Topological characteristics of the public transportation network in Hamburg

1. Introduction

In many cities, public transportation modes have been integrated in order to increase accessibility and reduce commute times. For the management of public transport networks (PTN), it is important to periodically evaluate the network's connectivity and ability to handle congestion and to adjust its spatial configurations in accordance with such evaluations. As the PTN systems have become vital for increasing the ridership, it would be of great interest to policy makers and planners to examine the topology of PTN in order to have more data-driven and graph-based knowledge of the structure and characteristics of PTN and to assess strategies for improving their performance.

Network analysis is an important aspect of transport geography because it describes the distribution of edges[1] and the arrangement of nodes as well as the relationships between them. Viewing transit systems as complex networks and applying the network-science perspective and the graph theory can have many benefits, including enabling a holistic view of the system, exploring network strengths and weaknesses and carrying out the comparisons between networks in different geographical locations (Derrible, 2012). Recommendations adopted by planners and designers for future plans and the optimization of performance of the transportation system can be developed based on the results of such an analysis (Hong, Tamakloe, Lee, & Park, 2019).

However, the complexities of urban public transportation networks make their examination rather challenging. This problem arises because transportation networks often consist of a large number of stations and links. Due to this complexity, research applying the graph[2]-based measures to systematically analyze and compare the networks remain to be quite limited.

---

1. "Edges" is a term in graph theory referring to the connection between a pair of nodes. In the current research of transportation networks, we also use "links" as an identical term to edges.
2. In graph theory, a network is considered as a graph. The basic units for constructing a graph are "nodes", which are represented by a point, and "edges", which are the connection between nodes and are represented by lines.

In the last few years, increasing availability of comprehensive data sources and powerful tools of information technology enable to combination of network science and computational geometry in order to accelerate the automatic retrieval and analysis of extremely large amounts of data (Boeing, 2017). The geodata can then be used to construct network graphs, visualize networks and analyze the attributes of networks by calculating the metric and topological indicators.

The objective of this chapter is to analyze the topological and structural properties of PTN in Hamburg metropolitan area by taking advantage of the recent developments of the network science tools, including the computer programming techniques, implementation of known algorithms and Big Data analysis. Based on the analysis of the topological characteristics, the current study examines whether the PTN in Hamburg possess the required features for advancing the efficient operation and the resilience to failure or disruption.

2 Literature review

Public transportation systems are typical examples of complex networks. They are made up of a set of nodes (i.e., stations) and links. The stations are distributed across the network and are connected by the links. The combination of several links between stations forms the routes or paths. Therefore, many ideas in graph theory and network analysis can be employed to explore the implications of the PTN structure, patterns and properties on public transportation (Shi, Wen, Zhao, & Wu, 2019) (Soh, et al., 2010) (Zhang, Li, Deng, & Wang, 2014) (Mishra, Welch, & Jha, 2012).

One of the important elements in understanding transport networks is the topological properties, which explain the physical arrangement, connection and relationship between the elements of a network (Zahedi, Mawengkang, Masri, Ramon, & Putri, 2019). The advantages of the application of topological properties include “simplifying complex scenarios and modeling numerous complex interactions within systems in order to allow the engineer to make modifications to the existing network for enhancement of services” (Hong, Tamakloe, Lee, & Park, 2019). In the current study, three different topological properties, i.e. geometric, small-world and scale-free properties, are explored and the indicators for studying these properties are introduced.

The first type of topological properties is the geometric properties, which can be measured by the attributes like the size and density of the network. Measurements for scrutinizing geometric properties include **number of nodes (*N*), number of links (*M*), link-node ratio, and degree of connectivity(γ)**.

The other types of topological properties, namely small-world and scale-free properties, have been identified in several studies as the two unique features required for efficient operation and network resilience (Zhang L. , Lu, Fu, & Li, 2019). First of all, a network has small-world properties if the following two conditions are satisfied:

1) the **average path length (*V*)** is close to that of a random network with the same number of nodes;

2) the **average Clustering Coefficient** is significantly higher than that of a random network with same number of nodes.

Secondly, if the **degree distribution (*p(k)*)** of a network follows the power law distribution, it is considered to have the scale-free property. Because this means that, compared to the exponential distribution, the degree distribution of a network reduces more gradually. This also indicate the possible presence of cluster that have many links and, therefore, have large degree (Rommel, 2014) (Humphries & Gurney, 2008).

