SMARTBUS: A Web Application for Smart Urban Mobility and Transportation

Sudha Ram
MIS Department, Eller College of Management
University of Arizona
sram@email.arizona.edu

Yun Wang
MIS Department, Eller College of Management
University of Arizona
yunw@email.arizona.edu

Faiz Currim
MIS Department, Eller College of Management
University of Arizona
currim@email.arizona.edu

Fan Dong
MIS Department, Eller College of Management
University of Arizona
fandong@email.arizona.edu

Ezequiel Dantas
Engenheiro Civil
Prefeitura Municipal de Fortaleza
University of Fortaleza
ezequieldantas@gmail.com

Luiz Alberto Sabóia
Secretario Municipal de Conservação e Serviços Públicos
Prefeitura Municipal de Fortaleza
luis.saboia@fortaleza.ce.gov.br

ABSTRACT

Systematic evaluation is crucial to the management and development of smart urban transportation, as it allows transportation planners to better understand the impact of their decisions and design targeted interventions to improve efficiency. Implementation of smart and adaptable public transportation is an important challenge in developing cities and newly industrialized economies where growth characteristics contribute to and can be impacted by factors like overcrowding and travel delays. In this paper, we focus on bus transportation, and present the design and implementation of a 3-layer web-based system for performance evaluation and decision support. This is part of a "Smart Cities" initiative, which is an international collaboration between academia and government. The first layer estimates fundamental indicators such as bus travel time and passenger demands by integrating heterogeneous data sources. A novel bus-stop network is then designed in the second layer, which enables the derivation of passenger patterns in public transit using network analysis. The third layer provides decision support by analyzing causal relationships between indicators. The proposed web-based system called SMARTBUS is being developed and validated with the city of Fortaleza in Brazil. We believe the use of generally available urban transportation data makes our methodology adaptable and customizable for other cities.

CCS Concepts

• Information systems Information systems applications • Applied computing Transportation.

Keywords: Smart City; Web Application; Urban Transportation; Network Analysis

1. INTRODUCTION

With the rapid growth of population and urban expansion, many cities are suffering from a struggling transportation system that cannot keep up with the ever-increasing demand for urban mobility. Such problems are more severe in cities of developing and newly industrialized economies that are often characterized by high urban density and challenges in public transport in terms of travel times, traffic congestion and passenger overcrowding [5]. To improve public transportation, a comprehensive analysis and evaluation system is beneficial for planners to identify aspects that are successful, uncover areas needing improvement, and also investigate the reasons behind these. The system can also be used to provide information to riders and support public awareness.

In addition to fundamental indicators such as passenger counts or travel speeds and times, an in-depth evaluation usually requires more sophisticated measurements. For example a common question regarding transportation efficiency could be whether or not the current transportation system meets the mobility needs of the city's commuters. To answer this question, the evaluation system should have appropriate indicators to measure urban mobility demands. As previous studies have suggested, human mobility exhibits simple and reproducible patterns [4]. In the context of urban transportation, a pattern could be regular use of certain terminals or bus stops.

Another important component of an evaluation system is providing planners with decision support. Urban transportation efficiency can be affected by a variety of factors such as passenger boarding / alighting volumes, time of day, service types, stop locations, weather conditions, road accidents and the interaction of these variables [7]. Measuring the extent to which each of these variables influences the outcome can help design targeted interventions for performance improvement.

Our research objective is to develop analytical metrics and methods using Big Data to support the design of solutions for smart urban transportation. We propose a three-layer system called SMARTBUS for evaluating and managing urban bus transportation in developing countries where buses are the primary public transportation choice. In the first layer, we compute bus travel times and passenger boarding location by combining Global Positioning System (GPS) measurements, passenger card scans, and open geographic information system (GIS) data. In the second layer, we create a bus-stop network based on information from the first layer. Innovative network metrics are designed to measure fulfillment of the current transportation needs. In the last layer, we implement analytical and visualization modules to measure and display the impact of variables related to the bus system. We demonstrate the implementation of this system as a Web application and show how it can be used for developing economies because it is cost-effective and scalable.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 describes how we design and

WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2888613

implement the three-layer application. Section 4 presents the potential use of the system in the city of Fortaleza in Brazil. Finally, Section 5 provides conclusions and future directions.

