Network Analysis of Indian Railways

Network Science Mini Project

Presented By

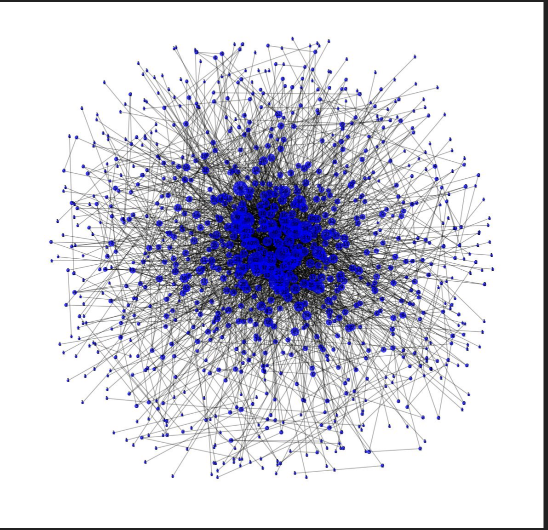
Mohammad Atif Quamar
2020523

Introduction

Transport networks are vital for a country's economic growth, and their analysis can reveal patterns to enhance efficiency and reduce costs and vulnerabilities.

Problem Statement

To analyze and visualize the Indian Railway Network Data and Identify Structural Patterns and Key Stations.



Deliverables/Pipeline

Network Creation → Network Analysis → Visualization → Infer Findings → Report Results

Network Creation

1. About the dataset

The data used for the analysis comes from the Indian Railways Timetable for trains of a specific date. The dataset was published by the Indian Government.

2. Parameters of Dataset

The dataset has 1,86,124 entries. Each entry has the following parameters:

Train No, Train Name, Station Code, Station name, Timings, Distance, Source and Destination Stations.

3. Strategy for network creation

The network was created with the following parameters:

- Source (Trains origin source station)
- Destination (Trains reaching destination station)
- Weights (Number of trains between source and destination stations)

Busy Railway Stations

PROBLEMATIC !!!!!!!

Methodology

Step 1: Data Preprocessing and Graph Creation

1. The first step was to preprocess the data to handle any missing values.

2. Removed redundant features such as Timings and Distance (not useful for us in creating the network).

3. Constructed a directed graph of railway stations using the NetworkX library. Firstly identified unique stations from the dataset to create nodes. Then, for each train, it iterates through its stations, creating edges between consecutive stations. Edges accumulate weights based on the number of trains connecting the stations.

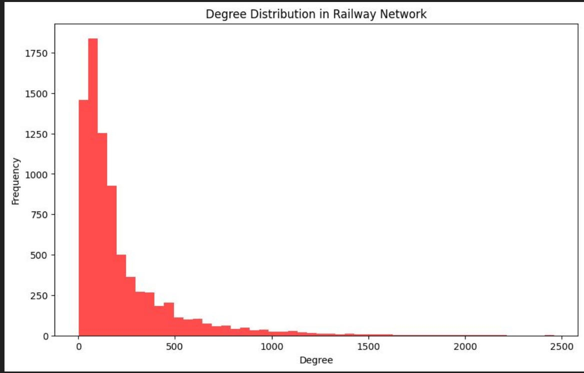
Step 2: Properties, Querying, Visualizations,

1. Found out various metrics of the network such as Degree Distribution, Centralities, CCs, Hubs, Connected Components.

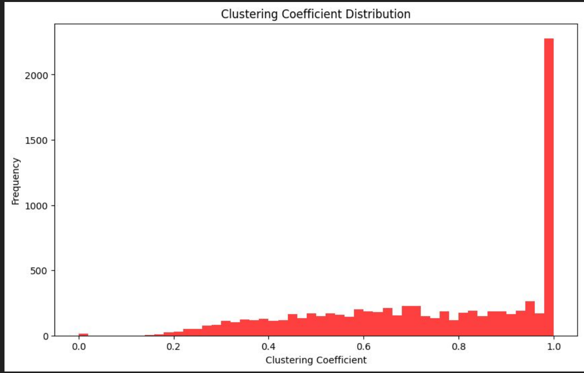
2. Visualized the network using different layouts. Visualized subgraphs of busy stations using the Fruchterman-Reingold Layout.

