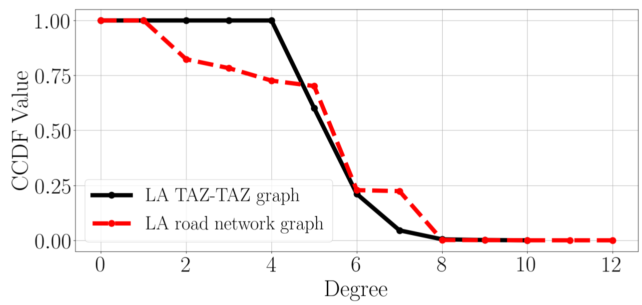


Figure 1: Comparing the degree distributions of the LA road network and the constructed TAZ-TAZ network.

them. We use the popular Louvain algorithm [4] as part of the Gephi graph visualization package [1] to find the communities in the LA area TAZ-TAZ network. Figure 2 shows a visualization of these communities along with the two communities selected for detailed analyses in both network form and spatial overlay. The first of the selected communities (C_1) has 91 TAZ nodes and 253 TAZ-TAZ edges whereas the second community (C_2) had 96 nodes and 516 edges. Not only did the two communities differ from the perspective of the average degree of the nodes (or the graph density metric), the amount of data available for these two communities differed by a lot, allowing us compare the performance of the optimization algorithm for an average case and a more extreme case. C_2 is nearly twice as dense as C_1 and has 3X more data available.

4 THE GRAPH-BASED MODEL AND THE OPTIMIZATION METHODOLOGY

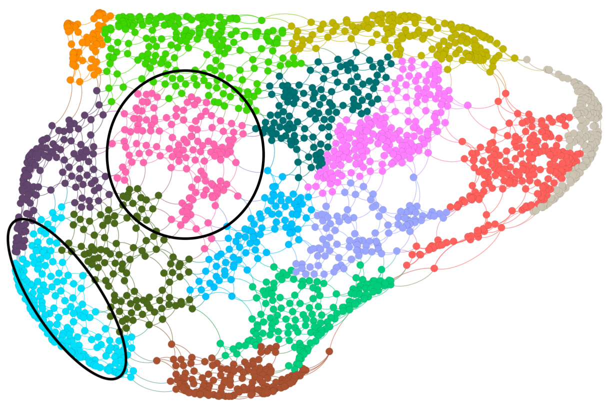
In this section we first propose a forward model for the travel time prediction and then show how the parameters of the model are estimated by means of an optimizer.

4.1 The Forward Model

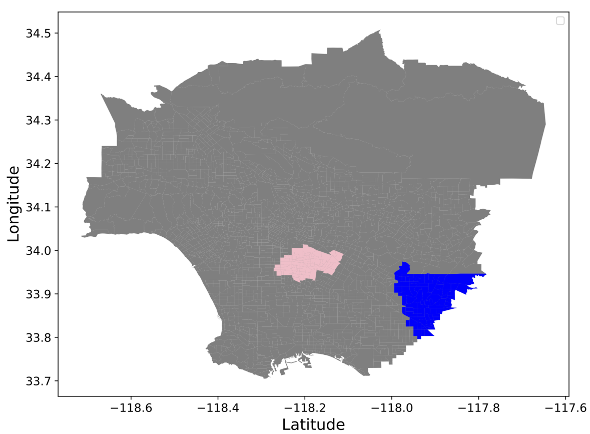
Our forward model is based on a simple weighted shortest-path routing [3, 16, 25] with modifications to account for the fact that we are working with the TAZ network. We first convert the undirected TAZ-TAZ network to a directed TAZ-TAZ network to account for the asymmetry in the travel times between two TAZs with respect to the direction. We also weight these edges initially with the Euclidean distance between the corresponding TAZ centers. We then route the path between a given pair of TAZs to be the weighted shortest path between them through the TAZ-TAZ graph. We also introduce slack variables, one per TAZ that models the average amount of time spent within each TAZ. Thus we express the average total transit time (T_{tr}) between two TAZs i and j as follows.

$$T_{tr}(i, j) = T^w(i) + \sum_{(k,l) \in P} T_{tr}(k, l) + T^w(j) \quad (1)$$

In Equation 1, $T^w(i)$ and $T^w(j)$ denote the average time spent within the TAZs i and j whereas $T_{tr}(k, l)$ denotes the average travel time across an edge in the TAZ-TAZ network and P denotes the set of all edges that form the shortest path between the TAZs i and j.



