Research Article

Complex Network Analysis of Pakistan Railways

Yasir Tariq Mohmand and Aihu Wang

School of Business Administration, South China University of Technology, Guangzhou 510640, China

Correspondence should be addressed to Aihu Wang; bmawang@scut.edu.cn

Received 14 December 2013; Accepted 16 February 2014; Published 18 March 2014

Academic Editor: Beatrice Paternoster

Copyright © 2014 Y. T. Mohmand and A. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the structural properties of Pakistan railway network (PRN), where railway stations are considered as nodes while edges are represented by trains directly linking two stations. The network displays small world properties and is assortative in nature. Based on betweenness and closeness centralities of the nodes, the most important cities are identified with respect to connectivity as this could help in identifying the potential congestion points in the network.

1. Introduction

In recent years there has been rapidly growing interest in investigating the statistical and dynamical properties of network systems, containing set of items called nodes or vertices and edges representing interactions between them. Examples include the Internet, the World Wide Web, social networks of acquaintance or other connections between individuals, organizational networks and networks of business relations between companies, neural networks, metabolic networks, food webs, distribution networks such as blood vessels or postal delivery routes, and networks of citations between papers.

Transportation networks are among the most important building blocks in the economic development of a country. The structure and performance of transportation networks reflect the ease of travelling and transferring goods among different parts of a country, thus affecting trade and other aspects of the economy. In the recent years, complex network analysis has been used to study several transportation networks. These include airport networks, for instance, the airport network of China [1, 2], airport network of India [3], US airport network [4], and the worldwide airport network [5, 6], urban road networks [7–9], and railway networks [10–14].

Railways are one of the most important modes of transportation around the world, with the topological properties of these railway networks attracting huge attention. Sen et al. [12] were amongst the first to apply complex network theory

to the railway network, while in the process of studying the statistical properties of the Indian railways the authors introduced a new topological representation, the P-Space topology, wherein stations or stops are identified as nodes and are connected if at least one train stops at both the stations. The authors introduced a new method to calculate the shortest distance between two stations. Based on these calculations, the small world properties and exponential degree distribution of the Indian railway network are identified. An extension to this was provided by Majima et al. [15] as the same topology was applied to the Japanese railway network and the same statistical results were obtained. While two different networks exhibited the same properties when illustrated using the P-Space representation, the Chinese railway network also displayed the small world properties of the shortest distance between stations and high clustering coefficient, however, with a power-law degree distribution [13]. In another attempt to explain the dynamic nature of the Chinese network, Guo and Cai [16] concluded that the network is a scale-free network when extracted in the L-Space topology. Similarly, Wang et al. [17, 18] represented the railway network of China in both L-Space and P-Space and successfully fitted a power-law distribution in both cases.

The PRN is a moderate railway network with over 620 stations and 7,791-kilometer track. Railways are the primary mode of intercity transportation in Pakistan and the network is responsible for transporting massive number of passengers and freight. Even though railways play an important role in shaping the transportation sector of Pakistan, no research has

FIGURE 1: Explanation of Space L & Space P.

been put forward into studying the complex nature of this network. To the best of our knowledge, this is the first study ever on the complex network theory application on PRN.

2. Network Construction

Before starting off with the analysis of PRN, it might be a good idea to define the proper network topology. Two methodologies exist in current literature for representing a network, Space L [8, 17] and Space P [8, 12, 18, 19] (Figure 1). Space L consists of nodes representing cities, bus, metro, train stops, and sea ports and a link between two nodes exists if they are consecutive stops on the route. Nodes in the Space P are the same as in the previous topology; here an edge between two nodes means that there is a direct bus, train, or metro route that links them. In other words, if a route A consists of nodes a_i , that is, $A = \{a_1, a_2, \dots, a_n\}$, then in the Space P the nearest neighbors of the node a_1 are $a_2, a_3, \dots, a_n$. The node degree k in this topology is the total number of nodes reachable using a single route and the distance can be interpreted as the number of transfers (plus one) one has to take to get from one stop to another, whereas the node degree k in the previous topology is just the number of directions one can take from a given node, while the distance equals the total number of stops on the path from one node to another [8, 12]. In this study, we use the Space P methodology to represent the PRN, as this has already been used to represent railway networks [2, 12, 14]. The network was constructed from the official “Pakistan railways time table,” kindly provided by Pakistan railways. The time table had complete details of railway stations, number of trains, and the arrival and departure of each train at/from each station.

