Quantifying the relative impacts of New York's L-train closure

Sebastian Bana, Fernando Melchor, Patrick Gitundu, Ian Stuart
NYU Center for Urban Science and Progress
New York, USA

***Abstract*—This report presents an analysis of the potential impacts of the closure of New York City's L Train. We identify the affected areas and characterize them using socioeconomic indicators. We then quantify the impacts on commuters in terms of likely travel time increases, and investigate whether certain communities will be disproportionally affected by the closure.**

Keywords*—*L Train; MTA; public transportation; social impact

I. Introduction

The L-train (14th Street-Canarsie Local) of the New York City Subway is slated to be shut down in early 2019 for a period of 18 months [1] in order to fix damage incurred during Hurricane Sandy. The closure of this line will significantly impact commuters and other riders of the L, who will need to find alternate transportation services. It will also affect businesses positioned at or near L-train stops and overall economic development in neighborhoods served by the line. In order to mitigate the adverse effects of the closure on citizens, the city needs to plan appropriate changes to other transit services. Increased bus service, using a combination of new routing and additional vehicles, will be needed to compensate for the L-train closure, particularly in less transit-dense areas of Brooklyn.

In this study, we developed a data analysis process to measure the impact of the closure and to help the city prepare. This process identifies and characterizes the affected areas using socioeconomic indicators, and approximates relative increases in commute times using the most frequent commuter destinations in Manhattan. The outputs of this process can likely help the city make more strategic decisions about where and how to organize alternate transit services, with the ultimate goal of achieving a minimum threshold of acceptable service at minimal added cost.

II. Literature Review

The main type of publications researched during this study were similar studies already performed by either the city or the academia [2]. The following where the main research points

A. General Information regarding the L train closure:

This included the information existing both the closure other interesting data that might further help deciding the type of data to use and analysis to perform.

B. Neightbour charactersitics.

Reports presenting relevant facts in relation to the characteristics of the neighbor that contributed to make decisions such as the selection of the socio-economic indicators used to characterize the area of impact.

III. Data and Methods

We used a number of publicly available data sources to build an integrated dataset containing the key variables for our analysis. We also used several open source APIs from Mapzen to help us define the affected area and calculate the results of alternate routing. These data sources and APIs along with their specific uses are described below:

A. US Census Bureau: 2010 US Census [3]

We used the 2010 US census socioeconomic-indicators to further characterize our units of analysis. The reporting unit of the census is the census tract. We selected a subset of three indicators to characterize the zip codes: Median age, Median Household Income and Percentage of White Population (As a proxy to measure ethnic diversity).

B. LODES (worker origin-destination data) [4].

The US Census Bureau maintains the Longitudinal Employer-Household Dynamics (LEHD) data products, which can be used to research and characterize workforce dynamics for specific groups [1]. We used the LEHD Origin-Destination Employment Statistics (LODES) data to capture commuter patterns for potentially affected areas.

C. Mapzen APIs [5]

We used Mapzen's Isochrone service to generate bounding polygons representing areas within a 15-minute walk of L-train stations. This area was used to determine which census tracts should be included in the initial study area. Mapzen's Turn-by-Turn API was used to generate routing data between study area tracts and associated commuter destination tracts in Manhattan.

D. MTA Data [6]

Spatial data made available by the MTA (including geometries for subway lines and stops and station turnstile data) were also used.

Using this data, we took the following approach to evaluating the impacts of the L-train closure:

A. Definition of the study area

We were interested in quantifying the impacts on commuters who use the L-train to commute to Manhattan. Using the 15-minute isochrone (a 30-minute isochrone was also generated), the study area was defined as those census tracts whose centroids fall within the bounds of the isochrone, excluding Manhattan census tracts. The study area was then further clipped to include only census tracts along the L between the Bedford and Broadway Junction stations (Figure I), under the assumption that the number of Manhattan commuters would be smaller toward the end of the line. People using L-train stops from Broadway Junction to Canarsie Rockaway Pkwy also have the option to use the 3, A-C or J-Z lines to commute to Manhattan.

