SYSTEMATIC STUDY OF INDIAN RAILWAYS SUBNETWORK: ZONE SPECIFIC ANALYSIS

H. P. Singh, S. Arora, A. Jain, N. Arora, A. Singh, R. Pal

Communicated by Shimpi Singh Jadon

MSC 2010 Classifications: 68R01, 68R10, 05C90.

Keywords and phrases: Network Structure Measures, Indian Railways, Average Path Length, Degree Distribution.

Abstract A country's transportation system is one of its most substantial economic growth measures. The empirical analysis of the Indian Railway Network (IRN) and statistical characteristics comprises 926 nodes i.e. train stations. The IRN consists of more than 8000 stations, whose study as one whole network is a herculean task. In this work, a small network scale is used which gives us the complete properties of the railway network with zone specific analysis instead of a generalized picture. On the basis of average degree, more than 8000 stations are reduced to 926 major stations and 16 zones are reduced to 4 zones. Detailed complex network analysis of the IRN network includes studies involving shortest-path length, and cumulative degree distribution accompanied by a degree of betweenness, closeness and clustering and the scale-free distributions weighted degrees, namely strengths and distributions of clustering coefficients. Several comparative statements are drawn co-relating the station structure measures, passenger footfall and network centrality measures for the 4 zones and the complete IRN. The comparative study can be extrapolated to affirm the small world characteristics of the 4 major zones considered in the research.

1 Introduction

The railway is among the most significant modes of transportation around the world. Francesco et al. [8] proposed a work of systematic methodology work on the European transport system. Their contribution proposes to analyze capacity and utility rates for the deployment of interconnected railway infrastructure. It provides not only the efficient theoretical approach of smallscale rail systems but examines possibilities and strategic analysis of large-scale international networks by using big data and open databases. Characteristics of Pakistan's railway network were studied by Mohmand et al. [18]. He considered the stations as the vertices and trains between the stations as the weighted edges. The network shows the characteristics of the small world. The congestion points in the network are recognized and have an important role in the network. Xing et al. [33] studied the weighted complex networks with topological and dynamic characteristics. The correlation between passenger flow and the topological structure provides theoretical analysis for urban transit planning and management. Studies on the public transportation system of Singapore [28], Boston subway [15], the India railway [25], the Chinese railway [16] and the world-wide airport network [1] have been carried out in detail. Studying the dynamical properties and keeping in mind the magnitude of interactions, it was found Singaporean transport systems show several similar features. Ghosh et al. [10] studied the Indian railways as a complex network. Taking it as a small world characteristic and an exponential distribution of nodes the possibility to improve the network performance was proposed. Regt et al. [23] investigated the public transport networks within Great Britain. By using the complex network, the robustness, and the efficiency of the network are looked into. Monechi et al. [19] introduced a fresh framework called the delay propagation model to assess the efficiency of spread over the network. Miglani et al. [17] proposed the notion of resource to resource hindrance matrix which captures the complexities of railway movements in a rail function. Cao et al. [5] modeled the Chinese Railway Network (CRN) and studied complex network theory. The significance of each node was characterized by data-driven integrated measures such as degree, strength, betweenness, and closeness. Chinowsky et al. [7] focused on identifying potential risks caused by climate change in the US rail network. A model-based method is used that combines predictions of climate change with present information. Many authors have been conducted for the complex network systems of Europe, Pakistan Railways, Shanghai Rail Transit System, etc. ([18] - [7]). All these studies have examined the networks as a whole whereas we have conducted the zone-wise study of the Indian Railway Network. Raicu et al. [21] contribute to the growth and knowledge of techniques to analyze network reliability in the formulation and implementation. The effect of high-speed rail on the network and circuit in China has been studied by Shaw et al. [26] and Hu et al. [13] respectively. Woodburn et al. [32] studied rail networks in context to freight under variable weather conditions. The importance of improving public transit networks is a challenging area and has been discussed by Jian et al. [14]. Further, we find various studies regarding the analysis and evaluation of complex networks in different modes ([6]-[22]).

