The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 4, August 2025
https://doi.org/10.33022/ijcs.v14i4.4920

Unraveling the Structure of India's Railway Network: Insights from Network Analysis

Fadil Al Afgani[1]
fadilalafgani@telkomuniversity.ac.id[1]
[1]School of Computing, Telkom University Bandung

Article Information

Received : 17 Jun 2025
Revised : 23 Jul 2025
Accepted : 3 Aug 2025

Keywords

graph, network, social networks, India's railway station.

Abstract

India's railway station network is a vast and complex system that plays a crucial role in the country's transportation infrastructure. In this analysis, we will explore the network of Indian railway stations using network analysis techniques. Network analysis is a statistical approach to analyzing the relationships between variables in a network. A network is a graphical representation of the relationships (edges) between variables (nodes). The study will involve constructing a network representation of the railway station network, where each station is represented as a node, and the connections between stations are represented as edges. This visualization allows us to identify the type of network, communities, overlapping communities, cascade failure, and heterogeneous information network within the network. Based on the analysis results, the formed network is a scale-free network. The community detection analysis using the Leiden algorithm shows that there are 23 clusters formed with a quality value of 0.97662. Overlapping communities are present when the value of $K \leq 3$, and there is the potential for cascade failure or an epidemic when the node with the highest degree is assigned the status of infected. The formed India's railway station network is a HINs (heterogeneous information network) as it consists of various types of entities with different characteristics.

A. Introduction

India's transportation system has been mostly dependent on railroads since they allow products and people to be moved across the nation across great distances. Comprising thousands of railway stations dispersed over India, India's railway network is among the biggest in the world. Examining and comprehending the complex dynamics of this railway station network is essential to maximising operations, raising efficiency and thus enhancing the general performance of Indian Railways given its size and complexity [1].

Recent studies have shown how important it is to grasp such complicated systems mostly by means of network analysis. Offering a strong approach for revealing the structure and linkages in large-scale transportation systems [2], network analysis studies the interactions between variables expressed as nodes and edges in a network [3]. Network analysis offers a visual and analytical approach to examine how stations interact and to evaluate aspects such connectedness and accessibility by depicting a railway system as a graph where nodes represent train stops and edges represent routes or connections between them [4]. Network analysis provides a visual and analytical means to study how different stations interact and to assess features such as connectivity and accessibility [5].

This approach has driven the increasingly used sophisticated network analysis techniques in recent years to better grasp the structure and effectiveness of railway networks. One research, for instance, assessed the Indian railway system by counting the trains directly linking two railway zones using a weighted graph where edge weights indicate the quantity of train connections [1]. This approach reveals important elements necessary to sustain operational efficiency and resilience over the system by viewing every zone as a network inside a bigger network.

This effort will model and investigate the Indian railway network using publicly accessible data and network visualization approaches, therefore doing a similar analysis [6]. Representing railway stations as nodes and links between them as edges, they will provide a complete graphical depiction of the network. By means of this visualization, one can detect network properties including network type, community structure, overlapping communities, cascading failures, and heterogeneous information networks, so offering more comprehensive understanding of the structure and performance of the Indian railway system.

B. Literature Review

1. Complex Network Analysis in Railway Systems

Understanding the structure and dynamics of railway networks now depends much on complex network analysis. Using sophisticated network concepts, Ishu Garg, Ujjawal Soni et al. (2022) identified important stations in the Indian railway network and found that community detection inherently divides nodes into continuous geographical areas [6]. Lamanna et al. (2025) also investigated train schedule connectivity in Norway using complicated network theory and the Infomap technique, therefore exposing significant secrets on the efficiency and leakage of the railway network [7].

2. Community Detection with Leiden Algorithm

In railway systems, community detection helps to find groups of highly connected stations. Understanding the modular architecture of transportation systems requires knowledge of community detection. Applying Random Walk Centrality (RWC), Jain and Manimaran (2023) suggested a fresh method to detect community structure in complicated networks this approach was then used to the Indian railway network to increase resilience and connectedness [8]. In this research, the Leiden Algorithm will be used. In complex networks, the Leiden algorithm is a community finding technique. Usually used for network community detection, this variation on the Louvain method is Maximizing the gap between the actual number of edges inside a community and the expected number of edges helps the Leiden algorithm to find well-connected communities inside a network. Based on modularity which measures the robustness of the community structure in a network this method [9].

Guaranturing well-connected communities, the Leiden algorithm enhances upon the modularity optimization approach applied in the Louvain algorithm. It has been created especially to effectively manage big and complicated systems [10]. By use of a resolution value, the Leiden algorithm provides fine-grained control over the size and granularity of the discovered communities. The Leiden algorithm is more adaptable in its application than some community detection systems that allocate every node to a single community as it can identify overlapping communities.

