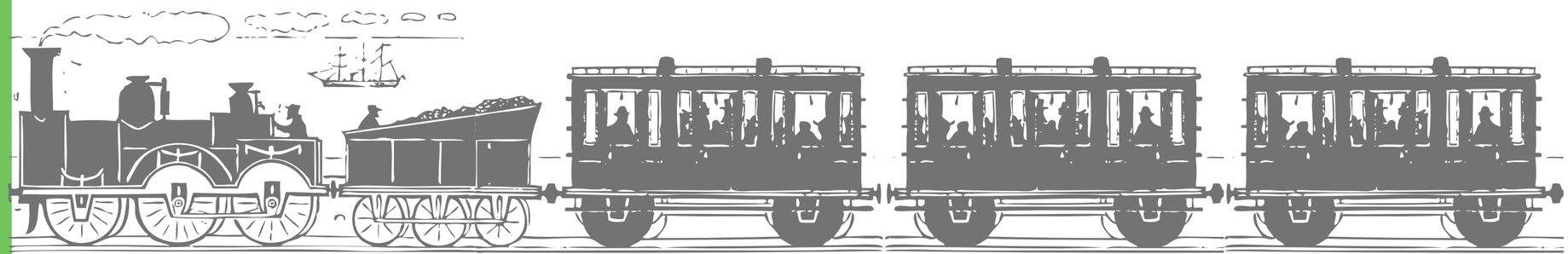
NETWORK SCIENCE PROJECT

Analysis of the Indian Railway Network

Harshita Pandey (MT19012), Rose Verma (MT19052), Suryank Tiwari (MT19019)



Introduction

MOTIVATION

- Transport networks are crucial to the economic development of a country.
- Analysis of these networks may lead to identifying patterns which increase the efficiency or reduce weaknesses and cost expenditures in the network.
- With rapid construction and investments into these ever evolving systems, need for proper analysis arises.

OBJECTIVES

- To extract properties and inferences out of the Indian Railway Network or IRN.
- To be able to query the network and draw out subgraphs and connections between regions.
- Compute various centrality measures, visualizations and network parameters.
- Arrive at inferences about the real world network.

DOMAIN

- We are extracting properties, features and inferences from a real world graph, which marks this problem to the domain of Network Science.
- Analysis and construction of graphs and networks has led to major socio-economic changes and we intend to contribute to this trend with a descriptive analysis of the IRN.
- Real world applications involve bringing down operating costs and preventing network clogging and failure.

Dataset

DATA

The data used throughout our analyses comes from the Indian Railways Timetable for trains available as on 01.11.2017 published by the Indian Government.

The dataset contains the following details for reach row: Train No, Train Name, Station Code, Station name, Timings, Distance, Source and Destination Stations.

TREATMENT

We consider each railway station to be a node.

An edge exists between two nodes if there exists a train that connects them both.

Hence, each station in a route must be connected to each other via a direct edge.

Edge Weights are considered on the basis of distance and on the number of trains between two nodes.

Methodology

GRAPH CREATION

We load the excel file, remove null values and then create directed weighted graphs based on two heuristics - train-count between two stations and the distance between the stations. Conversion to undirected graph is made as needed.

QUERYING AND VISUALIZATION

A subgraph querying and visualization system is generated to extract sub-regions of the graph and the regions between them. Information extracted for any number of nodes will contain the nodes, their neighbors and the shortest path connecting each node to every other node queried.