In the present work, the Indian Railway Network shows characteristics of a small-world network. We emphasized investigating the possibilities to reduce the dependence of IRN, which is one among the largest networks in the world. This work performs a complex network analysis of the IRN and extracts attributes based on central measures such as degree centrality, closeness centrality and betweenness centrality as well as average path length, coefficient clustering and degree distribution. IRN being so vast cannot be handled all at once, therefore we have decided to take 4 major zones based on the highest average degree out of 16 zones namely NR (Northern Railways), CR (Central Railways), WR (Western Railways), and SCR (South Central Railways). This work is divided into 4 sections. Section 1 gives the introduction and literature review of related works. Section 2 defines the network measures (network structure measures and centrality measures). Section 3 explains the network representation. Section 4 discusses the results of this study. Section 5 provides the conclusion of this work.

2 Network Measures

Complex network analysis has often been used to study and understand several transport networks such as airport networks, urban road networks, and railway networks. It should be noted that the transportation network has some characteristics that are never shared by other networks like their small size in comparison to large networks such as the Internet. This section serves the purpose of introducing the indices for assessing a network.

2.1 Network Structure Measures

Network Structure measures are used to categorize a network into a Regular Network, Random Network, Small-World Network, or a Scale-Free Network. The network measures used for categorization are described below.

Degree Distribution p(k): The degree distribution p(k), in a network with n nodes where m_k nodes have k degrees, is defined as the ratio of k degree nodes to the total nodes, i.e., m_k/n [4]. The cumulative degree distribution for degree k is denoted by P(k) and is represented as below

$$P(k) = \sum_{k'=k}^{\infty} p(k') \tag{1}$$

Average Path Length (L): It is defined as the average number of trains for the smallest number of hops or steps in all possible connections of stations [30].

where d_{ij} is the number of edges for the shortest path from *i* to *j*.

Clustering Coefficient (C_i) : It defines how complete the connection of a node is in the network. The implication of a higher C_i , the value is that the node has a more compact connection system with its neighbors. C_i of all nodes in a complete graph network is equal to 1. The clustering coefficient of the whole network C is written as:

$$C = \frac{1}{n} \sum_{v_i \in V} C_i \tag{3}$$

Here, V is the set of vertices (v_i) . Average path length, clustering coefficient, and cumulative degree distribution can collectively classify a network into categories. The values of these parameters for a Regular network are longer L, larger C, and point to point P(k) while that of a Random network are shorter L, small C, and Binomial or Poisson P(k). A Small-world network shows lower values of L, higher values of C, and Exponential P(k). Both Scale-free network and Real network have shorter L, larger C, and similar Power-law degree distribution [30].

2.2 Centrality Measures

In this paper, our objective is to quantify the connectedness and the influence of a railway station in the Indian Railway Network. The following three measures of centrality are sufficient to describe the characteristics of a station in the Indian Railway Network. In this section, three measures of centrality degree, closeness, and betweenness are studied. They are used for comparative studies of nodes i.e. stations in a network.

Degree Centrality (C_D) : It represents the number of edges shared by a node with others [9]. It can be written as:

$$C_D(i) = \sum_{j=1}^n a_{ij} \tag{4}$$

where element $a_{ij} = 1$ in a direct link exists between nodes *i* and *j* or else $a_{ij} = 0$. In this paper, the proposed network is a directed network and thus has two measures: in and out-degree.

Closeness Centrality (C_c): It is used to measure how close a node is to all other nodes and is written as [24]:

$$C_c(i) = \frac{n-1}{\sum_{(v_j \in V, i \neq j)} d_{ij}}$$
(5)

Betweenness Centrality (C_B): It measures how a node lies between other nodes in the network [9, 2]. The expression is given below:

$$C_B(i) = \sum_{k \neq i \neq j \in N} \sigma_{kj}(i) / \sigma_{kj} \tag{6}$$

where, σ_{kj} is the sum of all shortest paths between nodes v_k and v_j . $\sigma_{kj}(i)$ is the number of shortest paths that pass through v_i .

3 Network Narration

The network presented in this study represents the complete Indian railway network. The IRN consists of more than 8000 stations which makes it the fourth largest railway network by size. The study of such a large network is a herculean task therefore we have taken 926 major stations which could give us complete characteristics of the whole network [29]. It is divided into 16 major zones as of 2014 [20].

