The Characterization of Congestion in Cities Using Uber Movement Traffic Data

Sam Ephron, Saray Shai

Department of Computer Science
Wesleyan University, Middletown, CT 06459

Introduction

Congestion leads to massive changes in efficient traversal of road networks. When looking at edge weight as the time it takes to travel a road, traffic causes road networks to behave in a non-euclidean fashion. We used data from Uber movement to investigate congestion in Seattle, New York City, and Manhattan. Uber provides speed data for individual road segments. Each road only has a speed associated with it if there were five or more trips within the hour. The given speed is then the average of those trips. We used data averaged over the second quarter of 2019. We also developed a network-wide metric, C(h), to evaluate congestion on an hourly level within each city. This allowed us to find and distinguish between different states of the network. As the data from uber movement did not encompass every road in each city, we could use C(h) and speed distributions to assess the accuracy of a filling algorithm we created to give a sense of congestion across the entire city, not just only the roads we had data for.

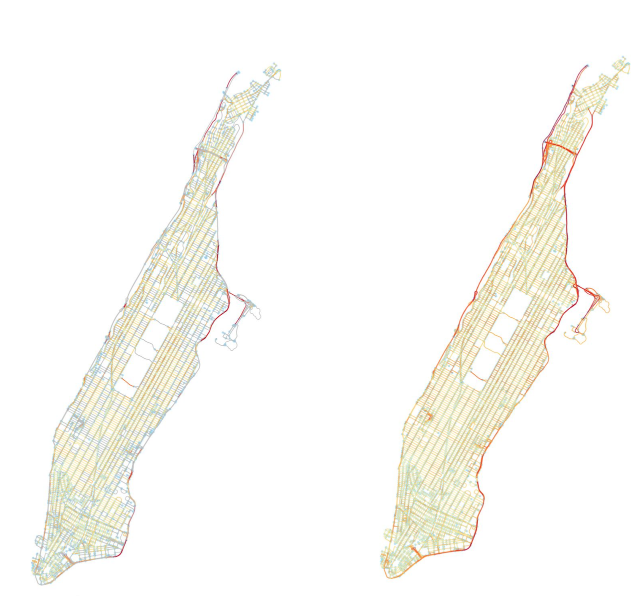
C(h)

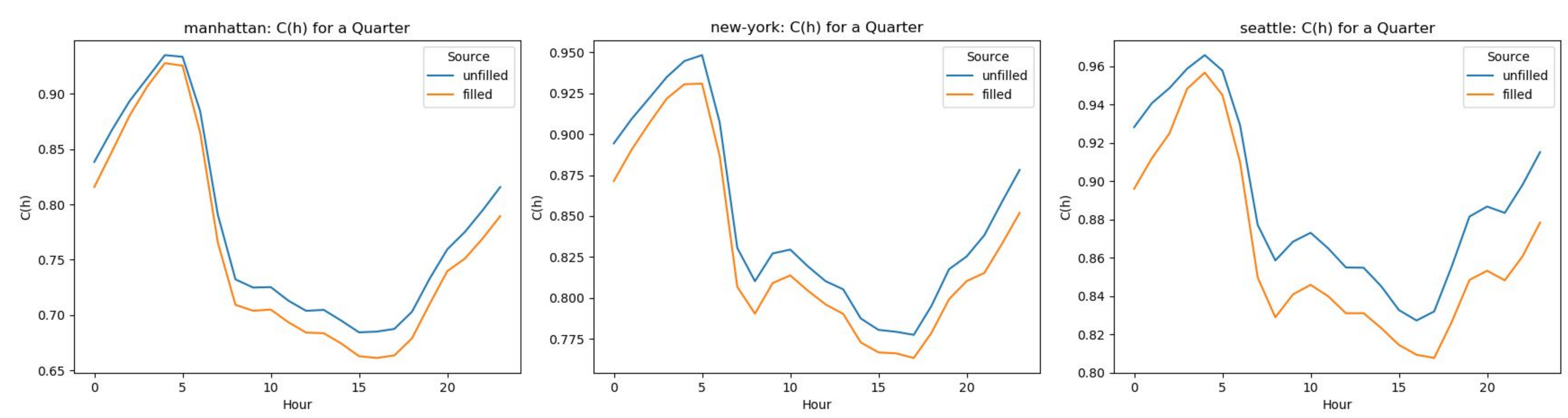
v_i(h) = speed on road i on hour h
x_i(h) = v_i(h)/max(v_i(h))
C(h) = sum(x_i(h))/n; where n is the number of roads represented

Thus, C(h) gives a sense of congestion on a network-wide level. While x_i(h) allows for a sense of congestion on individual roads.

Fill Algorithm

For each hour, we started with a queue of all known roads. Then, for each known road, we made a list of each unknown road going into or out of that known road. This guarantees that each of those unknown roads is adjacent to at least one known road. Every unknown road in this list is assigned the median speed of the adjacent known roads, and then added to the queue.

