Railway Accidents in India: by Chance or by Design ?

A thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology (Honours)

in

Computer Science and Engineering

by

Avishek Banerjee
07CS1041

advised by

Dr. Niloy Ganguly

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
May 2011

Certificate

This is to certify that the thesis entitled **Railway Accidents in India: by Chance or by Design ?** submitted by Avishek Banerjee (07CS1041) to the Department of Computer Science and Engineering is a bonafide record of research work carried out by him under my supervision and guidance. This thesis has fulfilled all the requirements as per the regulations of the institute and, in my opinion, has reached or exceeded the standard needed for submission.

Dr. Niloy Ganguly
Associate Professor
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
May 2011

Acknowledgment

I would like to express my gratitude towards Prof. Niloy Ganguly for his esteemed supervisory role he played to utmost perfection. Taking time out of his busy schedule, he ensured that the my project work was carried out methodically and meticulously. I especially thank him for his encouragement and his intuitive comments which were of critical importance for successful completion of this project, and am indebted to him for extending out all the necessary support throughout the duration of the project and for being a constant source of inspiration.

I would also like to thank Saptarshi Ghosh for his invaluable help and esteemed guidance for completing this entire project. I extend my thanks to Sanket Agarwal, Naveen Kr. Sharma and Arun Dobriyal for their help and enthusiasm. Their continuous support and encouragement has played a key role in the completion of this work to perfection.

Avishek Banerjee
07CS1041
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
May 2011

Contents

Chapter 1

Introduction

1.1 Transportation Networks

Transportation networks are amongst the most important building blocks in the economic development of a country. The structure and performance of transportation networks reflects the ease of travelling and transferring goods among the different parts of a country, thus affecting trade and other aspects of the country's economy.

In the recent years, Complex Network analysis has been used to study several transportation networks of different countries. These include airport networks (for instance, the airport network of China [11], airport network of India [3] and the world-wide airport network [7, 5]), urban road networks [14, 20] and railway networks [9, 18, 6, 19, 12].

Railways are one of the most prominent modes of transportation in many countries across the world and the complex topological properties of Railway networks of different geographical regions have attracted the attention of the research community. However, analysis of the structure of the Indian Railways(IR) has received considerably less attention, as compared to the railway networks of the European countries and China. In this situation, a detailed understanding of the network-structure and traffic-flow is essential to identify the possible problems in the IR.

In this thesis, various problems arising in Indian Railways(IR) have been addressed and analysed from complex networks perspective. The next 2 sections deal with the challenges that IR faces presently. This study can help in adopting effective extension policies in future, such as for more effective distribution of new trains and for better planning of the railway budget.

1.2 Indian Railway Network

The Indian Railway network (IRN) is one of the largest and busiest Railway networks in the world, handling massive numbers of passengers and quantities of goods daily. Railways are the most popular means of long-distance transportation in India, hence the IRN is often described as the backbone of this nation's economy.

The present scenario in the transportation sector in India gives further motivation for a detailed analysis of the IRN - it is a commonly voiced opinion among economists that the current transportation network in India is too weak to meet the demands of the country's rapidly growing economy [2]. For instance, factors such as the traffic between major cities exceeding the planned capacity [1] and over-utilized railway tracks are resulting in trains having to travel at reduced speeds and carry lesser amounts of freight, thus increasing the cost and time of transportation. Thus, the current structure and topology of IRN has been studied extensively in this thesis.

Moreover, Indian Railways have been established way back by British, and have evolved over the years to its present state. Since the railway network forms the backbone of connectivity in India, the evolution of the IRN can serve as an indicator of the economic growth of the country too. So, the evolution of IRN has also been studied in this thesis.

1.3 Recent Spate of Accidents

The IR has long served as the backbone of this nation's economy by being the most popular means of long distance transportation in India. However, the IR is facing several grievous problems in the recent years. More alarmingly, there has been a spate of Railway Accidents in India in the year 2010, leading to loss of a significant number of human lives and frequent disruption of traffic over large regions of the country [6]. Here we consider only those accidents that were caused due to collision among trains or derailment of trains and not due to terrorist activity or natural calamities like fire, floods. According to the Wikipedia page[9] enlisting the major rail-accidents in India, there have been 11 such accidents in 2010 alone as compared to 7 such accidents in the 5-year period of 2005-2009.

Thus, the Traffic patterns in the IRN have also been studied in this thesis to understand this repeated occurrence of accidents in a specific region in recent times. Analyzing the current IR traffic as well as the increase in IR traffic over the last two decades, we find that traffic in the said region has increased exorbitantly and it is quite probable that the present amount of traffic has exceeded the allowable safety-limits considering the IR resources (e.g. railway-tracks, signalling systems) available in the region.

Chapter 2

Literature Overview

Several transportation networks of different countries have been studied using the Complex networks tools. Some commonalities have been observed in the topological properties of almost all transportation networks, such as small-world properties. On the other hand, certain topological properties, such as the cumulative degree distribution, have been found to differ widely - power-laws for Indian airport network [3] and world-wide airport neworks [5], two-regime power-laws for the China airport network [11] and US airport network [13] as opposed to the exponential degree distributions of the railway networks of India [19] and China [6].

The fractal structure of the Railway network in Seoul was studied in [9] - the fractal dimension of the netork was found to increase with time; also a comparison between the fractal dimension of the ensemble of stations and that of the railway lines was proposed as a measure of the quality of the transportation system. The underground (subway) railway networks of Boston and Vienna were studied as bipartite station-train networks in [18] - several topological metrics of the networks were measured and compared with the corresponding theoretical predictions for random bipartite graphs using a generating function formalism. Various topological properties of the Chinese railway network have been studied in [6, 12], whereas [21] used a weighted representation of the Chinese railway network to propose a metric to quantify the dependence of a station on another.

To the best of our knowledge, the only study of the structure of the IRN from a Networks perspective was in 2003 by Sen et. al. [19], where the IRN was represented as a network of stations, two of which were linked by an edge if a train halted at both the stations. Hence the network considered in [19] was unweighted, and an edge simply indicated the presence of a train linking two stations. However, a transportation network is specified not only by its topology of connections between stations, but also by the dynamics of the traffic-flow taking place in the network. Such networks display a large heterogeneity in the capacity of the connections; for instance, a significantly larger number of trains can be expected to link two major stations compared to that linking less important stations. Thus, in order to get a complete description of transportation networks, it is essential to take into account the amount of traffic-flow along the connections. Representing the amount of traffic on different links by edge-weights can yield observations that might be undetected by metrics based on topological information alone, as was demonstrated for the world-wide airport network in [5]. Hence, in this thesis, we study the IRN as a *weighted* network of stations (nodes), where the weight of an edge indicates the number of trains linking two stations and study the relevant structural and topological properties of IRN.

It is to be noted that several different models have been used in literature to study transportation networks, and the observed topological properties often depend on the way the network is modeled. Most studies, including the ones referred to above, adopt a common network model where two nodes (airports or stations) are linked by an (undirected) edge if there exists a direct connection (flight or train) between the two nodes. 2 On the other hand, a *directed* network model was used in [12] to study the Chinese railway network, where the in-degree and out-degree of a node (station) were defined as the number of trains arriving at the station and the number of trains departing from the station respectively; the degree distribution of this network was observed to be a power-law. Transportation networks have also been modeled as bipartite networks (e.g. [18, 21]) and weighted networks (e.g. [3, 5]).

Another interesting feature of transportation networks is that they are evolving in nature, with new trains (flights) and new stations (airports) being introduced regularly. Though there have been studies on the evolving properties of several transportation networks (for instance, the airport network of China [11], Swiss road and railway networks [12], Indiana inter-urban network [13]), there have not been any prior study on the evolution of any transportation network in India, to the best of our knowledge. Here we study the evolution of the IRN using several snapshots of the network over the past two decades.

To the best of our knowledge, there has been no prior study on finding the cause of increasing Railway accidents in India, though there has been some previous work on understanding the Road accidents in the country. Infact no study on traffic analysis of Indian Railways have been carried out so far possibly due to lack of availability of Data.

Chapter 3

Data Collection

3.1 Present Railways Data

The IRN is a dense network where the total number of stations and train-routes are of the order of tens of thousands. In this study, we consider the IRN at a coarse-grained level - we consider only the 'express' train-routes and other long-distance train-routes (leaving out 'local' or suburban routes which traverse relatively short distances around major cities), and only those stations which are scheduled halts on at least one such train-route.

We crawled the data of Express Train-routes in the present IR and the stations on each route, from the official website of Indian Railways (www.indianrail.gov.in) in July 2010. The website hosts information of 2195 Express train-routes and 3041 stations, along with the scheduled time of each train reaching each station on its route. We consider each train-route to be a uni-directional path from the source station to the destination station because the train from source station A to destination station B often runs simultaneously with the train going in the reverse direction from station B to station A, hence both contribute to the amount of traffic at a given point of time. From the above data, we derived the Traffic scenario in IR for each individual day of the week, which gives the exact scheduled time of each train reaching any given station, on any given day of the week. We consider the traffic for each day of the week individually because several trains in IR are bi-weekly or tri-weekly ones, which are scheduled to run only on certain days of the week.

Along with the present traffic scenario, we also study the growth of traffic in the IR over the last two decades. For this, we collected the list of Express train-routes and the stations on each route for the years 1991, 1997, 2000, 2005 and 2009 from the "Trains At A Glance" (TAAG) Time-table published by the Indian Railways for the corresponding years. However, the scheduled time of the trains reaching each station could not be obtained for the older years; hence this "evolutionary data-set" has only been used to analyze the growth in volume of IR traffic over the years, and all temporal analyses uses the IR data of 2010 (as obtained from the IR website). It is to be noted that the data obtained from the IR website is much larger and more fine-grained (i.e. contains many more trains and stations) than the datasets obtained from the TAAG time-tables which list only the more important train-routes and stations, hence we have avoided comparing the data from the two different sources.

Limitations of the data:

- The data only contains the express train-routes, leaving out "local" or "suburban-routes" which traverse relatively short distances around major cities. It is to be noted that, since suburban trains usually travel over short distances at much lower speeds (compared to express trains), derailment / collisions are very rare among suburban trains, hence they can be ignored in our analysis.
- Also, the data does not include freight-trains which often travel over long distances and use the same railway-tracks with express passenger-trains (in fact, freight-trains have been involved in collisions with passenger-trains in some of the recent accidents). However, freight-trains in IR

usually run in an on demand- basis and consequently do not have any fixed schedule of travel, thus making it almost impossible to include them in the analysis.

3.2 Network Construction

Two different, but related, approaches have commonly been adopted in the literature to represent a railway network as a complex network. In the context of a railway network, a train-route is a sequence of stations at which a train following that route is scheduled to halt.

- A railway network can be represented as a *bipartite train-station network* [18, 21] with one set S of nodes representing stations and the other set T of nodes representing the train-routes; there is an edge between $s \in S$ and $t \in T$ if and only if station s is a scheduled halt in the train-route t.
- The more commonly used representation of a railway network is a network consisting of only station nodes, where two stations s_i and s_j are connected by an edge if there exists at least one train-route directly linking the two stations (in other words, if there exists at least one train-route such that both s_i and s_j are scheduled halts on that route). This representation is frequently used [6, 19, 3] to model different transportation networks since it directly captures some key facts on the connectivity of nodes (stations or airports) - for instance, the neighbours of a given station s_i are precisely those stations which can be reached from s_i by boarding a single train, while the shortest distance between an arbitrary pair of stations s_i and s_j is the minimum number of different trains that one needs to board to travel from s_i to s_j. In a weighted version of this *station-station network* representation, the weight of the edge between s_i and s_j is the number of train-routes on which both these stations are scheduled halts.

