README.md - Grip

Adjacency Matrices

These instructions are pertinent to the **programming project** of IS421: Enterprise Analytics for Decision Support, Semester 2, 2016.

This folder contains the adjacency matrices for the **Singapore road network**. Specifically, the following **two** CSV files that represent the adjacency matrices are contained:

- **edge_idx_matrix** -- adjacency of the nodes are represented by the non-zero edge indices (i.e., positive integers) between all pairs of adjacent nodes.
- **travel_time_matrix** -- adjacent nodes (i.e., the edges) are represented by the non-zero travel time (in minutes) between them.

The matrices are sparse (i.e., containing a lot of zeros) and of dimension 9,948 x 9,948 each. The first matrix represents the node adjacencies (i.e., the edges) via the edge indices, which are positive integers used to name the edges. The second matrix represents those edges via the (maximal) travel times (in minutes) through the road segments. Therefore, a zero entry in each matrix means that the respective row and column nodes are *not* adjacent. In addition, all the matrices are symmetrical and have zero diagonals. This means the road network is an **undirected graph**.

Each CSV file is a square matrix **with headers**. The rows and columns are exactly identical. The headers name the columns from 0 to 9947 (i.e., the number of columns - 1). This can be conveniently used to name or index the nodes. Otherwise, for most practical purposes, each header line (i.e., the first line of each CSV file) should be ignored when reading the file. Note that each of these CSV file, once expanded, can be quite large (~ 400 MB). It is thus essential that you write a program to read these files and construct an undirected graph out of them to represent the road network.

Finally, in addition to the adjacency matrices, there is an additional CSV file called **node_idx_id.csv**. This file is in fact a mapping from each **node_index** (each the row/column of the adjacency matrices) to a corresponding **node_id**. More specifically, the mapping is from the **idx** column of the CSV file to its **id** column. It is preferred to use the **node_id** (i.e., the **id** column) to name the nodes in the graph. Why is there such additional complexity? This is due to legacy code of the **Routing App v2** such that nodes must be named according to the **node_id** in order to be displayed on the map correctly. This function is *optional* and only useful if you want to display the input OD pairs and the taxis' original locations using the app. Otherwise, you can safely ignore this file (and the input function of the app.)