3. Topological Properties

Table 1 provides all computed network statistics, from basic network properties such as the number of nodes and edges to the more complex metrics such as clustering and assortativity.

3.1. Degree Distribution. The degree of a node, a measure of its connectivity, is defined as the fraction of nodes with degree k in a network. Degree is one of the measures of centrality of a node in a network and it symbolizes the importance of a node in a network. Commonly accepted rule is that

TABLE 1: Computed properties of Pakistan railways network.

Property	Value
Nodes, N	628
Edges, M	6078
Average path length, $\langle l \rangle$	3.15
Average clustering coefficient, C	0.97
Diameter, d	5
Average degree	19.36
Degree range	(2, 69)
Assortativity, r	0.34
Betweenness centrality	0.01
Closeness centrality	0.2
Efficiency	0.25

the larger the degree of a node is, the more important it becomes. The PRN is comprised of $N = 628$ nodes and $E = 6,078$ edges representing the direct link among stations. The average degree of the network is thus $2E/N = 19.36$ which indicates the average number of stations reachable from an arbitrary station via a single train.

The degree distribution $p(k)$ is an important feature that reflects the topology of the network and is defined as the fraction of nodes having degree k in the network. However, the cumulative degree distribution is usually preferred as 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 [14]. The cumulative degree distribution of the network is provided in Figure 2. As evident from Figure 3, the railway network of Pakistan is a moderately connected network, with majority of nodes having degrees of 29 or below, whereas a few stations share high degree connectivity and act as hubs. Karachi, Lahore, Hyderabad, Kotri, Rawalpindi, and Peshawar are the most connected stations; however, they also pose a threat to the operations of the railway network, as a failure of one of these major stations can cause a major portion of the network to crash down and halt. This has been the case in the past several times when failure at one major station caused a major halt of railway operations in Pakistan.

3.2. Small World Properties. Watts and Strogatz [20] proposed a model of small world network in the context of various social and biological networks. A small world network is categorized as a network in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of stations. Stated simply, a small world is a network having a small average shortest path length and a large clustering coefficient as compared to a random network with the same number of N . We apply the same method to see if the small world properties are present in PRN.

The average shortest path length (the minimum number of edges passed through to get from one node to another)

FIGURE 2: Cumulative degree distribution.

FIGURE 3: Average nearest neighbor degree.

from one node to all other nodes of the network is calculated using the following equation:

$$D = \sum_{s,t \in V} \frac{d(a,b)}{N(N-1)}, \quad (1)$$

where V correspond to the set of nodes in the network, $d(a,b)$ is the shortest path from a to b , and N is the total number of nodes in the network. A small average path length of two stops or stations ($D = 3.2$) means that there is connectivity among almost all the stations of PRN, regardless of geographical distance. The network also features small diameter (maximum path length of a network), $d = 5$.

Clustering coefficient (C_i) of a node i is defined as the ratio of the number of links shared by its neighboring nodes to the maximum number of possible links among them. The average clustering coefficient is defined as

$$\langle C \rangle = \frac{1}{N} \sum_{i=0}^N C_i. \quad (2)$$

Using the above equation, the average clustering coefficient (C) of the network is calculated to be 0.97, indicating that the PRN is a highly clustered network. This result is

substantially higher than the value of an equivalent Erdos-Rényi random graph [21], (C_{ER}) = 0.02. The clustering coefficient together with the small average path length (see above) indicates that the PRN is indeed a small world network.