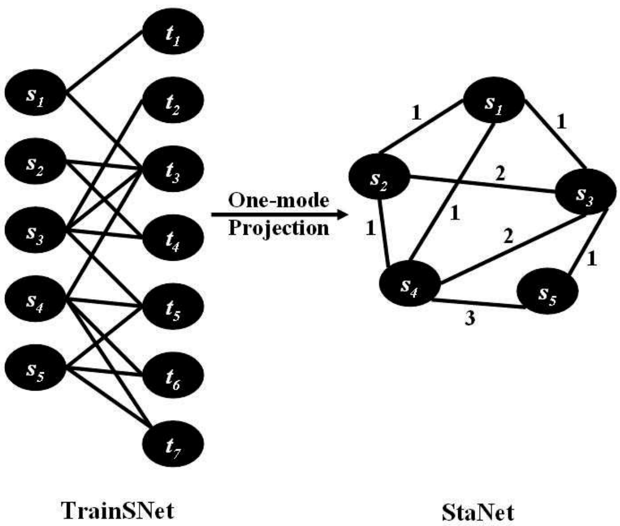

Figure 3.1: Obtaining a weighted station-station network (StaNet) by one-mode projection of bipartite train-station network (TrainSNet)

The station-station network representation can be derived from the bipartite train-station network by constructing a *one-mode projection* of the bipartite network over the station nodes, in which two stations $s_1, s_2 \in S$ are connected by an edge if they are linked to a common node $t \in T$ in the bipartite network. The weight of the edge linking s_1 and s_2 in the projection is thus the number of distinct nodes $t \in T$ to which both s_1 and s_2 are connected in the bipartite network (this is analogous to the number of train-routes on which both s_1 and s_2 are scheduled halts). This is the approach that has been used throughout this thesis to construct the weighted station-station network representation of the IRN.

Chapter 4

Topological Analysis

This chapter discusses the topological properties of the present-day IRN which is represented as a weighted station-station network of stations. The characteristics of IRN which have been investigated are Degree and Strength Distributions, Distribution of edge-weights, Strength-Degree Correlation, Weight-Degree Correlation, Degree-degree correlations and Clustering coefficient.

4.1 Degree and Strength Distributions

The degree distribution $p(k)$ of a network is defined to be the fraction of nodes having degree k in the network. Thus if there are N nodes in a network and n_k of them have degree k, we have $p(k) = n_k/N$. The cumulative degree distribution $P(k)$, defined as the fraction of nodes having degree at least k, i.e.

$$P(k) = \sum_{i=k}^{\infty} p(i)$$

is preferred for analysis in practice, because the degree distribution is often noisy and there are rarely enough nodes having high degrees to get good statistics in the tail of the distribution, whereas the cumulative distribution effectively reduces the number of statistical errors due to the finite network size [16].

The degree of a node in the station-station network is the number of stations that can be reached from the given station via a single direct train, hence the node-degree is a measure of the connectivity of a station. The cumulative degree distribution $P(k)$ of the station-station network of the IRN (Fig. 4.1) is observed to be an exponentially decaying distribution having the approximate fit $P(k) \sim exp(-\alpha k)$ with $\alpha = 0.0082$; however, it deviates from the exponential nature for larger k. This exponentially decaying nature of the degree distribution for the IRN agrees with observations in [19]. The deviation for large degrees can be attributed to the high cost of adding links in the station-station network (in order to link a given station to a new neighbour, a new train-route needs to be introduced or a new station needs to be introduced in an existing train-route).

It may be noted that in contrast to the exponential degree distributions of most railway networks, the degree distributions of most airport networks [11, 3, 5] have been observed to be power-laws which can be explained by the preferential attachment growth model [4]. There can be several explanations for this variation, some of which are as follows.

First, there exist significant differences between the architecture of railway networks and that of airport networks. In an airport network, if two airports are connected by an air-route, it is rare for there to be an intermediate airport in the route. However, in a railway network, even if most train-routes are plausibly introduced between major end-stations i.e. high-degree nodes (in agreement to the preferential attachment model), several smaller stations are present between the terminal ones along the train-route, thus raising the degrees of the smaller stations as well. This may result in exponential degree distributions which are known to be more homogeneous compared to scale-free distributions [8]. Second, the networks having power-law degree distributions are characterized by the presence of a few hubs which are very high-degree nodes. A railway station can handle only a limited number of railway-tracks and trains (which limits the degree of the corresponding node in the network), while it is relatively easier for an airport to have direct connections with a large number of others; thus hubs are more likely to be

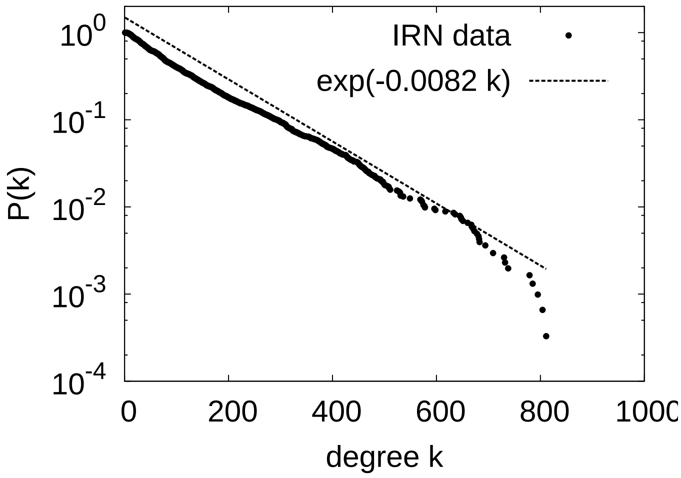

Figure 4.1: Cumulative degree distribution of the IRN (in semi-log scale, along with the exponential fits)

present in airport networks than in railway networks [1].

The *strength*, or weighted degree, of a node in a weighted network is defined as the total weight of the edges adjacent to the node [5]. In the station-station network representation, the strength of a node (station) represents the total number of different journeys that can be undertaken from that station (i.e. journeys to a different station or journey by a different train-route); hence, it is a measure of the available transportation from a station, which combines both the notions of connectivity and amount of traffic-flow (number of train-routes) through the station. For cities having large population and industrial production, the availability of transportation should match the high demands, hence the strength of such nodes should be high (along with high degree or connectivity). The distribution of node-strengths in the IRN (fig. 4.2) also exhibits an exponential nature similar to the degree distribution of the network.

[1]For instance, each of the metropolitan cities in India, which need to have high connectivity with all parts of the country, are served by *multiple* stations in order to share the high amounts of traffic; this limits the degree of the individual nodes (stations) in the network.

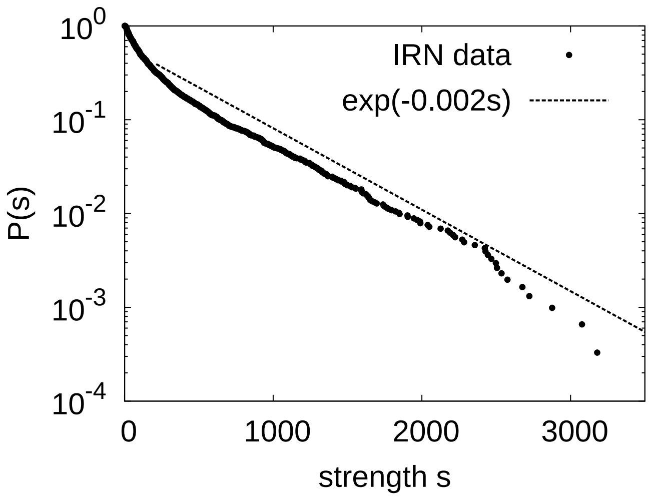

Figure 4.2: Cumulative strength distribution of the IRN (both in semi-log scale, along with the exponential fits)

4.2 Distribution of edge-weights

The edge-weights of the station-station network model the flow of traffic in the railway network - the weight w_{ij} of the edge between two station nodes i and j represents the number of train-routes which directly link both these stations; hence passengers (and freight) move more frequently along edges of higher weights. The analysis of edge-weights indicate a high level of heterogeneity in the traffic-flow in the IRN. The cumulative distribution of the edge-weights in the IRN (fig. 4.3) has an exponential fit $P(w) \sim exp(-\alpha w)$ with $\alpha = 0.12$.

4.3 Strength-Degree Correlation

To investigate the relationship between the degree and strength (weighted degree) of nodes, we plot the correlation between degree k and the average strength $s(k)$ of nodes having degree k in Fig. 4.4. $s(k)$ increases rapidly with k, following a power-law behaviour $s(k) \sim k^{\beta}$, with $\beta = 1.403$. In the absence of correlations between the edge-weights and the degree of adjacent vertices,

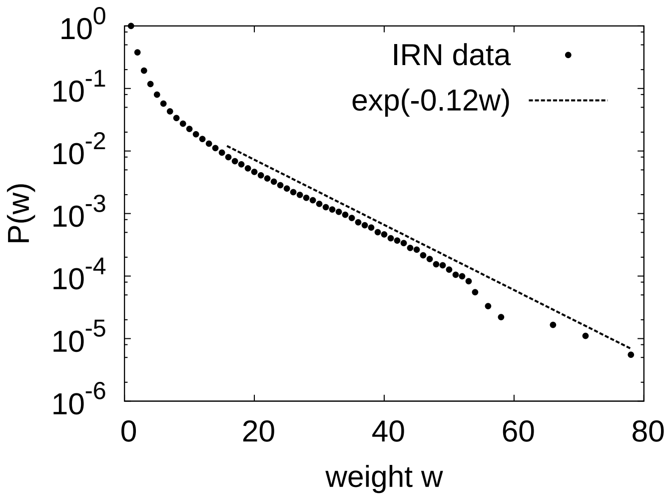

Figure 4.3: Cumulative distribution of edge-weights in the IRN, along with exponential fit (semi-log scale)

the strength of a vertex would be simply proportional to its degree, yielding $\beta = 1$ [5]. The higher value of β for the IRN implies that node-strengths are strongly correlated with node-degree in the IRN and the strength of nodes grow faster than their degrees. This indicates that introduction of new trains on existing routes (i.e. increasing the weights of existing edges, thus increasing the strength of nodes) is more common in the IRN compared to construction of new train-routes that link a station with new neighbours (i.e. increasing the degree of nodes). Similar trends have also been observed for the Chinese railway network [12].

4.4 Weight-Degree Correlation

The strength-degree relationship can also be characterized by the correlation of weight w_{ij} of the edge between nodes i and j, with the degrees k_i and k_j of the end-points i and j, as studied in fig. 4.5. It is evident that the links between high-degree nodes (important stations having high connectivity) tend to have high values of traffic in the IRN. Such high-traffic links between the major cities are generally referred to as *trunk routes*.