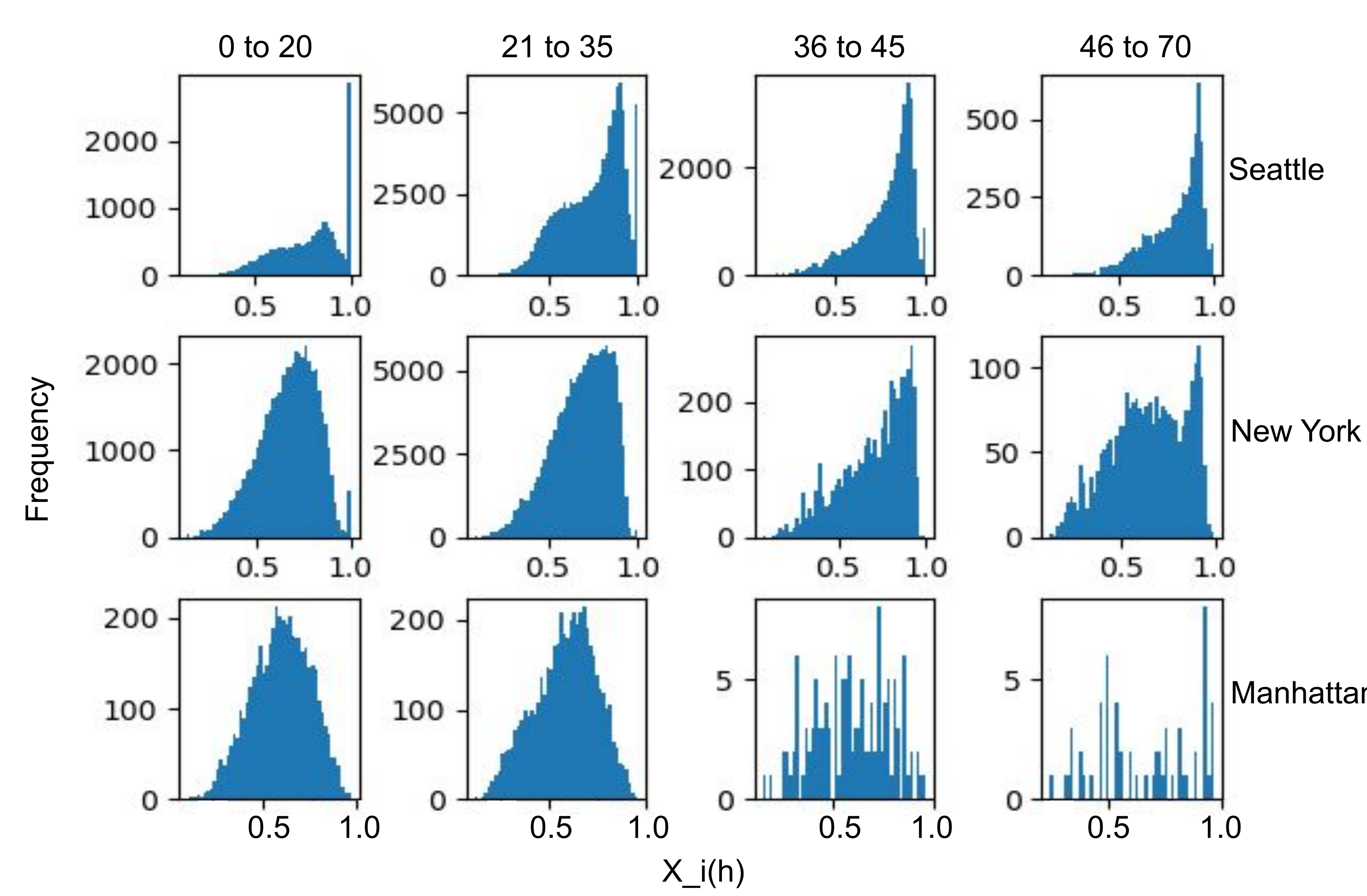




◢ These three figures show the filled and unfilled C(h) for the quarterly data.

◣ These are colored speed maps of each city using the filled quarterly data. The top image for each city is at 5:00 am, and the bottom at 4:00 pm

▼ This figure shows all the x_i's for each road binned roughly by speed limit for each city. The two fastest bins for Manhattan suffer from a lack of data, but otherwise it shows the similarities across speed limits.



Conclusion

Congestion follows relatively consistent trends on a network-wide scale. The levels of congestion are relatively similar, even across different types of roads. Additionally, the fill algorithm is able to complete the network using the quarterly data without compromising the network-level congestion. The next step is look at how the differing levels of congestion affect travel through the network, and look at data of individual days instead of averaged quarterly data.

Acknowledgements

Uber Technologies, inc. (2020). Quarterly Speeds Statistics by Hour of Day (Q2 2019) [Data file]. Retrieved from https://movement.uber.com/?lang=en-US