2. RELATED WORK

Traditional research on public transportation mainly focuses on the planning, operation, and control of urban transport systems based on buses [7]. Typical input information sources include transit network topology and characteristics, fare structure, service standards and constraints. These studies address the efficiency issues by looking at the problem of Transit Network Design and Timetabling, Frequency Setting, Scheduling Problem and Real-time control strategies [7].

With the development of technology for tracking such as, GPS signals, passenger card scans, AVL (Automatic Vehicle Location), APC (Automatic Passenger Counter), and traffic sensors, the amount and variety of available urban transportation data has greatly increased. This permits the use of data science techniques to analyze problems related to bus travel time. Previously proposed solutions for system management include investing in monitoring tools such as AVL and APC systems [3]. AVL-APC systems have been proposed to provide observations of indicators such as vehicle travel times, travel speeds and passenger boarding/alighting counts, as well as supporting studies on planning transport systems [6, 11], designing real-time control strategies [10] and predicting travel time [1, 12]. However, it is still a challenge to convert these research findings into real world solutions for developing cities where the access to expensive monitoring techniques on a system-wide scale is usually limited by resource constraints. When a comprehensive AVL-APC system is not available, it is necessary to estimate fundamental indicators from alternative data sources (e.g., fare collection records, bus GPS systems) that are commonly available and cost-effective. Besides on-board bus devices, smartphone applications installed on passenger mobile devices make it possible to collect tracking data via crowd-sensing [2]. This approach requires users to download and use the app and may suffer from bias due to self-reporting.

A related and important area that has not received sufficient attention in existing literature is the study of human mobility patterns. Prior research findings have revealed that human movement behaviors exhibit reproducible patterns [4]. For example, a recent study confirms spatiotemporal patterns through analyzing taxi-trace datasets [9]. We argue that transit patterns in bus systems are both observable and useful. Spatial (e.g., bus stop) and temporal (bus scheduling) information are likely to be more predictable for bus passengers (than taxi passengers). Therefore, in this study, we examine bus passenger patterns using both existing and new network analysis techniques to analyze boarding / alighting behavior.

3. METHODOLOGY FOR SYSTEM DESIGN AND IMPLEMENTATION

In this section, we first introduce the scope of this study followed by an overview of our proposed system called SMARTBUS and its architecture. Then we discuss the implementation details of each layer in SMARTBUS. The discussion emphasizes the mechanisms for processing input data and quantitative techniques to utilize the outcomes.

3.1 System Overview

The data used for this study were collected from the city of Fortaleza in Brazil during 2014-15. Fortaleza is the fifth largest city in Brazil with a population of about 3.6 million spread across about 120 square miles. Its bus transportation system has over 300 routes, spanning nearly 5,000 bus stops. Data from approximately 2,000 buses were in used for our analysis. Most of these buses have GPS tracking devices which transmit bus locations every 15-30 seconds. Automatic fare collection (AFC) provided data about passenger trips (the system supports both cash and smart card usage). Citywide, passengers made 30 million trips per month on average. During the period of study, dedicated bus lanes were introduced on several routes to improve the travel experience.

Table 1 summarizes the datasets and their attributes used in this study. Compared with an AVL-APC system which can cost between \$10,000 and \$20,000 per bus, combining GPS signals with fare collection records is a more economical alternative for traffic monitoring, since the cost for a GPS tracking device per bus is normally within \$200 per bus and AFC systems are widely available in many cities. However, using GPS and AFC systems requires additional computation for the derivation of bus travel times and passenger boarding / alighting counts. For example, a fare collection record indicates that passenger u_i boarded a bus b_j at time t_k but does not indicate the bus stop where the passenger boarded the bus. In order to extract such "location" information, we need to combine this data with the GPS repository by matching b_j and t_k. This search process is non-trivial, given that a typical month of GPS data contains 200 million records. We also include open source GIS data[1][2] to extract the route polyline between bus stops, and weather conditions (to improve the analysis accuracy).