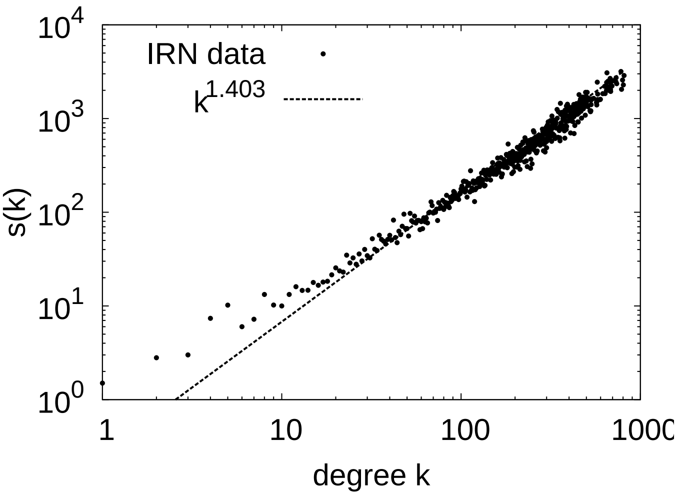

Figure 4.4: Average strength of % nodes having degree k, as a function of k, along with the power-law fit (log-log scale)

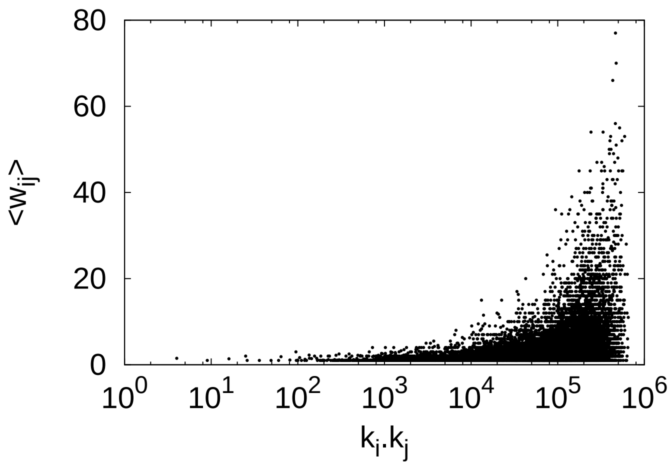

Figure 4.5: Correlation of edge-weights and product of end-point degrees in the IRN (semi-log scale)

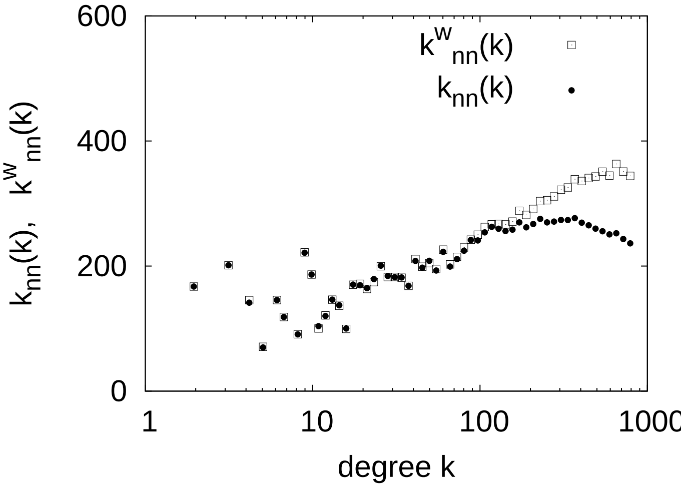

Figure 4.6: Average degree of nearest neighbours $k_{nn}(k)$ and average *weighted* degree of nearest neighbours $k^w_{nn}(k)$ of nodes having degree k, using logarithmic binning of degrees (log-log scale)

4.5 Degree-degree correlations

Another parameter used to investigate the network architecture is the correlation among degrees of neighbouring nodes, which can be observed from the average nearest-neighbour degree $k_{nn}(k)$ of nodes having degree k (fig. 4.6). It is observed that $k_{nn}(k)$ remains the same on the average over a significant range of degrees, implying the absence of major correlations among the nodes of different degrees. This behaviour of $k_{nn}(k)$ agrees with the results for the IRN in [19].

However, a completely different perspective is gained regarding the assortativity of the IRN by taking edge-weights into consideration. We use a weighted variant of the average nearest-neighbours degree, k^w_{nn}, as defined by Barrat et. al. in [5]. For a given node i, $k^w_{nn,i} > k_{nn,i}$ if the edges adjacent to i having the larger weights are connected to the neighbours (of i) having larger degree, and $k^w_{nn,i} < k_{nn,i}$ in the opposite case. Analogously, the behaviour of $k^w_{nn}(k)$ (the average weighted nearest-neighbour degree of nodes having degree k) indicates the weighted assortative or disassortative properties, taking into account the flow of traffic among the stations of the

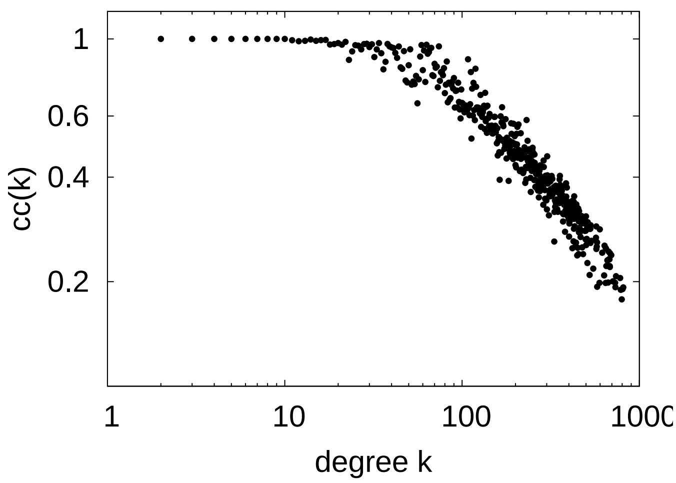

Figure 4.7: Average (unweighted) clustering coefficient $cc(k)$ of nodes having degree k, as a function of k

network.

Fig. 4.6 compares the variations of $k_{nn}(k)$ and $k^w_{nn}(k)$ with degree k (using logarithmic binning of k-values for better visibility); $k^w_{nn}(k)$ shows a pronounced assortative behaviour, implying that high-degree stations tend to connect with other high-degree stations, and the amount of traffic (weight) along such links between high-degree nodes tend to be high as well. Similar trends have also been observed for the world-wide airport network [5]. The topological assortativity coefficient, as defined by Newman [15], comes out to be 0.0813 for the IRN, indicating that the topology of the IRN is weakly assortative in nature. The definition by Newman was extended for weighted networks by Leung et. al. [10]; the weighted assortativity coefficient for the IRN is observed to be 0.2378, indicating that a pronounced assortative behaviour when the traffic-flow is taken into consideration.

4.6 Clustering coefficient

Fig. 4.7 plots the average clustering coefficient $cc(k)$ of nodes having degree k as a function of k; $cc(k)$ remains at a constant value close to unity for small

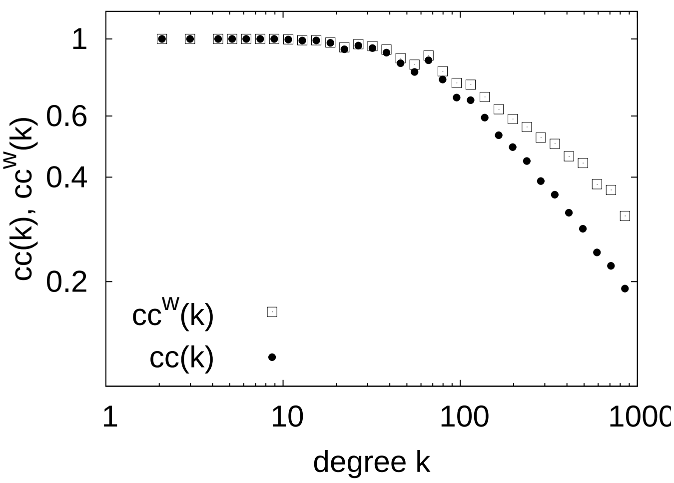

Figure 4.8: Average unweighted and weighted clustering coefficients as function of degree, using logarithmic binning of degrees (log-log plots)

k and then shows an almost power-law decay at larger values of k. This observation, which agrees with results in [19], can be explained as follows. All stations on the same train-route are linked to form a clique in the station-station network. The smaller stations (having low degrees) in the IRN are served by very few train-routes, hence they are linked only to other stations on these train-routes (other nodes in the clique), thus resulting in a clustering coefficient near to unity for the nodes with low degrees. On the other hand, major stations (having high degrees) are served by a large number of train-routes, hence these stations are linked with other geographically distant stations in diverse parts of the country, which themselves do not tend to be connected, thus lowering the average clustering coefficient for nodes with higher degree.

It has been shown [17] that a power-law decay of $cc(k)$ with degree k is an evidence of hierarchical organization in a network, which implies that low-degree nodes belong to interconnected communities. Thus an inherent hierarchy is evident from the structure of the IRN.

For a weighted network, the clustering coefficient has been re-defined [5] to incorporate edge-weights, in order to take into account the importance of the clustered structure based on the amount of traffic actually found in the cluster. Analogous to $cc(k)$, $cc^w(k)$ is defined as the *weighted* clustering coefficient averaged over all nodes of degree k. Fig. 4.8 compares the variations of $cc(k)$ and $cc^w(k)$ with degree k; both versions have similar values for low degrees, however $cc^w(k)$ lies consistently above the unweighted $cc(k)$ for intermediate and higher degrees, indicating that most of the traffic (i.e. edge-weights) in the IRN is accumulated on interconnected groups of high-degree nodes. Further, the variation of $cc^w(k)$ is much more limited in the whole spectrum of k compared to that of $cc(k)$, implying that high-degree stations have a tendency to form interconnected groups with high-traffic links (trunk routes), thus balancing the reduced topological clustering.

The clustering coefficient C of the network, which is the average of the clustering coefficients for all nodes, is 0.733, while the corresponding weighted clustering coefficient C^w comes out to be 0.789. $C^w > C$ again indicates that the major stations (high-degree nodes) form high traffic corridors among themselves.

From the above discussions, it is evident that considering the edge-weights in the station-station network of the IRN has led to a more complete reflection of the properties of the network, compared to what can be obtained from the network topology alone. This justifies our motivation of studying the IRN as a weighted network. The practical implications of the results obtained in this section in context of the IRN are discussed later in section **??**.

4.7 Identifying major stations in the IRN

In this section, we identify the major stations in the IRN from the station-station representation of the network. Since the node-degree is a measure of the connectivity of a station, the nodes with high degrees are evidently important in the network (this measure of importance of nodes is known as degree centrality). The top 10 stations in the IRN based on node-degree are listed in table 4.1. These stations can be classified into two groups based on

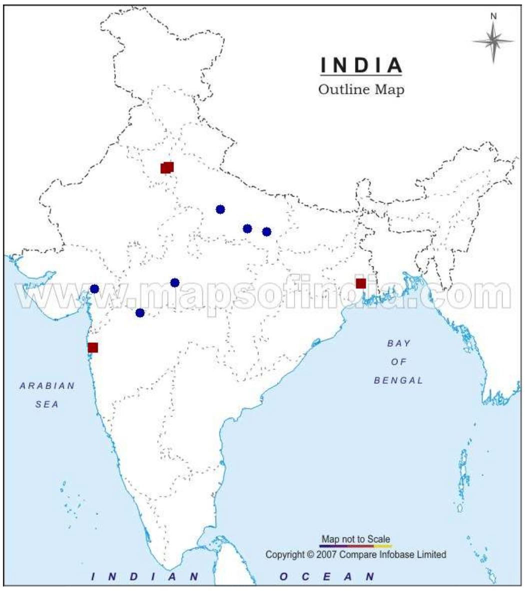

Figure 4.9: The top 10 stations in IRN based on *degree*

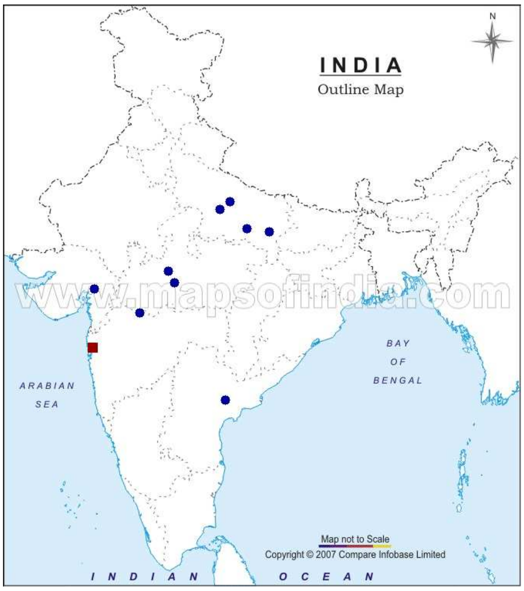

Figure 4.10: The top 10 stations in IRN based on weighted degree or *strength*;Stations in the vicinity of the metropolitan cities in India marked with (red) squares, other stations marked with (blue) circles

Top stations w.r.t. degree	**Top stations w.r.t. weights**
Kanpur Central	Itarsi
Howrah (*)	Vijayawada
Kalyan (*)	Kanpur Central
Ghaziabad (*)	Vadodara
Itarsi	Mughal Sarai
Varanasi	Kalyan (*)
Vadodara	Bhusawal
Allahabad	Lucknow
Bhuwsawal	Bhopal
Hazrat Nizamuddin (*)	Allahabad

Table 4.1: Top 10 stations in the IRN on the basis of node-degree and node-strength. The stations located in vicinity of metropolitan cities marked by (*).

their geographical locations, as shown in fig. 4.9:

- stations that are located in close vicinity to the metropolitan cities in India (e.g. Howrah near Calcutta, Kalyan near Mumbai)
- stations that are located in the central parts of the country or at the meeting points of railway lines connecting different zones (for instance, the left-most circle in fig. 4.9 is at Vadodara junction that is used by most train-routes linking the western zone of India with the southern, central and eastern zones)

Analogously, the nodes having high values of strengths (weighted degrees) are the ones which handle a high amount of traffic. Table 4.1 lists the top 10 stations in the IRN based on node-strength, and fig. 4.10 shows their geographical locations. Interestingly, almost all these stations (except one) are located in the central regions of the country or at the junction of railway lines connecting different zones. Though these stations handle large amounts of traffic, they often do not have as much resources (e.g. platforms, railway tracks) as the stations located in close vicinity of the metropolitan cities. For instance, Howrah, located near metropolis Calcutta and having the highest node-degree among metropolitan stations, has 23 platforms and 25 tracks, while the two stations with the highest node-strengths, Itarsi (located at the centre of the country) and Vijayawada (located on the lines linking south zone with east and north zones), have only 7 and 10 platforms respectively (as given in the Wikipedia articles on these stations). Hence these stations are potential points of congestion in the network.