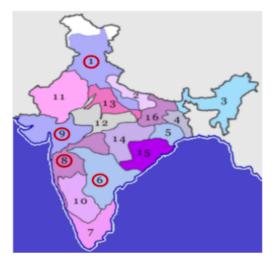


Figure 1: Geographical representation of 16 zones



Table 1: List of Indian railway zones

We have calculated centrality measures for all the 16 zones. Further, we have taken 4 zones according to the highest average degree and calculated P(k) for them. Figure 1 represents the 16 zones of the Indian Railway Network. The highlighted 4 zones are taken for in-depth study. Table 1 lists the names of all the zones.

The undirected graphs for 4 zones are made to determine the network where each node represents the station of the zone and each edge represents the connective of two stations via a direct train. The weight of the edges shows the total number of trains between the two stations. Total trains include each type of locomotive running on Indian railway tracks such as Diesel–Electric Multiple Unit (DEMU), Electric Multiple Unit (EMU), Mainline Electric Multiple Unit (MEMU), Rajdhani, Shatabdi, etc. The graph comes out to be a dense one and has been shown in figure 2.

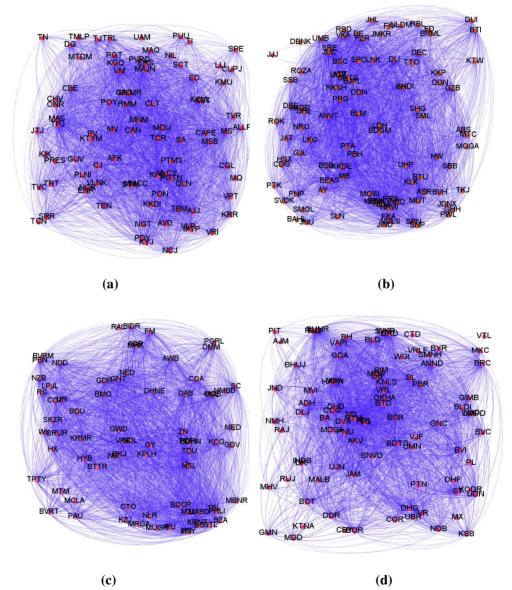


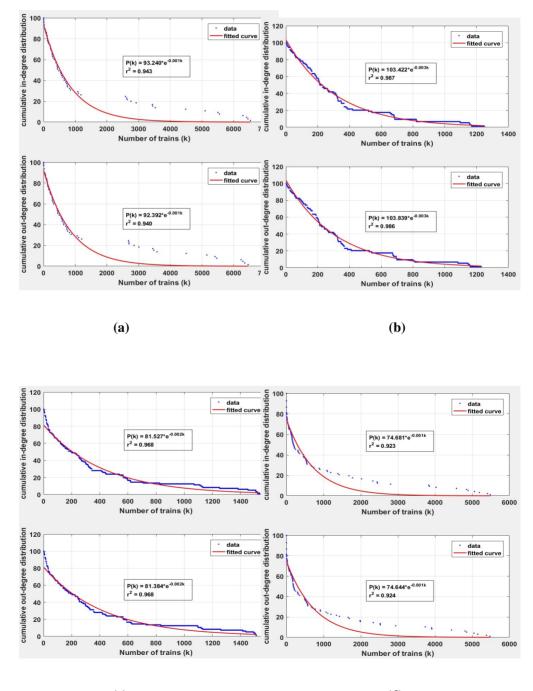
Figure 2: Four major Railway zones

Figure 2 (a) shows the interconnection of stations in the Central railway. There are a total of 65 nodes present in the representation. Figure 2 (b) shows the interconnection of stations in the Northern railway. A total of 96 nodes are represented in the graph. Figure 2 (c) depicts the connection in the South central railway. 74 nodes are present in the graph. Figure 2 (d) shows the interconnection of stations in the Western railway. It has a total of 97 nodes in the representation. Each connecting edge in the figure states the availability of a direct train between the connected nodes.

4 Network Analysis

This section presents the comparison of network structure measures and centrality measures between the overall IRN and the 16 individual zones with special emphasis on the four highlighted zones.

The cumulative degree distribution of 4 zones is presented in Figure 3. Figure 3 (a), (b), (c), and (d) are the graphs of P(k) for Central railway, South central railway, Northern railway, and Western railway, respectively. The figure shows the cumulative in-degree and out-degree distribution for the zones. The general trend for both cumulative distributions is that they are decreasing with a decreasing rate. The correlation r is found out to be 0.94, 0.98, 0.96, and 0.92 for CR, SCR, NR, and WR, respectively.