Transport networks with these two properties have been recommended for urban transit network because they possess desirable features, including robustness and resilience to failure or disruption (Chopra, Dillon, Bilec, & Khanna, 2016) (Wu, Gao, Sun, & Huang, 2004) (Huang, Grigolon, Madureira, & Brussel, 2018).

3. Data and computational tools

In this chapter we will use the above-mentioned indicators for the systematic analysis of transit networks using data from the General Transit Feed Specification (GTFS). This section introduces the background GTFS and describes the structure of the files in the dataset.

GTFS is an open platform and a data model initiated by Google and the Portland, Oregon, public transit agency (TriMet), in order to allow transit organizations to publish their schedules for routing and visualization purposes at a low cost (McHugh, 2013).

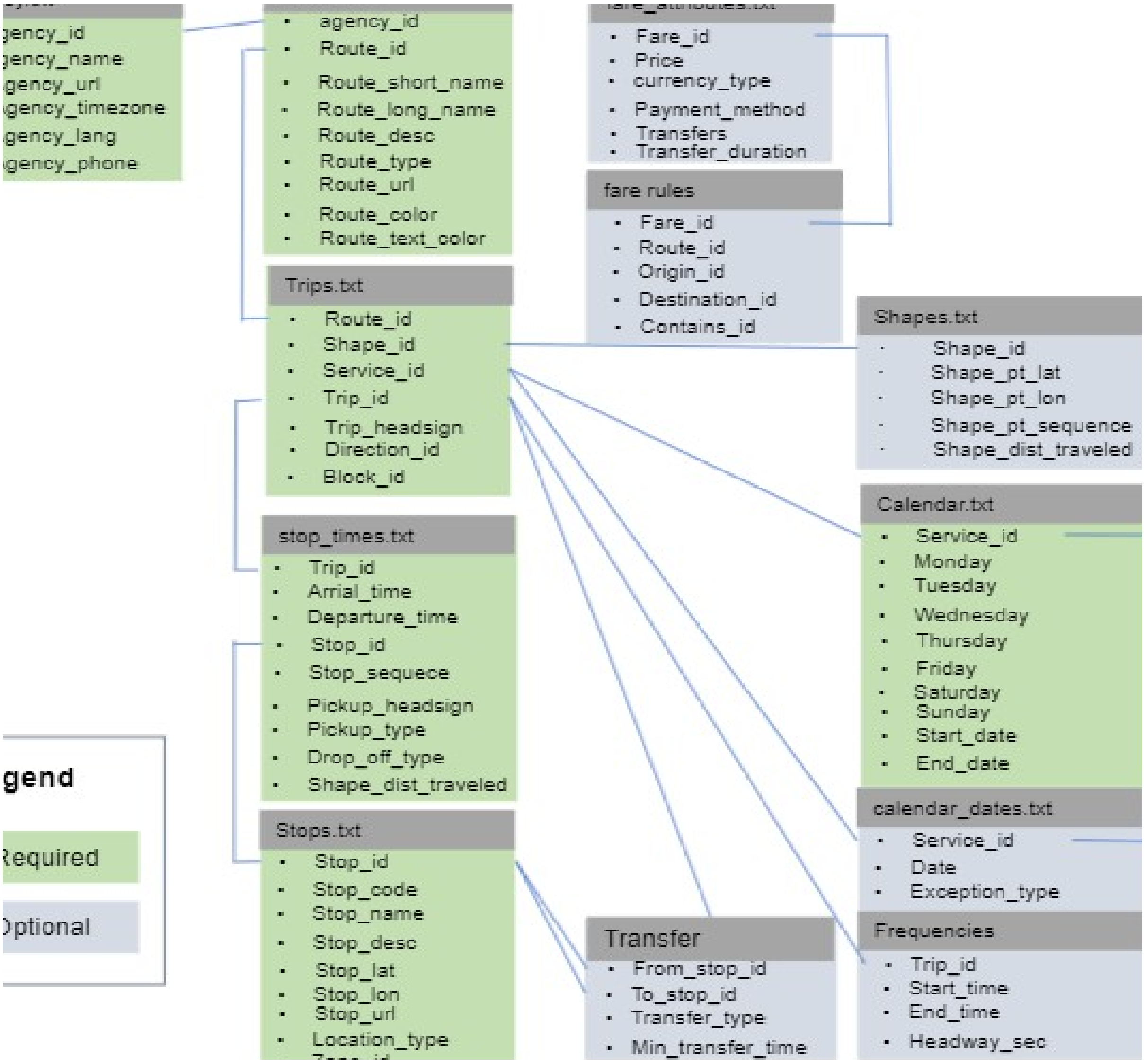

Figure 2.1 GTFS files and relationships.