The Leiden algorithm proceeds via the following phases:

a. Initialization: Each node is initially placed in a separate community.
b. Local node movements: Each node is sequentially examined and moved to its nearest neighbor which results in an increase in modularity.
c. Global community movements: Step 2 is repeated iteratively, changing the order of node examination to improve the algorithm's ability to find better partitions.
d. Modularity optimization: Additional optimization is performed on the resulting community partition to enhance modularity.
e. Termination: The algorithm stops when there are no significant node movements left.

These steps enable the Leiden algorithm to effectively detect communities within a network and optimize modularity.

3. Overlapping Community Detection with CPM

Overlapping communities are found using techniques including Clique Percolation Method (CPM), which offers understanding of the modular structure of the network and possible alternate paths in case of disturbances. The Clique Percolation Method (CPM) is an approach used for analyzing the overlapping community structure in networks. It is a popular method for detecting communities within networks where a node can belong to multiple communities. The concept of a network community refers to a group of nodes that are densely connected compared to connections with nodes outside the community. The Clique Percolation

Method defines a community as a maximal union of k-cliques, which can be reached from each other through a series of adjacent k-cliques. Two cliques are considered adjacent if and only if they overlap in K-1 nodes. By identifying these cliques and their connections, the CPM detects the overlapping communities in the network [11].

4. **Cascade Failure Simulation**

Railway system management depends critically on an awareness of network resilience to disturbances. According to a 2022 Garg et al. analysis, major stations and junctions in India are typically connected to nodes regarded as critical depending on network metrics. Moreover, the resilience of the Indian railway network to different disturbances, including operational faults and natural calamities, has been quantified using network science-based methods [1].

5. **Heterogeneous Information Network Modeling (Heterogeneous Information Network)**

For more thorough study, heterogeneous information network modeling helps to combine several kinds of data including train schedules, service types, and passenger information. This method provides more informed decision-making and helps to grasp the intricate relationships among several components in a railway system. Heterogeneous Information Networks (HINs) are a kind of network architecture including several kinds of nodes and relationships among them. HINs enable the description and analysis of complex real-world systems including several kinds of entities and their interactions, unlike conventional homogeneous networks where all nodes and linkages are of the same kind [12].

Among the numerous kinds of HINs are bipartite, star-schema, and multi-hub ones. Multiple hubs in the context of heterogeneous information networks (HINs) are those circumstances whereby several entities regarded as central or "hub" entities in the heterogeneous information network. Multiple-hub HINs feature more than one central entity with relationships to other entities acting as the hub of the network.

C. Research Method

This study aims to analyze the connectivity patterns of the railway station network in India using a network analysis approach. The study divides the research methodology into several systematic steps:

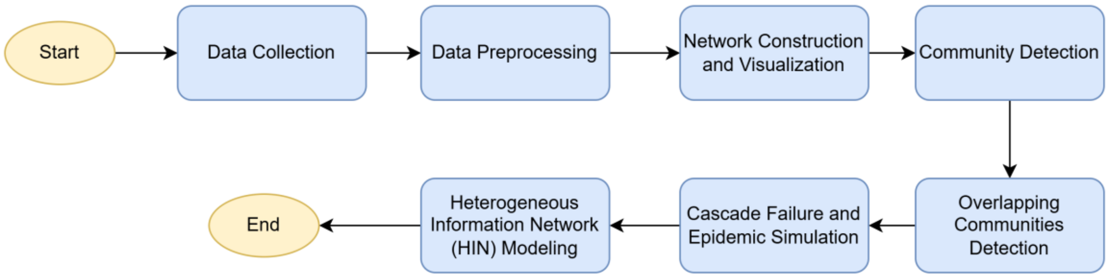

Figure 1. Research Methodology

1. **Data Collection**

   Data collecting from the freely accessible Indian train timetables comes first. Important data like train numbers and names, station codes and names, arrival and departure times, and station distances is contained in this dataset. The foundation of the network of railway stations is this information.
2. **Data Preprocessing**

   After the data is collected, a data cleaning process is carried out by removing entries that have empty values (NaN) and eliminating duplication. At this stage, data transformation is also carried out:
   - Station Code is used as the node ID.
   - Station Name is used as the node label.
   - Source Station and Destination Station are used to form edges.