(c) (d) Figure 3: Cumulative degree distribution of four major zones

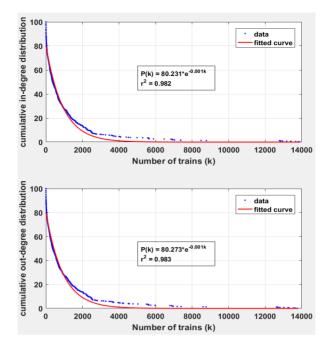


Figure 4: Cumulative degree distribution in the IRN

The overall network's Cumulative in-degree and out-degree distribution of the IRN are shown in Figure 4. It follows an exponential function as $P(k) = 80.27 * e^{-0.001k} (r = 0.98)$. The general trend is decreasing at a decreasing rate. Small-world networks follow such a distribution pattern hence, it can be said that the IRN shows the features of such a system. Further, it can be said that some major stations prevail in the system with a higher number of routes, and the number of routes to each station decreases at a fast rate and evens off towards small stations. This is in agreement with other previous research works [25].

The average path length of all the 16 zones is shown in Figure 5. It indicates the ease of commuting in a given network. Its value defines the minimum number of nodes traversal needed to reach the destination node. The clustering coefficient defines the degree of concentration i.e. the probability of fewer transfers when traversing the network and is shown in Figure 6.

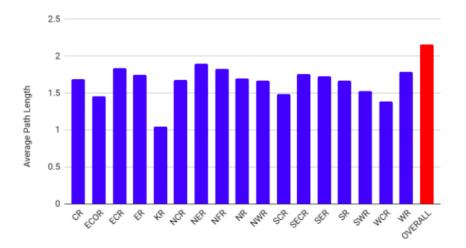


Figure 5: Average Path Length (L)

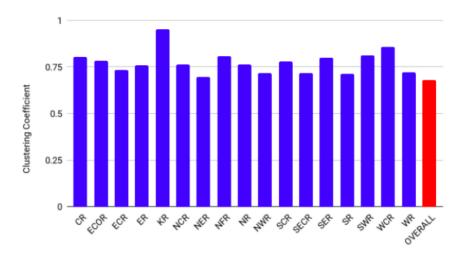




Figure 5 shows the average path length for all the 16 zones as well as the overall network. The average path length for the 16 zones lies between 1 and 2 while for the overall network it is 2.162. This shows that all the 16 zones have short average path length, which indicates that all the 16 zones, according to Table 2, depict the characteristics of small-world networks and the overall network, too, depicts the same. The clustering coefficients of the zones also depict the characteristics of the small-world network. Larger a clustering coefficient, the higher is the degree of concentration. Moreover, it also suggests a higher probability for commuting by making fewer transfers. C of the IRN is 0.679.

The average degree of a network tells the estimated number of edges originating and terminating on any node of the network. Figure 7 shows the average degree of all the 16 zones and the overall network. It is our criterion for selecting the major four zones that we are going to study. So, the top four zones according to the graph as shown in figure 7 are Central Railway (CR), Northern Railway (NR), South Central Railway (SCR) and Western Railway (WR).

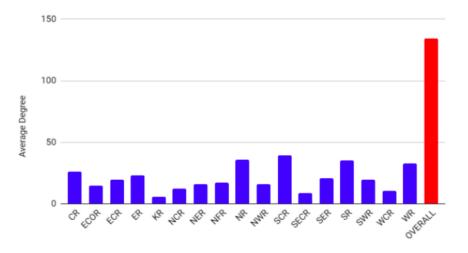


Figure 7: Average Degree

The centrality Measures of the four highlighted networks and the IRN are shown in Table 2. We can infer from Table 2 that there is a consistency between the ranks of degree and closeness. In the Central railway zone and Northern railway zone, the same stations appear for both indices. Nine stations are found common in the top 10 lists for both degree and closeness. The ranks of stations when determined by betweenness are found to be significantly different from the ranks determined by degree and closeness. Connectivity between a few stations is very high but they do not play a major role in transmissivity. Whereas, a few stations which are less connected play an important role and act as transfer hubs.

