Interim Project Report

Bisrat Moges and Jessica Ho

ELE381

1 Introduction

Over the past few years, ride-sharing has created an enormous impact on inter-city transportation. Studies have shown that yellow taxi usage has decreased significantly since 2012, as Uber, Lyft, and other ride-sharing companies have grown in popularity. Currently, any renovation in a modern city transportation system must take into account the new methods of transportation along with the changing preferences of people and how they choose to mobilize. It's no question that Uber and Lyft have dominant control on the ride-sharing economy - however, dozens of start-ups, like Zoox and Nuro, along with automobile giants like GM and Toyota, aim to revolutionize mass transit by achieving self-driving, ride-sharing, fully electric vehicles, in hopes of a cleaner, more efficient future that minimizes empty miles driven, eliminates unnecessary traffic congestion, and reduces carbon emissions from fuel due to idling and inefficient vehicle usage.

Such disruptive technology also has its limitations, particularly in its social reception and trustworthiness when applied to communities that must service people from a diverse range of socioeconomic backgrounds - from those who are used to cheap, public transit versus those who want safe, highly regulated and vetted cab drivers. Ride-sharing does provide a unique solution that comes at a cheaper price than taxi cabs but with the added convenience of a vehicle that can take you anywhere, beyond the stops designated by public transit stations. While technology consistently outpaces policy, government regulation of these ride-sharing vehicles has also recently increased. This has raised controversy surrounding the overall effect of these vehicles on congestion and vehicle usage, along with concern for their supplemental effects on the taxi industry.

The effect of ride-sharing on yellow taxi cabs has been especially studied in New York City, which has one of the oldest, most extensive public transit system in the country - along with the most traffic congestion. Ever since Uber was first introduced in May 2011, dramatic changes in taxi cab usage has taken effect. In July 2015, the city almost passed to impose a cap on the number of Ubers in the city, before dropping its plan on the condition that Uber conduct a traffic-study in the following months. Analyzing the net effect of ride-sharing, with respect to the most traditional form of transportation used - taxis - can provide a preliminary glance into what kind of supplementary effects are being triggered, and therefore in the long-run, which of these effects need to be guarded against by state policy.

2 Problem Statement

How have ride-sharing companies like Uber affected the decline in taxis, and what is the overall net effect on road traffic in NYC?

3 Methods

Our main analysis will come in the form of network comparisons among that of Uber, taxi cabs before Uber, and taxi cabs after Uber. We expect in general that the taxi cab usage after Uber to be significantly less, but to gain non-trivial insight to the overall problem statement, we break our analysis into a few sub-problems:

- How well does the amount of Ubers in the period 2014-2015 match the decrease of yellow taxis over that time period?
- Are there certain locations that experience more Uber vs taxi congestion (or vice versa)?
- Are there certain times of the day that experience more Uber vs. taxi congestion (or vice versa)?

3.1 Datasets

We have two main datasets to draw from:

- Uber: April - September 2014, January - June 2015 pick-up location date/timestamps
- NYC Yellow Taxi: 2009 - 2018, pick-up/drop-off locations date/timestamps, along with other information about payment (not necessary for our current purposes)

3.2 Network Specifications

In general, we will model the networks as directed graphs, with nodes as geographical pick-up and drop-off locations, and edges as the number of trips between those locations in a certain time period. Our specifications are as follows:

- Range of time: we are constrained to our smallest dataset (Uber) which provides trip information for two six-month periods in 2014 and 2015. Therefore, we will analyze a similar periods of time for our taxi cab data, possibly before the 2011 introduction of Uber, the current state in 2019, and periods in the middle to see when Uber peaked and started to really gain traction. We can vary the time period over which we average the total number of trips: per 6-months, per day, per times of day (i.e. morning/afternoon/evenings).
- Geographical location clusters: we plan to cluster the pick-up and drop-off locations based on geographically relevant NYC neighborhoods.

We make several assumptions due to the current limitations of our datasets:

- Uber drop-off locations: the Uber data set only specifies pick-up location and time, so we will randomize the drop-off locations/times using a uniform distribution across all the possible nodes, excluding the origin
- Number of passengers: we assume that each trip involves the same number of passengers and do not monitor for shared rides, Uber pool, Uber XLs, etc. Because we are concerned mainly about traffic congestion itself, i.e. the number of vehicles altogether on the road, and not passenger mobility itself, this should still provide insightful results.