   This stage produces a collection of nodes and edges that are ready to be used for network construction.
3. **Network Construction and Visualization**

   The Gephi tool then turns the processed nodes and edges into graph representation. Visualization of networks is designed to help to analyze network architecture including connection patterns, node degree distributions, and other graph properties.
4. **Community Detection**

   Leiden Algorithm is applied for community structure study. This stage aims to find groupings of stations that, in comparison to their links to other groups, are rather closely linked. This split of communities inside the network results from which one may better grasp the modularity of the network.
5. **Overlapping Communities Detection**

   Using the Clique Percolation Method (CPM), one can find overlapping communities. This detection is crucial since depending on the service path and connectivity, a single station in a real transportation network can belong to numerous operating communities at once.
6. **Cascade Failure and Epidemic Simulation**

   We investigate the network's susceptibility to systematic disturbances using a cascading failure scenario. The first node of infection is the node having the highest degree—hub. It is modeled over numerous timesteps how the failure spreads to other network nodes, therefore determining important nodes requiring priority in risk management.
7. **Heterogeneous Information Network (HIN) Modeling**

   Modeling heterogeneous information networks (HIN) with a multi-hub technique marks the last phase of the study. This model enables the depiction of several kinds of things (stations, arrival/departure times, station names) together with the interactions among them. HIN modeling seeks to depict the intricacy of the transportation system not amenable for one kind of node and edge.

D. Result and Discussion

1. Data Collection and Preprocessing

This study utilizes the Indian Railways timetable dataset provided by the Government of India. This dataset contains a total of 180.000 more data entries with 12 features. Due to limited resources, we chose 10,000 more data for this study. That encompasses various information related to train journeys. The data used includes Train No, Train Name, SEQ, Station Code, Station Name, Arrival Time, Departure Time, Distance, Source Station, Source Station Name, Destination Station, and Destination Station Name. Here is the dataset used:

TrainNo	TrainNam	SEQ	StationCo	StationNa	ArrivalTim	Departure	Distance	SourceStat	SourceStat	Destinatio	DestinationStationName
107	SWV-MAO	1	SWV	SAWANTW	0:00:00	10:25:00	0	SWV	SAWANTW	MAO	MADGOAN JN.
107	SWV-MAO	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTW	MAO	MADGOAN JN.
107	SWV-MAO	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTW	MAO	MADGOAN JN.
107	SWV-MAO	4	MAO	MADGOAN	12:10:00	0:00:00	78	SWV	SAWANTW	MAO	MADGOAN JN.
108	VLNK-MAC	1	MAO	MADGOAN	0:00:00	20:30:00	0	MAO	MADGOAN	SWV	SAWANTWADI ROAD
108	VLNK-MAC	2	KRMI	KARMALI	21:04:00	21:06:00	33	MAO	MADGOAN	SWV	SAWANTWADI ROAD
108	VLNK-MAC	3	THVM	THIVIM	21:26:00	21:28:00	51	MAO	MADGOAN	SWV	SAWANTWADI ROAD
108	VLNK-MAC	4	SWV	SAWANTW	22:25:00	0:00:00	83	MAO	MADGOAN	SWV	SAWANTWADI ROAD
128	MAO-KOP	1	MAO	MADGOAN	19:40:00	19:40:00	0	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	2	KRMI	KARMALI	20:18:00	20:20:00	33	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	3	THVM	THIVIM	20:40:00	20:42:00	51	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	4	SWV	SAWANTW	21:16:00	21:18:00	83	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	5	KUDL	KUDAL	21:38:00	21:40:00	104	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	6	SNDD	SINDHU D	21:54:00	21:56:00	114	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	7	KKW	KANKAVAL	22:18:00	22:20:00	132	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	8	VBW	VAIBHAVW	22:40:00	22:42:00	163	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	9	RAJP	RAJAPUR F	22:56:00	22:58:00	179	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	10	RN	RATNAGIR	23:52:00	23:57:00	244	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	11	SGR	SANGMESI	0:34:00	0:36:00	280	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	12	CHI	CHIPLUN	1:10:00	1:12:00	322	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	13	KHED	KHED	1:52:00	1:54:00	352	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	14	MNI	MANGAON	4:00:00	4:02:00	422	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	15	ROHA	ROHA	4:55:00	5:00:00	455	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	16	PNVL	PANVEL	6:20:00	6:25:00	532	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	17	KJT	KARJAT	7:40:00	7:45:00	560	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS
128	MAO-KOP	18	LNL	LONAVLA	8:30:00	8:33:00	588	MAO	MADGOAN	KOP	CHHATRAPATI SHAHU MAHARAJ TERMINUS

Figure 2. Dataset

Before utilizing the dataset, it is necessary to perform pre-processing or data preparation to enable visualization and analysis of the dataset (figure 2). The first step in the pre-processing process involves removing data with NaN (Not a Number) values and checking for duplicate values. Data with NaN values are removed because NaN values indicate missing or nonexistent data, which can affect the analysis and prevent data bias. The number of data entries in the dataset after removing NaN values and checking for duplicate values is 10,178.