Google defined a publishing standard and a common format for transit agency operational data, which include public transportation schedules and associated geographical information (e.g., stops, stop times, routes). The GTFS Feed provided by a public transit agency is a zip file containing six mandatory files with comma-separated values (CSV) and seven optional files. Each CSV file contains information about a part of the transport operator's system.

Figure 2.1 shows these files and their relationship. Combining these files together, we can describe the stops, routes, and schedules of an entire transit system and the outcome consists of many tables of a relational database (McHugh, 2013).

In this chapter, the Python library of peartree 0.6.4 and partridge 1.1.1 are employed to 1) check the Transit.Land API and query it for any and all operators that serves Hamburg; 2) retrieve the zip location of the original GTFS; 3) download the zip file to a local temporary directory; 4) convert the GTFS data into a directed multigraph network.

Through the entire thesis, OSMnx version 0.15.1 (Boeing, 2017) (Boeing, 2019), which is built on top of Python3 libraries of NetworkX 2.4, matplotlib 3.3.0, and GeoPandas 0.8.1, serves as an automatic, scalable, reproducible tool that performs the network graph analytics operations on GTFS data and calculates the indicator values. Pandas 1.0.5 and GeoPandas 0.8.1 are used in this research to carry out statistical analysis of the calculated indicator values. The figures in this research are generated using Matplotlib 3.3.0.

4. Indicator calculate and results: topological properties of the network

4.1 Case study area

The city of Hamburg in Germany is selected as the study area in the current research in order to exemplify the process of the proposed methodology framework. With the population of more than 1.84 million inhabitants and the size of 755.22 km^2, Hamburg is the second largest city in Germany.

The Hamburg metropolitan region, which includes the city of Hamburg and its neighboring district, has a population of more than 3 million people[3]. The Hamburg metropolitan region corresponds roughly to the service area of the Hamburger Verkehrsverbund (HVV), which can be translated as Hamburg Transport Association. The complete and integrated HVV transit network, including both railway public transport (U- and S-Bahn[4]) and bus networks in Hamburg metropolitan region, is shown in Figure 2.2, which is created by OSMnx using the retrieved GTFS data.

3 http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=urb_lpop1&lang=en, Retrieved 15 June 2021

4 U-Bahn, which literally translates as “underground railway” is the urban rapid transit or mass rapid transit (MRT) in German cities. S-Bahn is the German urban-suburban rail serving a metropolitan region and connecting the nearby districts and towns to the city.

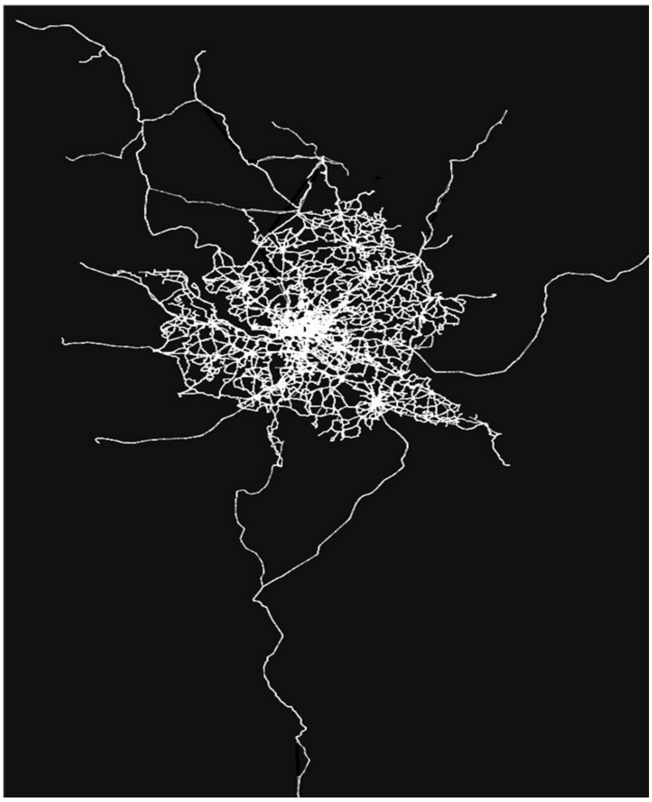

Figure 2.2 HVV network.