3.3. Degree-Degree Correlation. Another important topological characteristic of a network that is examined is the degree-degree correlation between connected nodes. A given network is said to be assortative if the high degree nodes have a tendency to connect to other high degree nodes. Similarly disassortative networks are where low degree nodes tend to connect to high degree nodes. Newman introduced a summary statistic for assortativity (r) in 2002 [22], defined as the Pearson correlation coefficient of the degrees at either end of an edge. Mathematically, this expression can be represented by the following equation:

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk (e_{jk} - q_j q_k), \quad (3)$$

where

$$q_k = \sum_j e_{jk}, \quad \sigma_q^2 = \sum_k k^2 q_k - \sum_k k q_k^2. \quad (4)$$

This statistic lies in between the range of $[-1, 1]$, where -1 indicates a completely disassortative network and 1 indicates a completely assortative network. For the PRN, the assortativity is measured to be 0.34 illustrating high degree nodes at one end of a link showing preference towards high degree nodes at the other end. To justify the result, the average degree of the nearest neighbor, $K_{nn}(k)$, for nodes of degree k , can be plotted using the following equation:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^N a_{ij} k_j. \quad (5)$$

If $k_{nn}(k)$ increases with k , the network is assortative. If $k_{nn}(k)$ decreases with k , the network is disassortative. Figure 3 represents the average degree of the nearest neighbor and it can be seen that the $k_{nn}(k)$ increases with degree k , consistent with a positive assortativity of 0.34.

3.4. Identifying the Major Stations in the PRN. To identify the stations with high traffic and congestion, betweenness and closeness centralities are used. Betweenness centrality of a node i can be defined as sum of the fractions of all-pairs shortest paths that passes through i . Mathematically,

$$c_B(i) = \sum_{s,t \in V} \frac{\sigma(s,t | i)}{\sigma(s,t)}, \quad (6)$$

where V is the set of nodes, $\sigma(s,t)$ is the total number of shortest paths, and $\sigma(s,t | i)$ is the number of shortest paths passing through i [23]. The top ten railway stations according to high betweenness centrality are given in Table 2. The station of Jacobabad leads the list as it acts as a link between three different provinces of Pakistan: Sindh, Punjab, and

TABLE 2: Betweenness centrality of top ten stations.

Betweenness centrality	Stations
0.41	Jacobabad
0.37	Kot Addu
0.28	Kundian
0.26	Rohri
0.25	Raiwind
0.24	Shahdara
0.22	Lodhran
0.17	Samasata
0.16	Larkana
0.16	Khushab

TABLE 3: Closeness centrality of top ten stations.

Closeness centrality	Stations
0.28	Kot Addu
0.26	Kundian
0.26	Jacobabad
0.26	Lodhran
0.26	Sher Shah
0.25	Raiwind
0.25	Rohri
0.24	Samasata
0.24	Khushab
0.24	Shahdara

Baluchistan. Similarly, the stations of Kot Addu, Kundian, Rohri, and Raiwind provide access to almost all of Pakistan as trains from different routes pass on through these stations.

Another studied parameter used to identify the major stations in PRN is the closeness centrality, defined as the average shortest distance from node i to all the other nodes, which reflects the closeness degree of the node with other nodes in the network. The mathematical expression is

$$C(V_i) = \frac{(N-1)}{\sum_{j=1}^n d(V_i V_j)}, \quad (7)$$

where $d(V_i V_j)$ is the shortest distance between V_i and V_j and is equal to the minimum stations from V_i to V_j in the network whereas $(N-1)$ is the normalization factor. Closeness centrality reflects the closeness degree from one station to all the other stations in the railway network, the larger the value is, the greater the influence is, and the wider range of service the station has. The top ten stations based on closeness centrality are listed in Table 3.

4. Conclusion

In this paper we have studied the PRN as an unweighted graph of railway stations. The network clearly displays small world properties and is assortative in nature. The betweenness and closeness centralities of the stations are also computed, wherein these stations are identified as potential

congestion points. As public transportation, especially railways, provides crucial mode of movement of passengers, the identification of possible congestion stations may serve an important role in identifying the limitations of the network. Although this study contributes a complex network analysis of the physical state of the PRN, given the availability of passenger/cargo flow data, it would also be interesting to study the weighted network as it could reveal a clearer picture of network dynamics in terms of passenger/cargo flow. Such a study would not only reveal the topological aspects but also provide a detailed insight into the network dynamics by identifying the stations with greater flow, the correlations of the edge weights with the degree of the vertices, and especially the eigenvector centrality where the quality of an edge also matters.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by 2011 Founded Project of National Natural Science Foundation of China (71171084), 2011 Research Fund for the Doctoral Program of Higher Education of China (20110172110010), and the Fundamental Research Funds for the Central Universities (2012, x2gsD2117850).