The next step is to create nodes and edges. From the dataset used, Station Code and Station Name are selected as nodes. Station Code and Station Name are chosen as they are the most important features in the dataset. Station Code is transformed into an ID, while Station Name is transformed into a Label to facilitate the network creation process (figure 3). The number of nodes formed is 214 nodes.

Subsequently, Source Station and Destination Station are chosen as edges since the focus of this research is to analyze the network formed by railway journeys. The selected features for the edges are Source Station (transformed into Source), Destination Station (transformed into Target), Train Name, Arrival Time, Departure Time, Distance, Source Station Name, and Destination Station Name (figure 4). The number of edges formed is 499 edges.

Here are the selected nodes and edges from the dataset used:

id	Label
SWV	SAWANTWADI ROAD
SWV	SAWANTWADI ROAD
SWV	SAWANTWADI ROAD
SWV	SAWANTWADI ROAD
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.
MAO	MADGOAN JN.

Figure 3. List Node

Source	Target	Train Nam	Arrival tim	Departure	Distance	Source Sta	Destination Station Name
SWV	MAO	SWV-MAO	0:00:00	10:25:00	0	SAWANTW	MADGOAN JN.
SWV	MAO	SWV-MAO	11:06:00	11:08:00	32	SAWANTW	MADGOAN JN.
SWV	MAO	SWV-MAO	11:28:00	11:30:00	49	SAWANTW	MADGOAN JN.
SWV	MAO	SWV-MAO	12:10:00	0:00:00	78	SAWANTW	MADGOAN JN.
MAO	SWV	VLNK-MAC	0:00:00	20:30:00	0	MADGOAN	SAWANTWADI ROAD
MAO	SWV	VLNK-MAC	21:04:00	21:06:00	33	MADGOAN	SAWANTWADI ROAD
MAO	SWV	VLNK-MAC	21:26:00	21:28:00	51	MADGOAN	SAWANTWADI ROAD
MAO	SWV	VLNK-MAC	22:25:00	0:00:00	83	MADGOAN	SAWANTWADI ROAD
MAO	KOP	MAO-KOP	19:40:00	19:40:00	0	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	20:18:00	20:20:00	33	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	20:40:00	20:42:00	51	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	21:16:00	21:18:00	83	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	21:38:00	21:40:00	104	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	21:54:00	21:56:00	114	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	22:18:00	22:20:00	132	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	22:40:00	22:42:00	163	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	22:56:00	22:58:00	179	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	23:52:00	23:57:00	244	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	0:34:00	0:36:00	280	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	1:10:00	1:12:00	322	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	1:52:00	1:54:00	352	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS
MAO	KOP	MAO-KOP	4:00:00	4:02:00	422	MADGOAN	CHHATRAPATI SHAHU MAHARAJ TERMINUS

Figure 4. List Edges

After data preparation, the node data and edge data are inputted into the Gephi application for analyzing the formed network

2. Visualization Graph

Based on the railway station data consisting of 214 nodes and 499 edges, visualized using Gephi, the network visualization results are as follows:

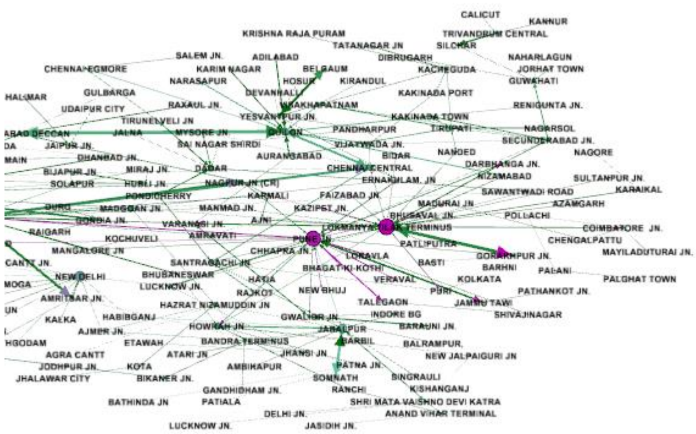

Figure 5. Visualization Graph

From the above graph visualization (figure 5), it can be observed that the network formed by nodes and edges can be used to determine the type of network based on average path length, degree distribution, and clustering coefficient.

3. Type of Network

After creating the graph visualization, the next step is to determine the type of the formed network. Several aspects can be calculated to determine the network type, such as average path length (table 1), degree distribution (figure 6), and clustering coefficient (figure 8). The average path length represents the average distance between one node and other nodes in the network [13]. It is defined as the average length of the shortest paths for all pairs of nodes in the network. A smaller distance indicates higher information efficiency [14]. Here are the results of calculating the average path length using the Gephi application:

Table 1. Average Path Length

Diameter	***Radius***	***Average Path Length***
9	0	4.0641

Next, we calculate the degree distribution of the network. Degree distribution refers to the probability distribution of degrees in a graph or social network. The degree of a node in a network is the number of connections or edges it has with other nodes. The degree distribution describes the probability distribution of these degrees across the entire network. Here are the results of calculating the degree distribution using the Gephi application:

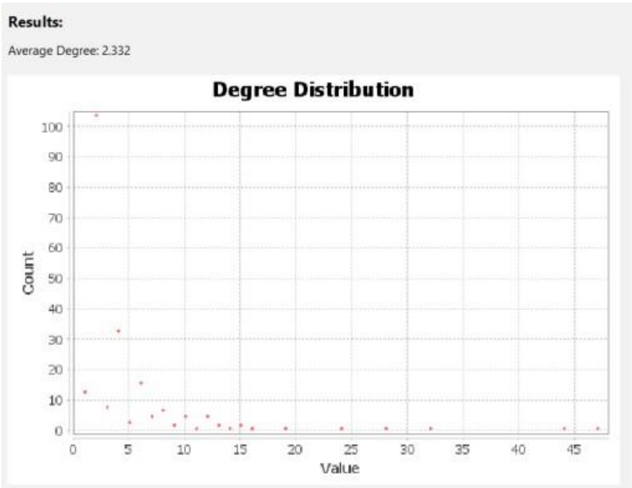

Figure 6. Degree Distribution

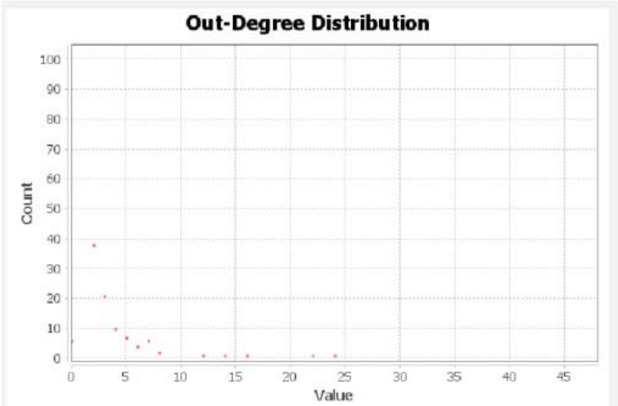

Figure 7. Out-Degree Distribution

After calculating the average path length (table 1) and degree distribution (figure 6), the next step is to calculate the clustering coefficient (figure 8). The clustering coefficient is a measure of how many nodes in a network tend to cluster together. It provides insights into the local connectivity patterns within the network and can indicate the presence of tightly knit groups or communities of nodes. Here are the results of calculating the clustering coefficient using the Gephi application:

Figure 8. Clustering Coefficient Distribution

Scale-free networks are a type of network structure where a few nodes have a significantly higher number of connections compared to the rest of the nodes. A network can be considered a scale-free network (table 2) if it has a relatively low average path length (most nodes have low connectivity degrees and only a few nodes have high connectivity degrees), a power-law degree distribution (short average path length but with a few very long distances), and a relatively high clustering coefficient with a value greater than 1/n (indicating strong local connectivity within the network).

Based on the calculations of average path length, degree distribution, and clustering coefficient, it can be concluded that the type of network formed is a scale-free network. There is a node with a significantly high degree, which is the LTT (LOKMANYA TILAK TERMINUS) Station with 24 connections. The network consists of 214 nodes and 499 edges, with 1/n being 0.00467. Here are the aspects of scale- free networks observed in the generated network:

Table 2. Aspects of Scale Free Network

Average Path Lenght	***Degree Distribution***	***Clustering Coefficient***
4.0641	Power-low	0.043

4. Community Detection Result

In the community detection stage, Community detection refers to the process of identifying groups or communities within a network where nodes or individuals have stronger connections with each other compared to those outside the community [15]. In this study, the Leiden algorithm is used for community detection. Here is the configuration in the Gephi application for community detection using the Leiden algorithm:

Figure 9. Configuration Leiden Algorithm

The following is an explanation of the Leiden Algorithm configuration (figure 9) used :

- The Leiden algorithm is used for community detection in networks using an iterative approach.
- The Constant Potts Model (CPM) is used as the quality function to measure the quality of community separation.
- The resolution parameter determines the level of detail in the resulting community partitions.
- The Leiden algorithm is executed for 10 iterations to maximize modularity.
- The Leiden algorithm is run once with a random seed of 0.