(a)



(b)

Figure 2: The top figure shows the network visualization of the TAZ-TAZ graph along with the two communities selected for further analysis while the bottom figure shows the spatial TAZ overlay with the two communities from the top figure, shaded.

Thus our model has variables that denote the travel time within each of the n TAZs and m travel times for each of the edges in the TAZ-TAZ network, giving us a total of $(m + n)$ model parameters that are non-negative that need to be estimated.

4.2 The Optimization Process

Assuming that for a specific time of the day, we have l pairs of TAZ-TAZ travel time (out of a possible $n(n-1)$ pairs) from the Uber Movement data, and assuming that the origin and destination TAZs are not the same, for each $(O - D)$ pair, upon computation of the (weighted) shortest path from O to D, we can write one equation that is similar to Equation 1. Thus we have a system of l equations with $(m + n)$ unknowns. Typically $l >> (m + n)$, giving us an overdetermined system. The solution is therefore obtained by minimizing the mean-squared error under the constraint that the unknown coefficients which indeed are the travel time variables are non-zero. This is illustrated by Equation 2.

$$\boldsymbol{T_{tr}} = \text{argmin}_{\boldsymbol{t}} \; ||\boldsymbol{S.t} - \boldsymbol{t_{tt}}||_2 \; ; \boldsymbol{t} \geq \boldsymbol{0} \tag{2}$$

T_{tr} denotes the vector of the various travel times as depicted in Equation 1, S denotes the matrix of shortest paths and the origin and destination nodes as required by Equation 1 and finally t_{tt} denotes the TAZ-TAZ travel times for each of the 'l' $(O - D)$ pairs. Given the non-negativity constraint, this least-squares problem is known as "Non-negative Least Squares" optimization. We employ its "nnls" implementation offered by the well-known Python scientific library "scipy". We employ an iterative scheme to refine the solution process in a manner similar to the algorithm suggested in [3]. Thus the initial iteration starts with the euclidean distance as the weights but then the subsequent iterations are based on a TAZ-TAZ graph weighted by the travel times estimated by the previous iteration which will be utilized for the shortest path calculations. For this TAZ-TAZ network, we don't employ any specific regularization method. We run the optimizer for 30 iterations or until convergence (as measured by the magnitude of the change in the solution vector between iterations) whichever happens earlier.

The full set of steps in the model building/estimation process is outlined below.

- Split the available data into training and testing data. We used a 75%-25% split.
- For each training instance, find the shortest path between origin and destination based on distance / estimated travel time. Set coefficients accordingly to produce one equation similar to Equation 1 for one training instance
- Prepare the (over-determined) system of equations for the full training set
- Solve for the travel times via a suitable optimization procedure, by minimizing the mean squared error (example: Non negative least squares, quadratic programming)
- Repeat steps (A) to (C) with travel times in step (C) as weights, until convergence
- Obtain model performance by predictions on the test data and calculating suitable error metrics

5 EXPERIMENTAL RESULTS

We used the Uber movement data for the Los Angeles area over the time period defined by the third quarter of 2016 as the data of choice. We also selected two different communities from the TAZ-TAZ graph as described in Section 3 and focused our analysis on them.

5.1 Convergence and Scaling

We first present the result related to convergence of the optimizer. Figure 3 shows how the magnitude of the difference vector (difference between the solution vector for a given iteration and it's immediately preceding iteration) changes as iterations proceed for the community C_1. As seen from the figure, the optimizer shows a clear convergence behavior.