3. From the various properties found out, infer the results of the Indian Railway Network

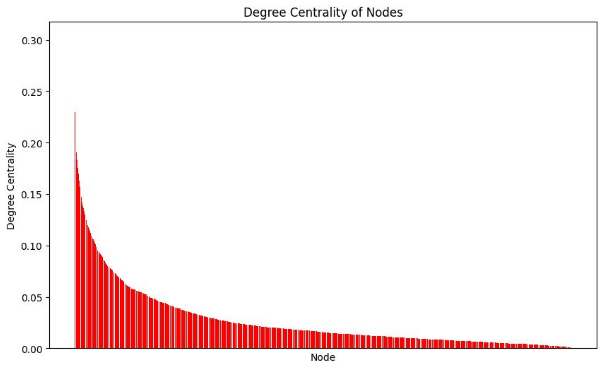
Findings



Follows power law -
Scale free distribution



Sparse Low to Mid Clustering- Indicating strong community structures



Vulnerable to targeted attacks - Some Nodes show much higher degree centrality

Statistics

Unique Source Stations = 921
Unique Destination Stations = 923
Number of nodes = 8147
Number of edges = 902602

Average degree = 110.78949306493188
Average CC = 0.73596
Most busiest station = HOWRAH (degree-wise)
Least busiest station = AJMER

Centrality Measures

	Node	Degree Centrality
0	HWH	0.302111
1	BZA	0.270439
2	CNB	0.263810
3	BSB	0.257672
4	GZB	0.252026
5	KYN	0.247974
6	ET	0.244414
7	LKO	0.243678
8	ADI	0.238522
9	MTJ	0.236312

	Node	Betweenness Centrality
0	HWH	0.034064
123	SDAH	0.021795
2	CNB	0.020931
1	BZA	0.015177
8	ADI	0.014302
82	YPR	0.014257
10	BRC	0.012968
3	BSB	0.012624
295	KOAA	0.011898
5684	PBE	0.011845

	Node	Closeness Centrality
0	HWH	0.512408
8	ADI	0.510766
10	BRC	0.506352
2	CNB	0.506288
3	BSB	0.505229
13	NDLS	0.505197
5	KYN	0.503664
14	MGS	0.503505
1	BZA	0.502141
6	ET	0.499810

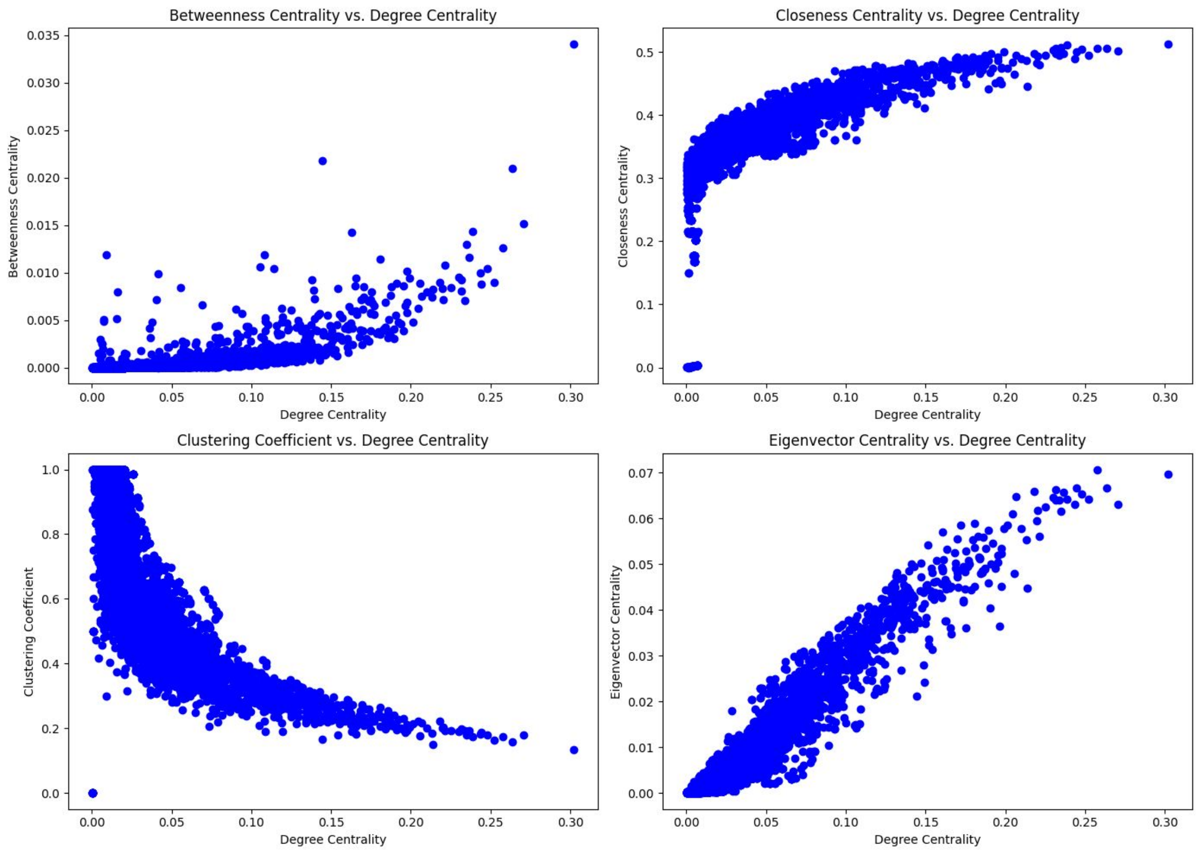
	Node	Eigenvector Centrality
3	BSB	0.070716
0	HWH	0.069651
2	CNB	0.066752
6	ET	0.066749
13	NDLS	0.066240
19	LDH	0.065925
9	MTJ	0.065690
5	KYN	0.065274
23	PNBE	0.064723
14	MGS	0.064570

