

[DRAFT]

On-demand public transport is making us mobile

RA

2021-06-18

1 Introduction

The TLC trip record data¹ records Yellow and Green taxi and other “for-hire vehicle” trips in New York City. In the earlier years in particular, Yellow and Green taxi records contain timestamps, trip distance, passenger count, fare, etc., but also approximate locations of pickup and dropoff. Based on those data we give a partial answer to the question

How many small buses could serve the same demand?

Conventions. The number in the margin refers to the corresponding code listed in §5.1.

2 Data preparation

2.1 Taxi trips²

We focus henceforth on May 2016 where ~ 12 M (Yellow) and ~ 1.5 M (Green) trip records are available. The ~ 11 M “for-hire vehicle” records bear no useful details for our purpose. We keep only the trips that begin and end in [Manhattan](#) with reported trip distance between 0.1 and 30 miles. We filter out records that lack geo-coordinates. See Fig. 1 for a summary #1
of the remaining trips.

2.2 Road graph²

We obtained the road network for Manhattan from the OpenStreetMap Overpass API³ and #3
filtered for roads that can plausibly sustain public traffic. It is represented as a digraph, #4
i.e. there are one-way roads and the routing $A \rightarrow B$ differs from $B \rightarrow A$. When modeling

¹<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

²The codes for this section were mostly written in 2019.

³https://wiki.openstreetmap.org/wiki/Overpass_API

individual trips, their reported pickup and dropoff locations are snapped to the nearest node of the road graph; we ignore about 10% of the trips where the discrepancy is over 20 m.

The map graphics are from MapBox.

#5

2.3 Traffic model

The trip trajectories are not available, only the pickup and dropoff locations (with potential GPS uncertainty of 10 to 100 m). We leverage the reported trip duration to infer plausible mean-field travel velocities (see Fig. 2c) throughout the road graph iteratively as follows. The velocities on all roads are initialized to 5 m/s. For a sample of trips that start at 18-19h the quickest trajectories are estimated. The velocities of the road bits participating in those trajectories are adjusted toward the reported trip duration. This is repeated.

#6

The adaptive routes we calculate below are the quickest trajectories w.r.t. this traffic model.

3 Optimization problem

We are facing a so-called vehicle routing problem with these main attributes:

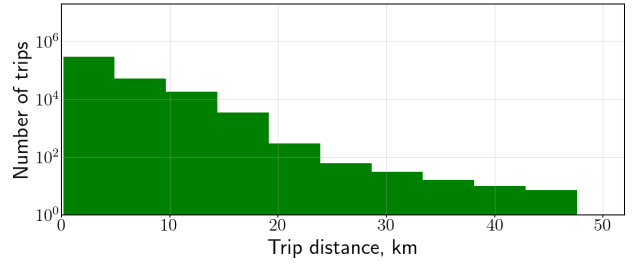
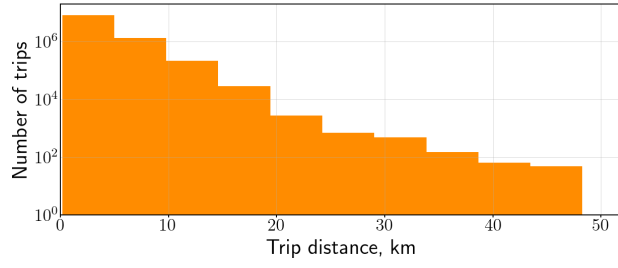
- Capacitated. There are N buses of maximal capacity of C passengers each.
- Time windows. Each passenger has to be picked up within $[-2\text{ min}, 5\text{ min}]$ of the recorded pickup time in the trip data (§2.1). The dropoff time window extends to 10 min after the recorded dropoff. To ensure feasibility, a passenger may be ignored at a certain penalty to the optimization objective. The buses are allowed to wait up to 10 min at any location.
- Depot. All buses start and finish at a certain location but have enough time to reach anywhere without compromising feasibility.

This is difficult in general. We use `ortools`⁴ to find reasonable solutions computationally. We can roughly assess optimality by allotting more time to the solver. We compare “customer satisfaction” across different fleet sizes N and bus capacities C .

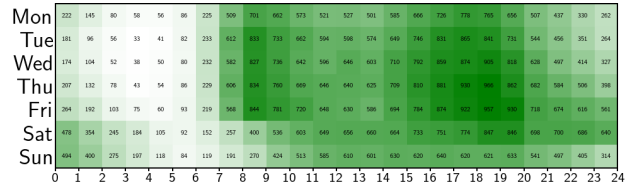
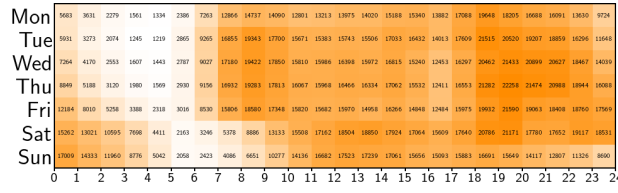
To obtain reasonable solutions on modest hardware we focus on a small slice of the trip data at a time, i.e. a few hundred passengers \times one hour \times a few square kilometers.

We pretend here that the demand and the traffic conditions are known in advance, whereas some 10 min in advance would be more realistic. Meanwhile, the density of requests in space and time is quite high. Thus, we believe our results remain informative.

⁴<https://developers.google.com/optimization/routing/vrp>



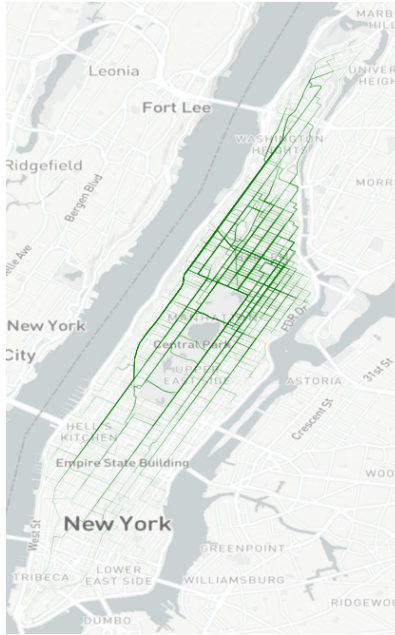
(a) Reported trip distance histogram.



(b) Pickup hour heatmap.

Figure 1: Summary of Yellow (\searrow) and Green (\nearrow) taxi trips filtered as in §2.1.

#2



(a) Sample Yellow taxi trips.

(b) Sample Green taxi trips.

(c) Inferred mean-field velocity.

Figure 2: A sample of shortest-path trajectories (§2.1/§2.2) and the traffic model from §2.3.

#7, #8

4 Case study

4.1 Times Square

We take all single-passenger trips with pickup and dropoff within $R = 1$ km of Times Square [#9](#) that are reported to start and end within 18:00–19:00 on May 1, 2016.

5 Appendix

5.1 List of codes

	page	https://github.com/numpde/optimum/blob/main/ ...
#1	p.1	code/data/20210610-NYCTLTC/a_download.py
#2	p.3	code/data/20210610-NYCTLTC/e_explore.py
#3	p.1	code/data/20210611-OSM/a_osm_download.py
#4	p.1	code/data/20210611-OSM/c_road_graph.py
#5	p.2	code/helpers/opt_maps/maps.py
#6	p.2	code/model/20210613-GraphWithLag/b_train.py
#7	p.3	code/data/20210611-OSM/e_explore.py
#8	p.3	code/model/20210613-GraphWithLag/b_train.py
#9	p.4	code/work/20210616-OPT1

TODOs: