Analysis of Airlines Network of India

Rahul Gokhale(05CS6008)
Sachin Kulkarni(05CS6029)

M.Tech(CS), 1st year,
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

1 Abstract

Transport Network of a country is one of the most important fundamental infrastructures, which indiate the development of the country. In this project, we tried to analyze various properties of Airlines Network of India as a Complex Network. All the domestic airports are considered as nodes and links between them represent number of planes between the two airports in one week duration. We find that this is a small-world network characterized by truncated power-law degree distribution. We also observed that it has disassortative mixing, which is a desired characteristic of a transportation network. We define and measure several other properties of this network and try to analyze the significance of the results.

2 Introduction

Transportation infrastructure plays very crucial role in the development of the country. Road, Rail and Airway are three major means of transport in India. Although Airtraffic constitutes very small portion of the transport traffic in India, it is growing at an accelerated pace. It is very important to analyze the growth of the airlines network and verify whether the connectivity provided by service providers is really helpful in managing air traffic better. Also this analysis can be helpful for administration, policy and efficiency.

We study the network for its topological structure and for its traffic dynamics. We study unweighted network to investigate its topoligical properties. Then we form weighted network by considering link weights as the number of flighs between the two airports per week. This network gives us the sense of actual traffic flowing between the network. Our analysis shows that Airlines Network of India is a small-world network. It is also found to be having disassortative topology, which is a good characteristic for a transport network.

3 Airlines Network of India

This network is formed by domestic airports as nodes. Edges between the node represent the presence of at least one flight between the two airports and weight of the edge is the number of flights flying between them per week.
Data is collected from the online timetable maintained by air-travel service providers. We have considered five major air-travel players in India, Indian Airlines, Air Deccan, Jet Airways, Kingfisher Airways, Air Sahara. Next subsection gives some of the basic statistics of this network.

3.1 Basic Statistics

Unweighted network of ANI is a directed network with number of nodes, $N = 106$ and number of directed links, $E = 393$. Let this adjacency matrix be $UW(N*N)$. Element uw_{ij} is 1 if there is at least one flight by any service provider from airport represented by node i and airport represented by node j on any one or more days of week.

Weighted network includes the information about the traffic flowing from one airport to another in terms of number of flights per week. Let this adjacency matrix be $W(N*N)$. Element w_{ij} stands for the number of flights per week from airport represented by node i and airport represented by node j. Majority of this matrix is symmetric. There are 44 such nodes whose indegrees differ from their outdegree. But this difference is not more than 2 for any of the nodes. In some analysis, we use the average weight of the edge.

4 Topological Analysis of unweighted network

This section discusses the result of analysis of unweighted network to find its topological characteristics.

4.1 Average degree of a node

Degree of a node is the number of nodes it is directly connected to. In directed network, indegree of a node i (K_i^in) is the number of nodes having a link to node i and outdegree of a node i (K_i^out) is the number of nodes, which node i has link to. Average degree of symmteric ANI is found to be 7.236.