Type	Name	Attributes
Major Datasets	Bus GPS Signals	<bus ID, latitude, longitude, timestamp>
	Fare Collection Records	<bus ID, timestamp, user ID, payment type, bus direction, route ID>
Complementary Datasets	Bus Stop Locations	<bus stop ID, latitude, longitude, rout ID>
	Dedicated Lanes	<length, location, effective date>
	Bus Schedule	<departure time, arrival time, trip interval>
Open source Datasets	Open GIS	<predecessor bus stop, successor bus stop, polyline>
	Weather	<date, weather condition, precipitation>

Table 1. Types of Data Sources

In SMARTBUS, we addressed this problem in the first layer where we implemented effective techniques for data integration and analyses. As illustrated in Figure 1, Layer-1 accepts input from GPS signals, fare collection records and bus stop locations to calculate fundamental performance indicators, e.g., bus travel time, bus segment speed and passengers' boarding locations. The outcomes feed into Layer-2 and Layer-3 for further analysis. In Layer-2 we extract and construct a bus stop network based on passengers' boarding locations. Metrics derived from a comparison between the bus stop network and transit network are then used to identify passenger needs that need to be fulfilled.

[1] OpenStreetMap: https://www.openstreetmap.org

[2] Google Maps: https://maps.google.com/

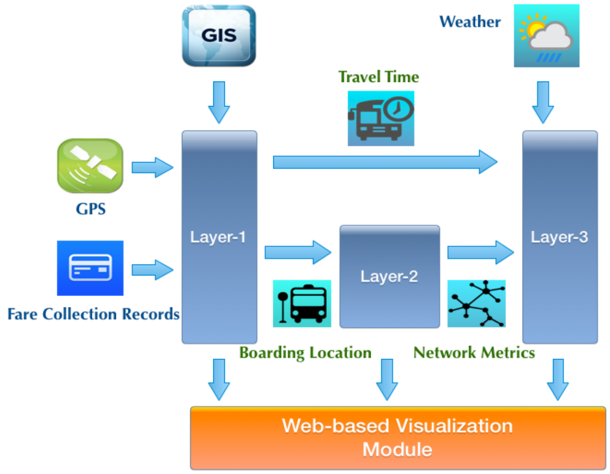

Figure 1. System architecture of SMARTBUS.

The indicators are used by analytic modules in Layer-3 to provide the users (city administrators and technicians) with insights into the current state of the system (both for infrastructural components performing normally and those that need additional understanding or intervention). In addition to the big data computations, each layer is connected to a web based visualization module which provides map-based interactive analysis to explore intermediate results from each step. For example, the dashboard can provide a transportation manager with metrics that point to the causes of bus delays (e.g., spikes in passenger demand which increased time spent at bus stops and, in turn, total trip time).

3.2 Estimating Fundamental Indicators

Travel time is one of the most important measures to evaluate the performance of an urban transportation system [8]. As mentioned in Section 2, one of the functions of Layer-1 is to extract bus travel time and passenger boarding location from GPS signals and fare collection records. By linking the GPS and AFC records, we are able to calculate start and end times for a bus on a route. We can also estimate the nearest bus stop (i.e., boarding location) for a passenger by comparing the time of a passenger card scan and the corresponding location from GPS records. One challenge here is to integrate and process very large volumes of data. We implement

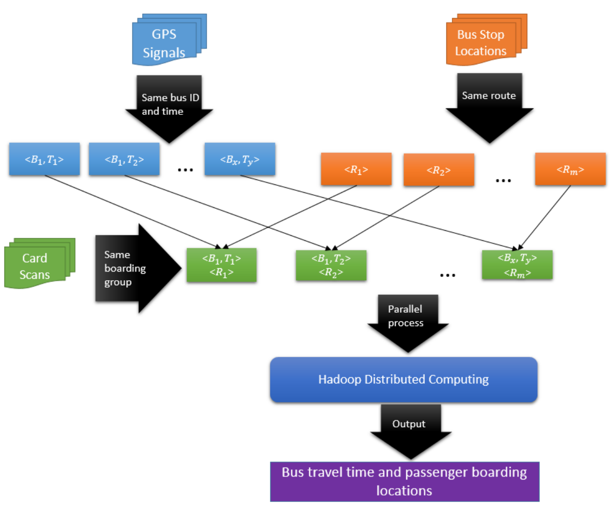

Figure 2. Data partitioning strategy and parallel computing

multiple techniques including data partitioning and Hadoop-based distributed processing to improve the running times.