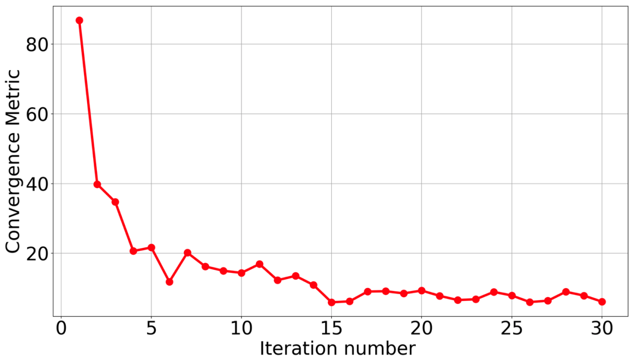


Figure 3: The convergence of the optimization process that is part of the travel time estimation flow.

Next, we conduct an analysis of the scalability of the method in order to inform the needs related to high performance computing (HPC) to solve for problems at the scale of cities and metropolitan areas. The overall execution time depends on the number of TAZ nodes, the number of TAZ-TAZ edges and the total number of data points in the training set (rows in the Uber data). As we incorporate larger portions of the overall network, we encounter an increase in all three of the above mentioned variables. As a result we observe a sharp increase in the run times for larger datasets. The scaling behavior of the flow is illustrated in Figure 4. It is very clear from the figure that when considering the full LA city network with 50k+ nodes and 130k+ edges, when making estimates at the physical road-network level, HPC based implementations and related optimizations become very important.

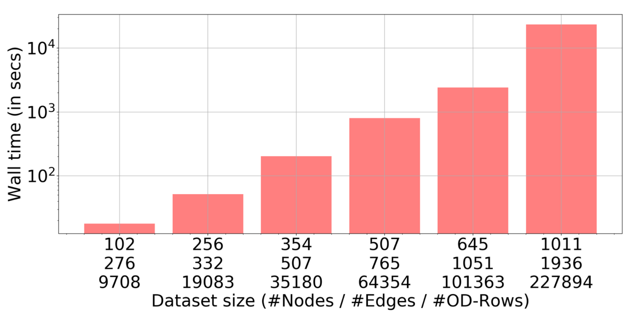


Figure 4: Scaling behavior of the modeling and optimization flow

5.2 Matrix Completion and Correlations

We now focus our attention on the $O - D$ travel time matrix completion. As explained in the previous section, once the travel times for the TAZ nodes and the TAZ-TAZ edges are obtained from the optimizer, using the training data, we then compute the travel times for the given TAZ-TAZ $O - D$ pairs in the test data using the forward model established in Equation 1. We then compare this with the ground-truth values and deduce an average error measure over the test dataset. Based on this reconstruction, we can complete the travel time $(O - D)$ matrices by estimating the missing entries from the forward model with the estimated parameters. The reconstruction process is visualized for two different communities and for two different time periods in Figure 6 We note that the model produces lower error (20%-25%) for the case with higher density of data while the sparse case (lower density) produces worse errors (40%-45%) even though the larger data case also meant that the number of unknowns was twice as large.

Finally, we consider a slightly different form of model validation. We look at $(O - D)$ pairs with a path length of two edges between them in the TAZ-TAZ network such that the travel times are present in the data for these pairs but the travel times for the two edges that constitute the path between these pairs of TAZs are not present, but rather estimated from our model and optimization procedure. We then do this estimation for several hours of the day (a separate model is run for each hour of the day) and then we plot the time variation of these travel times. We expect good correlation between these temporal curves given the geographical proximity. We found several examples with varying degrees of correlations and we show two such examples in Figure 5 below where the correlation was appreciable.