4.2 Shortest Path and Diameter

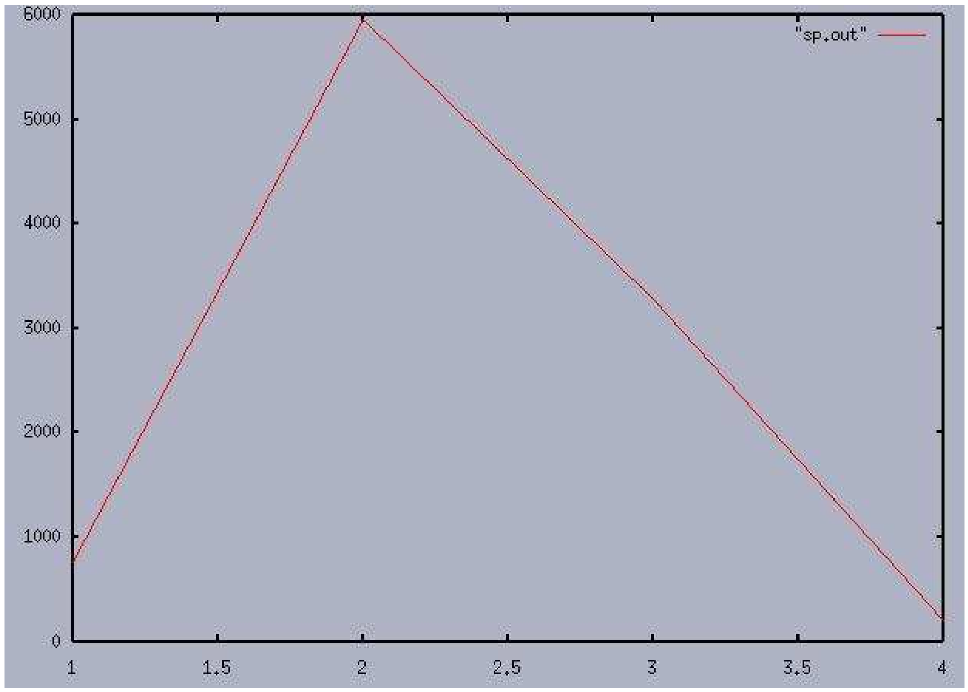

Figure 1: Shortest Path versus path length plot, x-axis:path-length, y-axis:number of paths of that length

Average shortest path length, L, is defined as,

$$L = \frac{1}{N * (N-1)} * \sum_{i,j=1, i \neq j}^{N} L_{ij} \tag{1}$$

where L_{ij} is the shortest path from node i to node j. In our network, average shortest path length was found to be 1.113.

Diameter of network D is defined as the maximum possible shortest path. Diameter of ANI is 4.

We also plotted the number of paths of length l, np_l versus l. Figure 1 shows this plot. From this plot we can observe that majority of the paths are of langth 2. Around 98% of the paths are of length $l \leq 2$, which indicates most of the time while traveling, traveller needs to change just one flight.

4.3 Degree Distribution

Degree is one of the measures of the centrality of the node. Higher degree nodes are always more important. Degree Distribution of a network gives the process of how network grows with time.

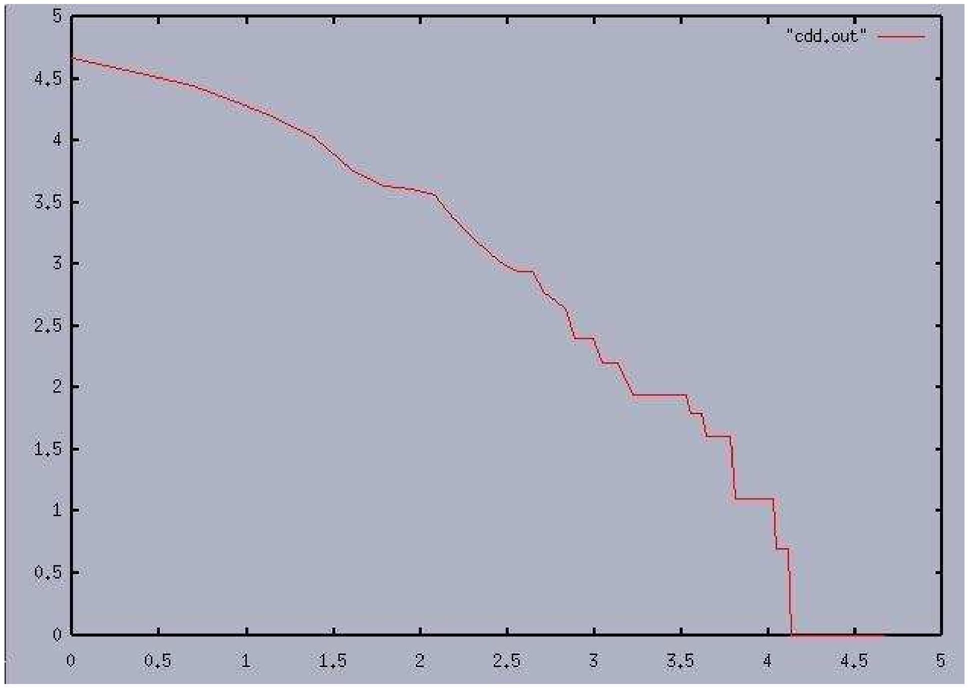

Figure 2: Cumulative Degree Distribution Plot,x-axis:log(degree),y-axis:log(number of nodes)

A scale-free networks (SFN) are characterized by power law decay of degree distribution function: $P(k) \propto k^{-\gamma}$. All scale-free networks are believed to be small-world networks but not vicecersa.
For ANI, we plot cumulative degree distribution, $P(>k)$ versus degree k on log-log scale. Figure 2 shows this plot. As seen in the plot, ANI follows power law as one can fit a straight line in that curve. Thus we can say Airlines Network of India has the characteristic of truncated power-law degree distribution and it is a scale-free network.