- Howrah (HWH) is a critical node across all centrality measures which tells its pivotal role in the railway network.

- Stations like BZA, CNB, and BSB are also significant but tend to have varied importance depending on the centrality measure, suggesting specialized roles in network connectivity or traffic management.

- The network appears to have several key nodes that are critical for its operation, and the loss of any of these major nodes could significantly impact the overall network performance.

Centrality measures vs Degree Centrality



Network Management

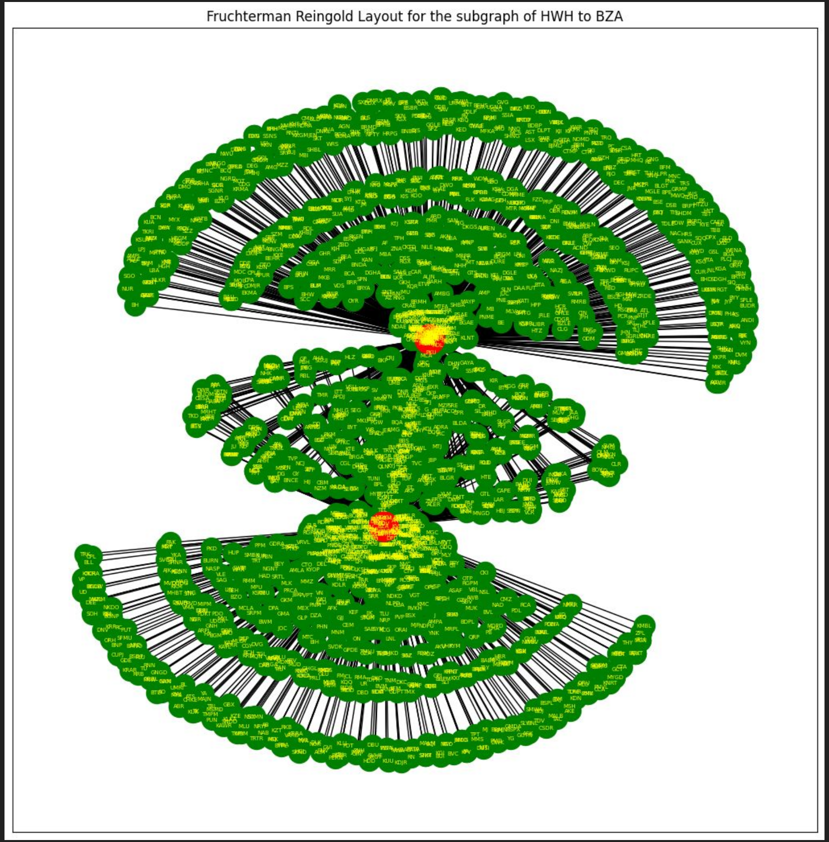
1. Risk and Resilience: Stations with high betweenness centrality are potential points of failure.