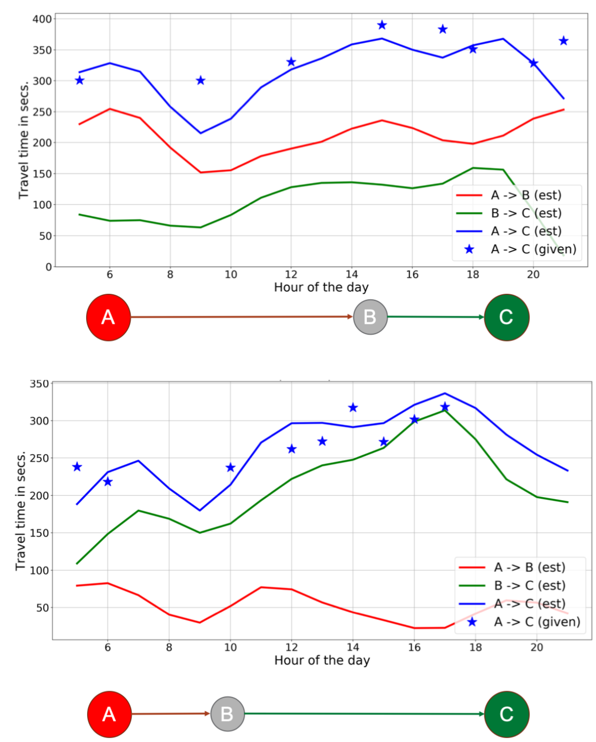
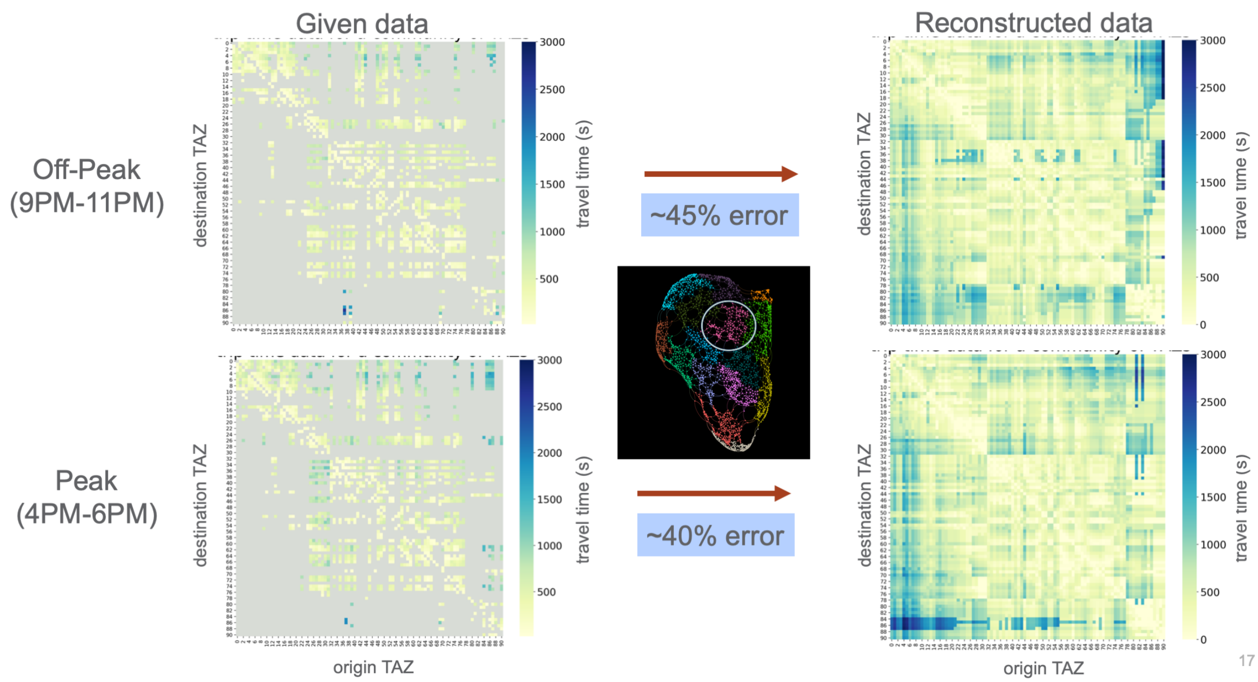
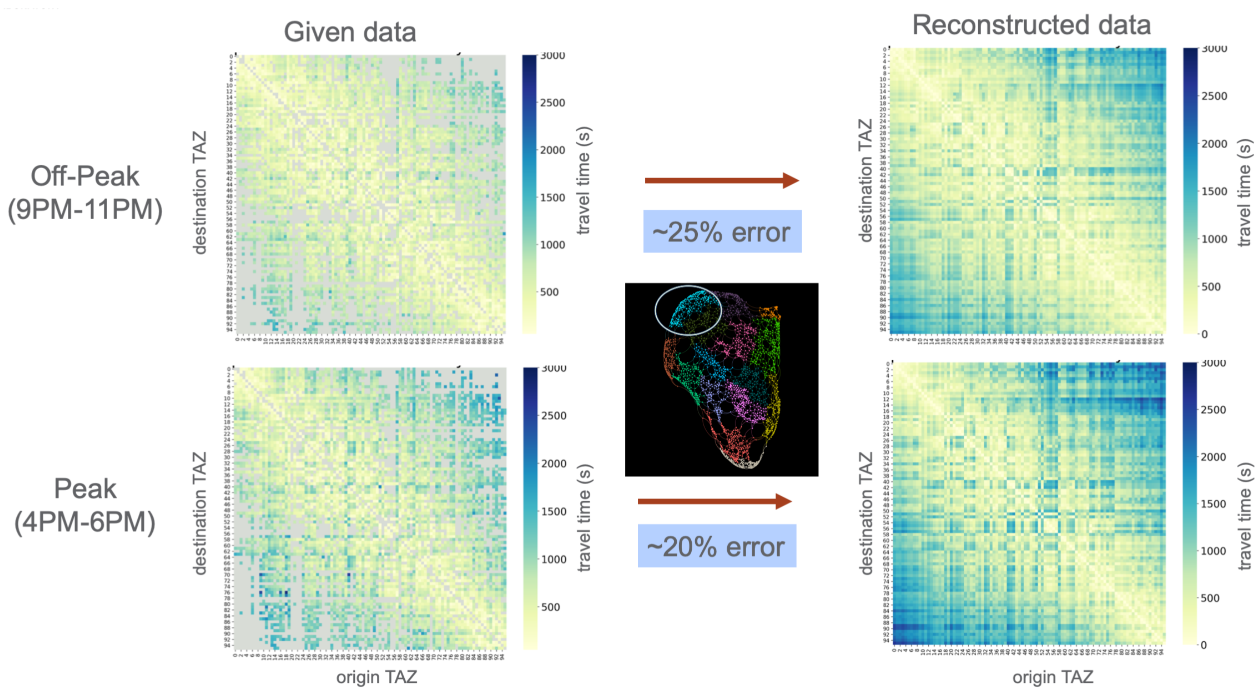


