Traffic Patterns In NYC

Richard Sowers (Professor)
Dan Work (Professor)
Vaibhav Karve (Graduate)
Derrek Yager (Graduate)
Xinyu Liu (Undergraduate)
Yicheng Pu (Undergraduate)
Gabe Shindnes (Undergraduate)
Ziying Wang (Undergraduate)
Yu Wu (Undergraduate)
Zeyu Wu (Undergraduate)
Dajun Xu (Undergraduate)

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Introduction

Goal

Provide generalized traffic trends in New York City depending on time of day and day of year and be able to specify anomalies in traffic due to unexpected events from taxi cab data.

The Data that we worked with was obtained through a Freedom of Information Law request from the New York City Taxi and Limousine Commission and contains hourly traffic estimates for each New York City road for the years between 2010 to 2013.

Tools

SNMF

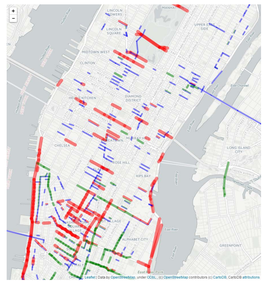
Graduate students Vaibhav Karve and Derrek Yager used a sparse non-negative matrix factorization algorithm to factorize the original time-spatial traffic flow matrix into two matrices: a time-signature matrix **(W)** and a signature-spatial matrix **(H)**. To balance computational power and error reduction they limited the total number of signatures (rank of the matrix) to 50 different trends.

Visualization Tools

Our group used the Python Folium package, a tool based off of Leaflet.js, to visualize the map and paths in New York City. Our plots are based on the geographical coordinates which are in the form of (Latitude, Longitude).

Video

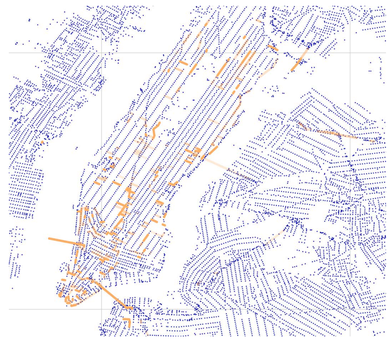
We also made a video showing the weights and distribution of the top three signatures changing over time. We decided to pick the three most popular signatures, and display the weights and distributions on different roads (links) over a one week period from January 8th, 2011 to January 14th, 2011.



Link: https://youtu.be/mbhEBgiJGw4

Visualization

The image below is a visual representation of one of the 50 signatures that we identified in our data (more specifically signature 6). The higher the opacity of the color on each link represents a higher weight.



Observations

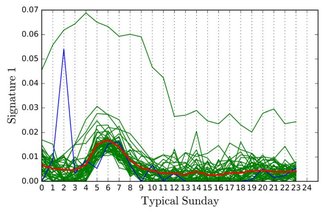
Researching current events during the time set of our data we could see how various events (both natural disasters and human-triggered events) impacted the flow of traffic in New York City.

Hurricane Irene

Background: Hurricane Irene lasted from August 27th, 2011 until August 29th, 2011. The effects of Hurricane Irene were some of the worst since Hurricane Agnes in 1972 and caused approximately $300 million dollars in damage over a 3 day period.
The usual peaks in signature 1 happen at around 5 am. However, the frequency of drivers on signature 1 was exceptionally higher than the average as can be seen in the above graphs through the 3 days of the hurricane.

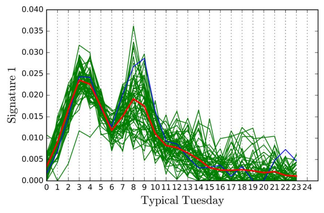
Daylight Savings

An unusual spike in traffic occured at 2am on March 13th, 2011 which coincided with the Daylight Savings, where we have one less hour of sleep.

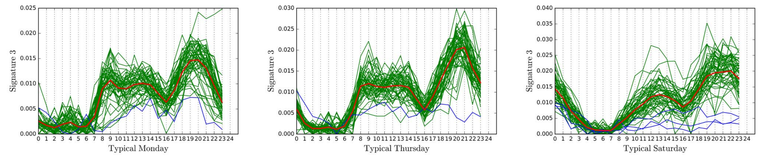


Earthquake

Traffic that is represented by signature 1 was not affected by the 5.8 magnitude earthquake that hit New York City on Tuesday, August 23rd, 2011.



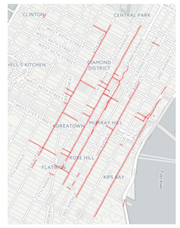
Holiday

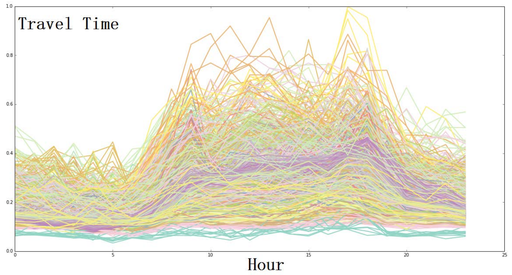


Monday: Independence Day
Thursday: Thanksgiving
Saturday: New Years Day, Hurricane, and Christmas Eve
For all of these major holidays we notice that traffic is considerably smaller than average. Note that Hurricane enlarged traffic flow clustered in signature 1 but decreased traffic flow clustered in signature 3.

Murray Hill

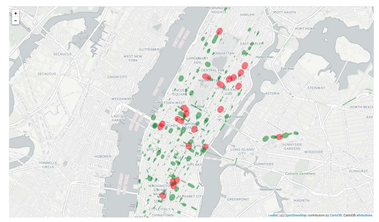
Almost all links in signature 15 are in this area centered by Murray Hill, which is noteworthy among 50 signatures. This picture shows the travel pace(travel time/street length) of each link, from which we can see the pace of these links in signature 15 are highly correlated.

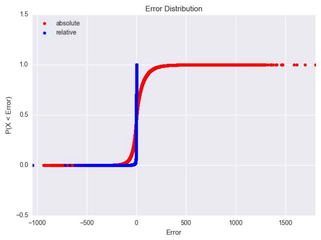




Error Analysis

To get a perspective of how accurate our trends were to real time data we took our real data and subtracted **W*H** to get a sense of absolute error. The one flaw with absolute error it tends to overemphasize error when there is a higher frequency of traffic. To counteract this we also took a look at relative error where we divide the absolute error by the actual amount of traffic. Both of these metrics are incredibly important for error analysis. On the map to the bottom-left we see the relative error distribution throughout NYC on New Years Day in 2011. We can clearly see that compared with relative error maps of normal days, the New Year's Eve estimate tends to have larger relative errors and those tend to be clustered around central park. In addition to this, we see that the links that are prone to large errors are usually comprised of only one signature. On the bottom right we see the overall distribution of relative and absolute error.





References
[1] DRAFT Multi-Jurisdictional Multi-Hazard Mitigation Plan.
http://www.co.orange.ny.us/filestorage/1162/975/2095/1008/10.2.29-*City of Port Jervis ID* 92337).*pdf*

Support for this project was provided by the Program for Interdisciplinary and Industrial Internships at Illinois, Illinois Geometry Lab, the Idea Lab, and the Department of Mathematics at the University of Illinois at Urbana-Champaign.