4.4 Clustering Coefficient

Clustering Coefficient of a node, C_i is defined as the ratio of the number of links amongst the neighbors of the node to the maximum possible links amongst them. It gives the probability that any two neighbors of the node have a direct edge between them. Average Clustering Coefficient of a network is defined as,

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{2}$$

where C_i is clustering coefficient of node i.
Average Clustering Coefficient of our network comes out to be 0.6178. We tried to find relationship of clustering coefficient of a node with its degree. For that, we plotted

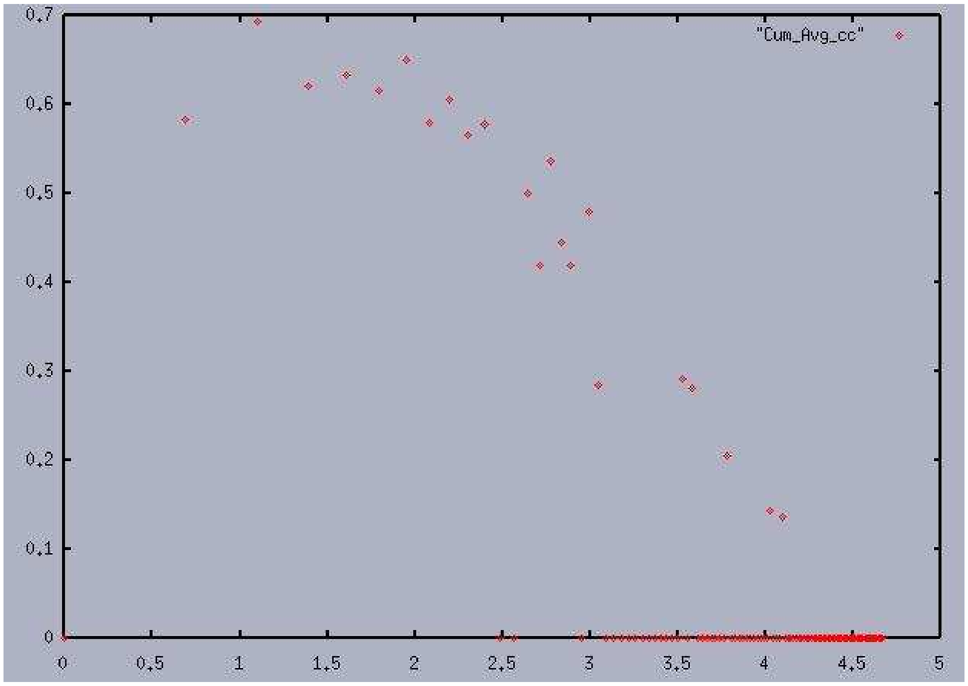

Figure 3: Average Clustering Coefficient per Degree Plot, x-axis:log(degree), y-axis:log(average clustering coefficient)

a graph of Average Clustering Ceofficient $accpd_k$, which is average of clustering coefficients of all the nodes of degree k. Figure 3 shows this plot of $accpd_k$ versus k on log-log scale. As we can see in this graph, nodes of lower degree exhibits high clustering coefficient where as nodes of high degree show low clustering coefficient. This indicates certain hierarchical structure in the Airlines Network.

4.5 Small World Network

Small World Networks are characterized by very small average shortest path length and high clustering coefficient.
In our network Average shortest path value (2.113) is smaller than that of a random network ($L_{rand} \approx ln(N)/ln(K) = 2.356$) of same size and same average degree. Also clustering coefficient value (0.6178) is significanly high. Along with this, ANI shows property of being a scale-free network. These results indicate that Airlines Network of India is a small-world network.