Further, fig. 4.10 shows that a majority of the stations with high strengths are limited to two specific regions - in the states of Uttar Pradesh and western parts of Madhya Pradesh. Comparing the locations of the metropolitan cities shown in fig. 4.9 with fig. 4.10, it is seen that these regions lie in between the metropolitan cities of India (between Calcutta and Delhi, and between Mumbai and Delhi respectively), and hence these regions contain several trunk-routes linking the metropolitan cities. Large amounts of resource and manpower are required in these regions for efficient management of the excessive traffic.

Chapter 5

Evolution of Indian Railways

This chapter studies the Evolution of the Indian Railway Network (IRN) by tracking the variation of different structural properties of the network over the past two decades. We obtained snapshots of the IRN from the time-table "Trains At A Glance" published in several years over the past two decades; the statistics of the snapshots of the IRN for each year is summarized in Table 5.1, along with some of the topological metrics of the station-station network of the IRN.

It is evident from the values in Table 5.1 that most of the topological metrics of the station-station network of the IRN, such as the Mean Shortest path length, Mean weighted clustering coefficient and Assortativity coefficient, remain stable with time. The average node degree, however, increases steadily over the years, thus implying that the IRN is becoming denser with time. In other words, the number of edges in the station-station network (i.e. the number of direct connections between stations) grows super-linearly in the number of nodes (stations). This densification of the IRN follows a power-law pattern $e(t) \sim n(t)^{\alpha}$, where e(t) and n(t) denote the number of edges and nodes in the network at time t and the exponent $\alpha = 1.5$, as shown in Fig. 5.1. Such a relation, that has been observed for several other evolving social and technological networks as well, is referred to as the densification power law [19]. Several networks obeying the densification power law have

Year	Number of trains	Number of stations	Metrics of the station-station network				
			Mean Node Degree	Mean shortest path length	Mean wt. clustering coeff.	Effective Diameter	Assort. coeff.
1991	134	1238	87.91	2.42	0.82	2.87	0.077
1994	200	1446	95.53	2.42	0.81	2.85	0.066
1999	378	2159	110.83	2.49	0.79	2.88	0.083
2002	460	2265	115.10	2.48	0.78	2.87	0.073
2005	594	2409	120.21	2.47	0.78	2.86	0.065
2009	898	2702	122.18	2.50	0.79	2.87	0.058

Table 5.1: Metrics of the IRN in different years

also been found to have an effective diameter that shrinks over time [19]. However, the effective diameter of the station-station network of the IRN is seen to remain remarkably stable over the past two decades (Table 5.1).

5.1 Degree Distribution

The cumulative degree distribution P (k) of the IRN (station-station network) for all the years are found to be exponentially decaying $P(k) \sim exp(\alpha k)$ in nature, as shown in Fig. 5.2. However, the absolute value of the exponent α decreases over the years, resulting in flatter distributions. A smaller value of α indicates a more homogeneous structure of the network with respect to degrees of nodes, i.e. a relative increase in proportion of stations with high degree. Hence it is evident that the connectivity of stations is consistently improving with time in the IRN.

5.2 Betweenness Centrality

Fig. 5.3 plots the average normalized betweenness centrality $c_b(k)$ of nodes having degree k as a function of k, for the years 1991, 1999 and 2009, along with the MLE (Maximum Likelihood Estimation) fits for the data of each

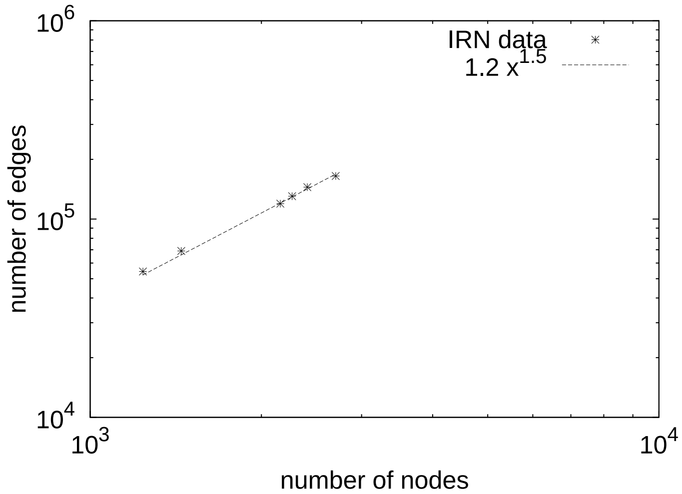

Figure 5.1: Number of edges e(t) vs. number of nodes n(t) in the station-station network of the IRN for different years (log-log)

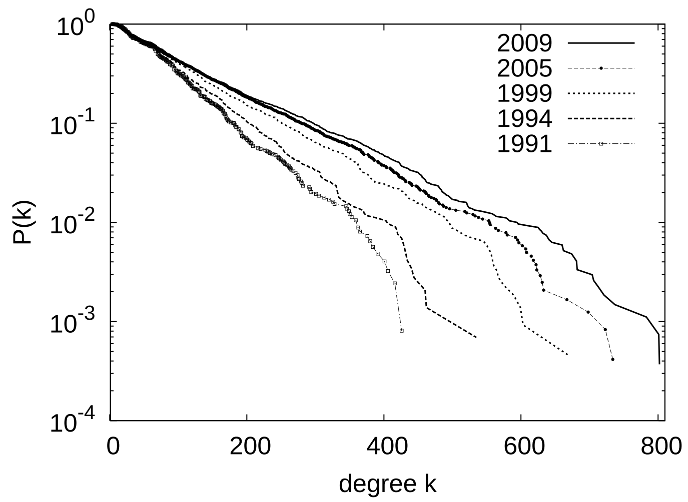

Figure 5.2: Evolution of cumulative degree distribution of station-station network of IRN (semi-log scale)

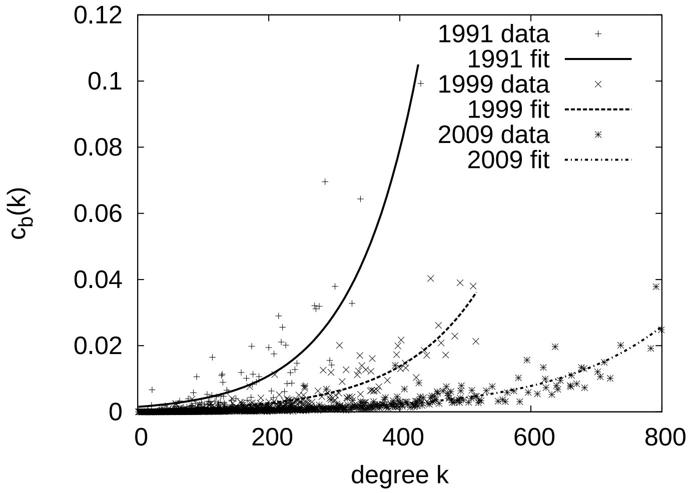

Figure 5.3: Average normalized Betweenness Centrality $c_b(k)$ of nodes having degree k, as a function of k, for the years 1991, 1999 and 2009

year. It can be observed that $c_b(k)$ follows an exponential distribution of the form $c_b(k) \sim exp(\alpha k)$ for each year, with the value of α decreasing with time. The exponential distribution implies that the betweenness centrality of nodes increase sharply with the node degree (number of directly linked stations) in the IRN. However, the maximum value of $c_b(k)$ is seen to fall with time - this indicates that given an arbitrary pair of stations(u,v), several different shortest paths between u and v are coming into existence in the IRN with time, and the fraction of these shortest paths passing through a particular node is getting reduced. This again indicates an improvement in the connectivity among stations (increasing number of shortest paths).

Chapter 6

Recent Spate of Railway Accidents

No doubt, Indian Railways has become the busiest railway networks in the world, but in recent years, it has been facing various challenges, amongst which frequent Railway accidents have become a common place. Alarmingly, there has been a spate of accidents in India in the year 2010 alone, leading to significant loss of human lives and disruption of traffic over large regions of the country. Here only those accidents that were caused due to collision among trains or derailment of trains have been considered though there are other factors as well like terrorist activity or natural calamities like fire, floods .

According to the Wikipedia page enlisting the major rail-accidents in India, there have been 11 such accidents in 2010 alone as compared to 7 such accidents in the 5-year period of 2005-2009. Details of the 11 accidents involving collision/derailment of trains in 2010 are listed in Table 1 while the locations of all such accidents in India since 2005 are indicated in Fig. 6.1.

Moreover, as shown in Fig. 6.1 and Fig. 6.2, as many as 8 out of the 11 accidents due to derailment collisions in 2010 have occurred within a spe-

Sl.No.	Date	Approximate Location	Description
1	Jan 2	Etawah, Uttar Pradesh	Lichchavi Express collided with Magadh Express (dense fog)
2	Jan 2	Panki, Uttar Pradesh	Gorakhdham Express and Prayagraj Express collide (dense fog)
3	Jan 3	Nij Bogaon, Assam	Arunachal Pradesh Express derailed
4	Jan 16	Tundla, Uttar Pradesh	Kalindi Express and Shram Shakti Express collide (dense fog)
5	Jan 22	Sathiyaon, Uttar Pradesh	freight train derailed
6	May 25	Naugachia, Bihar	Guwahati-Delhi Rajdhani express derailed
7	July 19	Sainthia, West Bengal	Uttar Banga Express collided with the Vananchal Express
8	Sep 20	Badarwas, Madhya Pradesh	freight train collided with Indore-Gwalior Intercity Express
9	Sep 21	Kanpur, Uttar Pradesh	freight train derailed
10	Sep 24	Kasganj, Uttar Pradesh	Rohilakhand Express derailed
11	Oct 4	Rasoiya, Uttar Pradesh	freight train derailed

Table 6.1: Railway accidents in India in the year 2010, involving derailment or collision among trains. Accidents numbered 1, 2, 4, 5, 9, 10 and 11 occurred in the upper Indo-Gangetic plain, while the accident numbered 6 occurred in the middle Indo-Gangetic plain

cific geographical region which is usually referred to as the "Indo-Gangetic plains" [8], comprising of the Indian states of Uttar Pradesh, Bihar and West Bengal and National Capital Territory of New Delhi. On the contrary, accidents during 2005-2009 were randomly distributed over the whole country. More specifically, the accidents numbered 1, 2, 4, 5, 9, 10 and 11 in 6.1 have occurred in the state of Uttar Pradesh (upper Indo-Gangetic plain) while the one numbered 6 occurred in Bihar (middle Indo-Gangetic plain).