References

- [1] W. Li and X. Cai, "Statistical analysis of airport network of China," *Physical Review E*, vol. 69, no. 4, Article ID 046106, 2004.
- [2] H.-K. Liu and T. Zhou, "Empirical study of Chinese city airline network," *Acta Physica Sinica*, vol. 56, no. 1, pp. 106–112, 2007.
- [3] G. Bagler, "Analysis of the airport network of India as a complex weighted network," *Physica A*, vol. 387, no. 12, pp. 2972–2980, 2008.
- [4] L.-P. Chi, R. Wang, H. Su et al., "Structural properties of US flight network," *Chinese Physics Letters*, vol. 20, no. 8, pp. 1393–1396, 2003.
- [5] R. Guimerà, S. Mossa, A. Turttschi, and L. A. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [6] A. Barrat, M. Barthélemy, and A. Vespignat, "Modeling the evolution of weighted networks," *Physical Review E*, vol. 70, no. 6, Article ID 066149, 2004.
- [7] H. Lu and Y. Shi, "Complexity of public transport networks," *Tsinghua Science and Technology*, vol. 12, no. 2, pp. 204–213, 2007.
- [8] J. Sienkiewicz and J. A. Hołyst, "Statistical analysis of 22 public transport networks in Poland," *Physical Review E*, vol. 72, no. 4, Article ID 046127, 2005.
- [9] Y. T. Mohmand and A. Wang, "Weighted complex network analysis of Pakistan Highways," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 862612, 5 pages, 2013.

- [10] K. S. Kim, L. Benguigui, and M. Marinov, "The fractal structure of Seoul's public transportation system," *Cities*, vol. 20, no. 1, pp. 31–39, 2003.
- [11] K. A. Seaton and L. M. Hackett, "Stations, trains and small-world networks," *Physica A*, vol. 339, no. 3–4, pp. 635–644, 2004.
- [12] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, "Small-world properties of the Indian railway network," *Physical Review E*, vol. 67, no. 3, Article ID 036106, 2003.
- [13] W. Li and X. Cai, "Empirical analysis of a scale-free railway network in China," *Physica A*, vol. 382, no. 2, pp. 693–703, 2007.
- [14] S. Ghosh, A. Banerjee, N. Sharma et al., "Statistical analysis of the Indian Railway Network: a complex network approach," *Acta Physica Polonica B, Proceedings Supplement*, vol. 4, no. 2, pp. 123–138, 2011.
- [15] T. Majima, M. Katuhara, and K. Takadama, "Analysis on transport networks of railway, subway and waterbus in Japan," in *Emergent Intelligence of Networked Agents*, pp. 99–113, Springer, Berlin, Germany, 2007.
- [16] L. Guo and X. Cai, "Degree and weighted properties of the directed China Railway Network," *International Journal of Modern Physics C*, vol. 19, no. 12, pp. 1909–1918, 2008.
- [17] R. Wang, J.-X. Tan, X. Wang, D.-J. Wang, and X. Cai, "Geographic coarse graining analysis of the railway network of China," *Physica A*, vol. 387, no. 22, pp. 5639–5646, 2008.
- [18] Y.-L. Wang, T. Zhou, J.-J. Shi, J. Wang, and D.-R. He, "Empirical analysis of dependence between stations in Chinese Railway Network," *Physica A*, vol. 388, no. 14, pp. 2949–2955, 2009.
- [19] M. Kurant and P. Thiran, "Trainspotting: extraction and analysis of traffic and topologies of transportation networks," *Physical Review E*, vol. 74, no. 3, Article ID 036114, 2006.
- [20] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [21] P. Erdos and A. Rényi, "On the strength of connectedness of a random graph," *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 12, no. 1–2, pp. 261–267, 1961.
- [22] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, 2002.
- [23] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136–145, 2008.