# Analysis of Indian Railway Networks


S Meghana Aparna
CS14B052
Computer Science and Engg

ASP Mounika
EE14B005
Electrical Engineering

M Anuja Reddy
EE14B093
Electrical Engineering


## I. Introduction

In this report, network analysis of Indian railways has been done using various measures and metrics. Railway networks have been modelled as simple graphs and hyper graphs and both kinds of graphs have been analysed. In simple graphs, stations are represented as nodes and two nodes have an edge if they are connected by atleast one train. In hyper graphs, a set of nodes are connected if they have a common train which passes through all these stations. For analysis, the projection of hyper graphs as a simple graph is considered in which the trains are nodes and two nodes are connected if they have station in common. The data required for this analysis has been collected from Government's Catalogue[3].

## II. ANALYSIS OF SIMPLE GRAPH

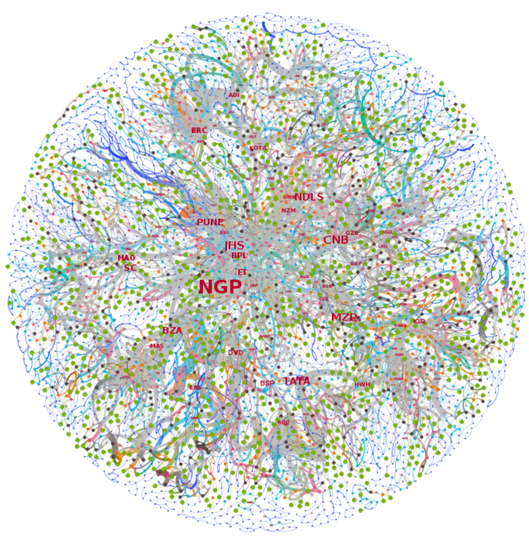


Fig. 1: Simple Graph representing railways network

In Figure 1, the size of station labels represents betweeness centrality of the nodes and the edge width represents the number of trains between the two stations.

### A. Betweeness Centrality

The betweenness centrality for each node is the number of shortest paths that pass through the node. Maximum betweeness centrality is observed for NGP(Nagpur). This implies that Nagpur is a very important station as many shortest paths pass through it.

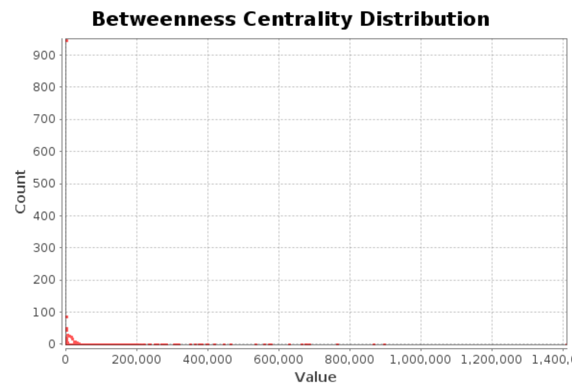


Fig. 2: Betweeness Centrality

### B. Diameter

The greatest geodesic distance between any pair of vertices is the diameter of the graph. Diameter of the graph is found to be 34. This implies that the maximum number of stations to be crossed in order to travel between any two stations in India is 34.

### C. Community Detection

The number of communities have been found out on the basis of modularity.We used the implementation of algorithm available in Gephi as mentioned in the paper[1].The number of communities have been found out to be 30 and the modularity is observed to be 0.865.

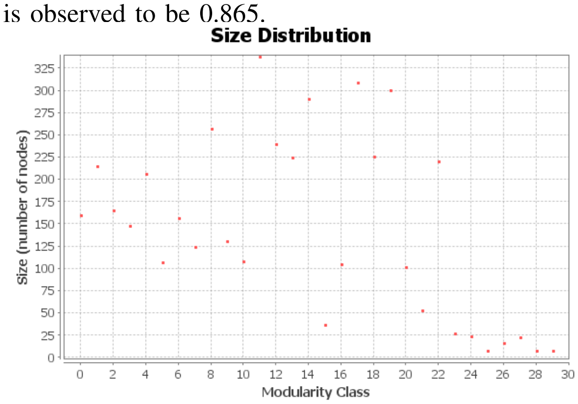


Fig. 3: Modularity Distribution

### D. PageRank

Pagerank is an extension of Katz centrality in which the importance of a node is inversely proportional to outdegree of its neighbours. The personalization matrix in computing pagerank is taken as the number of trains passing through that station. The calculation of Pagerank is done using NetworkX. The highest value of 0.0045 has been found for station BZA

(Vijayawada).This implies that station Vijayawada is significant as the average number of times a random train crosses Vijayawada station is higher than any other station.

$$\underline{x} = \alpha A D^{-1} \underline{x} + \beta \underline{1}$$

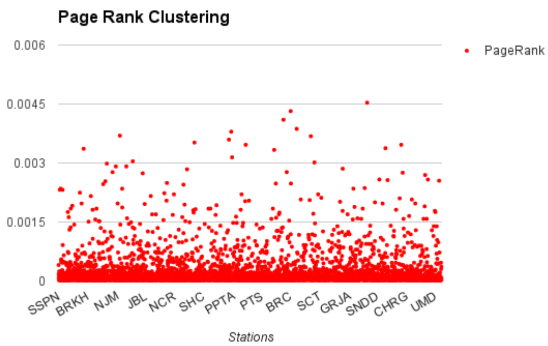


Fig. 4: Page Rank Clustering

### E. Clustering Coefficient

It is the fraction of paths of length two in the network that are closed[2].