EXTRACTING PROPERTIES

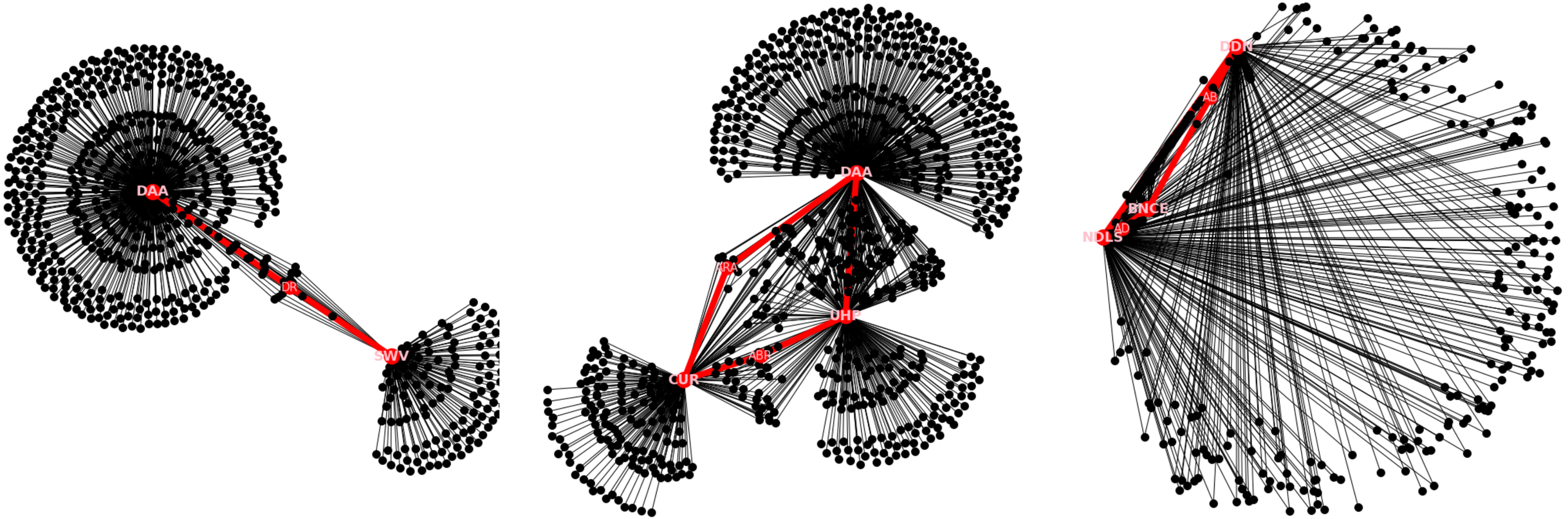
Analysis is performed to compute features like - degree centrality, betweenness centrality, closeness centrality, eigenvalue centrality, degree distributions, weighted degree distributions, degree correlations, cumulative strength distributions, clustering coefficient, path length distributions, small world properties, etc.

INFERENCES AND RESULTS

Out of the properties we arrived at our inferences to the IRN and presented the facts obtained in textual, visual, interactive plots/map form.

Querying System Visuals

RESULTS



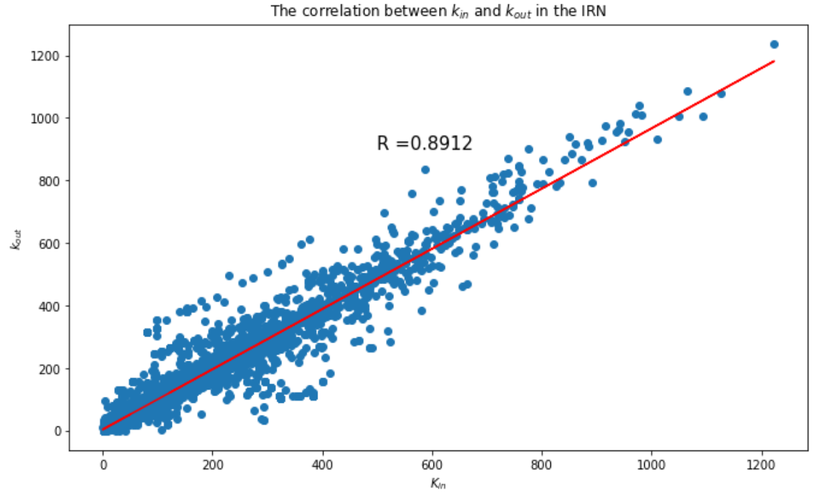
Graphs with edge weights based on number of trains between two stations. Queried nodes are printed in red with labels. Their neighbors are printed in black. Any nodes that connect the queried nodes with the shortest path are highlighted, labelled and the path is marked as red.

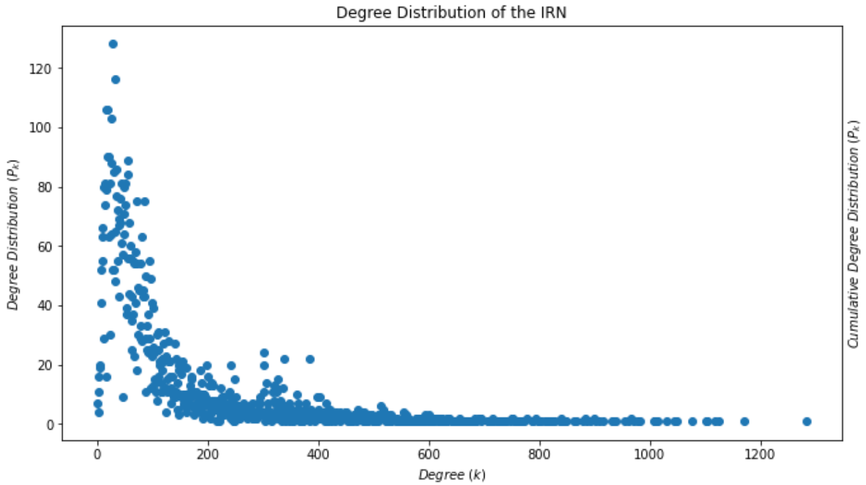
Graph with edge weight based on distance. The edge weight leads to neighbors that are far being printed farther away, rather than being clubbed in groups of number of train connections.

