Structure and Evolution of the Indian Railway Network

Saptarshi Ghosh, Avishek Banerjee, Naveen Sharma, Sanket Agarwal, Animesh Mukherjee, Niloy Ganguly

Department of CSE, IIT Kharagpur, Kharagpur 721302, India
Complex Systems Lagrange Lab., ISI Foundation, Torino 10133, Italy

SOLSTICE 2010
Outline

1. Current research on Transportation networks
2. Studying the IRN: Motivation
3. Network Construction
4. Topological properties of the present IRN
5. Conclusion
Current research on Transportation networks
Current Research on Transportation Networks

- Several recent studies on topological properties of
  - Airport networks (world-wide, China, India, ...)
  - Railway networks (China RN, Indian RN, Boston subway, Paris, ...)
  - Urban road networks (public transport systems of Poland, China, ...)

- Small-world properties reported for most transportation networks

- Degree distribution seen to vary: power-laws for most airport networks, exponential for Indian and Chinese RN

- Studies on evolution of transport networks: China airport network, Swiss road and railway networks, Indiana inter-urban network, ...
Studying the IRN: Motivation
Studying the IRN: Motivation

- One of the largest and busiest railway networks in the world, forming the backbone of transportation in India
  - Over 7000 stations and total route length of more than 64,000 km
  - Transports 20 million passengers and 2 million tonnes of freight daily

- Only previous study of topology of IRN from a complex networks perspective in [Sen, 2003]
  - IRN represented as unweighted network: nodes ⇔ stations, edge ⇔ existence of a train stopping at both nodes
  - Missing from the representation: dynamics of the traffic flow
Motivation (contd.)

- Essential to consider the amount of traffic flow along the connections of a transportation network
  - Represent as a **weighted network**: edge-weights $\Leftrightarrow$ amount of traffic on different links
  - Can yield observations that are undetected by metrics based on topological information alone

- **Evolution of IRN**: no previous study
  - Can serve as an indicator of the economic growth of the country

- Help in adopting effective extension policies e.g. more effective distribution of new trains, better planning of railway budget, ...
Network Construction
Collection of Data

- Data of present IRN collected from official website of Indian Railways (http://www.indianrail.gov.in/), in November 2009
  - Contains list of train-routes and stations at which each train-route halts
  - Only express trains and other long-distance train considered, and only those stations where at least one such train halts
  - 898 train-routes and 2702 stations

- Data for evolution of IRN:
Network Construction

Railway networks represented as
- Node $\Leftrightarrow$ station
- Edge $(u, v) \Leftrightarrow$ there exists a train-route that directly connects (i.e. halts at) both stations $u$ and $v$
- Edge-weight $w_{uv} \Leftrightarrow$ number of train-routes that halt at both stations $u$ and $v$

Undirected network, since all train-routes bidirectional

Edge-weights capture the dynamics of the traffic flow taking place in network
Topological properties of the present IRN
Degree Distribution

- Cumulative degree distribution decays exponentially
  \[ P(k) \sim e^{-0.0075k} \]
- Deviates at higher degrees \( \Rightarrow \) limited capacity of stations and high cost of adding new links
- Similar observations in [Sen, 2003]

Degree of node \( s \) \( \Leftrightarrow \) number of stations reachable from station \( s \) via a single direct train
Strength of a node: measure of transportation demands between this station and its neighbouring ones.

Cumulative strength distribution decaying exponentially with exponent 0.002.

Strength of node $s \Leftrightarrow$ total number of different journeys that can be undertaken from station $s$. 

Strength distribution $P(s)$ for the Indian Railway Network in 2009, with a cumulative decay of $\exp(-0.002s)$. 

The cumulative strength distribution decays exponentially with an exponent of 0.002.
Airport networks vs rail-station networks

- Power-law degree distributions reported for most airport networks e.g. world-wide airport network, airport network of India and China
- In contrast, exponential degree distribution of the Indian and Chinese railway networks
- Lack of preferential attachment in RN: trains stop at several smaller stations in their route between major end-stations
- Absence of few major hubs in RNs, whose failure might cause a complete breakdown in transportation
Airport networks vs rail-station networks

- Power-law degree distributions reported for most airport networks e.g. world-wide airport network, airport network of India and China

- In contrast, exponential degree distribution of the Indian and Chinese railway networks

- Lack of preferential attachment in RN: trains stop at several smaller stations in their route between major end-stations

- Absence of few major hubs in RNs, whose failure might cause a complete breakdown in transportation
Nodes with highest degree and strength

Top 10 stations w.r.t. node-degree

Top 10 stations w.r.t. node-strength
Greater flow of passengers along edges of higher weight