Understanding the cause of such localised Railway accidents require a good Traffic analysis of IR. The present analysis has been divided into the following sections. Section 6.1 divides the IRN into major Trunk segments and all the analysis has been carried out onto these segments only. Section 6.2 deals with the present traffic scenario of the above segments in terms of volume of traffic and headway analysis. Section 6.3 analyses the increase of volume of traffic over the last 2 decades. In section 6.4, a small experiment has been carried out to simulate the movement of the trains over the tracks in real-time and determining the most congested segments where there would be high probability of accidents owing to some technical glitch.

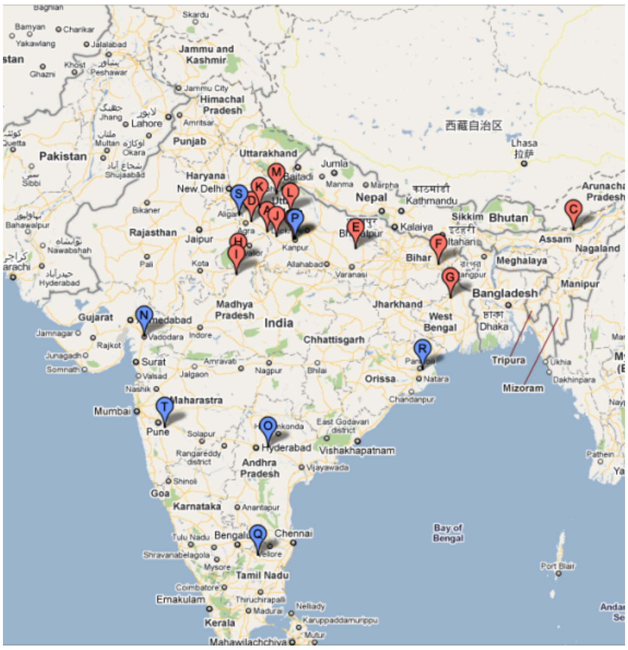

Figure 6.1: Sites of railway accidents involving derailment or collision among trains during 2005-2010. Sites of accidents in 2010 marked in red, those during 2005-2009 marked in blue

Figure 6.2: The Indo-Gangetic plain, showing its upper, middle and lower parts

6.1 Trunk-routes and trunk-segments

To analyze the IR traffic in various geographical regions, we consider the most important 'trunk-routes' in the Indian Railway Network (IRN), which are the high-speed, high-capacity routes (e.g. with replicated tracks) connecting major cities, and are mostly used by express train-routes; the trunk-routes are as indicated in the schematic map of the IRN (Fig. 6.3).

We divide the trunk-routes in the entire IRN into a set of 54 disjoint 'trunk-segments', where each trunk-segment is the portion of a trunk-route between two major junction stations (some examples of trunk-segments are given in Table 6.2). We consider a train-route to be using a given trunk-segment only if at least two stations within that trunk-segment are scheduled halts on the train-route. It is to be noted that more stringent conditions can be applied, e.g. a train-route can be considered to use a trunk-segment only if *all* stations on the segment are scheduled halts on the train-route. However, we desire to count those train-routes as well that travel along only a part of the trunk-segment. Moreover, several express train-routes in IR have scheduled halts at only a few major stations, and hence is not likely to stop at many stations in the same trunk-segment [1].

We study the characteristics of traffic in these trunk-segments, using the following metrics:

1. **Present Traffic**: The current traffic in the trunk-segments in analysed using the following metrics:

    (a) Average number of trains using a segment per day (section 6.2.1)

[1] This is actually the reason why we focus on the trunk-routes in our analysis, and not all possible edges / paths in the schematic map of the IRN (Fig. 6.3) - many of the express trains stop at few important stations only, and the IR web-site / TAAG time-tables list only the scheduled stoppages in a particular train-route (*not* all stations through which a train passes). Hence, even if we consider all edges in the network, the data-set will grossly under-estimate the number of trains travelling along the relatively less important edges. We find that 1561 out of the 2195 train-routes in our dataset (i.e. 71%) use at least one trunk-route. Hence it seems sufficient to focus on the traffic in the trunk-segments.

   (b) Headway, which is a metric of the distance or time-interval between consecutive vehicles using the same route in a transportation system (section 6.2.2)

   (c) Average Run-time delay of trains travelling along a segment (section 6.2.3)

2. **Growth in Traffic over the last two decades**: We study increase in traffic in different trunk-segments (section 6.3).

3. **Simulation of traffic-flow:** We simulate the flow of rail-traffic on different trunk-segments according to the IR schedule (section 6.4), in order to find how frequently trains travelling on a trunk-segment come in close spatial proximity with one another.

Some points to note in analysis of IR traffic:

There are about 45 'privileged' train-routes in IR (named "Duronto" and "Rajdhani"), that do not have scheduled halts for long distances within the route. For instance, the high-speed "Duronto" express trains which have been introduced since 2009 to serve as non-stop point-to-point links between metropolitan cities, have no scheduled intermediate stoppages; hence the IR website specifies only the source and destination stations for such trains. Consequently, our analyses (which considers a train to use a trunk-segment only if the train halts at at least 2 stations within the segment) is likely to under-estimate the number of trains for some of the trunk-segments. We attempt to reduce such cases by selecting only the major junction stations (at which most trains are likely to halt) to define each trunk-segment.

Another point to note is that most of the metropolitan cities in India are served by multiple railway stations (e.g. the capital Delhi is served by 7 stations, metropolis Kolkata is served by 3 stations, and so on) and trains starting from (going to) these cities can start (end) at any of these stations. However, all trains travelling from (to) a particular metropolitan city use the same trunk-routes except for few tens of kilometres in the immediate

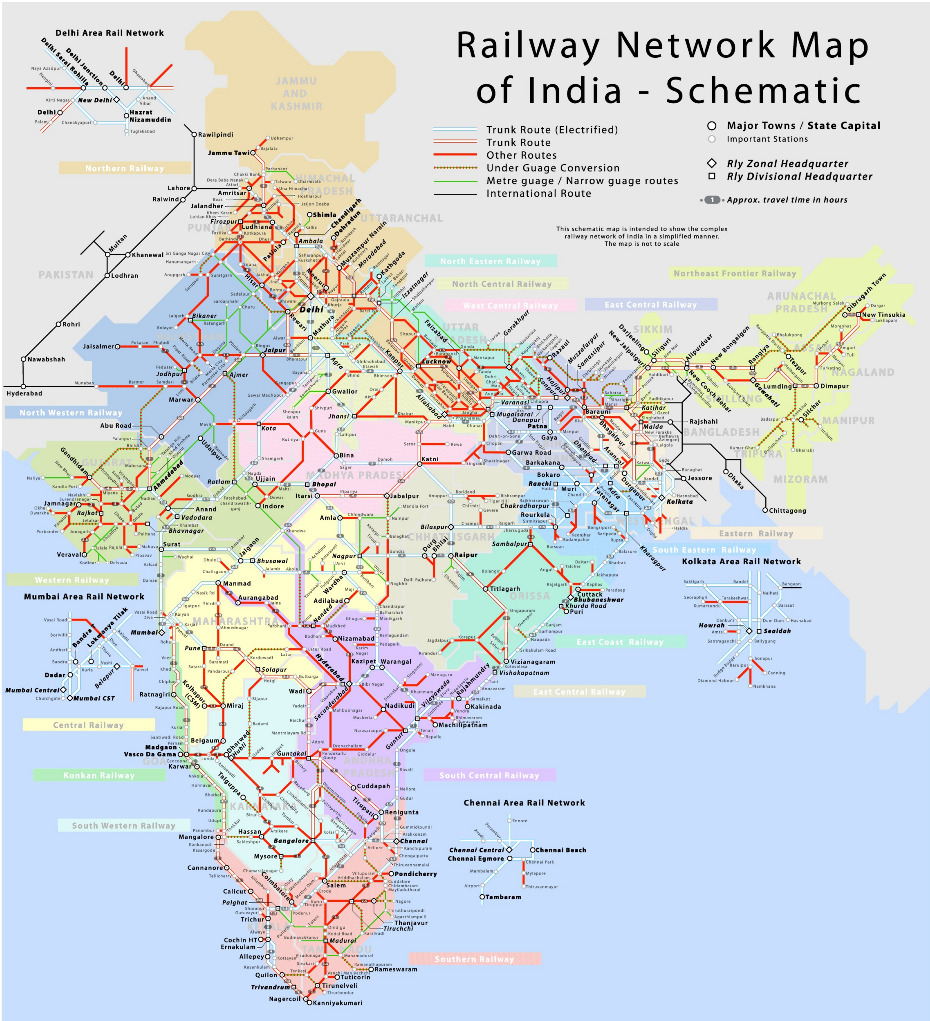

Figure 6.3: Schematic map of Indian Railway Network showing trunk-routes

proximity of the city (e.g. almost all trains travelling to any of the 7 stations serving Delhi from the city of Kanpur use the Delhi-Kanpur trunk-segment, and branch towards a particular station within Delhi only after they reach within few tens of kilometres from the city of Delhi). Hence in our analysis, we represent each metropolitan city as a single station while measuring the traffic on the trunk-segments linked to metropolitan cities.

A third point to note is that there exist a large number of *daily* train-routes in IR which require more than 24 hours to reach the destination station; for such routes, there may be *multiple physical trains* running on the same route at a given point of time. For example, consider a daily train-route that starts from the source station at 12:00 on each day and reaches the destination station at 17:00 on the next day. At 14:00 on any given day, there will be two trains running on this route, one having started the previous day and nearing the destination, and the other having started on the given day. Our datasets for IR traffic on a particular day consider all such trains individually (since all of them contribute to the traffic in some trunk-segment). However, two physical trains running on the same train-route are always separated from each other by a large geographical distance at any given point of time (since they start at an interval of 24 hours), hence our analysis will count at most one physical train out of these, as using a particular trunk-segment on a given day of the week.

6.2 Analysis of present IR traffic

6.2.1 Volume of traffic on trunk-segments

We measure the number of trains using each trunk-segment individually for all 7 days of the week, and identify the 22 trunk-segments which are used by the highest average number of trains per-day in Table 6.2. 13 out of these 22 trunk-segments are distributed over two specific geographical regions, as described below.

1. **Trunk-segments in the middle and lower parts of the Indo-Gangetic plain:** As many as 8 out of the 22 trunk-segments handling highest per-day IR traffic are in the Indo-Gangetic plain, as indicated in Table 6.2. From the view-point of IR, this region is one of the most important ones in the country, since it connects the eastern parts of India and the metropolis Kolkata with Delhi, the capital city of India. In particular, it can be noted that the trunk-segment having by far the highest average per-day traffic (Delhi-Tundla-Etawah-Kanpur) is in the

middle Indo-Gangetic plain, and as many as 3 of the accidents in 2010 have occurred along this segment.

2. **Trunk-segments in the central parts of India:** 5 out of the 22 trunk-segments in Table 6.2 are located in central India. Due to their central location, these trunk-segments handle a large number of trains going from southern parts of India to the northern parts, and from western to eastern parts of the country. Several trunk-routes connecting metropolitan cities pass through these trunk-segments, some examples being the Delhi-Mumbai route through the Jhansi-Bina-Bhopal (ranked 6 in Table 6.2) and the Agra-Gwalior-Jhansi segment (ranked 9), the Kolkata-Mumbai trunk-route through the Katni-Jabalpur-Itarsi segment (ranked 8), and so on.