Figure 5: Temporal correlation between given travel times for TAZ pairs with path length of 2 and estimated travel times along the constituent edges.



(a) Travel time heat maps at two different times of the day for the community C_1



(b) Travel time heat maps at two different times of the day for the community C_2

Figure 6: Travel time heat maps for both the given data (left in all 4 cases) and the reconstructed data (right in all 4 cases) for two different communities with different densities (or sparsities) of data, for two different times of the day (off-peak and peak) are shown.

6 CONCLUSIONS AND FUTURE WORK

In this work we explored the Uber movement data and leveraged it as a cost-effective surrogate for arterial travel time estimation. Specifically, in this work we restricted ourselves to a coarse-level TAZ-TAZ network. We described how the network was constructed and provided summary network statistics on the same. We then presented a forward model for the travel time estimation between any origin-destination TAZ pair. The same model was used in conjunction with an optimizer on the Uber TAZ-TAZ trip time dataset to estimate the unknowns and eventually enable the completion of the $(O - D)$ travel time matrix. We presented results related to convergence, scaling and different forms of validation. Our future work is focused on extending the results to the physical road network to estimate travel times along the road segments, reduce the error via a combination of better models and optimizers and use HPC resources to scale the flow to city-scale networks.

REFERENCES

[1] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks.

[2] Edward Beimborn et al. 2006. A transportation modeling primer. (2006).