As illustrated in Figure 2, GPS signals are partitioned by a pair of bus ID and time <$\boldsymbol{B_x}, \boldsymbol{T_y}$>. Bus stop locations are partitioned by route ID <$\boldsymbol{R_m}$>. Card scans are segmented so that passengers boarding the same bus at same time are in the same group. Using this partition strategy, card scan segments are effectively linked with relative GPS signals and bus stop locations. Linked data are processed simultaneously in our Hadoop distributed computing system.

3.3 Network Analysis of Urban Mobility

Passenger boarding / alighting data is key to understanding urban mobility demands, as boarding/alighting pairs reveal passenger transit patterns. However, unlike boarding information which can be estimated from fare collection records, passenger alighting data is harder to estimate. Therefore we introduce a new network analysis approach to learn urban mobility patterns using only the passenger boarding information.

3.3.1 Bus Stop Network Definition

Our extraction of a bus stop network is motivated by the fact that over 75% of passengers have at least two boarding records in a day. Boarding at different locations can reflect a transit pattern, e.g., return trips or transfers. For example, a passenger may exit a bus close to where they board next. From a human mobility perspective, a pattern of passenger movement in urban transit can indicate an underlying background characteristic. For example, students might share a same first leg from their school to a terminal and from there they might take different buses to their own home. In this study, our assumption is that that the higher the number of times passengers board at a pair of bus stops, the more likely there is an urban transit need between the two areas. Based on this assumption, we created a bus stop network using the number of shared boarding passengers (SBP) among bus stop pairs.

Since urban mobility might exhibit seasonal patterns, we generate these networks on a monthly basis. Let B be the set of bus stops, the network of month p is denoted as $N^p(V^p, E^p)$ where $V^p \subseteq B$ is a subset of bus stops and E^p is a set of undirected edges that connect bus stops. For bus stops $b_i, b_j \in B$, if more than τ people board at both b_i and b_j in a day (in month p), we add nodes b_i, b_j to V^p and e_{ij} to E^p. In our current implementation, we empirically set τ as 20. After connecting all bus stops that meet the minimum SBP, we normalize edge weights to measure the strength of connections. In the example in Figure 3, the number attached to each node indicates the number of passengers boarding at that stop. SBP between bust stop b_1 and b_2 is 20 and SBP between bus stop b_1 and b_3 is 50. A *global maximum* normalization will result in the weight of $e_{12} = 0.4$ and $e_{13} = 1.0$. However, b_3 has a larger passenger base than b_2. On the other hand, 50% of the passengers who boarded at b_2 also boarded at b_1. Therefore, the normalization should take into consideration passenger counts at each stop so that connection e_{12} has a larger weight than e_{13}. To do so, we applied the normalization method used in [13] which adds a neutralizing step before the global maximum normalization. In particular SBP value between two bus stops will be divided by a product of the number of passengers boarded in each bus stop. Using this method, weight of $e_{12} = 1.0$ and $e_{13} = 0.2$.

3.3.2 *Urban Mobility Patterns from Network Measurements*

In this section, we discuss how the extracted bus stop networks can help understand urban mobility patterns. These measurements are further categorized by the different aspects they capture.

The first category of measurements is based on community detection results. A community is partitioned so that the intra-community-connection is maximized while inter-community-connection is minimized. If we interpret the community detection results in the context of urban mobility, a community defines a living area for a group of people. Hence, we can use the following methods to characterize patterns.