Considering the overall network, the centrality measures are calculated for each zone and are correlated to the number of passengers traveling in each zone. The values in Table 3 show a very similar trend between the closeness centrality of the zone-wise network and the overall network (IRN).

CR			NR			SCR			WR		
B/W	CLS	Degree									
NRL	KYN	KYN	PTK	NDLS	NDLS	KCG	SC	BZA	BRC	ADH	BVI
TNA	PNVL	PNVL	ASR	DLI	DLI	SC	HYB	SC	ANND	BVI	ADH
KYN	TNA	TNA	VKA	DEE	DEE	HYB	KCG	KCG	ADI	BSR	BSR
PNVL	DR	DR	JUC	DEC	DEC	FM	FM	FM	UMN	ADI	BA
LTT	CSTM	CSTM	MB	NZM	NZM	BIDR	BZA	HYB	ND	BDTS	BDTS
CSTM	LTT	LTT	DLI	ANVT	ANVT	BZA	TPTY	KZJ	BH	BCT	BCT
DR	PUNE	PUNE	NDLS	UMB	UMB	CCT	RU	WL	AKV	BRC	DDR
PUNE	DD	DD	DEE	JUC	JUC	COA	KZJ	GDR	ADH	ANND	MDD
DD	MMR	MMR	DEC	LDH	LDH	GNT	WL	RU	BSR	ST	GMN
SUR	BSL	BSL	ANVT	SRE	SRE	NZB	PBN	TPTY	BVI	UDN	CCG

B/W: Betweenness CLS: Closeness

Table 2. Top 10 stations by degree, closeness and betweenness

Zone	Number of	N	Passengers			
	Stations	Degree	Closeness	Betweenness	Travelling (in millions)	
CR	65	1361.954	0.589	42.677	1716	
ECOR	27	209.481	0.711	12	90	
ECR	60	179.383	0.522	46.233	254	
ER	83	491.241	0.583	61.41	1156	
NCR	42	66.429	0.959	18.429	171	
NER	52	115.929	0.392	43.788	165	
NFR	53	109.288	0.525	43.019	87	
NR	96	208.491	0.573	62.146	637	
NWR	37	350.427	0.562	22.568	153	
SCR	74	125.568	0.583	35.351	357	
SECR	34	321.081	0.693	15.765	256	
SER	51	321.081	0.413	36.255	125	
SR	86	213.667	0.607	57	806	
SWR	44	468.674	0.619	21.795	191	
WCR	18	175.682	0.636	5.778	139	
WR	97	217.056	0.646	60.887	1625	
IRN	926	986.54	0.431	1008.7	7928	

Table 3. The relation between Centralities and Passengers Travelling

This suggests that the closeness of stations does not scale with the size of the network. The drastic rise in the value of betweenness centrality suggests that the connection of high traffic stations with other zones significantly improves its betweenness centrality. A high value of degree centrality also suggests the same i.e. when studying the whole network, the trains which connect stations of different zones account significantly in improving the degree of a station. The values of the centrality measures for all the railway zones are given in Table 3. These values are used to plot the figures 8-10.

The relationship between the centrality measure betweenness and the passenger volume is given in Figure 8. Although the graph contains some outliers, when the data points are fitted into an exponential model, the trend comes out to be increasing. Assuming Central railways to be an outlier of the system, one can clearly see that the zone having higher passenger volume, in general, have a high value of betweenness among nodes i.e. the stations in zones having high footfall are well connected.

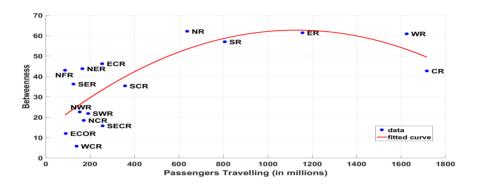


Figure 8: Betweenness vs Passenger Volume

Figure 9 shows the relationship between closeness and passenger volume. Similar to the previous centrality, this graph also contains outliers. The general trend of the graph shows a constant relation i.e. no interdependency among the taken parameters. This can be interpreted as the distance between stations in a zone is not dependent on the number of passengers traveling in that zone.