The rest of the trunk-segments carrying high traffic are distributed in different regions - these link metropolitan cities to different parts of the country. For instance, the Amritsar-Jalandhar-Ambala (ranked 5) and Ambala-Panipat-Delhi (ranked 10) segments link the capital city Delhi with northern parts of India; the Ahmadabad-Vadodara-Surat (ranked 2) and Surat-Mumbai (ranked 16) trunk-segments link metropolis Mumbai with northern / north-western India while the Bhusaval-Manmad-Kalyan segment (ranked 3) links Mumbai with central India; the Kolkata-Kharagpur segment (ranked 17) links metropolis Kolkata with south-eastern parts of India; Vishakhapatnam-Vijayawada (ranked 14) and Vijayawada-Guntur-Chennai (ranked 19) segments link metropolis Chennai with eastern parts of India.

6.2.2 Headway analysis for traffic trunk-segments

In this section we perform a more fine-grained temporal analysis of the IR traffic in various trunk-segments by analyzing the distribution of traffic over time within a day. For this, we use the concept of 'headway' which is defined as the time-interval (or distance) between two consecutive vehicles travelling along the same route. Headway is measured with respect to a reference point

- conceptually, a timer is started when a vehicle passes the reference point and the time elapsed until the next vehicle passes the reference point is measured.

A smaller headway signifies a more frequent service, e.g. metro rail systems operate with headways on the order of 1 to 5 minutes, and cars on a highway have as little as 2 seconds headway between them. For a railway system, the headway is usually measured in parts of an hour. The desired (or safe) headway for a transportation system may be decided by various safety criteria, but the essential idea is to allow sufficient time to a vehicle to safely stop behind the vehicle in front of it in case the vehicle in front has to stop unexpectedly at some point. For example, in a railway system, if a train stops at an unexpected spot, then a train travelling behind it on the same track must be stopped at a signal; hence the the safe headway of a particular track is decided by factors like the distance between consecutive signals along the track, the distance required by a train to stop after the brakes are applied, and so on.

To compute the headway for traffic on a particular trunk-segment s, we selected a station T within that segment (as the reference point) and considered the sequence of time-instants at which trains using segment s pass through station T. The intervals between these time-instants thus give the headway for traffic on segment s at different times within a day. We measure the average headway for segment s over a day by averaging these time-intervals. We also plot the variation of headway over a day for a few trunk-segments. We repeated this analysis individually for each day of the week, and observed that the statistics of headway are almost similar for each day across all the segments. Hence we report the headway statistics for a particular day (Monday).

Table 6.3 shows the top 20 segments ranked in increasing order of average headway over a day. Most of the segments with low headways are again from the Indo-Gangetic plain or from central parts of India (3 out of the top 5 are from the Indo-Gangetic plain).

Further, it can be observed from Table 6.1 that most of the accidents in 2010 due to collisions between trains have occurred in early hours of the day. Considering this, we also report the average headway between midnight and 7 a.m. of the trunk-segments in Table 6.3, and the segments for which the headway between midnight and 7 a.m. is lesser than the average headway are marked. Out of the 5 segments located in northern India and upper Indo-Gangetic plain in Table 6.3, 3 have lesser than average headway between midnight and 7 a.m. Since the northern regions of India experience dense fog in the early hours of the day during the winter months (which greatly reduces visibility), it can be detrimental for trains to travel with small headway in these segments in the early hours of the day. In specific, the Delhi-Tundla-Etawah-Kanpur segment has the lowest headway between midnight and 7 a.m., and there have been 3 accidents along this very segment in January 2010 (the ones numbered 1, 2 and 4 in Table 6.1), the reported cause being that the engine-drivers did not react to signals amidst dense fog conditions.

Fig. 6.4 shows the variation of headway over a day for some of the trunk-segments (those marked with (*) in Table 6.3). In these figures, the x-axis represents the time of day measured in minutes elapsed since 00:00 (i.e. midnight), and the y-axis represents the headway in minutes. Each impulse is drawn at the time (x-value) at which a train using segment s passes through the reference station for the segment, and the height of the impulse denotes the time in minutes before the next train using segment s passes through the reference station. In other words, the point (x, y) denotes that a train using segment s passes the reference station at time x, and the next train using segment s arrives at the reference station after y minutes (hence the next impulse is drawn at time $x + y$). From the plots, it's evident that for Delhi-Kanpur and Dhanbad-Howrah Segments (both in Indo-Gangetic Belt), the lower Headways are more skewed in morning and late evening hours, which is not acceptable since lower visibility enhances the chances of accidents.

Fig. 6.5 shows the cumulative distribution of headway for the segments in Fig. 6.4. Interestingly, here also fraction of times with lower headways is maximum in Delhi-Kanpur and Ahmadabad-Surat segments.

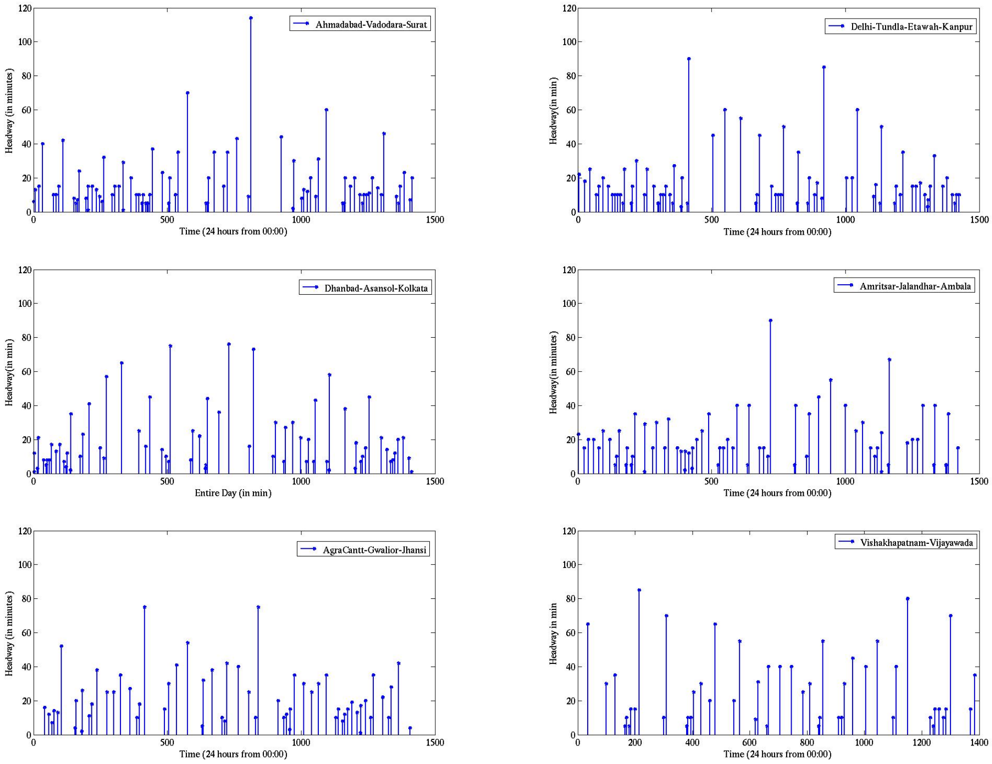

Figure 6.4: Plots showing variation of headway with time-of-day (measured in minutes since 00:00) for few selected trunk-segments: Ahmadabad-Surat (western India), Delhi-Kanpur (upper Indo-Gangetic plain), Dhanbad-Kolkata (lower Indo-Gangetic plain), Amritsar-Ambala (northern India), Agra-Jhansi (central India) and Vishakhapatnam-Vijayawada (southern India)

6.2.3 Run-time delay over trunk-segments

The average run-time delay for the train-routes in IR can be known from the web-site `www.indiarailinfo.com` (though not the official web-site, this web-site is one of the most comprehensive sources of data on Indian Railways). For some of the trunk-segments handling large volumes of traffic, we collected the delay information for all the train-routes that use these segments (as identified in Sec 6.2.1), and we compute the average delay of all train-routes that use a particular segment. Table 6.4 reports the average delay for some

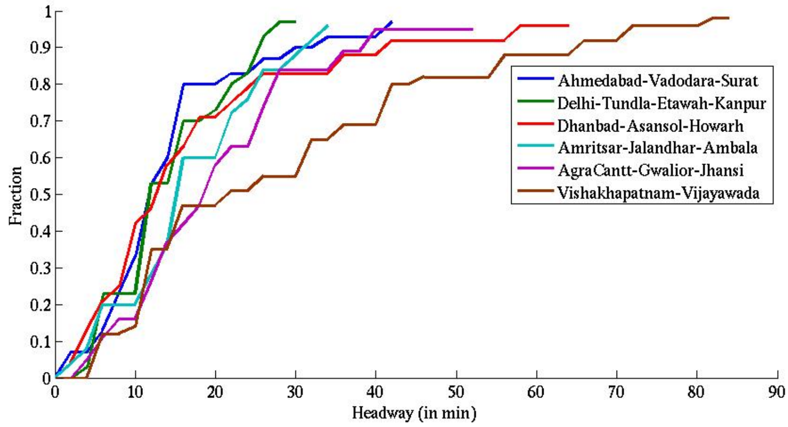

Figure 6.5: Cumulative distribution of headway for the same trunk-segments as in Fig. 6.4; x-axis shows headway values (in minutes) and y-axis shows the fraction of times when headway values are greater than x-value.

of the segments. The average delay is highest for the Delhi-Kanpur segment, which, along with the fact that this segment also handles the maximum volume of traffic (see Table 6.2), indicates heavy congestion in this segment.

6.3 Evolution in IR traffic over the last two decades

As stated in Section 3.1, we collected the list of express train-routes and stations on each route in the years 1991, 1997, 2000, 2005 and 2009 from the "Trains At A Glance" (TAAG) time-tables of the corresponding years. The number of train-routes and stations in the dataset for each year is summarized in Table 6.5. Since the time / day of week for each train to reach any given station could not be obtained for the older years, we cannot compute per-day statistics for all years. Hence, in this section, we only consider the total number of train-routes using a given trunk-segment in each of these years.

6.3.1 Increase in newly constructed Tracks

The increase in the number of stations over the years (as noted in Table 6.5) may be due to (i) introduction of new intermediate stations in *existing* express train-routes, or (ii) construction of *new railway tracks* in new geographical regions. Since we could not collect any statistics on the construction of new tracks in Indian Railways, we estimate the increase in the number of *tracks* in the period between two years y_{prev} and y_{later} as follows.

Evidently, any increase in the number of stations in those train-routes which exist in the data-set for both years can be ignored (i.e. these stations do not reflect construction of new tracks). Among the new train-routes that appear in the data-set for y_{later} (and do not exist in the data-set for y_{prev}), we consider a train-route to run over a newly constructed track if and only if the route includes at least two consecutive stations that did not exist in the data-set for the year y_{prev}. Table 6.6 shows the increase in number of train-routes and the estimated increase in number of train-routes that use new tracks over the last two decades. It is evident that even with this liberal definition of a new track, only a small fraction of the newly introduced train-routes appear to be using newly constructed tracks, moreover this fraction is decreasing with time (though it has improved in 2005-2009 compared to 2000-2005). This shows that while the number of train-routes in IR has increased rapidly, there has been relatively very little construction of new railway-tracks. This is leading to over-utilization of existing railway-tracks, resulting in congestion and trains having to travel at reduced speeds [1].

The IR authorities have also admitted that "since 1950-51, route-kilometers has increased by just 18% and track-kilometers by 41% even though in the same period freight and passenger traffic had gone up by more than 12 and 11 times respectively

6.3.2 Increase in traffic through trunk-segments

To estimate the increase in IR traffic along a given trunk-segment s over the last two decades, we measure the percentage increase in the number of train-routes using s in the year 2009, as compared to that in 1991. More specifically, for segment s we measure the increase in traffic between the years 1991 and 2009 as

$$\Delta T_{1991,\,2009}(s) = \frac{TR_{2009}(s) \;-\; TR_{1991}(s)}{TR_{1991}(s)} \times 100\% \tag{6.1}$$

where $TR_y(s)$ is the number of train-routes using trunk-segment s in the year y (i.e. as obtained from the TAAG of year y).