[3] Dimitris Bertsimas, Arthur Delarue, Patrick Jaillet, and Sébastien Martin. 2019. Travel Time Estimation in the Age of Big Data. *Operations Research* 67, 2 (2019), 498–515.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[5] Geoff Boeing. [n. d.]. A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science* ([n. d.]).

[6] Geoff Boeing. 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65 (2017), 126–139.

[7] Xingyi Cheng, Ruiqing Zhang, Jie Zhou, and Wei Xu. 2018. Deeptransport: Learning spatial-temporal dependency for traffic condition forecasting. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[8] Yew-Yih Cheng, Roy Ka-Wei Lee, Ee-Peng Lim, and Feida Zhu. 2015. Measuring centralities for transportation networks beyond structures. In *Applications of social media and social network analysis*. Springer, 23–39.

[9] Paolo Crucitti, Vito Latora, and Sergio Porta. 2006. Centrality in networks of urban streets. *Chaos: an interdisciplinary journal of nonlinear science* 16, 1 (2006), 015113.

[10] Linton C Freeman. 1977. A set of measures of centrality based on betweenness. *Sociometry* (1977), 35–41.

[11] PJ Gundaliya, Tom V Mathew, and Sunder Lall Dhingra. 2008. Heterogeneous traffic flow modelling for an arterial using grid based approach. *Journal of Advanced Transportation* 42, 4 (2008), 467–491.

[12] Ryan Herring, Aude Hofleitner, Pieter Abbeel, and Alexandre Bayen. 2010. Estimating arterial traffic conditions using sparse probe data. In *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 929–936.

[13] Aude Hofleitner, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. 2012. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. *IEEE Transactions on Intelligent Transportation Systems* 13, 4 (2012), 1679–1693.

[14] Bin Jiang. 2009. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science* 23, 8 (2009), 1033–1048.

[15] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJiHXGWAZ

[16] Eric Hsueh-Chan Lu, Chia-Ching Lin, and Vincent S Tseng. 2008. Mining the shortest path within a travel time constraint in road network environments. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 593–598.

[17] Kamaldeep Singh Oberoi, Géraldine Del Mondo, Yohan Dupuis, and Pascal Vasseur. 2017. Spatial Modeling of Urban Road Traffic Using Graph Theory. In *Spatial Analysis and GEOmatics 2017*.

[18] OpenStreetMap contributors. 2019. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[19] Sergio Porta, Paolo Crucitti, and Vito Latora. 2006. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design* 33, 5 (2006), 705–725.

[20] Rami Puzis, Yaniv Altshuler, Yuval Elovici, Shlomo Bekhor, Yoram Shiftan, and Alex Pentland. 2013. Augmented betweenness centrality for environmentally aware traffic monitoring in transportation networks. *Journal of Intelligent Transportation Systems* 17, 1 (2013), 91–105.

[21] Hesham A Rakha, Hao Chen, Ali Haghani, Xuechi Zhang, Masoud Hamedi, et al. 2015. *Use of Probe Data for Arterial Roadway Travel Time Estimation and Freeway Medium-term Travel Time Prediction*. Technical Report. Mid-Atlantic Universities Transportation Center.

[22] Zhao Tian, Limin Jia, Honghui Dong, Fei Su, and Zundong Zhang. 2016. Analysis of urban road traffic network based on complex network. *Procedia engineering* 137 (2016), 537–546.

[23] Uber Technologies, Inc. 2019. Data retrieved from Uber Movement, (c) 2019 . https://movement.uber.com.

[24] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. 2014. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43 (2014), 3–19.

[25] Lingkun Wu, Xiaokui Xiao, Dingxiong Deng, Gao Cong, Andy Diwen Zhu, and Shuigeng Zhou. 2012. Shortest path and distance queries on road networks: An experimental evaluation. *Proceedings of the VLDB Endowment* 5, 5 (2012), 406–417.

[26] Xuechi Zhang, Masoud Hamedi, and Ali Haghani. 2015. Arterial travel time validation and augmentation with two independent data sources. *Transportation Research Record* 2526, 1 (2015), 79–89.