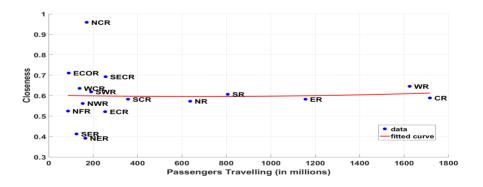


Figure 9: Closeness vs Passenger Volume

Figure 10 represents the relationship between the degree of stations among different zones to the passengers traveling in that zone. This graph also has some outliers but the general trend of the graph is increasing. This trend validates the hypothesis that the higher the number of trains between stations in a given network, the more the number of passengers will travel in that zone.

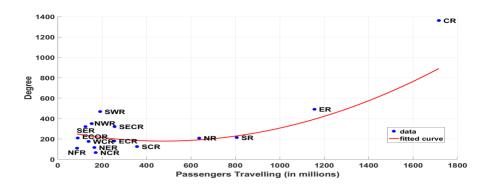


Figure 10: Average Degree

Figures 8-10 represent the study of 16 different zones of the Indian railway network. The comparison of the overall network with the 16 smaller networks as well as the results of the nodal analysis of 16 zones and the whole IRN are given in Table 3. For each zone, different centralities have been calculated. These centralities were plotted against the passenger volume in the above three figures to showcase their physical significance.

5 Conclusion

In this paper, we have used complex network analysis tools and theories to examine the Indian Railway Network. It has been concluded that Indian railways show small-world characteristics such as most stations can be reached from every other station by the nearly straightforward route. The degree distribution presented in this work also affirms the exponential nature of the function thus signifying major dominance of large stations in the region. For the whole network average path length came out to be 2.162, the clustering coefficient came out to be 0.679 and the average degree came out to be 134.269. These values re-affirms that IRN shows similar properties to that of a small-world network. The present study focused on four major railway zones of the Indian Railway Network. The study can be extended to incorporate other factors such as population to see the effect on the network system, the usage of solar energy in the Indian railways and cost analysis of the network.

Declaration of Interests: The authors declare that there is no conflict of interest.

References

- Amaral, L. A. N., Scala, A., Barthelemy, M. and Stanley, H. E., Classes of small-world networks, *Proceedings of the National Academy of Sciences* 97(21), 11149–11152 (2000).
- [2] Anthonisse, J. M., *The rush in a directed graph*, Stichting Mathematisch Centrum. Mathematische Besliskunde (1971).
- [3] Bagler, G., Analysis of the airport network of India as a complex weighted network, *Physica A: Statistical Mechanics and its Applications* **387(12)**, 2972–2980 (2008).
- [4] Barabási, A. L. and Albert, R., Emergence of scaling in random networks, *Science* 286(5439), 509–512 (1999).
- [5] Cao, W., Feng, X., Jia, J. and Zhang, H., Characterizing the structure of the railway network in China: A Complex Weighted Network Approach, *Journal of Advanced Transportation*, 1–10 (2019). https://doi.org/10.1155/2019/3928260
- [6] Cats, O., Topological evolution of a metropolitan rail transport network: The case of Stockholm, *Journal of Transport Geography* **62**, 172–183 (2017).
- [7] Chinowsky, P., Helman, J., Gulati, S., Neumann, J. and Martinich, J., Impacts of climate change on operation of the US rail network, *Transport policy* 75, 183–191 (2019).
- [8] Francesco, R., Gabriele, M. and Stefano, R., Complex railway systems: capacity and utilisation of interconnected networks, *European Transport Research Review* 8, 1–21 (2016).
- [9] Freeman, L. C., A set of measures of centrality based on betweenness, *Sociometry*, 35–41 (1997).

