Hierarchical Organization of Railway Networks

Praveen R,[*] Animesh Mukherjee,[†] and Niloy Ganguly[‡]
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur, India – 721302
(Dated: May 21, 2008)

I. INTRODUCTION

In the recent years, networks have proved to be highly useful in modeling and characterizing complex systems that occur in various branches of science. In fact, almost any large-scale system, be it natural or man-made, can be viewed as a network of interacting entities, which is complex, irregular, and usually changing over time. One of the most important observations that researchers have repeatedly made in the past is that complex networks often exhibit hierarchical organization [2, 4–7, 10]. The nodes in the network can be divided into groups, which can be further subdivided in smaller (and possibly more cohesive) groups and so forth. In most of the cases these groups have been shown to correspond to certain functional units such as friendship communities in social networks [2], modules in biochemical networks [4, 5, 10] and sound patterns in consonant and vowel inventories [6, 7].

Railways are perhaps one of the most important means of transportation for any nation. In this context, it becomes extremely necessary to study the structural properties of **Railway Networks** (henceforth RN). The RN can be conceived of as nodes representing stations and a link between two nodes representing a direct train connection between them. The number of direct trains between two stations define the weight of the edge between the nodes representing those stations. Although there have been some attempts to explore the small-world properties of RNs for different countries [3, 9], none of them investigate the hidden structural patterns of these networks. Since railways play a very crucial role in shaping the economy of a nation it is important to study the above patterns, which in turn can not only be used for a more effective distribution of new trains but also for a better planning of the railway budget. Therefore, the primary objective of this work is to analyze the hierarchical organization of two RNs (for which we could collect data) namely, the **Indian Railway Network** (IRN) and the **German Railway Network** (GRN). Apart from the standard small-world properties like clustering coefficient, average path length, diameter we conduct a detailed community structure analysis of the two networks in order to capture the basis of their organization. Since the networks are weighted, we employ the Modified Radicchi et al. algorithm (henceforth MRad) (see [6–8] for references) to detect the community structures. We observe that the pattern of community formation reflects the underlying economic geography of the area/region with each community including within it a few hub nodes (i.e., important junctions). Furthermore, the inter-community hubs are well connected with each other.

The rest of the paper is laid out as follows. In section II we formally define the two networks, outline their construction procedure and present the basic structural properties. In the next section, we briefly review the important steps of the MRad algorithm and present the results of the community analysis for the two networks. We conclude in section IV by summarizing our contributions, pointing out some of the implications of the current work and indicating the possible future directions.

II. DEFINITION, CONSTRUCTION AND STRUCTURAL PROPERTIES OF THE TWO RAILWAY NETWORKS

Definition: RN can be represented as a graph $\mathrm{G} = \langle\, V_S,\ E\,\rangle$ where V_S is the set of nodes labeled by the stations and E is the set of edges connecting these stations. There is an edge $e \in E$ between two nodes if and only if there exists a direct train between the stations represented by these nodes. The weight of the edge e (also *edge-weight*) is the number of direct trains between the nodes (read stations) connected by e. We shall call this network of stations as **Sta**tion-Station **Net**work or **StaNet**. Consequently, the Indian RN is $\mathrm{StaNet}_{\mathrm{IR}}$ and the German RN is $\mathrm{StaNet}_{\mathrm{GR}}$. Figure 1 presents a hypothetical illustration of $\mathrm{StaNet}_{\mathrm{IR}}$ and $\mathrm{StaNet}_{\mathrm{GR}}$.

Construction: In order to construct $\mathrm{StaNet}_{\mathrm{IR}}$ we collected the data from *http://www.indianrail.go.in*, which is the official website of the Indian railways. The website hosts information of around 2764 stations and approximately 1377 trains halting at one or more of these stations. Consequently, the number of nodes in $\mathrm{StaNet}_{\mathrm{IR}}$ is 2764. For the German railways, we could collect only a small amount of data consisting of 80 stations (equivalent to the number of nodes in $\mathrm{StaNet}_{\mathrm{GR}}$) and around 100 trains passing through them from the regional timetable archival in the form of a compact disk – *Deutsche Bahn Electronic Timetable* (see *http://www.hacon.de/hafas_e/ce_promo.shtml* for further information).