B. Socioeconomic analysis of the study area at the census tract level

1) Census data: From the three variables used to characterize the area, we only found significant disparities for the geographic distribution in terms of ethnic diversity: The are clusters of census tracts that are predominantly white (more than 70% of the population) in the area of Williamsburg, at the beginning of the Brooklyn segment of the line, vs. less than 40% of white population, towards the end of the line. When we look at this distribution and we merge these results with the ones on the map with the percentage increase in commute, we realize that the most affected commuters are on their majority white.

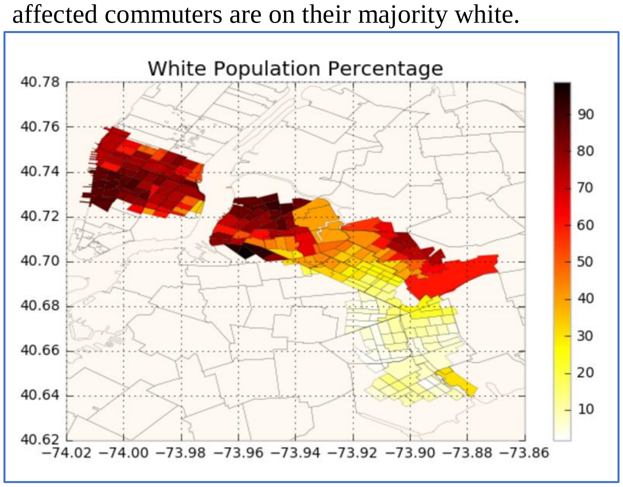

Figure I: L-Train Affected Area. Census Tracts white population percentage.

2) Turnstile data: Turnstile data was used to find the stations with the highest ridership numbers. This data is key in order to forecast the demand for alternate services in particular segments of the affected area. The following map shows the affected area along with the stations number of weekly weekday riders.

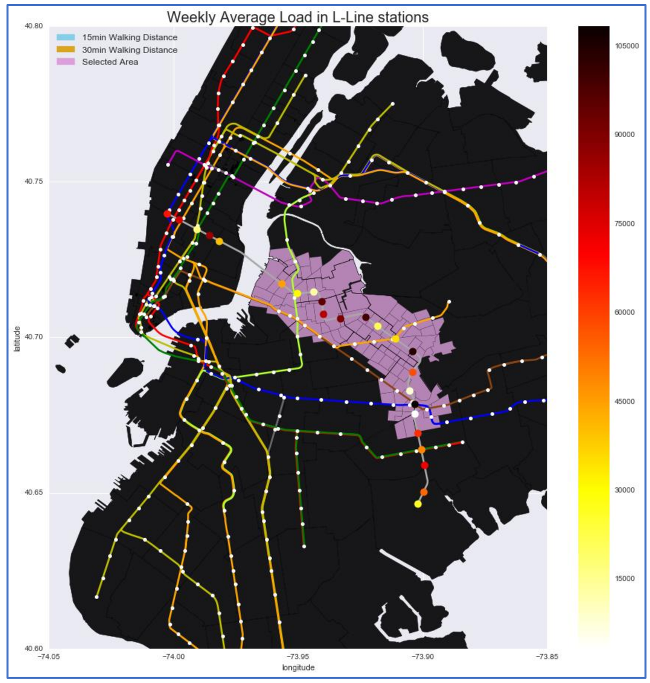

Figure II: L-Train Affected Area. Census Tracts within 15-minute walk of L-train and station ridership.

3) LODES data: LODES provide the most complete picture of commuter dynamics that is available, offering counts of commuters by origin-destination census block pairs. While some jobs are not accounted for, the data is generated from various federal records and not from a survey, and provides a reasonable estimation of relative and absolute commuter numbers (Figure III). The Mapzen API limits make retrieving alternate route times for each destination tract in Manhattan a prohibitively lengthy process, so using the LODES dataset (aggregated to the census tract level) we extracted the top ten destination tracts, in terms of number of commuters, for each origin tract. This new dataset then became the input for route generation.

