Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

DDS FAU

Project Seminar: Data and Cases Information

SS 2024

Frauke Liers, Jorge Weston, Arthur Miehlich
Friedrich–Alexander Universität Erlangen–Nürnberg, Department of Data Science

Test cases

- The data set consists on speed forecasts using real data collected by bus drivers in the city of Chicago.
- The network can be represented as a graph with 538 nodes and 1287 edges.
- Each edge is associated with the coordinates of its start and end node, its weight, and its name.

Data Files

The data set consist on 3 .csv files:

- **data-all.csv**
- **graph.csv**
- **route-all.csv**

General Network Information

File: graph.csv

This table consists on the following columns:

- Edge number
- Edge name
- Node A
- Node B
- Geographic X coordinate of node A
- Geographic Y coordinate of node A
- Geographic X coordinate of node B
- Geographic Y coordinate of node B

Instances Information

File: data-all.csv

This table contains the information of the different instances.

- The first row contains the **edge names**, and three empty columns at the end.
- Each of the subsequent row represents a different instance.
- For each instance the weights of the edges can be found in their respective column. Additionally, the last three columns are the day of the month (1-31), the month (Mr, Apr, or Mai), and the time stamp (hh_mm_ss), respectively.
- The edge's weight represent the speed that is estimated to be used in the edge, as it is measure as mph.

Route Information

File: route-all.csv

This file contains the route used in each instance. Each row represents an instance, and contains the sequence of nodes that form the used route. Also, the last three columns are the day of the month (1-31), the month (Mr, Apr, or Mai), and the time stamp (hh_mm_ss), respectively.

The route to be analyzed starts in node **94** and ends in node **162**

Some considerations

- The data set need to be clean, i.e., it is necessary to check for its completeness and decide (in a justified way) how to clean or complete the data set.
- The data set needs to be separated into two.
  1. **Training set**: The part of the data set that is considered as historical data. This set is formed by all the instances whose route is known.
  2. **Test set**: This part of the test set include all the instances whose routes are still unknown, and therefore, the instances that need to be optimized.

Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

DDS FAU

Project Seminar: Data and Cases Information

SS 2024

Frauke Liers, Jorge Weston, Arthur Miehlich
Friedrich–Alexander Universität Erlangen–Nürnberg, Department of Data Science