Table 2.1 Computed properties of Hamburg integrated transit network.

	Property	Hamburg	Seoul
Topological characteristics	Number of nodes, *N*	14416	12873
	Number of links, *M*	17914	19354
	Link-node ratio	1.24	1.5
	Degree of Connectivity,	0.000172	0.000233
Small-world properties	Average path length (in network distance), *V*	11.396	28.169
	Average Clustering Coefficient, *C_avg*	0.038	0.028
Scale-free network properties	Average degree, *k_avg*	2.485	3.007

Table 2.1 provides the results of the computed network statistics. In order to evaluate the HVV network, we also include the previous research of the public transportation network in Seoul (Hong, Tamakloe, Lee, & Park, 2019) as a reference for comparison. In the following sections, we explain the relationship between these indicators and the level of network connectivity.

4.2 Geometric properties

1) Size

In the current research, the size of the networks is measured by the **number of nodes (*N*)**, i.e. stations, and the number of **links (*M*)**. Naturally, there exists a positive correlation between the size of the network and the level of accessibility. If the covered area remains constant, a higher number of stations and links (which also means a higher number of connections and routes) provides higher accessibility to destinations.

A higher number of nodes does not necessarily mean that the network also has a higher number of links than another comparable network. For example, as one can see in Table 2.1, the total number of stations in the integrated system (including the U-Bahn, the S-Bahn and the bus networks) is 14,416, and the number of links is 17,914. In comparison with the PTN in Seoul, the HVV network has more stations but less links. As this example shows, it is not possible to evaluate the connectivity by using only the number of nodes or only the number of links. Rather, such evaluation requires the use of a complex indicator, namely the **link-node ratio**.

2) Link-node ratio

The link-node ratio is an index of connectivity which divides the total number of links by the total number of nodes in a catchment area (Cerin, et al., 2013) (Molaei, Tang, & Hardie, 2021). This ratio formally expressed by the following formula:

$$link\text{-}node\ ratio = N \,/\, M \tag{1}$$

where N is the total number of nodes and M is the total number of links in the network. Naturally, higher values indicate better connectivity. Ewing (1996) suggests setting the threshold for high street connectivity at the link-node ratio of 1.4. In the HVV network case, the link-node ratio is 1.24, which is lower than the threshold score and therefore falls short of reaching the high connectivity level. In comparison with the Seoul, HVV is also lower than the score of the PTN in Seoul, which is 1.5. Therefore, in terms of the link-node ratio, the connectivity of the Seoul PTN is better than that of the HVV network.

3) Degree of Connectivity, γ

The Degree of Connectivity (γ) describes the relationship and interaction between the elements, i.e. nodes and links, in a network system. In this research we use the gamma index to measure the Degree of Connectivity, which is defined by dividing the observed number of links by the possible maximum number of links in the network (Rommel, 2014). It can be calculated as follows:

(2)

where N is the total number of nodes in the network, M is the total observed number of links in the network and M_{max} is the possible maximum number of links in the network. The value of gamma lies between 0 and 1, where the value of 1 indicates a completely connected network. This means that each node is connected with one link to all other nodes (Czerkauer-Yamu, 2012). In practice, such a perfectly connected network would of course be extremely unlikely.

In the HVV network case, the gamma index (i.e., the Degree of Connectivity) is 0.000172, which is smaller than the Seoul network's value of 0.000234. Based on the results of the underlying basic network properties displayed in Table 2.1, we infer that the HVV network does not provide higher connectivity compared to the theoretical reference and the transportation network in Seoul.

4.3 Small-world properties

A network has the small-world properties, proposed by Watts and Strogatz (1998), if it possesses the following features: 1) most nodes are not neighbors of one another; 2) most nodes can be reached by a small number of nodes by small number of steps (Mohmand & Wang, 2014). From the perspective of transportation research, investigating the small-world properties of a transportation network can be very useful for the researchers evaluating a transportation network. First of all, due to its short average path length, a network with the small-world properties means it has high capability and connectivity to link the nodes (stations) more effectively. Therefore, analyzing the small-world properties informs us about **how efficient in communicating the network is**. Secondly, a small-world network is structurally more resilient to failure or disruptions (Chopra, Dillon, Bilec, & Khanna, 2016). Therefore, analyzing the small-world properties can be very useful when evaluating **how robust a network is**, i.e. how good it will perform if one of its nodes (stations) stops functioning for some reason.