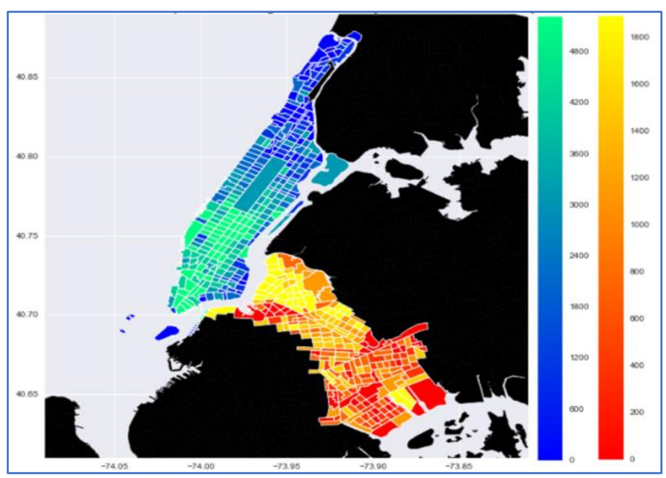

Figure III: L-Train Affected Area. Census Tracts within 15-minute walk of L-train and station ridership.

C. Alternate route data generation for representative commutes

The Mapzen Turn-by-Turn API allows for the generation of the fastest route between two points using any combination of public transit modes. It also allows for the exclusion of specific transit lines or stops. Using these features, we were able to generate a pair of routes for each origin-destination census block pair: one route assuming the L-train is operational (though not requiring that it be used, if another route is faster), and the other explicitly excluding the L. Therefore, for each tract in the study area, 10 route pairs were generated – one pair for each of the ten most populous destination tracts (Figure IV).

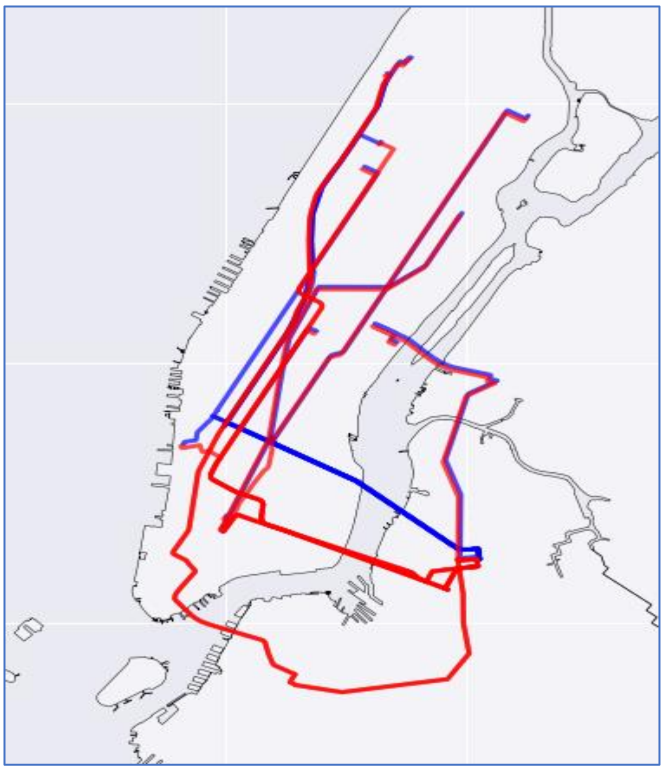

Figure IV: Alternate Route Generation Sample: Blue Lines include the L Train, while Red Lines Exclude it.

D. Identification of the most impacted census blocks based on average commute time increase

The difference in trip duration for each route pair was used to calculate a percentage change in travel time for each origin-destination pair, which was averaged over the 10 pairs for each origin tract to get a single value quantifying the impact on commuters. Percentage change in travel time was zero if the original fastest route did not utilize the L-train. A simple comparison of average increase in travel time was used to compare impacts across the study area. These values were plotted on a map to compare impacts by geographic location. This map is presented in the final section of this report.

All the code written to retrieve and manipulate the data, perform the analysis and build the visualizations was organized into a single github repository in a series of Jupyter Notebooks with the intention of allowing to reproduce the analysis or perform a similar analysis for a different subway line.

