ORIE 5129: e-Logistics
Homework Assignment 1

Due Friday February 26 at 11:59am Eastern. (Note: Just before lunchtime, Eastern time!)

Please submit a single PDF document formatted to print and show all your work clearly. Feel free to scan and submit handwritten work. Do not spend too much time on wordprocessing your answers. Please also submit any code.

Question 1

In this question, we are going to work with travel-time data in New York City. There are two files you will work with. The first file is "nyc_nodes.csv", which tells you the name, latitude/longitude (latlong) and Mercator coordinates of the points in which we are interested. The second file is "nyc_links.csv", where each row corresponds to a road segment and tells you the start node, end node and average travel time of that road segment from 7pm-12am on a weekday. For simplicity, we assume all roads are two-way.

(a) Compute the length of each road segment. (You can use the haversine() function in python to compute distances when you have latlong. There is no need to report the length.) Generate a scatter plot where the length of the segment is on one axis and the time required to travel the segment is on the other. Fit the length-time data to a straight line and report its slope. According to your observations, what is the unit of time used in the data? Justify your answer.

(b) Color each road segment on the map by the speed on that road segment. (You can work from the sample code named "map.py." It is ok to not provide a legend, but be sure to explain in words which color corresponds to which speed range) and submit the graph. Also looking at the scatter plot you drew in part (a), does anything look strange? What? Should we adjust these strange data points? Justify your answer. (There is more than one correct answer to this question). If your answer is yes, adjust the strange data points by adjusting the travel time and plot the adjusted length-time scatter plot.

(c) Ambulances usually travel faster than regular traffic speeds when traveling at "Lights and Sirens" speeds on their way to an incident. Let us assume these travel times are 0.9 times the travel time we have in the data for regular traffic. Compute the travel time for the ambulances on each road segment. Don't hand in the adjusted travel times; just show us your code.

Question 2

In this question, we want to model how emergency calls for ambulance service, as received by the Fire Department of New York (FDNY), are distributed across the city. The way we model these calls is not accurate, but should be enough for our work in this class. (True data would require us to sign a non-disclosure agreement.) We will use "nyc_population.csv", which tells you the population in each region within NYC (not the whole city) and the label of a representative point matching that region.

(a) There were $1,531,870$ calls in total in 2019. Assuming that the calls arrive at a constant rate per hour (they don't; typically they are concentrated during the day with peaks in the morning and afternoon), what is the number of calls per hour across the entire city?

(b) We divide NYC into a little more than 1000 different regions and assume that the probability that the next call comes from a particular region is proportional to the population of the region. (In reality they're more likely to come from poorer areas.) Use "nyc_population.csv" to compute the probability that the next call comes from each region. (Assume the probability sums up to 1 exactly). Don't hand in your probabilities; just describe how you got them.

(c) Generate a map where each region is indicated on the map at its representative point, colored to indicate the region's population. (You can work from sample code named "map.py." It is ok to not provide a legend, but explain in words which color corresponds to which population range.)

(d) Compute the smallest ambulance travel time needed between each pair of representative points. (In python, you may use single_source_dijkstra() function from package networkx. For the first argument of this function, consider constructing the graph by from_pandas_edgelist() function from package networkx). There is no need to report these travel times, but do report the travel times from Grand Central Station to the site of the World Trade Center, and also the travel time from Cornell Tech to Columbia University.

Question 3

Solve the integer program for figuring out the smallest fraction of the popular vote that leads to a win in the Electoral College as discussed in class on Feb 9. You'll need to get information on the population/voting population and the number of electoral votes for each jurisdiction from somewhere. Cite your sources and explain any adjustments you make to the data to get your inputs for your optimization. Compare the answer you obtain to the estimate of 0.25 found in class. Explain any difference. Would modeling Maine and Nebraska's breakdown by congressional districts make any difference? (There is no need for calculations on this last point but explain your answer.)

Question 4

The National Basketball Association (NBA) championship is a series of up to 7 games between 2 teams. The first team to win 4 games wins the series. Suppose that the probability that Team A beats Team B in any given game is p, independent of all other games. Find a formula that gives the probability that Team A wins the NBA championship in terms of p, and plot it as a function of p. Explain the intuition behind the shape of the plot. Also, explain why your formula equals $P(X \geq 4)$, where X is a binomial random variable with parameters 7 and p. Comment on how this model might be extended to capture the reality that the game locations move back and forth between the home cities of the two teams (but do not implement your extension).