1) Examine connections ranked higher in terms of their weight.
2) Measure maximum and average distance between connected bus stops within the community to infer living area diversity.
3) Add neighborhood information such as schools, shopping malls or residential blocks to label the regions in a community to further explain factors that characterize the community.

The second group of metrics reflects temporal dynamics of the network by measuring node and edge persistence rates across time. A high overlap of nodes and edges indicates a stable demand for urban mobility. In contrast, low rates of persistence of node or edges may be due to a changing population of bus passengers.

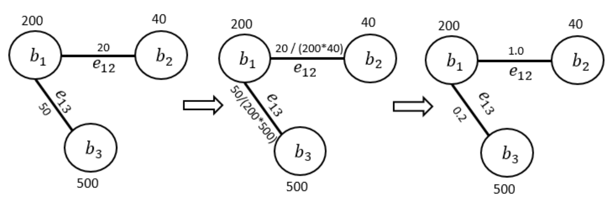

Figure 3. Example of bus stop network normalization

Many interesting indicators can be measured by combining our bus stop network with the transit network. Traditionally, transit network efficiency can be measured by the shortest path between stops. Let $N^p(V^p, E^p)$ be the bus stop network for month p, d_{ij} be the distance of shortest path between bus stops b_i and b_j in the transit network, w_{ij} be the weight of b_i and b_j in N^p; then a weighted efficiency is measured as:

$$\frac{1}{|E^p|}\sum\frac{w_{ij}}{d_{ij}}\ for\ e_{ij}\ \in E^p$$

The motivation is that distance of shortest path affects the overall efficiency more on pairs of bus stops that have higher weights in the bus stop network. This value can be used to measure the overall efficiency of the bus transit system.

Similarly, we can examine a pair of bus stops that has a larger weight with the combination of its transit network attributes, such as:

1) Polyline distance (distance the bus needs to travel). If the polyline distance is too long, a city planner may consider designing a shorter route to reduce travel time and improve passenger experience.

2) Number of bus routes. For high-demand transportation needs, if passengers have more alternatives, it is less likely that overcrowding will occur.

3.4 Analysis of Bus Delays

A key requirement for Layer-3 is to analyze how different factors affect the health of the bus network. For example, an important measure is trip delay (i.e., difference between the estimated travel time and the scheduled time).

The interactions between factors and their impacts on bus delay time may change under different circumstances (e.g., different routes). At the city level, some interventions were carried out in Fortaleza to reduce travel time, e.g., introducing dedicated bus lanes on some parts of routes. Therefore, performing route-level regression analysis can help planners understand the impact of their decisions and better identify the reasons that caused delays.

To implement Layer-3, the system requires comprehensive and accurate input measures. . For example, the following input factors are useful in determining bus network health.

1) Number of boarding /alighting passengers
2) Road conditions (e.g., route length, proportion of dedicated lanes and number of traffic lanes)
3) Environment variables (weather conditions, date & time)