Here are the results obtained in the Gephi application for community detection using the Leiden algorithm:

Results	
Quality	0.9766214953271019
Number of clusters	23

Figure 10. Result of Leiden Algorithm

The following is an explanation of the results of the Leiden Algorithm (figure 10) :

- Quality indicates the level of quality in the community separation achieved by the Leiden algorithm. A higher quality value indicates a better separation of communities.
- Number of clusters refers to the number of clusters or communities identified in the network using the Leiden algorithm. In this case, 23 clusters were found.

  Here are the visualization results of the Leiden algorithm in the Gephi application :

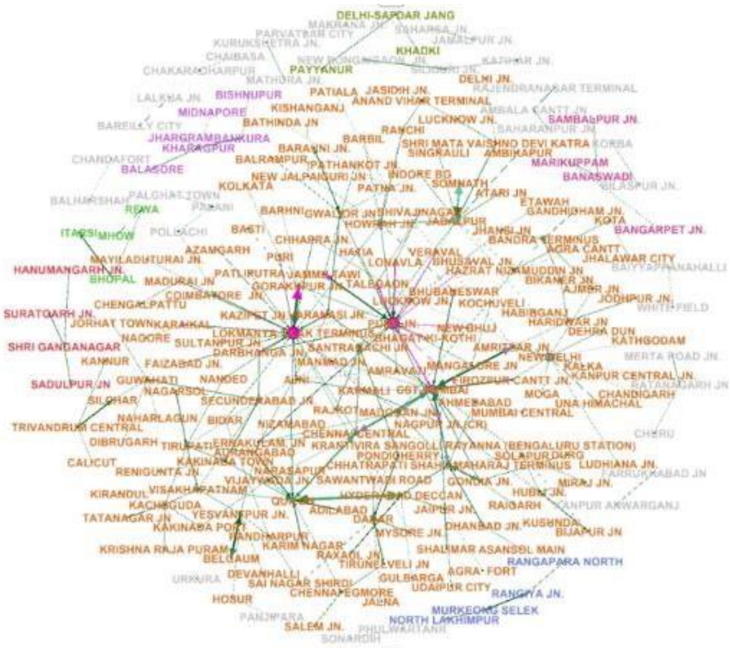

Figure 11. Visualization Result of Leiden Algorithm

5. Detection of Overlapping Communities

   Overlapping communities refer to a phenomenon in network analysis where nodes in a network can belong to multiple communities simultaneously. A community in a network is a group of nodes that exhibit strong interconnections or similar characteristics. Traditionally, community detection algorithms assign each node to a single community, assuming that nodes are mutually exclusive. However, in real-world networks, nodes often participate in multiple communities, reflecting the complex and overlapping nature of relationships [16].

In this study, the Clique Percolation Method (CPM) is used to detect overlapping communities. Here are the results of detecting overlapping communities using Python:

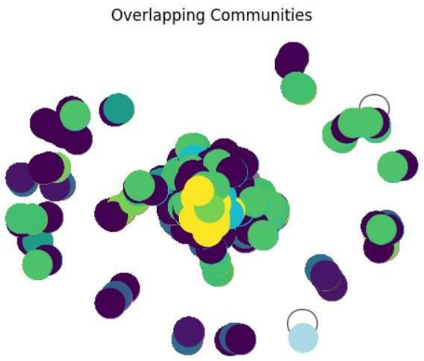

Figure 12. Overlapping Communities with K = 1

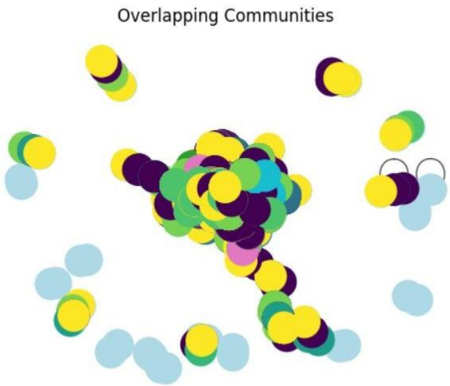

Figure 13. Overlapping Communities with K = 2

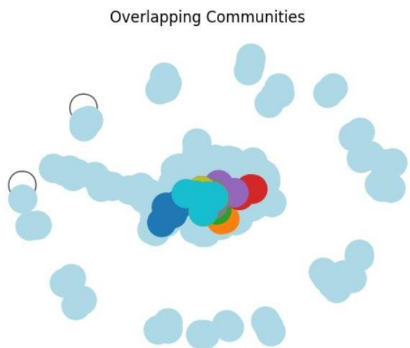

Figure 14. Overlapping Communities with K = 3

Figure 15. Overlapping Communities with K = 4

From the above results of detecting overlapping communities (figure 12, figure 13, figure 14, figure 15), it indicates that if the value of K is greater than or equal to 4, there are no overlapping communities, whereas if K is less than or equal to 3, there are overlapping communities.

