Data 620 – Web Analysis – Week 4 Assignment Centrality Measures

I. Network Analysis with Bike Sharing Dataset

By Enid Roman and Jose Rodriguez

II. Table of Contents

III. Abstract

The bike-sharing network analysis project aims to gain a holistic understanding of the city's bike-sharing program through advanced network analysis methodologies. The project encompasses a range of objectives, including evaluating transportation efficiency, studying user behaviors, assessing environmental impacts, and analyzing the economic viability of the bike-sharing initiative. The study leverages a diverse dataset that includes ride information, station details, and user profiles to create a comprehensive network model. By employing graph theory, centrality measures, clustering algorithms, and other network analysis techniques, the project seeks to uncover patterns, optimize infrastructure, and provide actionable insights for urban planners and policymakers. The project's significance lies in its potential to contribute to more efficient and sustainable urban transportation systems, promoting healthier and eco-friendly commuting alternatives.

A. Background

Bike sharing is like a modern version of renting bikes. It's super easy – you sign up, grab a bike from one spot, and when you're done, you can drop it off somewhere else. There are more than 500 of these bike-sharing programs worldwide, with over 500,000 bikes. People are really into them because they help with traffic, are good for the environment, and promote healthier living.

B. Objectives of the Network Analysis Project

1. Analysis Questions:

* What is the overall topology of the bike-sharing network, and how are stations interconnected?
* What patterns exist in user behavior, and how do they vary between member and casual users?
* How do ride patterns differ between am and pm?
* What are the most common starting and ending points for different user groups?

2. To address the analysis questions, the project will adopt the following approach:

* Understanding Network Structure:
  + Analyzing the topology of the network to understand how nodes (stations) are connected through edges (rides).
  + Identifying patterns, clusters, or communities within the network.
* User Behavior Analysis:
  + Studying user behavior based on ride types (member vs. casual), time of day, or day of the week.
  + Identifying popular starting and ending points for different user groups.

V. Methodology

A. Data Collection

The dataset contains data from the company that operates bike sharing services in Chicago city. The data contains 13 different datasets for each month for the year 2021. Each dataset contains 13 columns 49623 to 822411 rows each. Following are the columns in the dataset and what they represent:

1. Types of Data (Graph, Node, Edge Attributes)

* Attributes
  + ride\_id : the unique id to refer to each trip
  + rideable\_type : the type of the bike used for the trip
  + started\_at : date-time for when the trip started
  + ended\_at : date-time for when the trip ended
  + start\_station\_name : name of the station from where the trip started
  + start\_station\_id : unique id of the station from where the trip started
  + end\_station\_name : name of the station where the trip ended
  + end\_station\_id : unique id of the station where the trip ended
  + start\_lat : latitude of the start station
  + start\_lng : longitude of the start station
  + end\_lat : latitude of the end station
  + end\_lng : longitude of the end station
  + member\_casual : member denotes the users who have subscribed to annual membership, and casual denotes the users who haven't

2. Sources of Data

* Website where data was collected.
  + Kaggle.com
* Libraries
  + Pandas
  + Numpy
  + Matplotlib
  + Seaborn
  + Networkx
  + Community
  + Itemgetter
  + Bipartite
  + Operator
  + StringIO

3. Data Preprocessing

* Merge all 12 dataset with pandas.concat.
* Cleaning and Formatting Network Data
  + Remove rows that have missing information drop.na.

B. Graph Representation

1. Node and Edge Attributes

2. Graph Traversal Methods

C. Network Analysis Techniques

1. Centrality Measures:

* Degree Centrality
* Betweenness Centrality
* Closeness Centrality
* Eigenvector Centrality

2. Community Detection

* Modularity-based Methods
* Louvain Method
* Girvan-Newman Algorithm

3. Clustering Algorithms???

* K-means Clustering
* Hierarchical Clustering

4. PageRank Algorithm

* Application for Ranking Nodes

5. Link Prediction

* Common Neighbors

6. Routing Algorithms

* Shortest Path Algorithms

D. Network Visualization

1. Visualization Techniques

2. Tools and Platforms

VI. Expected Results

A. Anticipated Insights

B. Impact on the Targeted Issues

References

https://www.kaggle.com/datasets/nabajithazowary95/bikeshare-data?resource=download-directory