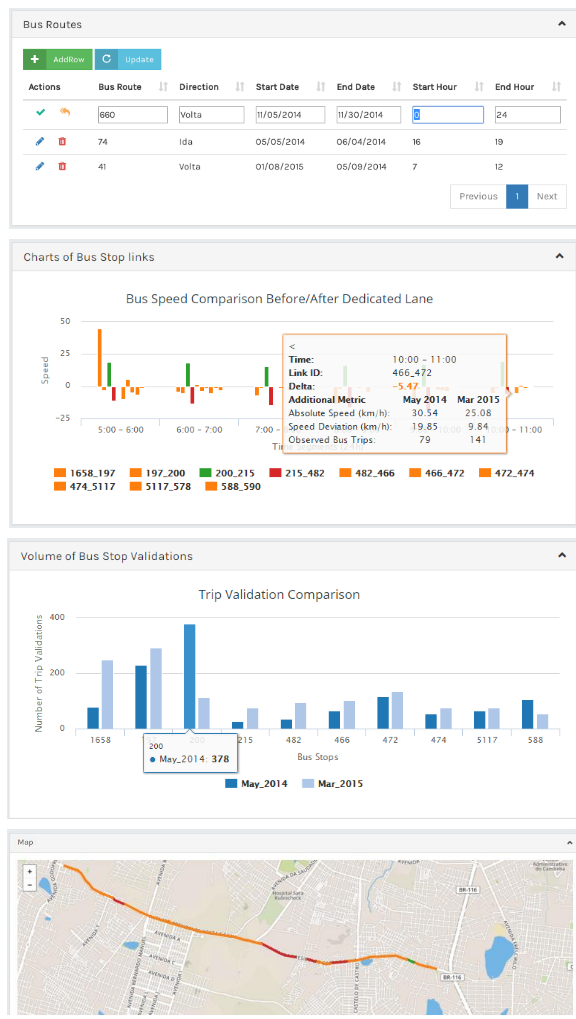

Figure 4. User interface for dedicated bus lane evaluation

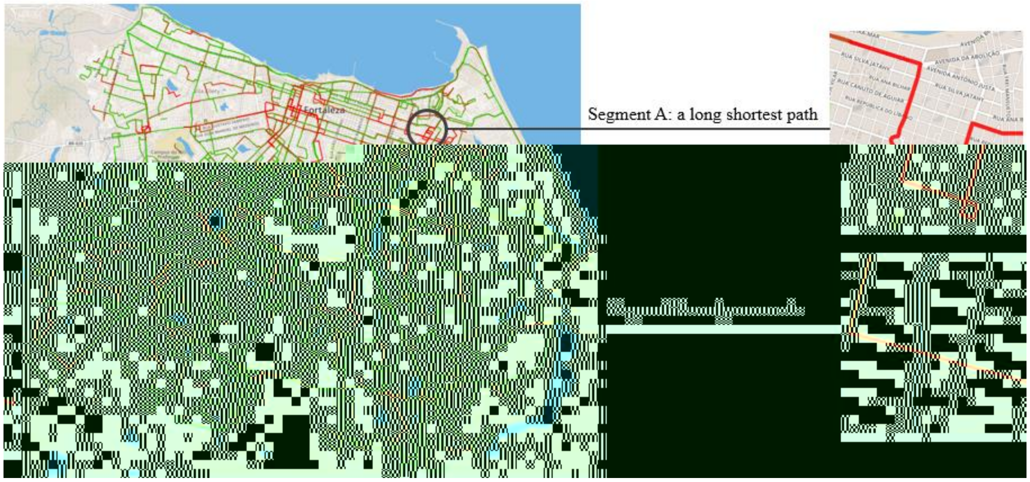

Figure 5. Use bus stop network metrics to identify problematic route segments

4) Fleet operator characteristics (company, number of fleets)
5) Road accidents

The data comes from multiple sources including the city government, fleet operators, weather sensors, etc. This in turn leads to challenges in data integration.

4. USING SMARTBUS FOR DECISION MAKING

In this section we describe several cases where SMARTBUS is useful for evaluating an urban transportation system and supporting its redesign. Each layer in the system is connected to a web-based visualization module which provides interactive tools for transport planners and managers to explore the results.

4.1 Visual Analysis of Indicators

In Layer-1, we estimate bus travel time and the number of boarding passengers. Visualizing these fundamental indicators on the map can reveal information on when and where severe bus delays and overcrowding happen. For this purpose, we used open source map data to extract polylines between bus stops. Using the polylines, we calculated the travel distance between bus stops and then converted travel time into average travel speed, which is a generalizable indicator when distances between bus stops vary.

As illustrated in Figure 4, the visualization tool allows users to select bus routes, time periods and other conditions to examine the state of the system using fundamental indictors such as bus travel time, travel speed and passenger boarding volume. Bus routes are color coded on the map to help explore spatial patterns of the metrics that are being examined. One example of Layer-1 visualizations is to examine the effectiveness of dedicated bus lanes by comparing bus speeds before and after the dedicated lane is introduced. Besides a speed comparison, the system also provides a comparison of scheduled bus trips and passenger volume. This is useful to understand the cases where a dedicated lane does not result in a significant speed boost, perhaps due to increased number of passengers in the corresponding bus route.