4.6 Node Between Centrality in unweighted network

Node Betweenness is a measure of the influence a node has over the spread of information through the network.
we measured the Betweenness Centrality of a node by the number of contributions from

essentially all paths between the nodes, not just shortest paths. we used the washall's algorithm to find the closure of the unweighted network. Betweenness Centrality of a node is computed using the loss of connectivity in the closure of the network because of removal of that node.
Following table shows the top 5 nodes, with highest betweenness centrality:
When we computed the loss due to each node, it was observed that there are very few nodes, which when removed generate huge loss of traffic. For majority of the nodes, traffic loss is almost same and significantly less than that of those top 5 nodes (shown in the table). This indicates the presence of hubs in the network, which handle majority of the air traffic.

Table 1: Nodes with high Betweenness Centrality

No.	Airport	Total Path Loss	% Path Loss
1	Mumbai	1558	13.87
2	Delhi	1138	10.13
3	Kolkata	1086	9.67
4	Chennai	793	7.06
5	Banglore	504	4.49

5 Analysis on Weighted Network

This section presents the analysis over the wight Airlines Network.

5.1 Traffic Plot

We studied the variation of flow of traffic with the degree of the node. We measured the strength of a node i (S_i) as,

$$S_i = \sum_{j=1}^{N} uw_{ij} * w_{ij} \tag{3}$$

Strength of the airport represents total traffic flowing through it per week. We plotted the Strength of node (Si) versus Degree of node (K_i). One can see this plot in Figure 4. We observed that S_i increases with K_i and the relationship is greater than linear, $S_i \propto K_i^{\beta}$, where β is greater than 1.
This indicates that strength of a node is strongly corelated to its degree. Larger the airport, more the traffic it handles and rate of increase of traffic is greater than that of degree.
One more interesting observation is that there are quite a few nodes with high degree

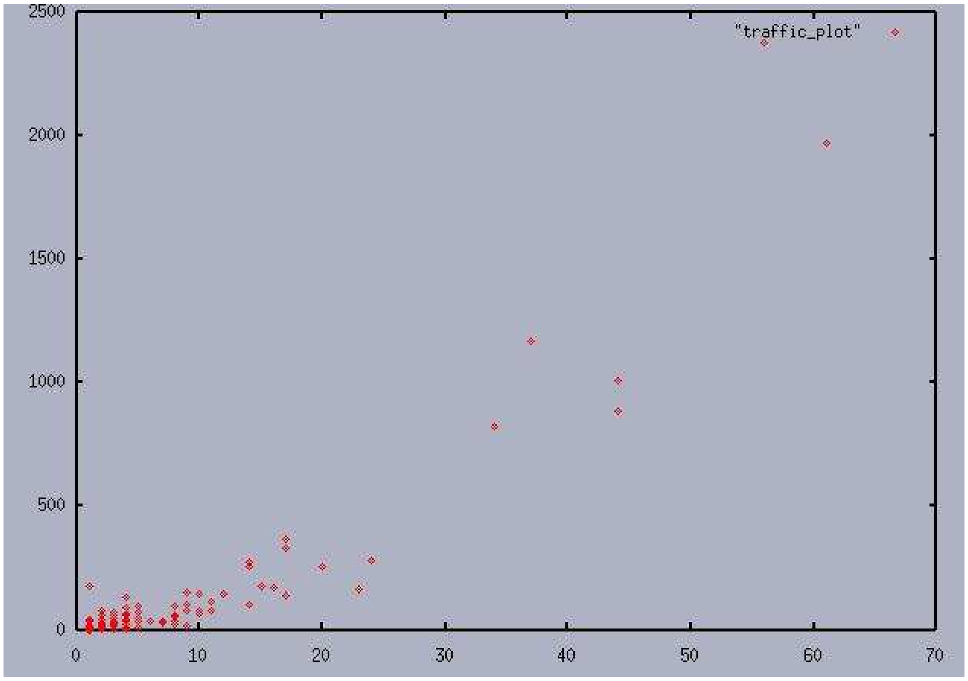

Figure 4: Strength of the node versus degree of node plot, x-axis:degree,y-axis:strength of node

and high strength and in the low degree-low strength region, we can see huge number of nodes. These are high degree-high strength nodes are the nodes with high betweenness centrality.

5.2 Degree Corelation

Degree Corelation is studied to see whether the degrees of any two directly connected nodes have any relationship between them. For this, we measure Average Degree of Nearest Neighbors, $k_{nn}(k)$, for nodes of degree k. This Average Nearest Neighbor Degree (ANND) is defined as,

$$k_{nn}(k) = \sum_{l} l * P(l|k) \tag{4}$$

where $P(l|k)$ is conditional probability that given vertex of degree k is connected to a vertex of degree l.