6. Potent of Cascade Failure or Epidemic

Cascade failures refer to a phenomenon where the failure of one component or node in a system triggers a chain reaction of subsequent failures in other interconnected components or nodes [17]. In network analysis, Cascade Failure refers to the spreading of diseases/viruses/failures in a network that causes a domino effect, while Epidemic refers to the rapid spread of information or viruses in a network with high speed and coverage. To check the Cascade Failure and Epidemic Potential, the first step is to determine the Initial node by assigning "healthy" status to all nodes, except for the LTT node which is given the "infected" status. This is because LTT is a node with a high degree. Then, the status is updated by checking all nodes. If a node is detected to have the "infected" status, its neighbors will also be checked. Neighbors of an "infected" node that have the "healthy" status will be assigned a new status, which is "infected". This process is repeated until there is no possibility of infected nodes. Here are the results of checking the cascade failure and epidemic potential using Python with 10 timesteps:

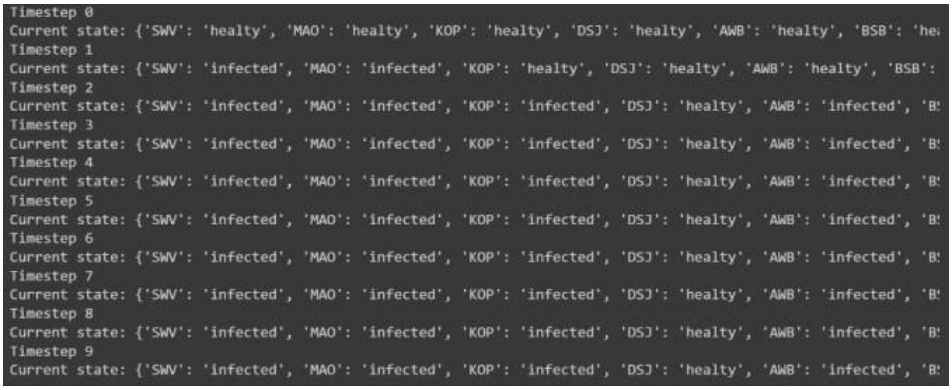

Figure 16. Result of Checking the Cascade Failure

From the results above (figure 16), it can be seen that the majority of nodes change their status to infected when the node with the highest degree is given the infected status.

7. Multiple-Hub Scheme on Heterogenous Information Network
   In this research, the multi-hub type is used to create HINs (Heterogeneous Information Networks). In a Multiple-Hub, there are multiple core entities that act as hubs within the network and have connections with other entities. Here is the HINs formed with the multi-hub type:

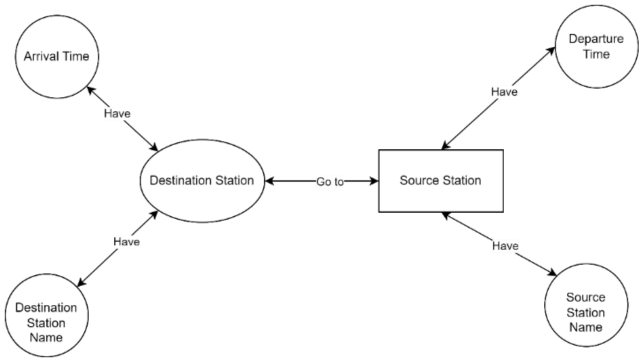

Figure 17. Multi-Hub

In the above HINs image (figure 17), there are several connected entities. The "destination station" and "source station" entities act as central hubs that are interconnected. The "destination station" entity has connections with "arrival time" and "destination station name" entities, while the "source station" entity has connections with "departure time" and "source station name" entities.

E. Conclusion

Based on the analysis results, the type of network formed from the Indian railway's timetable dataset is a scale-free network. This is because it has a relatively low average path length, a power-law degree distribution, and a relatively high clustering coefficient with a value greater than 1/n. The community detection using the Leiden algorithm shows the formation of 23 clusters with a quality value of 0.97662. Overlapping communities are identified using the CPM method, where overlapping communities exist when the value of K is less than 4, while no overlapping communities are found when K is greater than 4. The cascade failure or epidemic potential check is performed by assigning an infected status to the node with the highest degree for a timestep of 10, and the result shows that some nodes will change their status to infected. The formed network is a HINs network as it consists of various types of entities with different characteristics.

F. Acknowledgment

The author would like to express his deepest gratitude to Telkom University for the support, resources, and conducive research environment so that this research can be completed.