Structural Properties: The study of the standard

[*]Electronic address: `praveen.321@gmail.com`
[†]Electronic address: `animeshm@cse.iitkgp.ernet.in`
[‡]Electronic address: `niloy@cse.iitkgp.ernet.in`

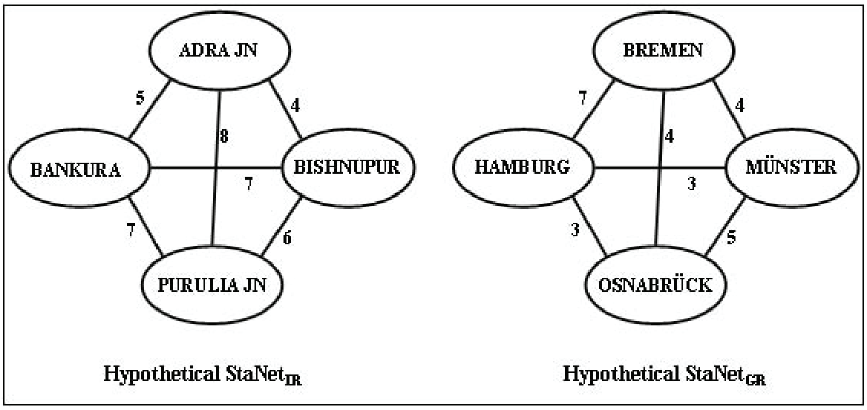

FIG. 1: A hypothetical illustration of the nodes and edges of StaNet$_{\text{IR}}$ and StaNet$_{\text{GR}}$. The labels of the nodes denote the names of the stations. The numerical values against the edges represent their corresponding weights. For example, the number of direct trains from "BANKURA" to "BISHNUPUR" is 7.

Properties	StaNet$_{\text{IR}}$	StaNet$_{\text{GR}}$
Weighted CC	0.79	0.75
Avg. Path Length, Diameter	2.43, 4.00	1.76, 3.00

TABLE I: Structural properties of the two networks. CC: Clustering Coefficient.

small-world properties, i.e., the clustering coefficient, the average geodesic path length and the diameter for the two networks indicate that our results are in agreement with the observation made by the earlier researchers in [9] and [3] respectively.

Clustering Coefficient – The clustering coefficient for a node i is the proportion of links between the nodes that are the neighbors of i divided by the number of links that could possibly exist between them. For instance, in a friendship network it represents the probability that two friends of the person i are also friends themselves. For weighted graphs such as StaNet$_{\text{IR}}$ and StaNet$_{\text{GR}}$, this definition has been suitably modified in [1]. Table I reports the values of the weighted clustering coefficient for the networks. Since these values are significantly high it implies that the neighboring stations of a station are also highly connected via direct trains for the railway system of both the nations.

Average Geodesic Path Length and Diameter – The average length of the geodesic (or shortest) path between any two arbitrary nodes is a measure of how well the network is connected. The results in Table I imply that both StaNet$_{\text{IR}}$ and StaNet$_{\text{GR}}$ have a very high connectivity. Furthermore, the diameter, which is the length of the "longest shortest path" between any two nodes is also small for both the networks (see Table I). This indicates that any arbitrary station in the network can be reached from any other arbitrary station through only a very few hops.

The above results collectively show that the RNs of both the nations exhibit small-world properties, which in turn means that in practice, a traveler has to change only few trains to reach an arbitrary destination.

Nevertheless, these results do not shed much light on the hidden structural patterns present in these networks. Therefore, we attempt to unfurl these patterns through the community structure analysis of the networks. In the next section, we review the algorithm for community detection and present the results obtained by applying it to the networks.

III. THE MRAD ALGORITHM

The MRad algorithm for detecting community structures in weighted networks has been introduced by us in [6]. For the purpose of readability, we briefly recapitulate the idea here.

The original algorithm of Radicchi et al. [8] (applied on unweighted networks) counts, for each edge, the number of loops of length three it is a part of and declares the edges with very low counts as inter-community edges.

Modification for Weighted Network: For weighted networks, rather than considering simply the triangles (loops of length three) we need to consider the weights on the edges forming these triangles. The basic idea is that if the weights on the edges forming a triangle are comparable then the group of stations represented by this triangle are highly and equally well-connected with each other, thereby, rendering a pattern of connection. In contrast, if these weights are not comparable then there is no such pattern. In order to capture this property we define a strength metric S for each of the edges of StaNet as follows. Let the weight of the edge (u,v), where u, $v \in V_S$, be denoted by w_{uv}. We define S as,

$$S = \frac{w_{uv}}{\sqrt{\sum_{i \in V_C - \{u,v\}} (w_{ui} - w_{vi})^2}} \qquad (1)$$