4.2 Route Segment Analysis

The bus stop network we developed in section 3.3 provides new ways to analyze existing bus routes. Given a bus speed map, a transit planner can first identify problematic bus route segments based on its bus stop network weight, and then explore the potential reasons leading to congestion. For example, in Figure 5, we see two bus segments (in red) where the bus speed is slow. Both of them have normalized weights close to 1.0, which indicates a strong transit pattern between the two ends of the segments. Segment A has 12 bus stops that are used by different routes. We can adjust one of the routes so that it connects the two ends of segment A in a more direct way to reduce the length of shortest path. Segment B is another problematic route segment where only one bus route connects the two ends of the segment. Therefore, planners can adjust other routes to serve this segment or increase the fleet size or schedule frequency for this route.

4.3 Factors causing Bus Delays

In the previous sections, we examined the performance of separate bus segments. In this section, we analyze bus delays at the route level. We compare the actual time that a bus took to finish a trip (from origin to the destination) with the scheduled time. Computed bus delays can be used in two ways: 1) a visualization tool for exploring relationship between bus delays and selected dependent variables; 2) a regression model to quantitatively measure factors contributing to bus delays.

4.3.1 Bubble Chart for Bus Delay Analysis

In the visualization tool, we implemented a dynamic bubble chart in which routes are represented as bubbles in different colors. Bubble size and X-axis position can be used to represent dependent variables. Figure 6 visualizes the number of passengers boarding each hour on the X-axis, while number of buses on the route is indicated by the size of the bubble. Every frame in the animation demonstrates a relationship between bus delays, the number of buses, and the number of boarding passengers. Figure 6 indicates that bus delays are strongly related to the number of boarding

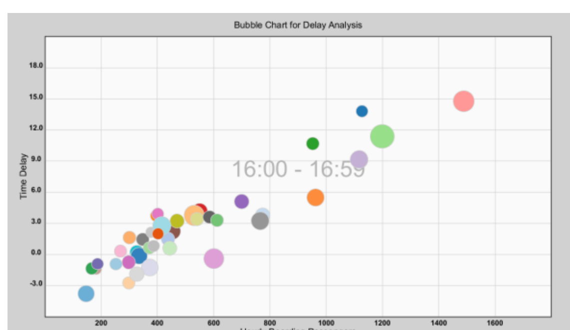

Figure 6. Dynamic bubble chart for bus delay analysis

passengers. For routes that have high demands, longer delays and fewer buses (shown as small bubble size), increasing the fleet on the route might reduce the delay.

4.3.2 A Regression Analysis of Bus Delay

To better understand how different factors affect bus delays, we developed a regression model. Results are shown in Table 2 for a popular route, using comparison data from one month in 2014 and the corresponding month in 2015. Every bus trip during the period is an instance. A subset of independent variables were used for the analysis based on data availability.

The regression results lead to several conclusions: 1) Weekday and morning rush hour trips are more likely to be delayed. 2) The impact of weather was not significant for the analyzed months in part because persistent heavy rains were not experienced during the analysis period (i.e., 8 rainy days across the analyzed months). 3) Afternoon rush hour had "positive" impact on delay owing to the fact that the schedule budgeted 15-20 minutes more time than for the morning rush hour period. Based on this finding, we suggest increasing the scheduled time in the morning or using higher capacity buses. Correspondingly, a reduction in scheduled trip time can be proposed for the afternoon to allow more buses on this route in afternoon. 4) The dedicated bus lane program was found to be generally effective, which suggests that routes with similar demand (and road) conditions can consider introducing dedicated bus lanes.