Local clustering coefficient : $C(i) = \binom{k_i}{2}^{-1} T(i)$

Average Clustering Coefficient : $\sum_{i \in V} C(i)/n$

Average clustering coefficient of 0.4 is observed which is the number of triangles in which a node is involved with respect to the ones it could be involved in.

### F. Components

The maximal subset of vertices such that each vertex is reachable by some path is a component of a graph.The number of connected components in Indian Railway Network is found to be three of which one is much larger(4327 nodes) compared to the other two(8 each). This implies that almost all the places in India can be reached by means of railways.

## III.ANALYSIS OF HYPER GRAPH

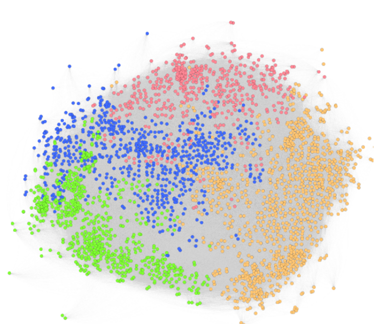

Fig. 5: Projection of hyper graph as simple Graph

### A. Diameter

The diameter of hypergraph was found to be 4. This implies that a maximum of 4 trains need to be changed to travel between any two stations. This reveals more information than a simple graph in which the diameter is 34. This also implies a "Small World" phenomenon of Indian Railway Network.

### B. Community Detection

Community detection in hypergraphs is done using its projection as simple graphs. Though 58 communities have been reported, only 4 major communities have been observed. This is in contrast to 34 communities observed in case of simple graphs. In Figure 5, colour of node is based on the community it belongs to. The modularity index of hypergraphs is 0.431.

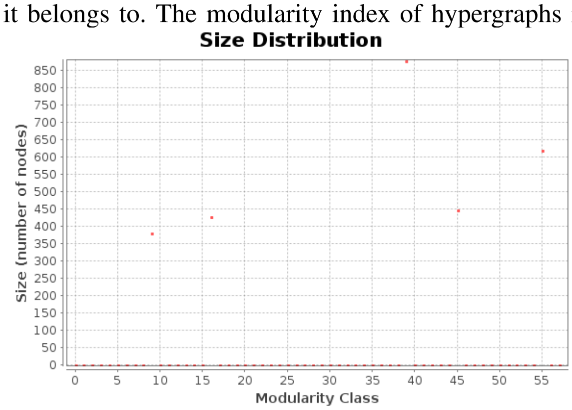


Fig. 6: Modularity Distribution

### C. PageRank

Some trains are used extensively compared to others as they connect many stations. Taking this as the personalization matrix, the PageRank is computed. The highest value of page rank is found for about 90 trains. This implies that if any random station is to be reached, then the chances of travelling in these trains is higher than others.

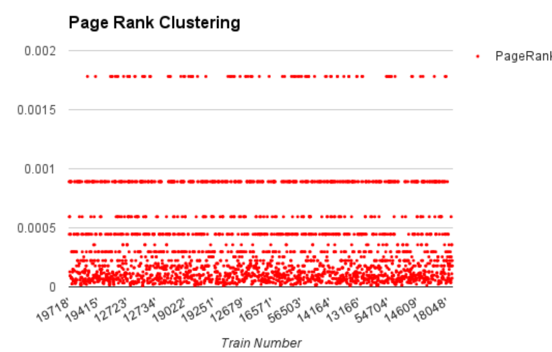


Fig. 7: Page Rank for Hyper Graphs

### D. Components

The number of components was reported as 54, but only 1 major component was observed and many other small sized components. All small sized components contain only one train which implies that stations with these trains cannot be reached via any other train. We can also observe that almost all the places in India can be reached by one or more trains.

## REFERENCES


[1] S. N. Satchidanand, S. K. Jain, A. Maurya and B. Ravindran, "Studying Indian Railways Network using hypergraphs," 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), Bangalore, 2014, pp. 1-6.

[2] Clustering Coefficient: https://en.wikipedia.org/wiki/Clustering_coefficient

[3] Railways Dataset: https://data.gov.in/catalog/indian-railways-train-time-table-0

[4] Github link for code, Gephi files and data: https://github.com/cs14b052/CNA_RailwayNetworks