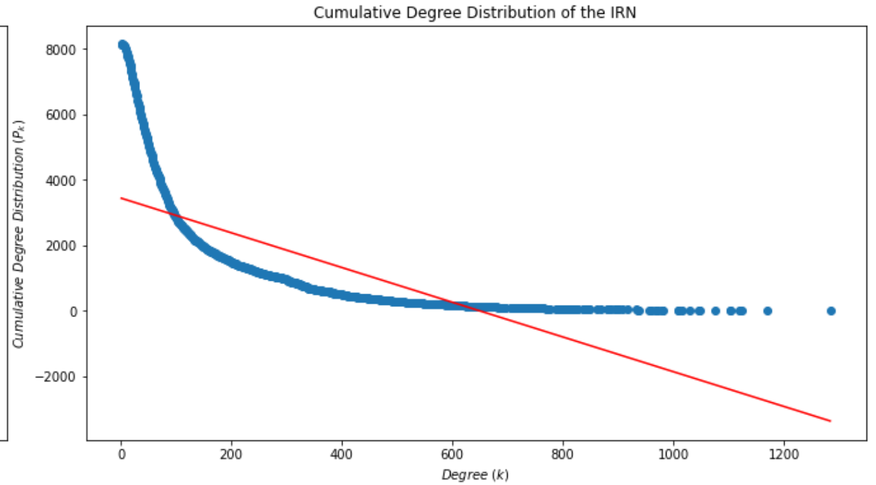
Graph property results

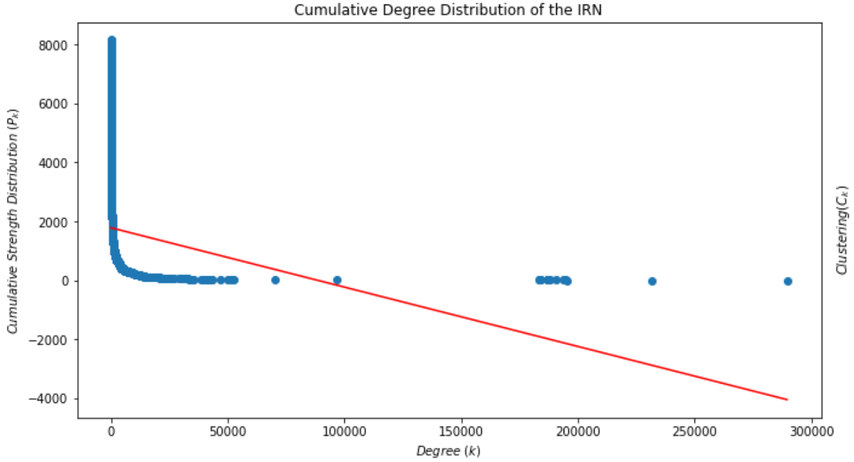
RESULTS

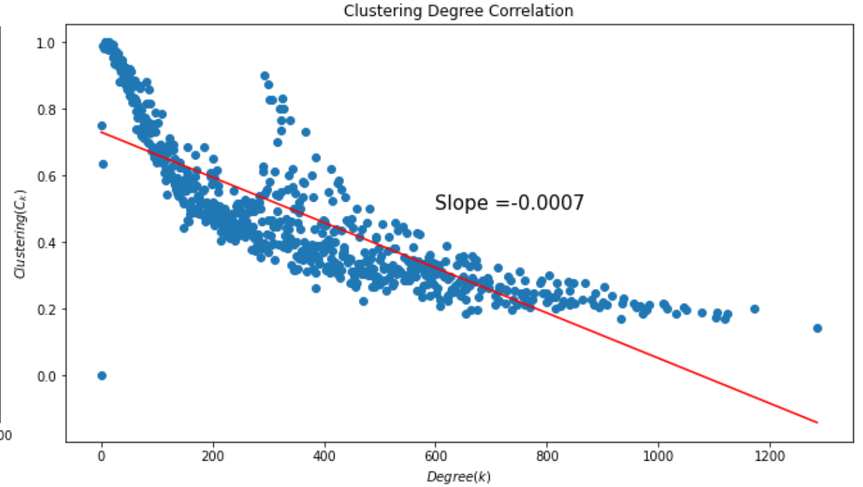
Computed degree correlations, degree distributions, cumulative degree distributions, cumulative degree correlations and path length distributions along with other metrics,