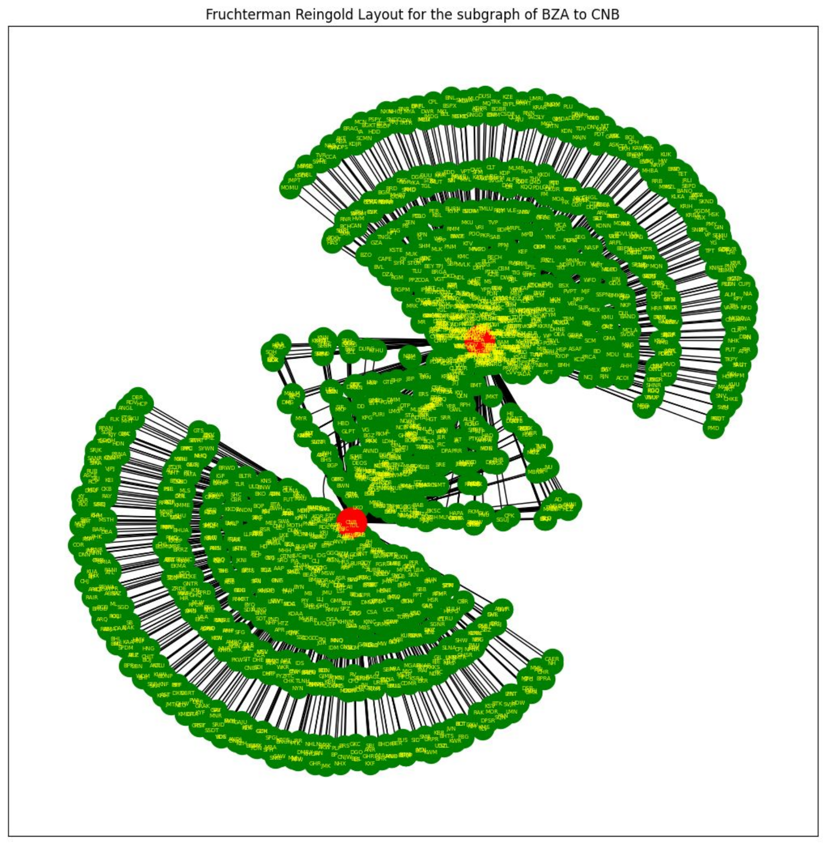
2. Infrastructure Investment: Stations exhibiting high closeness and eigenvector centrality are prime candidates for infrastructure improvements.

