Empirical Analysis of Crowdsourced Google Travel Time Data for Sustainable Transportation Applications

September 2016

The National Center for Sustainable Transportation Undergraduate Fellowship Report

Tului Gantulga, Undergrad Research Assistant

Miguel Jaller, Assistant Professor

National Center for Sustainable Transportation

ITS UCDAVIS INSTITUTE OF TRANSPORTATION STUDIES

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; University of California, Riverside; University of Southern California; California State University, Long Beach; Georgia Institute of Technology; and University of Vermont. More information can be found at: ncst.ucdavis.edu.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the United States Department of Transportation's University Transportation Centers program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Acknowledgments

This study was funded by a grant from the National Center for Sustainable Transportation (NCST), supported by USDOT through the University Transportation Centers program. The authors would like to thank the NCST and USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project.

Empirical Analysis of Crowdsourced Google Travel Time for Sustainable Transportation Applications

A National Center for Sustainable Transportation Research Report

September 2016

Tului Gantulga, Department of Civil and Environmental Engineering, University of California Davis

TABLE OF CONTENTS

Introduction **1**

Methodology **1**

Analysis of Design Factors **1**

Introduction

Congestion affects the movements of people and goods. Overall, congestion negatively impacts the efficiency and competitiveness of urban areas. For freight, where the value of the cargo, reliability and other logistic performance metrics are very important, congestion impacts are even more acute. Moreover, there are no free congestion data available for many urban areas, and the results of local travel demand models are not accessible.

In Latin American, for instance, just a handful of studies have provided indications of average speeds in some cities, and public agencies rely on expensive data collection efforts to have a snapshot of the traffic conditions and calibrate their travel demand models.

The objective of this research is to take advantage of google traffic crowdsourced data and develop a methodology to capture congestion levels and traffic conditions in study areas. In doing so, the author used the Google Maps API as the source of distance and traffic related data. This project, exploratory in nature, pilot tested the methodology and conducted a preliminary assessment of the impacts of selecting various design factors.

Methodology

Randomly generated, 2500 origin-destination (OD) points are sent to Google within specified square sizes at random times during predetermined 10 segments of hours. Data returned is analyzed to determine average speed, max speed, min speed, coefficient of variation, and congestion impacts on speed. Then, a sensitivity analysis is done to determine the optimal square size that is the biggest and most accurate. This is due to the daily limit of 2500 requests set by Google. Once the most optimal square size is determined, more data can be sent to Google to create an offline map and a directions program that optimizes routes based on previous stated parameters.

Analysis of Design Factors

Downtown Los Angeles, Anaheim, Lakewood, Pasadena, Torrance, and Sao Paolo, Santiago, and Barranquilla are chosen for analysis. The results are compared with other traffic data collected form a different source, which shows that the received Google data is close to the other data. Following are the graphical representations of the analyzed results:

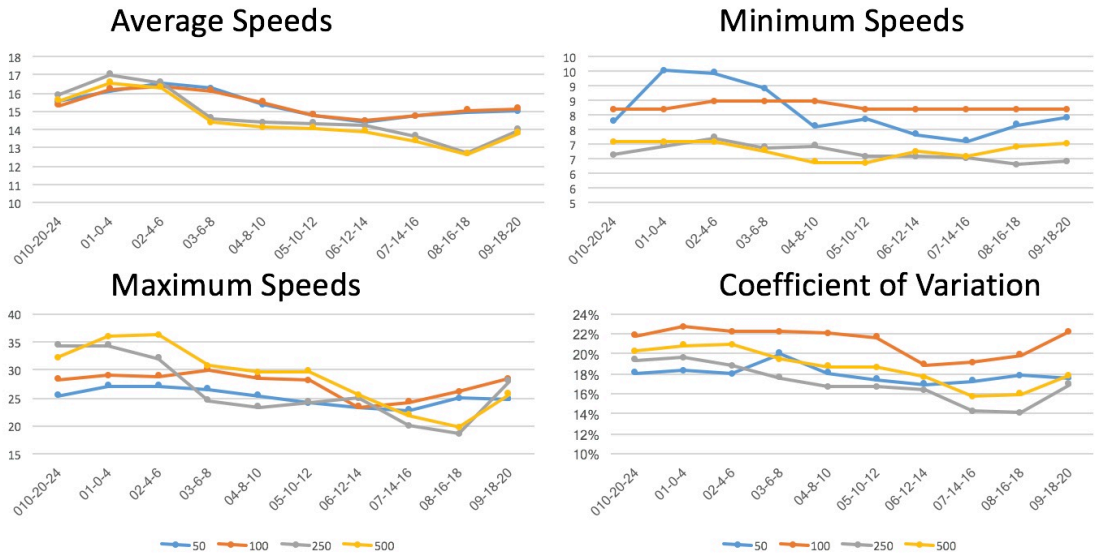

Figure 1. Analysis of Downtown LA throughout the day

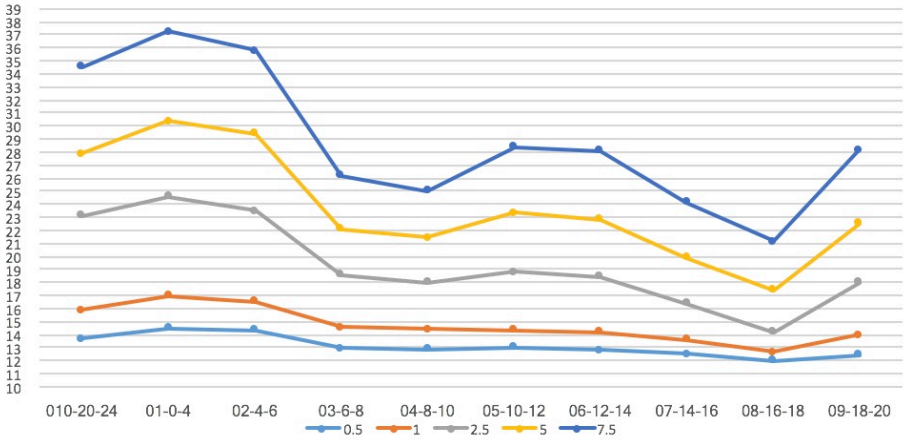

Figure 2. Downtown LA average speeds with different square sizes on Monday

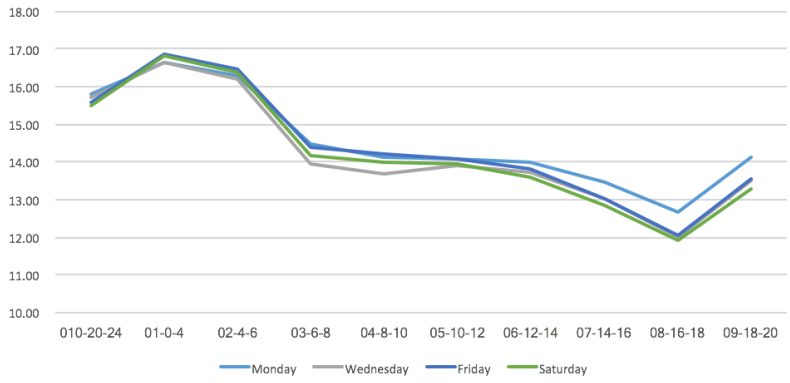

Figure 3. Average speeds of 1 sq. mile Downtown LA throughout the day

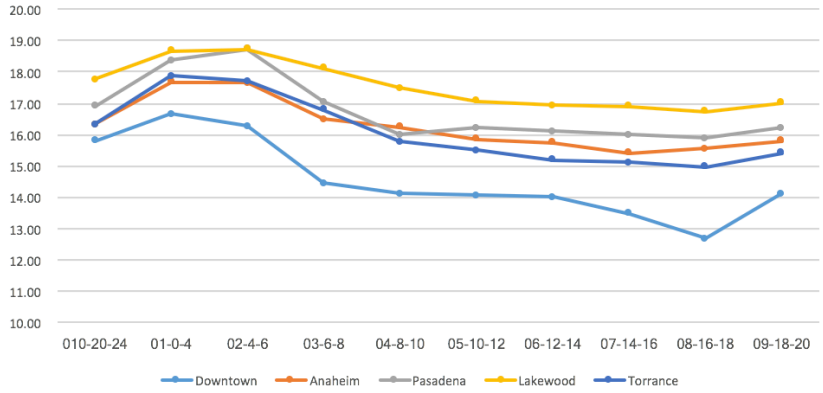

Figure 4. Average speeds of LA Area throughout the day

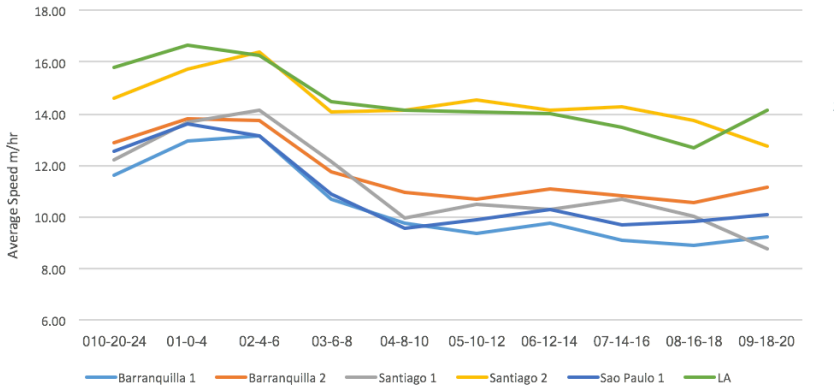

Figure 5. Comparison of average speeds of Latin American cities and LA

Figure 1 is the analysis of different number of OD points. This is part of the sensitivity analysis to determine the most optimal number of OD points. To determine, the coefficients of variation are compared with each other and the number of OD points with the least variation is chosen. In this case, 250 OD points is the most optimal one. Of course this is only for Downtown LA, and it might be different for other parts of the country and the world, but for this research, 250 OD points is used.

Figure 2 shows that as the size of the square increases, so does the average velocity in each segment throughout the day. This is partially due to the availability and usage of freeways for longer distances. Also, the average velocities are at their maximum from midnight to 4 AM because there are much less cars during that time of the day and it starts to increase as the day begins and people go to work. It hits its minimum between 8 AM and 10 AM because that's when most people go to work and starts to increases, but decreases as people get off work and go home around 4 PM and 6 PM. The square with 0.5-mile side length is closest to the actual

average velocities, however, 1-mile side length is close to it, and everything else deviates far from it. Thus, a square with 1-mile side length is the most optimal size for analysis.

Figure 3 shows the average speeds throughout the day on Monday, Wednesday, Friday, and Saturday. They are consistent with each other.

Figure 4 is the comparison of places around LA with each other. As expected, Downtown LA has the slowest traffic and the other places have faster traffic.

Figure 5 compares Downtown LA with Latin American cities. LA has faster traffic than these cities probably due to the availability of highways and freeways, population density, car density, and other factors. Latin American cities are analyzed for the purpose of comparing it with other GPS data to make sure that they are accurate.

In Conclusion, 250 OD points for 1 square mile is the most optimal size and number of ODs.