4 Current Progress

Our work so far has involved preparing the data for analysis. This has involved searching for appropriate tools for the analysis of Geo-spatial data and exploring trade offs and benefits of using any particular tool.

4.1 Work Towards Solution Implementation

Data from Uber was obtained thanks to a Freedom of Information Law request that had been filed by FiveThirtyEight. Data concerning yellow taxis was downloaded via an API that had been created by various agencies in New York City.

We selected Networkx, a python package made for the analysis of Geo-spatial data in order to draw inferences from our data sets. In learning to use it we found that Networkx offered a lot of tools for exploring the structure of networks. There are two weaknesses of this particular package that we have planned solutions for. First, the package was made to analyze networks in their most abstract form and not for spatial data. In other words, there is no notion of a point or metric with this particular package. A lot of our preliminary work involved finding a suitable representation of our longitude and latitude coordinates in a way that was compatible with this package while also being able to refer to each node with its particular geographic location. This wasn't as direct as simply mapping the longitude and latitude to a Cartesian plane because many points were differentiated by a value less than one which is understandable given that from a global perspective, many of the points are tightly packed together. To get around this we created a new mapping from the existing points to a plane in which the points were distributed in a way that would allow for easier analysis.

The second weakness of the Networkx package is that the focus on functionality meant that this package doesn't have robust methods for visualizing networks which will be important for getting some intuition about travel patterns as well as for presentation purposes. In this case, there is existing software to create attractive visuals.

4.2 Preliminary Results

We implemented some very preliminary code to check the feasibility of our methods and goals. Figure 4.1 is based off an even 10x10 grid-like division of nodes based on their latitude and longitude and 75 data points. It shows that from the first 100 points, there is one popular spot in particular where many Ubers seem originate from. However, obviously much has to be done to better cluster the geographical locations and eventually analyze the entire data

set. Its possible that we should zoom in on the area with a lot of pick-ups and that the geographical range is too large due to outliers. We should also work on finding a visually meaningful way to represent the edge weights.

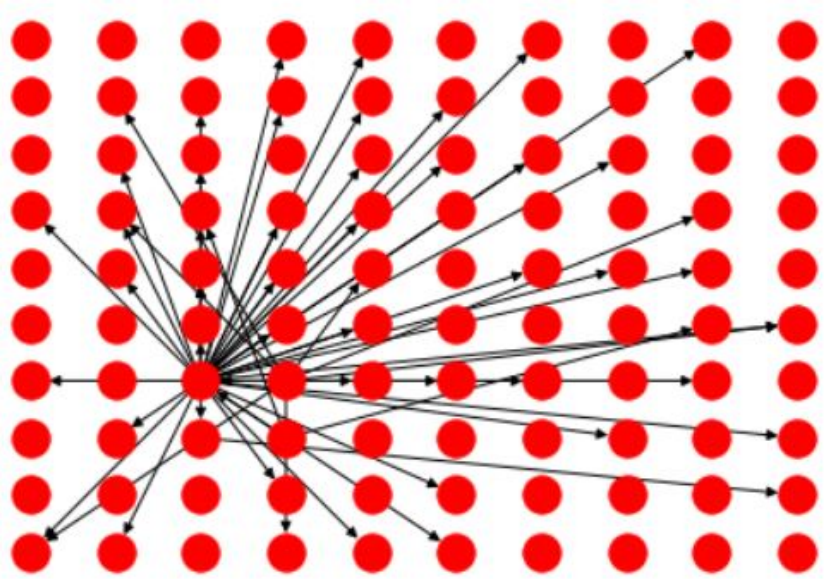

Figure 4.1: Preliminary network of 75 samples from Uber data spread over a 10x10 grid

5 Next Steps

After researching several different methods of analysis and adjusting the feasibility and scope of our project goals, our next main goal is to model the networks in a more accurate way. This will involve:

- Getting smarter latitudinal / longitudinal ranges of NYC neighborhoods and mapping our nodes onto those
- Visualizing the network on a geographical map of NYC, along with edge weights
- Processing the entirety of the Uber and Yellow Taxi datasets
- Conducting geographical analysis among networks
- Averaging number of trips over time period, rather than calculating a raw total
- Conducting temporal analysis among networks
- Draw conclusions based on observed patterns (if any) and evaluate against similar existing studies to provide guidance for further network analyses, if time permits