- [10] Ghosh, S., Banerjee, A., Sharma, N., Agarwal, S., Ganguly, N., Bhattacharya, S. and Mukherjee. A., Statistical analysis of the Indian railway network: A complex network approach, *Acta Physica Polonica B Proceedings Supplement* 4(2), 123–138 (2011).
- [11] Guida, M. and Maria, F., Topology of the Italian airport network: A scale-free small-world network with a fractal structure?, *Chaos, Solitons & Fractals* **31(3)**, 527–536 (2007).
- [12] Guimera, R., Mossa, S., Turtschi, A. and Amaral, L. N., The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, *Proceedings of the National Academy of Sciences* 102(22), 7794–7799 (2005).
- [13] Hu, X., Huang, J. and Shi, F., Circuity in China's high-speed-rail network, *Journal of Transport Geogra*phy 80, 102504 (2019). https://doi.org/10.1016/j.jtrangeo.2019.102504
- [14] Jiang, R., Lu, Q. C., and Peng, Z. R., A station-based rail transit network vulnerability measure considering land use dependency, *Journal of Transport Geography* 66, 10–18 (2018).
- [15] Latora, V. and Marchiori, M., Is the Boston subway a small-world network?, *Physica A: Statistical Mechanics and its Applications* **314**, 109–113 (2002).
- [16] Li, W. and Cai, X., Empirical analysis of a scale-free railway network in China, *Physica A: Statistical Mechanics and its Applications* **382(2)**, 693–703 (2007).
- [17] Miglani, D., Dogra, A., Belur, M. N., and Rangaraj, N., Railway junction simulation and analysis for mixed rail traffic, *Proceedings of the Symposium on Advanced Train Control and Safety Systems for Indian Railways (ATCSSIR)*, 1–7 (2017).
- [18] Mohmand, Y. T. and Wang, A., Complex network analysis of Pakistan railways, Discrete Dynamics in Nature and Society, 1–5 (2014). http://dx.doi.org/10.1155/2014/126261
- [19] Monechi, B., Gravino, P., Clemente, R. D. and Servedio, V. D., Complex delay dynamics on railway networks from universal laws to realistic modelling, *EPJ Data Science* 7, 1–18 (2018).
- [20] Pereira, V. E., Managing people in the world's largest commercial employer: An exploratory study on Indian Railways, *International Journal of Indian Culture and Business Management* 10(2), 136–156 (2014).
- [21] Raicu, R. and Taylor, M. A. P., Assessing rail transport network performance and reliability, WIT Transactions on The Built Environment 88, 75–84 (2006).
- [22] Rajput, N. K., Badola, P., Arora, H. and Grover, B. A., *Complex Network Analysis of Indian Railway Zones*, arXiv:2004.04146 (2020).
- [23] Regt, R. D., Ferber, C. V., Holovatch, Y. and Lebovka, M., Public transportation in Great Britain viewed as a complex network, *Transportmetrica A: Transport Science* **15**(**2**), 722–748 (2019).
- [24] Sabidussi, G., The centrality index of a graph, Psychometrika 31(4), 581-603 (1966).
- [25] Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G. and Manna, S. S, Small-world properties of the Indian railway network, *Physical Review E* 67(3), 036106 (2003).
- [26] Shaw, S. L., Fang, Z., Lu, S. and Tao, R., Impacts of high speed rail on railroad network accessibility in China, *Journal of Transport Geography* 40, 112–122 (2014).
- [27] Small-world network, Available at: *http://www.scholarpedia.org/article/Small-world_network* (Accessed 2 September 2019).
- [28] Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G. K. K., Hung, T. G. G., Di, P., Prakasam, S. and Wong, L., Weighted complex network analysis of travel routes on the Singapore public transportation system, *Physica A: Statistical Mechanics and its Applications* 389(24), 5852–5863 (2010).
- [29] Station Code Index, Indian Railways 2014-2015, Available at: http://www.indianrailways.gov.in/railwayboard/uploads/directorate/coaching/pdf/Station_code.pdf (Accessed 2 September 2019).
- [30] Watts, D. J. and Strogatz, S. H., Collective dynamics of 'small-world' networks, *Nature* 393, 440–442 (1998).
- [31] Woodburn, A., Effects of rail network enhancement on port hinterland container activity: A United Kingdom case study, *Journal of Transport Geography* 33, 162–169 (2013).
- [32] Woodburn, A., Rail network resilience and operational responsiveness during unplanned disruption: A rail freight case study, *Journal of Transport Geography* 77, 59–69 (2019).
- [33] Xing, Y., Lu, J. and Chen, S., Weighted complex network analysis of Shanghai rail transit system, Discrete Dynamics in Nature and Society, 1–8 (2016). http://dx.doi.org/10.1155/2016/1290138

Author information

H. P. Singh, S. Arora, A. Jain, N. Arora, A. Singh, R. Pal, Cluster Innovation Centre, University of Delhi, Delhi-110007, India.

E-mail: harendramaths@gmail.com, profswatiarora@gmail.com* (*Corresponding Author), aayushjain0829@gmail.com, namrata.arora1997@gmail.com, arvindsingh18798@gmail.com, rishi580pal@gmail.com