Cumulative distribution of edge-weights has exponential fit
\[ P(w) \sim \exp(0.12w) \]

Deviates from exponential fit for both very small and very large values of weight

Weight \( w_{uv} \leftrightarrow \) number of trains that directly connect stations \( u \) and \( v \) (halts at both stations)
Strength-Degree Correlation

- $s(k)$ increases rapidly with $k$ as a power-law
  - $s(k) \sim k^\beta$, $\beta = 1.383$
- Introduction of new trains on existing routes is more prevalent compared to construction of new routes (i.e., linking stations to new neighbours)

Correlation between degree $k$ and the average strength $s(k)$ of nodes having degree $k$
**Clustering coefficient**

- $cc(k)$ close to unity for small $k$: smaller stations served by very few train-routes, all stations on these routes linked to each other

- **power-law decay of $cc(k)$ for higher $k$:** major stations linked with several geographically distant stations in diverse parts of the country, which themselves are mostly not connected

- Similar results in [Sen, 2003]

- Average c.c. $cc(k)$ of nodes having degree $k$, as a function of $k$
Weighted Clustering Coefficient

- $cc^w(k)$: average weighted clustering coefficient of nodes of degree $k$ [Barrat, 2004]
- $cc^w(k)$ lies consistently above $cc(k)$: most of the traffic (edge-weights) accumulated on interconnected groups of stations
- Lower variation of $cc^w(k)$ compared to that of $cc(k)$: high-degree stations form interconnected groups with high-traffic links, thus balancing the reduced topological clustering

Variation of $cc(k)$ and $cc^w(k)$ with degree $k$, using logarithmic binning of $k$-values
Degree-degree correlations

- $k_{nn}(k)$ stable on the average over a significant range of degrees $\Rightarrow$ absence of major correlations among the nodes of different degrees
- Newman’s assortativity coefficient 0.058 $\Rightarrow$ weakly assortative
- Resilient to targeted attacks such as removing high-degree nodes (malfunctioning of stations due to natural disaster or terrorist activity)

$k_{nn}(k)$: average degree of neighbours of nodes having degree $k$, as a function of $k$
- $k_{nn}^w(k)$: average weighted degree of neighbours of nodes having degree $k$, as function of $k$

- $k_{nn}^w(k)$ shows pronounced assortative behaviour $\Rightarrow$ high-degree stations connect with other high-degree stations, and the amount of traffic (weight) along such links tend to be high as well

Variations of $k_{nn}(k)$ and $k_{nn}^w(k)$ with degree $k$, using logarithmic binning of $k$-values.
Evolution of IRN
Station-station network over the years

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

- Most metrics remain stable over the years
- Average node degree increases steadily
- Number of edges grows super-linearly in number of stations, following densification power-law $e(t) = n(t)^{1.5}$
Evolution of Degree Distribution

- Exponential cumulative degree distribution for all years: \( P(k) = \exp(-\alpha k) \)
- Absolute value of exponent \( \alpha \) decreases over the years, resulting in flatter distributions
  - A more homogeneous structure of the network w.r.t. node-degrees
  - Connectivity of stations improving consistently with time in IRN

Cumulative degree distribution for different years
Betweenness Centrality

- Exponential distribution $c_b(k) \sim \exp(\alpha k)$ for each year, with the value of $\alpha$ decreasing with time
- Betweenness centrality of nodes increases sharply with the node degree (number of directly linked stations)

Average normalized betweenness centrality $c_b(k)$ of nodes having degree $k$ as a function of $k$, for the years 1991, 1999 and 2009
Betweenness Centrality (contd.)

- Maximum value of normalized $c_b(k)$ falls with time
  - Given an arbitrary pair of stations $(u, \nu)$, several different shortest paths between $u$ and $\nu$ are coming into existence IRN with time
  - The fraction of these shortest paths passing through a particular node is getting reduced
  - Indicates improvement in connectivity among stations

Average normalized betweenness centrality $c_b(k)$ of nodes having degree $k$ as a function of $k$, for the years 1991, 1999 and 2009
Conclusion
Conclusion

- **Consistent advancement of the IRN**
  - Network exhibits an assortative topology, thus making it resilient against failure of nodes (stations)
  - IRN becoming more homogeneous w.r.t. node-degrees, implying improved connectivity among stations

- **Bottlenecks**
  - Construction of new connections between stations has been significantly less than the introduction of new trains along existing connections
  - High values of node-strengths indicate possibilities of congestion at major stations due to excessive traffic-densities
  - Need to identify nodes which are susceptible to congestion, and increase resources around such nodes
Bibliography

The architecture of complex weighted networks

Small-world properties of the Indian Railway Network
Phys. Rev. E 67(036106), 2003
Thank You