Typical features of a spatial structure with the small-world properties are, compared to a random network with the same number of nodes, its <u>small average shortest path length</u> and <u>large Clustering Coefficient</u>. Previous research (Humphries & Gurney, 2008) (Mohmand & Wang, 2014) (Telesford, Joyce, Hayasaka, Burdette, & Laurienti, 2011) (Chopra, Dillon, Bilec, & Khanna, 2016) has proposed to compare the **average path length *(V)*** and **average Clustering Coefficient *(C_avg)*** of the real network to an Erdos–Renyi (E–R) random network constructed with the same number of nodes as the real network.

The steps are as the following.

i) Calculate the average path length, V_{real}, and the Clustering Coefficient, C_{real}, of the real network.

ii) Calculate the average path length, V_{random}, and the Clustering Coefficient, C_{random}, of a random network with the same number of nodes.

iii) Calculate the normalized shortest path $\lambda = V_{real}/V_{random}$ and $\gamma = C_{real}/C_{random}$.

iv) If λ and γ fulfil the criteria, $\lambda \approx 1$ and $\gamma > 1$, the network can be identified as a small-world network.

In the following, this method is applied to see if the small world properties are present in the HVV network.

1) Average Path Length (*V*)

In network analysis, the shortest path length is the minimum number of links along the shortest path in order to travel from the origin node to the destination node. The average path length[5], V, indicates the average number of links passed through the shortest paths out of all possible pairs of nodes. It is an indicator measuring the progression of a network in time, i.e. how easy it is to move from one node to another node. It can be formally expressed as

(3)

where N is to the total number of nodes in the network, V is the average path length and d_{uv} corresponds to the length of the shortest path between stations u and v. A small average path length indicates that there is good connectivity and efficient communication among the stations in the network, regardless of geographical distance.

In the case of the HVV network, the average path length is 11.396, which is smaller than the value of 28.169 in the case of Seoul. This means that, on average, a passenger has to transverse 28.169 links in Seoul but only 11.396 links in Hamburg in order to travel to his or her destination.

2) Clustering Coefficient (C) and average Clustering Coefficient (C_avg)

The Clustering Coefficient, C, determines the connectivity among neighbors of a node-u. Assuming that the node u has k_u neighbors, the Clustering Coefficient of node-u, C_u, can be calculated by the following equation

(4)

where E_u is the number of links between node-u's neighbors and $k_u (k_u - 1)/2$ is the normalization factor, which is the maximum number of links that can possibly exist among the neighbors of the node-u. Further averaging C_u by dividing it by the total number of nodes, N, the aggregated level of clustering within the network can be

5 The length here refers to number of links (i.e., the network distance), rather than the number of meters (i.e., the metric distance). In other words, the measurement unit for the path length is the number of links.

defined by

(5)

As shown in Table 2.1, in the case of the HVV network, the average Clustering Coefficient is 0.038, which is larger than the value of 0.028 in the case of Seoul. In terms of average Clustering Coefficient, there are more "hubs" and "clusters of stations" in HVV network than in the case of Seoul.

As shown in Table 2.1, the average path length, *V*, of the HVV network constitutes 11.396 and its average Clustering Coefficient, *C_avg*, is 0.038 for HVV network.

According to *step iv* in the previous section, the values of average path length and average Clustering Coefficient need to be compared to the corresponding values of an E–R random network with the same number of nodes. The results of such a comparison show that:

- $\lambda = V_{real} / V_{random} = 6.3$, which does not meet the criteria of $\lambda \approx 1$
- $\gamma = C_{real} / C_{random} = 5.28$, which does not meet the criteria of $\gamma > 1$

Based on this comparison, we conclude that the small-world properties cannot be identified in the HVV network. This indicates that the connectivity and robustness of the HVV network are not very resilient to failure or disruptions. If one of the stations stops functioning for some reason, it is possible that there will be a significant adverse effect on the average speed and ease of achieving the destination. In other words, the degree of fault-tolerance of the HVV network is rather low.

4.4 Scale-free properties

1) Node Degree (*k*) and average Node Degree (*k_avg)*

Node Degree, *k*, is the number of links connected to the node. It shows how accessible a node is and is used to characterize the local features of the network (Wang & Fu, 2017). It is both the **connectivity measure,** which is why it is a basic parameter for examining the scale-free properties of the network, and the most straightforward **centrality measure**, which indicates the importance of a node in the network.