If there is any relation between $k_{nn}(k)$ and k then there is some corelation among nodes' connectivity and there is possibly some hierarchical structure in network topology. If $k_{nn}(k)$ increases with k, then it means that nodes with high degree (high value of k) have tendancy to get connected to high degree nodes on an average (as Average Neighboring Degree is high). Such networks are said to have assortative mixing. In

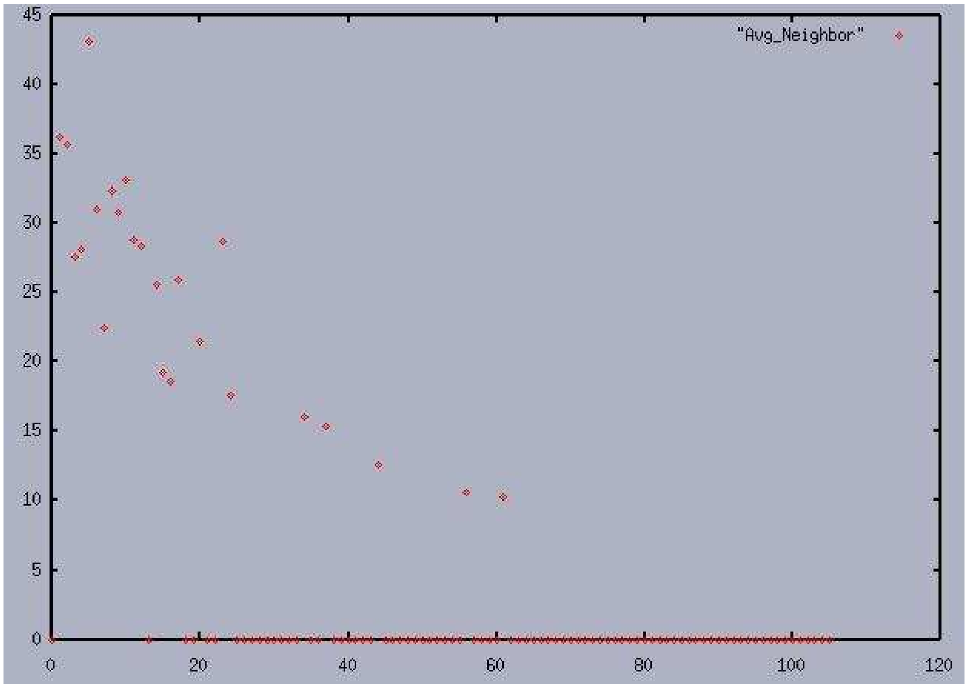

Figure 5: Average Nearest Neighboring Node Degree,x-axis:degree,y-axis:average neighboring node degree of node

disassortative mixing, high degree nodes have their average neighboring degree low indicating their tendancy to get connected to low degree nodes.
For weighted network, Average Nearest Neighboring Degree is calculated as,

$$k^w_{nn,i} = \frac{1}{S_i}\sum_{j=1}^{N} uw_{ij} * w_{ij} * s_j \qquad (5)$$

Figure 5 shows the plot of Average of Average Nearest Neighboring Degree, $k_{nn}(k)$ for all nodes of degree k versus degree of node k. Figure 6 shows the cumulative average of average nearest neighbor degree. In Figure 5, we can see that Average Nearest Neighbor Degree of low degree nodes is very high and that of high degree nodes is very low. This trend indicates the weighted disassortative nature of the Airlines Network. There are few major hubs in the ANI network and they can be considered as the backnbone providing connectivity to the overall network. Hubs are high degree nodes and they are connected to mostly low degree nodes. This feature is consistent with nature of airport networks of few other countries such as Airport Network of China.