Table 6.7 shows the increase in IR traffic for the 22 trunk-segments carrying the most traffic in the present IRN (as identified in Section 6.2.1). It is evident that some of the trunk-segments in the upper and middle Indo-Gangetic plain have experienced phenomenal increase in traffic.

Fig. 6.6 shows the percentage increase of traffic (i.e. number of train-routes using a segment) in the years 1997, 2000, 2005 and 2009 with respect to the traffic in 1991 for a few selected trunk-segments. We report results for two segments each from five different zones: the Indo-Gangetic plain (shown in red), western India (shown in green), central India (in blue), Southern India (in black), northern India (in orange). In agreement with Table 6.7, the trunk-segments in the Indo-Gangetic plain show the highest increase of traffic.

6.4 Simulation of traffic-flow

In this section, we simulate the actual flow of traffic in the trunk-segments according to the Indian Railways schedule, in order to find how frequently trains travelling on the same trunk-segment come in close spatial proximity with one another (which would indicate higher congestion and higher probability of collisions in case of human errors such as failure of the engine-driver

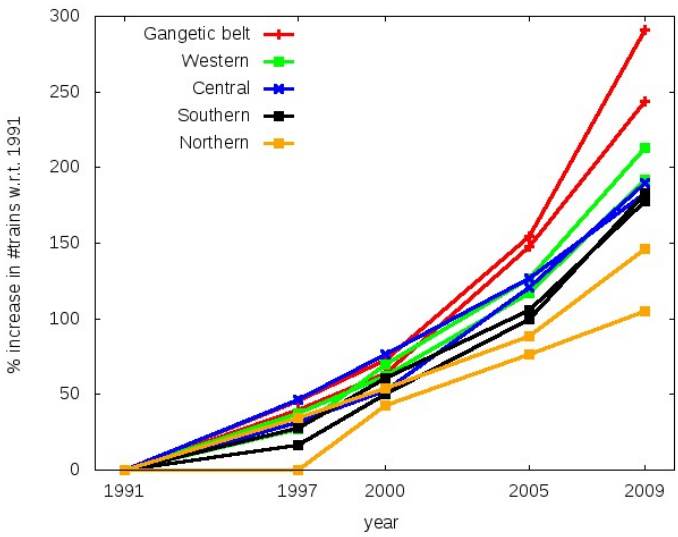

Figure 6.6: Percentage increase in IR traffic in 1997, 2000, 2005 and 2009 with respect to traffic in 1991, for the segments marked with (*) in Table 6.7.

to react to signals). We simulate the traffic-flow according to the 'Block System' of train movement used in IR, which has been explained below.

6.4.1 Block System in Railways

Railways in India and several other countries as well follow the 'Block System'in which a railway track is considered as a series of 'block sections' (or simply blocks) such that when one train is occupying a block, no other train is allowed to enter that block (in order to maintain a safe headway between trains travelling on the same track). At each end of a block, there are stations or signals which control the traffic entering into the block from that end - when a train has entered a block, no other train is allowed to enter the block until it has left the block.

Historically, a block used to be the section of a track between two consecutive stations. In the present times, a long stretch between two stations is usually fragmented into two or more blocks (called "Intermediate Block Sections" or IBS) in order to increase track utilization, and signals are placed between every pair of consecutive blocks to control the traffic entering the blocks. In the present days, the signals are controlled almost entirely by the movement of trains on the blocks as detected by track circuits, and this system is known as "Automatic Block Signalling". In Indian Railways, block lengths are of the order of 4 to 8 kilometres under normal circumstances

It is to be noted that there can be one or more rail-tracks in a block - a single-line block has just one track, which is used for traffic in both directions. In a double-line block, there are two tracks of which each may be dedicated to traffic in one direction or (more commonly), both the lines can be used for bidirectional traffic. Sections with heavy traffic can have more than two tracks as well.

When a train approaches an automatic signal that is in the 'stop' state, the train must come to a standstill to the rear of the signal. But in most cases, the rule is that after waiting at the signal for some time (normally 1 or 2 minutes), the train may pass the signal at a low speed-limit (typically restricted to 15 km/hr) *even if the signal remains in the 'stop' state*, with the engine-driver alert for other vehicles on the track This rule has been imposed in order to prevent unnecessary delay in the flow of traffic. However, drivers often accelerate beyond the safe speed-limit assuming that the track is clear of other trains, and this may lead to collisions if a train is stopped at some unexpected spot, and another train moving on the same track crosses a signal at a speed higher than the safe-limit.

6.4.2 Simulating traffic-flow using the block system

We simulate the flow of traffic in the trunk-segments assuming the block system described above. In the simulation, we assume that a train is allowed to continuously proceed according to its schedule (i.e. without being stopped by any signal), and we study how frequently trains would have come in close spatial proximity with one another (i.e. multiple trains in the same block) if all trains would move exactly according to their schedule.

It is to be noted that in reality, if a block is occupied by a train, other trains on the same track would normally be stopped from entering the block by signals (as described above). However, failure of drivers to react to signals may lead to collisions between trains running on the same track. Hence this simulation of traffic-flow in a particular trunk-segment aims to estimate how frequently multiple trains would be present in the same block if each train travels according to its schedule. A higher number of such instances indicates a higher probability of collisions in the event that a driver fails to react to a signal.

For a given trunk-segment s, we know its length l_s (in kilometres) and the trains which use the segment. From the IR time-table, we note, for each train T using s, the exact time of day when T enters s and the time at which it exits s. Hence we know the period of time t_s^T (measured in minutes) during which each train T is scheduled to run in segment s. We assume a fixed block-length l_b (for instance $l_b = 6$ kilometres) and consider segment s as a sequence of $n_s = \frac{l_s}{l_b}$ number of blocks. We further assume that each train T travels at a uniform speed throughout, which implies that T requires $t_b^T = \frac{t_s^T}{n_s}$ time-units (minutes) to traverse each block in the segment. It is to be noted that a train may not traverse the whole length of segment s. Since a segment is a sequence of stations, a train can enter s at an intermediate station and / or leave s at an intermediate station. In such cases, we consider the train to be traversing a fraction of the blocks in the segment.

The simulation of traffic on a particular trunk-segment s proceeds as follows. Each time-step is considered to be 1 minute. For each train T using segment s, T enters the segment at its scheduled time, and sequentially traverses the blocks in the segment (i.e. T stays at each block for t_b^T time-steps before going to the next block) until it reaches the end of the trunk-segment. The simulation continues until all trains using segment s have completed their traversal of the segment. The total number of time-steps in the simulation of traffic for some selected trunk-segments is reported in Table 6.9. We count the number of time-steps during which there are more than a given number (say, k) of trains in at least one block. Here k captures the notion of the number of parallel tracks in a block - a block having k

tracks can accommodate up to k trains simultaneously, but at least one train has to be stopped (at a signal) from entering the block if more than k trains are scheduled to be in the block at a certain point of time.

We performed the simulation individually considering the traffic for each day of the week, and observed that the results are almost similar for each day. Hence we report the results for traffic on a particular day (Monday). Table 6.9 gives the results of the simulation for traffic on a few selected trunk-segments, one each from different geographical regions - we select that trunk-segment which has the minimum headway among all trunk-segments in the said geographical region (see Table 6.3). The number of time-steps during which at least one block is scheduled to have more than k trains is reported for $k = 2$, 3, 4, assuming a block length $l_b = 6$ kilometres. Experiments using block-lengths $l_b = 4$ and 8 also produced similar trends.

It is evident that if all trains were to travel strictly according to the IR schedule, then the trains running on the trunk-segments in the Indo-Gangetic plain would come into close spatial proximity with one another much more frequently as compared to trains running on trunk-segments in other geographical regions. This implies a higher probability of collisions between trains running on the segments in the Indo-Gangetic plain in the event of an engine-driver failing to react to a signal.

In all the tables 6.8, 6.9, 6.10, Indo-Gangetic segments (like Delhi-Kanpur and Dhanbad-Asansol-Kolkata) has the maximum number of frequent overlapping blocks indicating that these segments are very congested with trains being too close to eachother occupying consecutive blocks. In these circumstances, faulty breaking of signals by a single train may cause a devastating rail accident, which has also been observed to be true.

Rank	Trunk-segment	Average daily traffic	Geographical Location
1	Delhi-Tundla-Etawah-Kanpur	104.286	upper Indo-Gangetic plain
2	Ahmadabad-Vadodara-Surat	86.286	western India
3	Bhusaval-Manmad-Kalyan	81.286	western India
4	Delhi-Mathura-Agra	80.571	upper Indo-Gangetic plain
5	Amritsar-Jalandhar-Ambala	79.143	northern India
6	Jhansi-Bina-Bhopal	74.286	central India
7	Dhanbad-Asansol-Kolkata	73.286	lower Indo-Gangetic plain
8	Katni-Jabalpur-Itarsi	67.714	central India
9	Agra-Gwalior-Jhansi	64.714	central India
10	Ambala-Panipat-Delhi	60.857	northern India
11	Kanpur-Allahabad	59.429	middle Indo-Gangetic plain
12	Mughalsarai-Ara-Patna	59.143	middle Indo-Gangetic plain
13	Ujjain-Bhopal-Itarsi	56.857	central India
14	Vishakhapatnam-Vijayawada	52.571	southern India
15	Sonpur-Barauni-Katihar	52.143	middle Indo-Gangetic plain
16	Surat-Mumbai	46.857	western India
17	Kolkata-Kharagpur	44.857	eastern India
18	Allahabad-Mughalsarai	41.714	middle Indo-Gangetic plain
19	Itarsi-Bhusaval	41.714	central India
20	Wardha-Kazipet-Hyderabad	41.714	southern India
21	Vijayawada-Guntur-Chennai	40.714	southern India
22	Lucknow-Varanasi	39.143	middle Indo-Gangetic plain

Table 6.2: Top 22 trunk-segments in IRN, in decreasing (non-increasing) order of average per-day number of trains using the segment. Trunk-segments are indicated by the two end-stations and intermediate stations in some cases to resolve ambiguity.

Rank	Trunk-segment	Avg. headway (minutes)		Geographical location
		over a day	before 7 a.m.	
1	Ahmadabad-Vodadara-Surat (*)	17.925	13.96 ↓	western India
2	Delhi-Tundla-Etawah-Kanpur (*)	19.1	**13.67** ↓	upper Indo-Gangetic plain
3	Delhi-Mathura-Agra	20.22	26.42	upper Indo-Gangetic plain
4	Dhanbad-Asansol-Kolkata (*)	20.44	17.41 ↓	lower Indo-Gangetic plain
5	Amritsar-Jalandhar-Ambala (*)	20.77	16.52 ↓	northern India
6	Agra-Gwalior-Jhansi (*)	22.44	19.63 ↓	central India
7	Bhusaval-Manmad-Kalyan	23.04	20.39 ↓	western India
8	Katni-Jabalpur-Itarsi	23.81	26.33	central India
9	Jhansi-Bina-Bhopal	25.08	25.62	central India
10	Ambala-Panipat-Delhi	26.471	28.20	northern India
11	Mughalsarai-Ara-Patna	26.63	31.53	middle Indo-Gangetic plain
12	Vishakhapatnam-Vijayawada (*)	28.265	26.30 ↓	southern India
13	Kanpur-Allahabad	28.33	20.52 ↓	middle Indo-Gangetic plain
14	Ujjain-Bhopal-Itarsi	28.47	33.33	central India
15	Wardha-Kazipet-Hyderabad	30.51	37.50	southern India
16	Surat-Mumbai	31.35	27.50 ↓	western India
17	Kolkata-Kharagpur	34.14	39.10	eastern India
18	Itarsi-Bhusaval	35.13	37.50	central India
19	Sonpur-Barauni-Katihar	37.73	58.40	middle Indo-Gangetic plain
20	Delhi-Moradabad	38.05	28.84 ↓	northern India