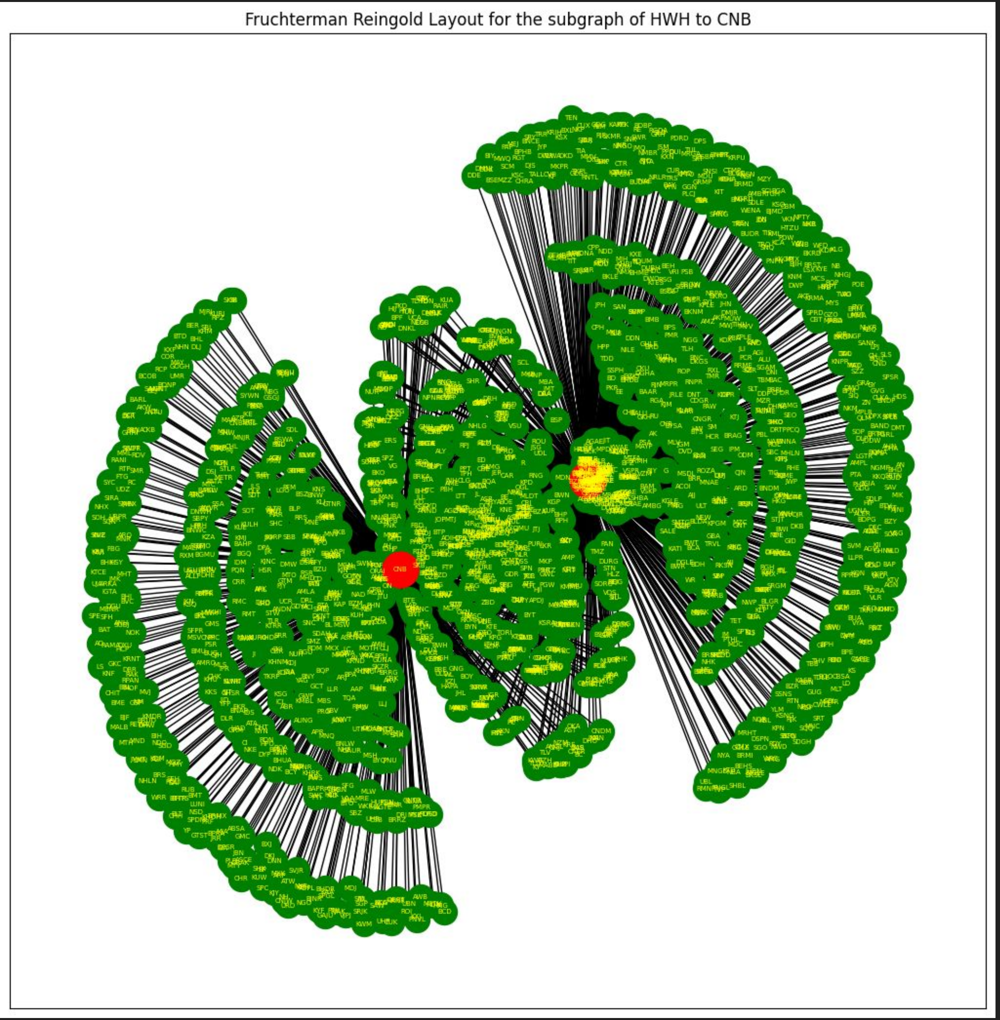
3. Service Optimization: The graphs suggests potential over-saturation at some major hubs. Optimizing the number and frequency of train services through these hubs could improve overall network efficiency and service quality.

4. Regional Services: Understanding clustering among stations with fewer connections can help in tailoring local services better and improving regional connectivity.

Visualizations (Subgraphs for Top 3 busiest stations)







Key Findings:

The top 3 busiest stations were found out to be:

1. HOWRAH = 2461 trains
2. VIJAYAWADA JN = 2203 trains
3. KANPUR CENTRAL = 2149 trains

Strategic Importance of HWH

Both subgraphs involving HWH highlight its central role in the network, underpinning its importance for both regional and long-distance connectivity.

Regional Importance of BZA

The connectivity patterns around BZA suggest its importance as a regional hub, crucial for connecting various smaller stations and facilitating inter-regional traffic.

Network Density and Resilience

The density of connections, especially in the HWH subgraphs, points towards a robust network structure with high resilience but also potential vulnerability due to the centrality of HWH. Disruptions at this hub could impact a wide range of connections.

Results

- According to the Centrality Measures of the network, stations like HOWRAH, VIJAYAWADA JN, KANPUR CENTRAL and BHARATADA can be regarded as **hubs** in the network as they handle a large number of trains. These stations can result in the busyness of the IRN.
- The degree distribution follows a **power law distribution** where a small number of nodes have high degree centrality while a large number of nodes have relatively low degree centrality. This implies that the **network is vulnerable to targeted attacks.**
- Most of the nodes have low degree centrality, which indicates that they have fewer direct connections. This might suggest that these are smaller stations or nodes that connect primarily through the major hubs.
- The Indian Railway Network has a total of 7 connected components, which implies that it is impossible to reach some of the nodes if we start from some nodes.
- stations starting from specific stations in the Network.
- The network has an average clustering coefficient of **0.73** suggesting a **high degree of clustering** and **robust local connectivity among nodes.** This indicates **strong community structures** within the network, enhancing its resilience and operational efficiency.
- This network can be classified as a **small-world network** due to its high clustering coefficient and likely short average path lengths.

THANK YOU !