Communities from $\text{StaNet}_{\text{IR}}$	Region	η
Adra Jn., Bankura, Midnapore, Purulia Jn., Bishnupur	West Bengal	0.42
Ajmer, Beawar, Kishangarh	Rajasthan	0.42
Abohar, Giddarbaha, Malout, Shri Ganganagar	Punjab	0.50
Communities from $\text{StaNet}_{\text{GR}}$	**Region**	**η**
Bremen, Hamburg, Osnabrück, Münster	North-west of Germany	0.72
Augsburg, Munich, Ulm, Stuttgart	Extreme south of Germany	0.72
Diasburg, Düsseldorf, Dortmund	Extreme west of Germany	1.25

TABLE II: Some of the communities obtained from $\text{StaNet}_{\text{IR}}$ and $\text{StaNet}_{\text{GR}}$.

if $\sqrt{\sum_{i \in V_C - \{u,v\}} (w_{ui} - w_{vi})^2} > 0$ else $S = \infty$. The denominator in this expression essentially tries to capture whether or not the weights on the edges forming the triangles are comparable. If the weights are not comparable then this denominator will be high, thus reducing the overall value of S. StaNet may be then partitioned into clusters or communities by retaining only those edges that have S greater than a predefined threshold η.

We apply the MRad algorithm to $\text{StaNet}_{\text{IR}}$ and $\text{StaNet}_{\text{GR}}$ in order to extract the communities. Some of the example communities obtained by varying η are presented in Table II. Note that these results are representative. The results indicate that geographic proximity is the basis of the hierarchical organization of the RNs. The geographically distant communities are connected among each other only through a set of hubs or junction stations. Some of the results are presented on the maps (or partial maps) of the two countries in Figure 2 for a better visualization.

Not only does the above results help us to determine the basis of the hierarchical organization of the RNs but also can be useful while planning the distribution of new trains. For instance, although "Bharatpur" is a district headquarter in West Bengal it seems that it is not sufficiently well-connected with the neighboring stations like Farakka and Maldah and therefore, is not within the community formed by these stations (see top left map in Figure 2). A similar example is "Hanover" in Germany that should have been connected to the community of "Hamburg"/"Bremen" (refer to the map at the bottom in Figure 2).

IV. CONCLUSION

In this paper, we analyzed the structural properties of the RNs of India and Germany. Some of our important findings are

1. The RNs of both the nations indicate small-world properties, which is in agreement with the observations made by the earlier researchers;
2. The networks exhibit hierarchical organization and the hierarchy is formed on the basis of geographic proximity;
3. Community analysis of the RN of a nation can be helpful in planning the distribution of new trains.

A question that still remains unanswered is that how the two different nations with completely different political and social structures can have exactly the same pattern of organization of their transport system. This is possibly because of the fact that transportation needs of humans (across geography and culture) are to a large extent similar. For instance, short-distance travel for any individual is always more frequent than the long distance ones. Furthermore, on a daily basis, a much larger bulk of the population do short-distance travels while only a small fraction does long-distance travels. Therefore, there is always a pressure on having more trains within a region than across regions. Thus universal needs for transportation perhaps renders universal patterns in the RNs.

The aforementioned argument can be verified by having a single growth model that, subject to some parameters, can accurately explain the structural properties of the empirically obtained networks. We look forward to develop such a model in future.

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

[2] A. Clauset, M. E. J. Newman and C. Moore. 2004. Finding community structure in very large networks *Phys. Rev. E* **70**, 066111.

[3] L. Cui-mei and L. Jun-wei. 2007. Small-world and the growing properties of the Chinese railway network. *Frontiers of Physics in China* **2**(3), 364–367.

[4] R. Guimera and L. A. N. Amaral. 2005. Functional cartography of complex metabolic networks. *Nature* **433**, 895.