G. References

[1] N. K. Rajput, P. Badola, H. Arora, and B. A. Grover, “Complex Network Analysis of Indian Railway Zones.” 2020. [Online]. Available: https://arxiv.org/abs/2004.04146

[2] Q. Hu, L. Bian, J. Lin, and M. Tan, “Multi-Objective Optimization for High-Speed Railway Network Based on ‘Demand–Supply–Management’ Model,” *Applied Sciences*, vol. 13, no. 20, 2023, doi: 10.3390/app132011406.

[3] W. Wang *et al.*, “Analysis of the Chinese railway system as a complex network,” *Chaos, Solitons & Fractals*, vol. 130, p. 109408, 2020, doi: https://doi.org/10.1016/j.chaos.2019.109408.

[4] A. Kianinejadoshah and S. Ricci, “Capacity Assessment in Freight-Passengers Complex Railway Nodes: Trieste Case Study,” *Infrastructures*, vol. 7, no. 8, 2022, doi: 10.3390/infrastructures7080106.

[5] A. Šperka, J. Čamaj, M. Dedík, and Z. Bulková, “Evaluation Methodology of the Railway Stations Using the AHP Method in the Transport Hubs from the Freight Transport Point of View,” *Infrastructures*, vol. 8, no. 12, 2023, doi: 10.3390/infrastructures8120177.

[6] I. Garg, U. Soni, S. Agrawal, and A. Shukla, “Application of Complex Network Principles to Identify Key Stations in Indian Railway Network,” in *Proceedings of International Conference on Computational Intelligence*, R. Tiwari, M. F. Pavone, and R. Ravindranathan Nair, Eds., Singapore: Springer Nature Singapore, 2023, pp. 197–204.

[7] F. Lamanna, M. Prisma, and G. Medeossi, “Unveiling connectivity patterns of railway timetables through complex network theory and Infomap clustering.” 2025. [Online]. Available: https://arxiv.org/abs/2504.09214

[8] A. Jain and P. Manimaran, “Detecting communities via edge Random Walk Centrality.” 2023. [Online]. Available: https://arxiv.org/abs/2309.05614

[9] E. W. Ambarsari, A. Paramita, and Desyanti, “Clustering of YouTube Viewer Data Based on Preferences using Leiden Algorithm,” *International Journal of Informatics and Data Science*, vol. 1, no. 2, pp. 55–63, 2024.

[10] S. H. Hairol Anuar *et al.*, “Comparison between Louvain and Leiden Algorithm for Network Structure: A Review,” *Journal of Physics: Conference Series*, vol. 2129, no. 1, p. 012028, Dec. 2021, doi: 10.1088/1742-6596/2129/1/012028.

[11] A. Baudin, M. Danisch, S. Kirgizov, C. Magnien, and M. Ghanem, “Clique Percolation Method: Memory Efficient Almost Exact Communities,” in *Advanced Data Mining and Applications*, B. Li, L. Yue, J. Jiang, W. Chen, X. Li, G. Long, F. Fang, and H. Yu, Eds., Cham: Springer International Publishing, 2022, pp. 113–127.

[12] Y. Sun and J. Han, “Mining heterogeneous information networks: A structural analysis approach,” *SIGKDD Explorations*, vol. 14, pp. 20–28, Jan. 2012, doi: 10.1145/2481244.2481248.

[13] I. E. Yanti, O. N. Pratiwi, and R. Y. Fa’rifah, “Analisis Respon Masyarakat Terhadap Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Social Network Analysis,” *eProceedings of Engineering*, vol. 8, no. 5, pp. 8988–8999, 2021.

[14] A. Utami and A. Alamsyah, “Analisis Karakteristik Strategi Konten E-commerce Indonesia menggunakan Pendekatan Social Network Analysis

untuk Market Intelligence," *Jurnal Manajemen Bisnis (Performa)*, vol. 15, no. 2, pp. 116–123, 2018, doi: https://doi.org/10.29313/performa.v0i0.3473.
[15] S. Souravlas, A. Sifaleras, M. Tsintogianni, and S. Katsavounis, "A classification of community detection methods in social networks: A survey," *International Journal of General Systems*, vol. 50, pp. 63–91, Jan. 2021, doi: 10.1080/03081079.2020.1863394.
[16] V. da F. Vieira, C. R. Xavier, and A. G. Evsukoff, "A comparative study of overlapping community detection methods from the perspective of the structural properties," *Applied Network Science*, vol. 5, no. 1, p. 51, Aug. 2020, doi: 10.1007/s41109-020-00289-9.
[17] R. Burkholz, H. Herrmann, and F. Schweitzer, *Explicit size distributions of failure cascades redefine systemic risk on finite networks*. 2018. doi: 10.48550/arXiv.1802.03286.