In the transportation research, Node Degree is an indicator that express the accessibility level of the stations in the transit network. A station with higher Node Degree means that its connectivity is higher, and it is, therefore, more accessible and important in the network. Because this indicates that the station has a higher chance to increase its capacity for receiving many passenger flows. However, the stations with higher Node Degree are also the most vulnerable to attacks or other disruptions because the failure of these most connected stations can cause a failure of a large portion of the network.

Another indicator, the **average Node Degree, *k_avg***, of a network is a measure that shows the overall level of connectivity of the network because it indicates how connected each node in the network is. The network-wide average Node Degree, *k_avg*, can be expressed as

$$k\_avg = 2 * M / N \tag{6}$$

where *M* refers the total number of links and *N* refers to the total number of nodes.

As shown in Table 2.1, in the case of Hamburg, the HVV network has *N* = 14416 nodes and *M* = 17914 links among stations. The average Node Degree of the HVV network, *k_avg*, is, therefore, 2 **M / N* = 2.485, which indicates that, on average, a station is directly connected to 2.485 other stations. This value is smaller than 3.007 in the PTN in Seoul.

2) Node Degree Distribution, *p(k)*, and Cumulative distribution of Node Degree

The **Node Degree Distribution, *p(k),*** of a network helps us to understand the structure and the topology of a network and to identify whether or not the network is scale-free. A network has the scale-free properties when it is insensitive to the change of scale. In other words, irrespective of whether or not the network size increases, its underlying structure remains unchanged (Wang & Chen, 2003).

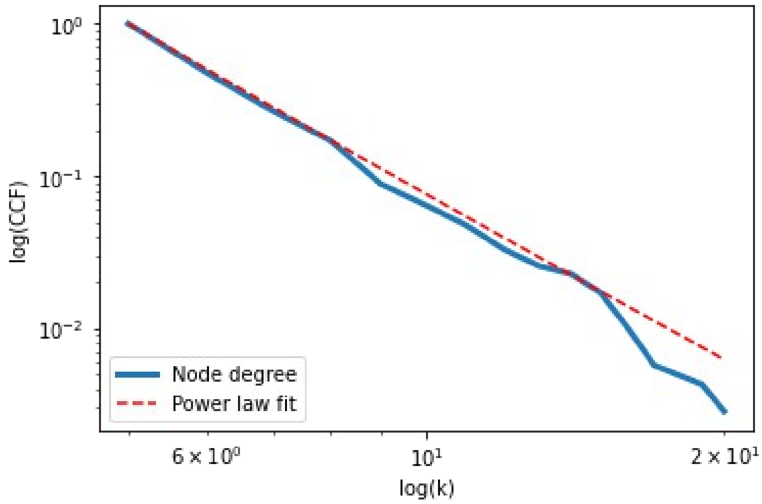

Figure 2.3 Log–log cumulative distribution of station nodes of HVV network.

If a network possesses the scale-free properties, its cumulative Node Degree Distribution should follow the power law distribution. There are two steps in evaluating how well our network data fit the power law distribution. Firstly, by plotting the cumulative Node Degree distribution of a network in a logarithmic scale, like the blue line in Figure 2.3, we can examine how it fits with the line of power law, indicated by the red dash line. If the distribution follows the power law, the blue line would be characterized by a more gradual fall. This means that there is a large number of nodes with only a few links to other nodes, while there are relatively few nodes that are connected to many other nodes. As shown in Figure 2.4 Node Degree Distribution, p(k), of stations in HVV network.Figure 2.4, which presents the Node Degree Distribution of the HVV network, it is shown that the stations in the HVV network have a minimum degree of 1 and a maximum degree of 20. The majority of the stations have 2.5 links connected to them. Again, the nodes with a larger number of links have a stronger influence on the entire network's structure and dynamics. Because the study of scale-free networks helps us to determine whether a network has a large number of important nodes, it is often used for determining how resilient or robust a network is (Wu, Gao, Sun, & Huang, 2004).