Table 2. Multiple linear regression model of bus delay on a popular bus route

				β
Weekday	Binary	-12.196	2.704	-.069***
Morning Rush Hour	Binary	-20.840	1.790	-.165***
Afternoon Rush Hour	Binary	10.056	1.888	-.077***
Rain	Binary	-.290	1.868	-.002
Number of Passengers	Numeric	-3.68	.018	-.306***
Dedicated Lane	Binary	5.311	1.427	.052***

***: significant at $p < 0.001$

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an innovative web application called SMARTBUS to support the evaluation and redesign of urban transportation systems for transportation planners, managers, and technicians. In the first layer, the system addresses the challenge of developing useful performance indicators by computing bus travel time and passenger boarding volumes from raw bus GPS data and fare collection records. The second layer uses passenger boarding information to create a bus stop network for the identification of bus passengers' transit patterns. Combining these patterns with the current bus transit network helps identify design issues that can be addressed. Layer three is designed to analyze the impact of different factors on route-level bus delays and provides suggestions based on the results. We implemented and evaluated the SMARTBUS system to evaluate and improve the bus transportation system of the city of Fortaleza in Brazil. While the primary users of the web-based application are local government technicians and managers, the system can provide transportation metric information (e.g., average transit times or speeds) for the public, and can be used to promote awareness of city initiatives and results (e.g., show how bus network or dedicated bus lanes are improving travel times or overcrowding).

We are continuing to extend this work in several directions. We are implementing a predictive model to estimate the probability distribution of passenger's alighting location given a boarding location. The passenger alighting information can expand our current network analysis. We also are extending the predictive model to incorporate related data such as road accidents, road and traffic congestion conditions. Future work involves integrating the bus system with other forms of urban transportation such as bike sharing.

6. REFERENCES

[1] Bai, C., Peng, Z., Lu, Q. and Sun, J. 2015. Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes. *Computational Intelligence and Neuroscience.* 2015, (2015), 1–9.

[2] Farkas, K., Feher, G., Benczur, A. and Sidlo, C. 2015. Crowdsending based public transport information service in smart cities. *IEEE Communications Magazine*. 53, 8 (Aug. 2015), 158–165.

[3] Furth, P.G., Hemily, B., Muller, T.H.J. and Strathman, J.G. 2003. Uses of Archived AVL-APC Data to Improve Transit Performance and Management : Review and Potential. *Transportation Research Board.* 23, June 2003 (2003).

[4] González, M.C., Hidalgo, C. A., and Barabási, A.-L. 2008. Understanding individual human mobility patterns. *Nature*. 453, 7196 (Jun. 2008), 779–82.

[5] Gwilliam, K. 2003. Urban transport in developing countries. *Transport Reviews*. 23, 2 (Jan. 2003), 197–216.

[6] Hadas, Y. and Shnaiderman, M. 2012. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. T*ransportation Research Part B: Methodological*. 46, 8 (2012), 1068–1084.

[7] Ibarra-Rojas, O.J., Delgado, F., Giesen, R. and Muñoz, J.C. 2015. Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*. 77, (Jul. 2015), 38–75.

[8] Kieu, L., Bhaskar, A. and Chung, E. 2012. Bus and car travel time on urban networks : integrating Bluetooth and bus vehicle identification data. *25th ARRB Conference - Shaping the future: Linking policy, research and outcomes* (2012), 23–26.

[9] Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W. and Wang, Z. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science in China*. 6, 1 (2012), 111–121.

[10] Sun, A. and Hickman, M. 2005. The Real-Time Stop-Skipping Problem. *Journal of Intelligent Transportation Systems.* 9, 2 (Apr. 2005), 91–109.

[11] Tétreault, P.R. and El-Geneidy, A.M. 2010. Estimating bus run times for new limited-stop service using archived AVL and APC data. *Transportation Research Part A: Policy and Practice*. 44, 6 (Jul. 2010), 390–402.

[12] Zhang, C. and Teng, J. 2013. Bus Dwell Time Estimation and Prediction: A Study Case in Shanghai-China. *Procedia - Social and Behavioral Sciences*. 96, Cictp (2013), 1329–1340.

[13] Zhang, K., Ram, S. and Bhattacharya, S. 2015 Large Scale Network Analysis for Online Social Brand Advertising. *forthcoming in MIS Quarterly, Special Issue on Transformational Issues in Big Data and Analytics for Networked Business*, 2015.