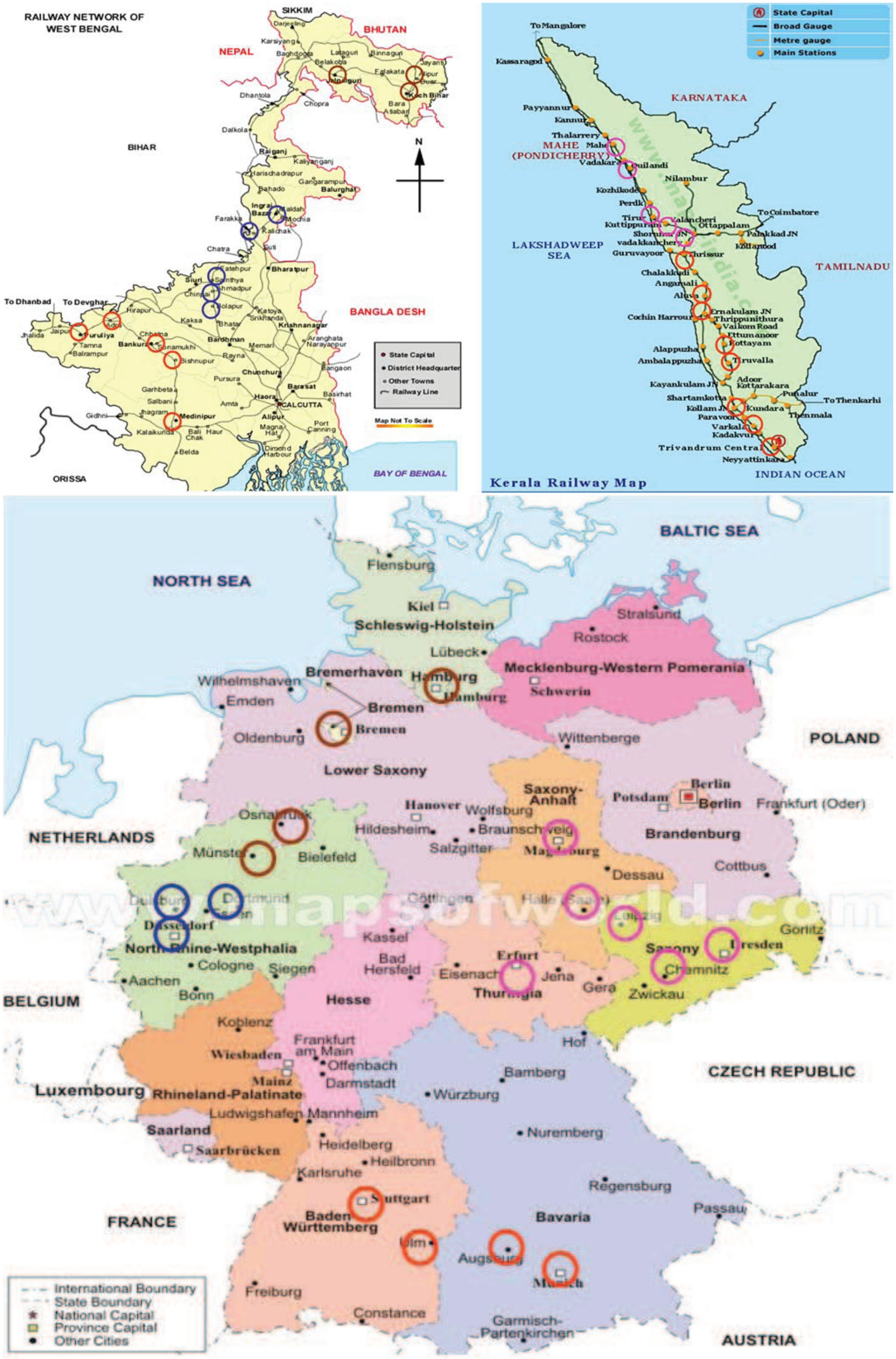

FIG. 2: Some of the communities shown on the maps (partial maps) of the two countries. Top left map is of West Bengal, which is an eastern state of India. At the top right is the map of Kerala, which is a state in south India. The map at the bottom shows the whole of Germany. The circles of a single colour denote those stations that are in the same community. Courtesy: *http://www.mapsofindia.com* and *http://www.mapsofworld.com*.

[5] M. C. Lagomarsino, P. Jona, B. Bassetti and H. Isambert. 2001. Hierarchy and feedback in the evolution of the Escherichia coli transcription network. *PNAS* **104**, 5516.

[6] A. Mukherjee, M. Choudhury, A. Basu and N. Ganguly. 2007. Modeling the co-occurrence principles of the consonant inventories: A complex network approach, *International Journal of Modern Physics C*, **18**(2), 281–295.

[7] A. Mukherjee, M. Choudhury, S. RoyChowdhury, A. Basu and N. Ganguly. (to appear). Rediscovering the Co-occurrence Principles of Vowel Inventories: A Complex Network Approach, *Advances in Complex Systems preprint: http://arxiv.org/abs/physics/0702056*.

[8] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto and D. Parisi. 2003. Defining and identifying communities in networks. *PNAS*, **101**(9), 2658–2663.

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

[10] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A.-L. Barabási. 2002. Hierarchical organization of modularity in metabolic networks, *Science* **30**, 1551.