Table 6.3: Top 20 trunk-segments in increasing order of average headway over a day (based on Monday traffic). Also shown are the average headway in the first 7 hours of the day (i.e. between midnight and 7 A.M.), segments having lesser than average headway in the first 7 hours are marked by ↓. (*) indicates those segments for which variation of headway with time of day has been plotted in Fig. 6.4

Rank	Trunk-segment	Avg. Delay of routes using segment	Geographical Location
1	Delhi-Tundla-Etawah-Kanpur	1 hr 52 min	Upper Indo-Gangetic plain
2	Jhansi-Bina-Bhopal	1 hr 10 min	Central India
2	Ahmadabad-Vadodara-Surat	29 min	Western India

Table 6.4: Trunk-segments in IRN, in decreasing (non-increasing) order of Average Delay

Year	# train-routes	# stations
1991	750	548
1997	920	561
2000	1104	608
2005	1444	622
2009	1918	680

Table 6.5: Number of train-routes and stations for different years (as obtained from the "Trains At A Glance" time-tables)

Duration	# new train-routes	# new train-routes using new tracks
1991 - 1997	168	70
1997 - 2000	186	70
2000 - 2005	340	20
2005 - 2009	474	50

Table 6.6: Increase in number of train-routes and estimated increase in the number of train-routes that use new tracks

Rank	Trunk-segment	% increase in traffic wrt 1991	Geographical Location
1	Lucknow-Varanasi (*)	290.909	middle Indo-Gangetic plain
2	Delhi-Tundla-Etawah-Kanpur (*)	244.0	upper Indo-Gangetic plain
3	Sonpur-Barauni-Katihar	243.75	middle Indo-Gangetic plain
4	Kanpur-Allahabad	231.25	middle Indo-Gangetic plain
5	Ahmadabad-Vadodara-Surat (*)	213.333	western India
6	Bhusaval-Manmad-Kalyan (*)	191.667	western India
7	Katni-Jabalpur-Itarsi (*)	189.474	central India
8	Jhansi-Bina-Bhopal (*)	183.333	central India
9	Wardha-Kazipet-Hyderabad (*)	183.333	southern India
10	Mughalsarai-Ara-Patna	180.0	middle Indo-Gangetic plain
11	Vishakhapatnam-Vijayawada (*)	177.778	southern India
12	Itarsi-Bhusaval	158.824	central India
13	Vijayawada-Guntur-Chennai	156.522	southern India
14	Allahabad-Mughalsarai	152.632	middle Indo-Gangetic plain
15	Amritsar-Jalandhar-Ambala (*)	146.154	northern India
16	Ujjain-Bhopal-Itarsi	130.769	central India
17	Dhanbad-Asansol-Kolkata	125.0	lower Indo-Gangetic plain
18	Delhi-Mathura-Agra	119.355	upper Indo-Gangetic plain
19	Surat-Mumbai	116.667	western India
20	Kolkata-Kharagpur	110.0	eastern India
21	Ambala-Panipat-Delhi (*)	104.762	northern India
22	Agra-Gwalior-Jhansi	103.571	central India

Table 6.7: Top 22 trunk-segments in IRN, in decreasing (non-increasing) order of % increase in IR traffic between 1991 and 2009. (*) indicates those segments for which the increase in traffic is plotted in Fig. 6.6

Trunk-segment	k=2	k=3	k=4	Geographical location
Delhi-Tundla-Etawah-Kanpur	189	5	0	upper Indo-Gangetic plain
Ahmadabad-Vadodara-Surat	83	3	0	western India
Dhanbad-Asansol-Kolkata	60	1	0	lower Indo-Gangetic plain
Vishakhapatnam-Vijayawada	58	2	0	southern India
Amritsar-Jalandhar-Ambala	35	0	0	northern India
Agra-Gwalior-Jhansi	33	0	0	central India

Table 6.8: Simulation of rail-traffic on few selected trunk-segments assuming **block length = 4 km**: number of time-steps during simulation when at least one block contains more than k trains (see text for details)

Trunk-segment	k=2	k=3	k=4	Geographical location
Delhi-Tundla-Etawah-Kanpur	270	21	0	upper Indo-Gangetic plain
Dhanbad-Asansol-Kolkata	203	15	0	lower Indo-Gangetic plain
Ahmadabad-Vadodara-Surat	112	8	0	western India
Vishakhapatnam-Vijayawada	98	0	0	southern India
Amritsar-Jalandhar-Ambala	89	0	0	northern India
Agra-Gwalior-Jhansi	49	1	0	central India

Table 6.9: Simulation of rail-traffic on few selected trunk-segments assuming block length 6 km: number of time-steps during simulation when at least one block contains more than k trains. Column 2 shows the total number of time-steps in the simulation (see text for details)

Trunk-segment	k=2	k=3	k=4	Geographical location
Delhi-Tundla-Etawah-Kanpur	332	27	0	upper Indo-Gangetic plain
Dhanbad-Asansol-Kolkata	293	17	0	lower Indo-Gangetic plain
Amritsar-Jalandhar-Ambala	246	36	2	northern India
Ahmadabad-Vadodara-Surat	243	29	0	western India
Vishakhapatnam-Vijayawada	190	5	0	southern India
Agra-Gwalior-Jhansi	78	9	0	central India

Table 6.10: Simulation of rail-traffic on few selected trunk-segments assuming **block length = 8 km**: number of time-steps during simulation when at least one block contains more than k trains (see text for details)

Chapter 7

Conclusion

7.1 Topological Study

In this thesis, we studied the Indian Railway Network as a weighted complex network of stations, where the edge-weights represent the amount of traffic between two stations. We observed that the IRN has exponential distributions of node-connectivity and traffic-flows. Also the major stations (high-degree nodes) tend to be linked among themselves and most of the traffic in the IRN flows among these high-degree nodes.

Our analysis brings out certain drawbacks in the IRN, which are as follows. The node-strengths (weighted degree) grow faster compared to node-degrees in the IRN (fig. 4.4) implying that the construction of new links between stations has been significantly less than the introduction of new trains along existing links. Considering the limited capacity of links to handle trains, this shows the need for construction of new links among stations [1].

The correlation of edge-weights with the degrees of the adjacent nodes (fig. 4.5) corroborates another reported cause for concern in the present-

[1]This has been recognized by the Indian Railways authority as well, and it has been announced [1] that 25,000 kilometres of new railway-tracks would be constructed by 2020, which is far greater than the average rate of construction of tracks till now.

day IRN - traffic on the trunk-routes between the large cities far exceeds the planned capacity, which means that trains have to travel more slowly and the railway tracks wear out faster than intended [2, 1]. Hence new train-routes can be introduced to connect the larger cities; also, the links in the existing trunk-routes should be replicated to handle the large amounts of traffic. We also identify some of the stations that handle large amounts of traffic (fig. 4.10). The infrastructure at these stations should be improved to ease the congestion in the network. Thus the study provides several guidelines for improving the performance in the IRN.

7.2 Study on Accidents

Our analysis on Railway accidents brings out the fact that the there has been unbalanced rise in rail-traffic over several years in the Indo-Gangetic regions whereas less emphasis has been given on constructing newer routes/infrastructure in these areas leading to tremendous pressure or over-saturation of rail tracks.

It is to be noted that the Indo-Gangetic plain is one of the most densely populated regions of the country, hence the demand for transportation of passenger and freight traffic is also huge in this region, and is likely to rapidly increase with the growing population. Hence possibly the increase in the IR traffic is required and justified is some way. However, some regional / political bias may also have been at work, specially since the ministers in charge of Indian Railways have been mostly from the states of West Bengal, Bihar and Uttar Pradesh (i.e. the states in the Indo-Gangetic plains) over the last two decades, so a large number of new trains have been introduced every year in this region. This may have led to unplanned increase in traffic which has now exceeded the 'safe' limit considering the available resources in the region. Though the immediate cause for accidents can be various, unbalanced rise in traffic is one of the primary contributing factors. For example, high amounts of traffic, lower headways, enhanced Block-occupancy etc cause the IR-employees (e.g. driver, those in charge of signalling, maintaining the rail-tracks, etc) to be over-burdened, thus raising the chances of human error which may lead to accidents.

Again, dense fog in the early hours of the day (specially in winter) often reduce visibility, hence trains running on the same track with very low headway are more probable to ram into one another if the engine-driver overlooks a signal. This is the reported immediate cause of the accidents in the Delhi-Tundla-Kanpur trunk-segment which handles the highest number of trains per day and has the smallest headway between trains in the early hours of the day according to our analyses. It is also reported that an automated signaling system was experimentally launched in the Delhi-Tundla-Etawah-Kanpur trunk-segment in order to cope with the high amount of traffic on this segment, and malfunctioning of this signaling system was the immediate cause for most of the recent accidents in this segment.

On an optimistic note, the Indian Railway authorities have realized the urgent need to improve and increase the resources, and invest in safety-mechanisms. For instance, it has been declared [1] that 25,000 kilometres of new railway-tracks would be constructed by 2020, which is far greater than the average rate of construction of tracks till now. Prompted by the intolerable number of accidents in 2010, the IR authorities have decided to introduce centralized electrical and electronic interlocking systems on all important routes to check recurrence of accidents due to signal failure or overshooting of signals by drivers. The IR has also recently decided to install the indigenously developed "Anti-Collision Device" technology. These devices, attached to all engines, use the GPS satellite system for position updates, and network among themselves to take decisions for timely auto-application of brakes to prevent collisions.

Bibliography

[1] *Indian Railways Vision 2020.* Ministry of Railways, Government of India, Dec 2009.

[2] *Have India's clogged rail lines stalled progress?*, Jun 2010. `http://www.ndtv.com/news/india/indias-clogged-rail-lines-stall-economic-progress-31977.php`.

[3] G. Bagler. Analysis of the airport network of India as a complex weighted network. *Physica A*, 387(12):2972–2980, May 2008.

[4] A.-L. Barábasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[5] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *PNAS*, 101:3747–3752, March 2004.

[6] L. Cui-mei and L. Jun-wei. Small-world and the growing properties of the Chinese railway network. *Frontiers of Physics in China*, 2:364–367, July 2007.

[7] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *PNAS*, 102(22):7794–7799, May 2005.

[8] H. Hu and X.-F. Wang. Unified index to quantifying heterogeneity of complex networks. *Physica A*, 387:3769–3780, 2008.

[9] K. S. Kim, L. Benguigui, and M. Marinov. The fractal structure of Seouls public transportation system. *Cities*, 20(1):31–39, February 2003.

[10] C. C. Leung and H. F. Chau. Weighted assortative and disassortative networks model. *Physica A*, 378(2):591–602, May 2007.

[11] W. Li and X. Cai. Statistical analysis of airport network of China. *Phys. Rev. E*, 69(4):046106, April 2004.

[12] W. Li and X. Cai. Empirical analysis of a scale-free railway network in China. *Physica A*, 382(2):693–703, August 2007.

[13] C. Li-Ping, W. Ru, S. Hang, X. Xin-Ping, Z. Jin-Song, L. Wei1, and C. Xu. Structural properties of US flight network. *Chinese Physics Letters*, 20(8):1393, 2003.

[14] H. Lu and Y. Shi. Complexity of public transport networks. *Tsinghua Science & Technology*, 12(2):204–213, April 2007.

[15] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20):208701, October 2002.

[16] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[17] E. Ravasz and A. L. Barabasi. Hierarchical organization in complex networks. *Phys. Rev. E*, 67(2):026112, February 2003.

[18] K. Seaton and L. Hacket. Stations, trains and small-world networks. *Physica A*, 339(3-4):635–644, August 2004.

[19] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. Small-world properties of the Indian railway network. *Phys. Rev. E*, 67:036106, March 2003.

[20] J. Sienkiewicz and J. A. Holyst. Statistical analysis of 22 public transport networks in poland. *Physical Review E*, 72:046127, Oct 2005.

[21] Y. Wang, T. Zhou, J. Shi, J. Wang, and D. He. Empirical analysis of dependence between stations in Chinese railway network. *Physica A*, 388(14):2949–2955, July 2009.