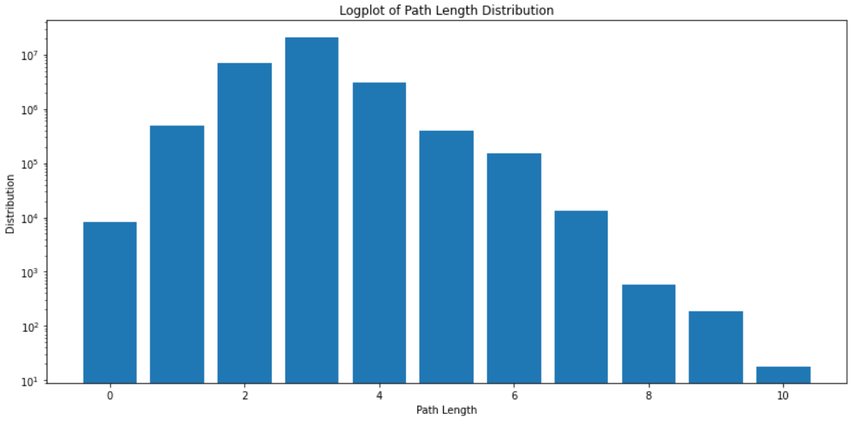






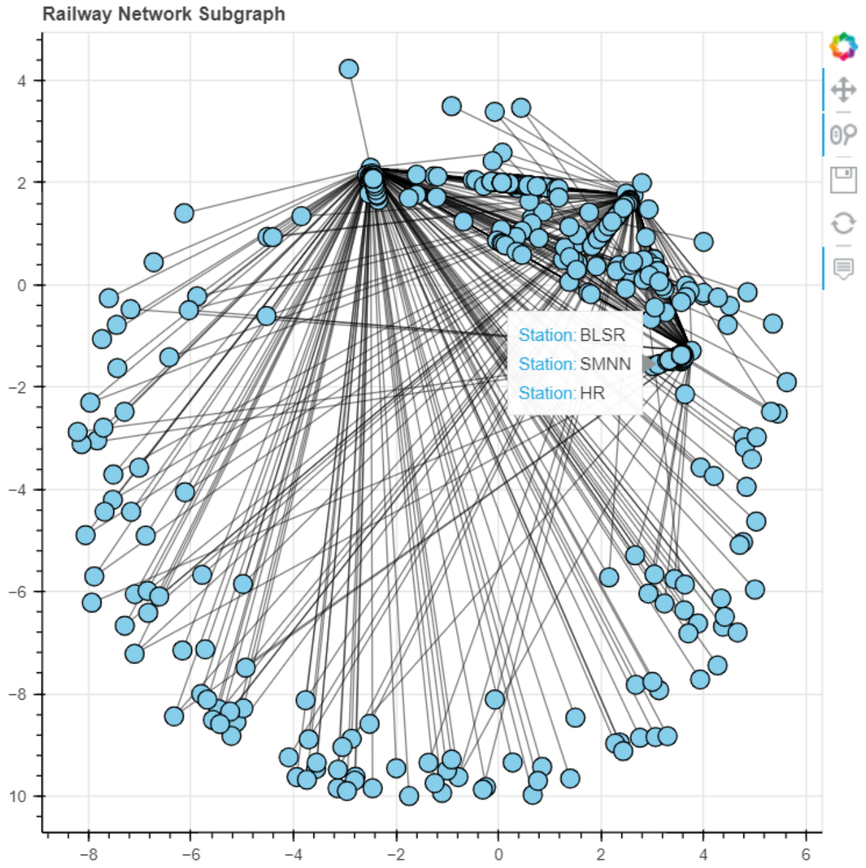


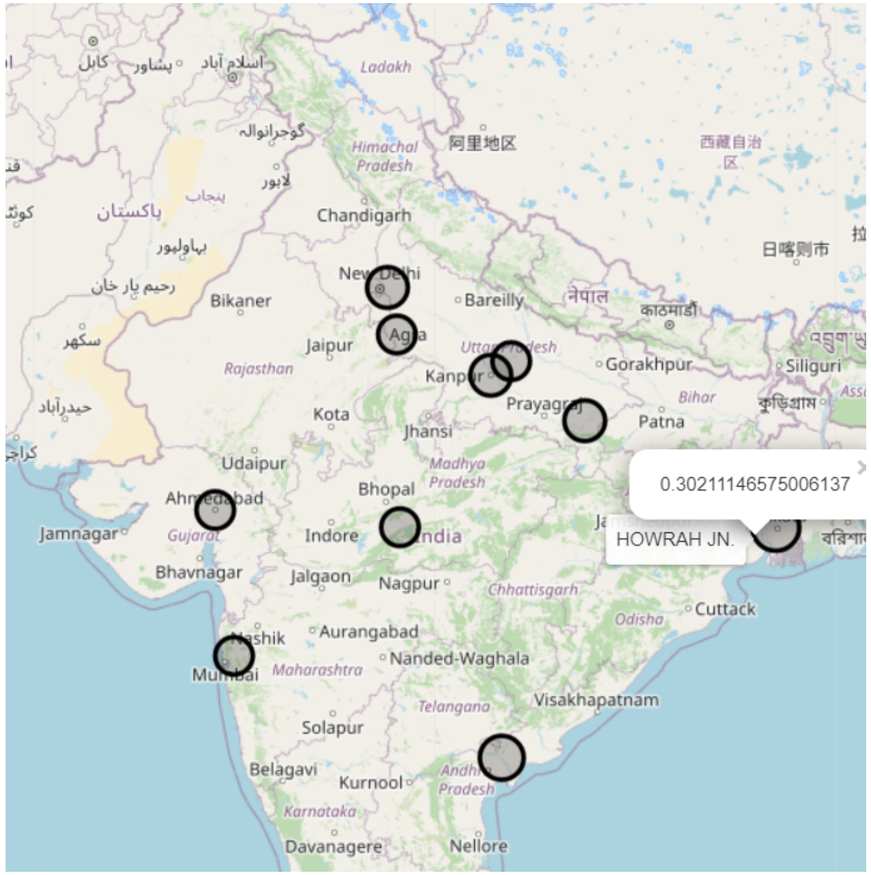




Further Visualizations

RESULTS





We added interactive plots and maps with hover tooltips to allow further study of the metrics/subgraphs that are extracted/queried.

Inferences

Most "popular" nodes:

- HOWRAH JN.
- VIJAYWADA JN
- KANPUR CENTR
- VARANASI JN.
- GHAZIABAD JN

Bridge nodes:

- HOWRAH JN.
- SEALDAH
- KANPUR CENTR
- VIJAYWADA JN
- AHMEDABAD

Most influential nodes:

- VARANASI JN.
- HOWRAH JN.
- KANPUR CENTR
- ITARSI
- NEW DELHI

Facts

RESULTS

Nodes (Stations): 8147
Trains: 7580
Edges: 902602

Average path length: 1.379
Maximum path length: 10
Diameter: 122.845
Clustering Coefficent: 0.7635

Average Degree: 122.845
Weighted Average Degree: 1513.201
Connected components: 7

Longest train route: 4260 km
Shortest train route: 1 km
Maximum distance between any two consecutive stations: 1301 k
Minimum distance between any two consecutive stations: 1 km
Average total train route distance: 439.7 km
Average distance between consecutive stops: 17.91 km

Final thoughts and Conclusion

- Inference based on Geographic Location- Stations like Vadodara Junction, Vijaywada Junction handle a high amount of traffic as denoted by degree centrality and are potential points of congestion in the network as they lack the resources as compared to the stations near metropolitan states.
- The strong indegree and outdegree correlation and the high reciprocity parameter R of 0.8912 both indicate that IRN is a symmetrical network.
- IRN is not completely connected, it has a total of 7 connected components. That is, it is impossible to reach some stations starting from specific stations in the Network.
- Maximum path length of IRN is 10, i.e. there are 10 train changes required to reach from a specific station to another.
- Comparing with the Barabasi Albert Model of same size (Clustering Coefficient: 0.02830 and Average Path Length: 2.5275), the Indian Railway Network has a high clustering coefficient (0.61295) and short average path length (1.4792), thus IRN can be regarded as a small-world network.

Novelty

- Region connectivity calculation and visualization.
- Inference analysis.
- Interactive visualization.