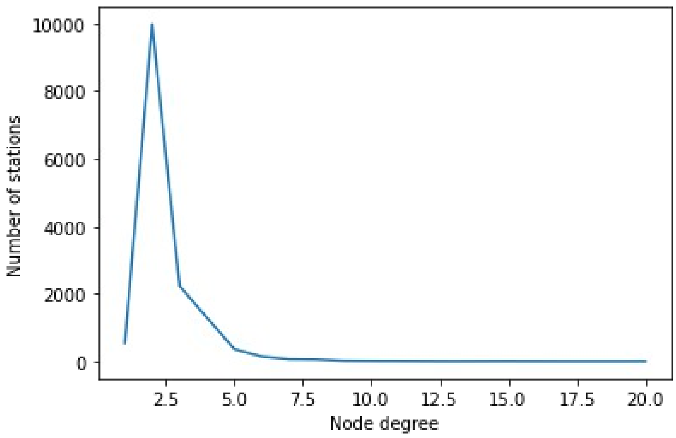

Figure 2.4 Node Degree Distribution, p(k), of stations in HVV network.

Secondly, in order to evaluate whether the power-law distribution itself is a statistically plausible model, we employ the Kolmogorov–Smirnoff tests and compute the *p*-value for the fitted power-law model to test how well the data (i.e. the blue line in Figure 2.3) fit the power law distribution. Since the p-value of our data is 0.365 and the *p-value* > 0.1, we have relatively strong support for a conclusion that the Node Degree Distribution follows the power-law distribution. And the probability of finding a station with *k* connections is proportional to $k^{-1.28}$, which means that there are stations with very high degree in the network. The results indicate that the HVV network is vulnerable to significant failures or disruptions if one of the stations with a large number of links stops operating for some reasons (e.g., due to construction or maintenance works).

5. Conclusion and discussion

In this chapter we have examined the topological properties of Hamburg metropolitan area's public transportation network. First of all, the geometric properties have been measured by the number of nodes, number of links, link-node ratio and the Degree of Connectivity. Secondly, the small-world properties have been measured by the average path length and Clustering Coefficient. Thirdly, the scale-free properties have been measured by average and cumulative distribution of the Node Degree. Based on the analysis of the topological properties, the current study examines the level of robustness and resilience to failure or disruption and whether any unique features required for advancing the efficient operation exist in the network.

Wherever possible, results of theoretical models or previous research have been provided, serving as reference for comparison when investigating the characteristics and determining the levels of connectivity, performance and resilience of the network.

Our examination of the spatial configuration of the HVV network has found that the network has low fault tolerance. First of all, in terms of geometric properties, the HVV network does not facilitate easier movement between stations than the theoretical reference network or the Soul transportation network. Furthermore, the small-world properties cannot be identified in the HVV network. This indicates that the connectivity and robustness of the HVV network have low level of resilience in case of failures or disruptions. In other words, if one of the stations stops functioning for some reason, the speed and ease of getting to a destination can be severely affected. The examination of scale-free properties also reveals that the HVV network is vulnerable to significant failures or disruptions if one of the stations with a large number of links stops operating due to, for example, construction or maintenance work.

In this chapter, we have shown that analysis of the structural composition of the PTN can help us to evaluate the level of fault tolerance based on the topological properties and spatial configuration of the network. The existence or absence of the agglomerated community structure can be used to determine the robustness of the network and to identify structural defects that require to be improved. These results are imperative for the future enhancement of mobility, such as strategically constructing new stations, relocating the existing stations, creating more links.

In future research, the analysis in this chapter can be used as a basis for comparing multiple PTNs in different cities. Further investigation can also consider the effects of other factors, including the population density or the size, structural pattern and shape of the catchment area. One can also compare the small-world properties of the entire network with that of the sub-network in the inner city, where the network density is high. By identifying the clusters of sub-networks within the entire network, it might also be useful to explore various characteristics of these sub-networks and examine how these characteristics affect the connectivity or resilience of the entire network.

Another interesting investigation would be testing the scalability of the network shape and structure. This means examining whether the existing network can scale both in size and connectivity. “A network’s shape dictates its scalability and connectivity. A good shape and structure allow a network to scale with growth and still maintain the focus of its vision and purpose through connectivity. If the topology of the network discourages scalability and connectivity or has an unhealthy obsession of one over the other, then it may not be shaped for movement" (Yang, 2018). For example, Hamburg has recently planned the new route, U5 (https://www.hamburg.de/u5/). Further investigation can test to what extend an additional route improves the connectivity of the entire network. If adding one additional route significantly improves the connectivity of the network, the shape and structure of the HVV network can be considered to have good scalability.