5.3 Assortativity Measure

We also calculated Normalized Corelation Function (r), defined in a paper "Assortative Mixing in Networks" by M. Newman. This function is defined as,

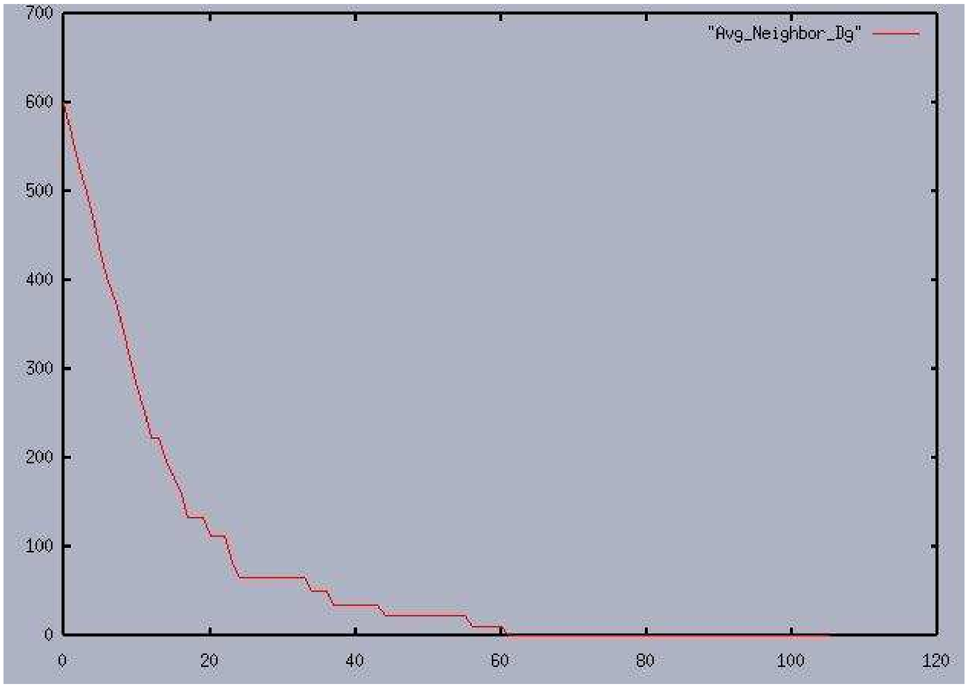

Figure 6: Cumulative Average Nearest Neighboring Node Degree,x-axis:degree,y-axis:cumulative average neighboring node degree of node

$$r = \frac{1}{\sigma_q^2} \sum_{jk} j * k * (e_{jk} - q_j * q_k) \tag{6}$$

where, $\sigma_q^2 = \sum_k k^2 * q_k - [\sum_k k * q_k]^2$ is equal to variance of the distribution of q_k and e_{ij} is the joint probability distribution of the remaining degrees of the two vertices at either end of randomly chosen edge.

Range of this function is $-1 \leq r \leq 1$. For random graphs, value of r is 0. For graphs with assortative mixing, r becomes positive and for diassortative graph, it is negative. Magnitude of r indicates how strongly graph shows assortative ir disassortative mixing.

For Airlines Network, value of r comes out to be -0.4133. This shows that Airlines Network of India is very strongly disassortative as magnitude 0.4133 is very high compared to magnitudes of other disassortative networks (Food web: -0.276, neural network: -0.163, Internet:-0.189, Indian Railway Network:-0.033).

6 Conclusion

We analyzed Airlines Network of India as complex network and found some interesting features of it. It is a network with diameter 4 and has very low average shortest path, 2.113. This shows that on an average, traveller needs to change just one flight to reach from any airport to any other.

Airlines Network of India shows small-world charactertistics and also follows truncated scale-free degree distribution. When plotted the traffic (weighted) flowing through the airports, we observed that rate of increase of traffic on an airport increses with the degree and this rate is faster than linear rate.
Analysis of Node betweenness centrality shows that there are very few nodes in the network, which act as hubs handling majority of the air traffic and they provide the backbone of Airlines Network of India. One more useful characteristic found was Airlines Network of India folllows disassortative mixing, which is a desired property of a transportation network. This indicates that the hubs in the Airlines Network have tendancy to get connected to low degree nodes and which helps reducing the average shortest path and makes the network small-world.

References

[1] P. Sen, S. Dasgupta, P. Sreeram, G. Mukharjee, S. Manna **Small-world Properties of Indian Railway Network**

[2] G. Baglar **Analysis of the Airport Network of India as a complex weighted network**

[3] M. E. J. Newman. **Assortative Mixing in networks**