DATA620 Week 4 Assignment: Centrality Measures

I have an interest in urban bike-sharing systems. The data I would analyze is the open data provided by the Citibike system [https://www.citibikenyc.com/system-data].

At a basic level, the bike stations can be viewed as graph nodes connected by trip data (trips starting at one station and ending at another) represented by edges. The edges can be assigned weights calculated by the number of trips within a certain time period (per month is the obvious choice as the system data is grouped by month in ZIP archives).

Accessing the data is simply a matter of extracting the ZIP data and appending the CSV files to obtain the desired time period. Each row records a single trip, the columns containing start/stop times, start and stop station IDs, street locations, and global coordinates. Rider-specific data include sex, year of birth, subscriber type (monthly or not), and the ID of the bike that was used.

Trip data would be loaded into a dataframe with edge weights calculated from trip counts between nodes (undirected for this case), and an adjacency matrix generated. From this matrix the weighted graph structure can be generated.

The nodes have no particular categorical classification, but one can be generated from the global coordinates. I propose classifying the nodes into neighborhoods, such as ‘SOHO’ or ‘East Village’ or ‘Murray Hill’, or some other arbitrary area designation, to study the flow of traffic from one city neighborhood to another. Classifying the nodes into neighborhoods using lat/long coordinates would be somewhat involved, as another API or GIS source needs to be queried, but is conceptually straightforward.

A hypothetical outcome that could be predicted from comparing degree centrality across neighborhood groups could be the identification of node groups that have higher traffic flows for the purposes of bike supply allocation, the installation of new nodes, and the strategic rebalancing of finite bike supply between nodes with high degrees of connectedness.

Frank de Olivera