IV. Results

Figure VI shows the primary results of this study, which are the percentage changes in commuter travel time for each census tract in the study area, approximated using the top 10 destination census tracts in Manhattan for each origin tract. From the map we can see that there is an obvious spatial pattern to the results. Tracts closer to Manhattan and right along the L or to the north of the L see the greatest percentage change in commute times. Access to other subway lines in these areas is limited. Tracts along the southern edge of the study area see little to no impact on travel times, due to their access to other Brooklyn subway lines.

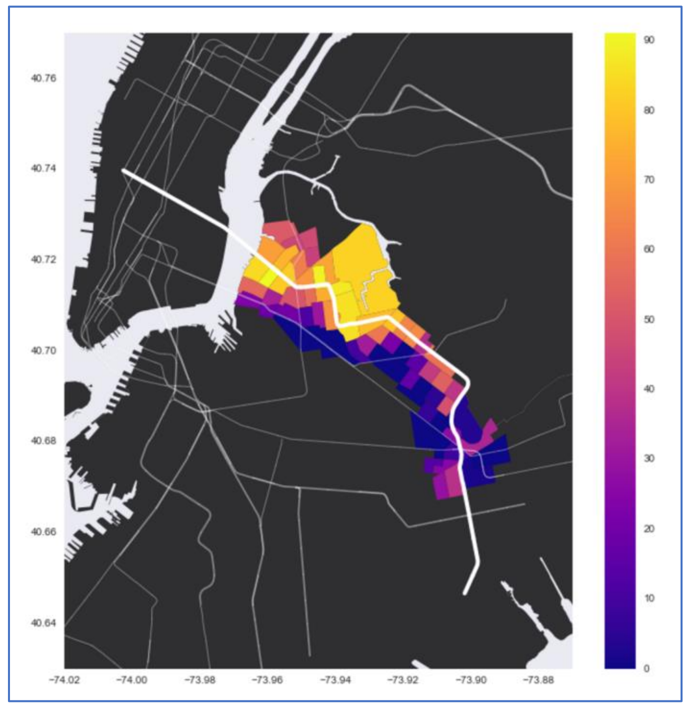

Figure VI: Percentage increase in commute time for affected census tracts.

It is clear that areas where the L-train is the only feasible subway option will be hit the hardest, which one might expect. What may not have been as obvious is that many of these areas could see their commute times nearly doubled. This impact will likely be unacceptable to many commuters and could have serious economic effects on workers and businesses. It would behoove the MTA to focus its efforts on these hardest-hit areas.

V. Conclusions and Discussion

In this study, we applied geo-spatial and statistical analysis to assess the impact of the closure of the L train in NYC population. We identified the affected area and analyzed its socio economic characteristics. Based on the information obtained from the LODES dataset and the calculations on alternate routes we were able to quantify the impact in terms of percentage of increase of commute time.

A series of tools were developed to perform this analysis. These tools enabled data extraction and manipulation, different types of geospatial and statistical analysis and visualizations.

Moving forward, an area of interest would be to looking into the implications of providing citizens of the affected census tracts, with alternatives to improve their commutes and the impact of these alternatives on the other public transportation services.

VI. REFERENCES

[1] MTAs Press Release: L Line's Canarise Tunner Closure. July 2016. Consulted March 2017. http://www.mta.info/press-release/nyc-transit/l-line%E2%80%99s-canarsie-tunnel-close-2019-18-months-undergo-massive

[2] Mitchell L. et Al.: L Train Closure and Mitigation. NYU Rudin Center for Transportation. August 2016. Consulted March 2017. http://wagner.nyu.edu/rudincenter/wp-content/uploads/2016/08/NYU-Rudin-L-Train.pdf

[3] US Census Bureau: 2010 US Census. Accessed through https://www.census.gov/2010census/data/. Consulted March 2017.

[4] US Census Bureau: 2010 LEHD Origin-Destination Employment Statistics (LODES). Accessed through https://lehd.ces.census.gov/data/ . Consulted March 2017.

[5] MAPZEN: ADP Documentation. Accessed through https://mapzen.com/documentation/overview/ Consulted March 2017.

[6] MTA: MTA Developer Resources: MTA Turnstile Data. Accessed through http://web.mta.info/developers/turnstile.html Consulted March 2017.