AN APPLICATION OF A NETWORK SCIENCE TOOL FOR EXAMINING AND ANALYSING THE STRUCTURE AND TOPOLOGICAL PROPERTIES OF PUBLIC-TRANSPORT NETWORKS: A CASE STUDY

Dr. Dimitrov S. PhD.[1], Prof. Ceder A. PhD.[1], MEngSt. Mathieson G. [1], BSc. Victor R.[1]
Transportation Research Centre, University of Auckland, New Zealand[1]

sdim492@aucklanduni.ac.nz

Abstract: *More than one century network science (NS) has been extensively used in numerous studies of different research fields, including public transport (PT). This work presents an application of a NS tool enabling to explore complex public-transport networks (PTNs). This tool explores networks for small-world, scale-free and random network characteristics, in a case study to examine and analyze, by using NS concepts, two PT systems - Washington DC's Metro Network (WMN) and Oslo's Metro Network (OMN). The performed analyses focused on the structure and the topological properties of the examined networks. As the networks have longer average path lengths, compared with random networks, and because of not being clustered, these metro networks demonstrated they are not small-world networks. The analyses also show that in contrast to the OMN the WMN has certain characteristics associated with scale-free networks; that is, a small number of highly connected nodes (hubs) and node degree distribution that can be represented by a power-law function. Nevertheless, it still cannot be considered as a pure scale-free network because of its empirical distribution which is better approximated by an exponential rather than a power-law function. The metro WMN cannot be considered a random network because of having hubs. Thus, it is concluded that the examined WMN is an evolving complex network.*

Keywords: NETWORK SCIENCE, PUBLIC-TRANSPORT NETWORK, NETWORK TOPOLOGICAL PROPERTIES, NETWORK ANALYSIS, NODE-DEGREE DISTRIBUTION, AVERAGE SHORTEST PATH LENGTH, NETWORK'S EFFICIENCY

1. *Introduction*

Network science [1] or the science of networks has revealed more opportunities for researchers and scientists who work in different research areas. One of the first attempts for application of network science as a tool is made by Euler [2] in the 18th century in resolving the well-known Koenigsberg seven bridges problem and De Solla Price [3] who pictured the network of scientific papers through linking each published paper to the papers that are directly associated with it. While network science is based on graph theory, it is also comprised of methods and approaches rooted in other research fields. Basically, network science is a result of the convergence of many fields, such as network analysis [4], social network analysis [5,6,7,8] as well as physical [9] and biological sciences [10]. Thus, by providing new tools, mechanisms, and improved methods, NS has enabled exploring, examining, analyzing, modelling and simulating the existing variety of complex networks through evaluating their characteristics and properties in different ways. Recent findings show that NS made it possible to understand complex networks' structure, topological properties and dynamics [11,12]. In this way, by revealing networks' strengths and weaknesses, and thereafter, establishing directions to improving networks' structure, NS has facilitated overcoming problems that have yet to be solved by applying existing methods. Theories related to NS and the findings of complex systems are a good foundation for practical studies of the structure of real-world networks (including PTNs) through an evaluation of topological properties such as average network clustering coefficient (appearing below only as clustering coefficient, CC), average network shortest path length (appearing below as an average path length APL), node degree distribution (NDD) and network efficiency (local and global) in different research fields, including public transport network design and operations [13,14].

2. *Prerequisites and means for solving the problem*

The limitations of the existing approaches and tools for analyzing real-life complex networks, such as PTNs, place a number of new challenges for researchers and scholars in terms of the data required, methods to be applied, and software to be used. Recently Dimitrov and Ceder [15] developed a method enabling to examine and analyse complex public-transport networks (PTNs) incorporating computer programming, large-scale network- and statistical data analysis techniques. The method was tested on a case study.

The objective of this study, which is practically oriented, was to examine and analyse the structure and topological properties of real-world public-transport networks on a case study by using an existing NS tool which uniquely combines computer programming techniques, implementation of known algorithms, network- and data analysis. A network-science perspective was applied, while seeking to reveal network strengths and weaknesses.

In order to achieve the objective, the following essential tasks have been completed:

(i) Collecting, processing and analyzing data on a comparative basis of a case study;

(ii) Examining the structure of the observed PTNs and evaluating their topological properties and characteristics by performing network- and statistical data analysis using the NS tool;

(iii) Drawing reasonable conclusions and outlining the possibilities for a further research work for improving public-transport networks' operation.

A flow-chart of the tool applied for processing the general transit feed specification (GTFS) [16] data used, the calculations made, further data processing as well as all types of analyses performed, is shown in Fig. 1.

3. *Solution of the examined problem*

3.2. Data collection and data processing

The objects of examination in this case study were the subway (metro) networks in Washington DC (USA) and Oslo (Norway). The first of the observed public-transport metro networks in Washington DC, WMN, (Figures 2a and 2b) is composed of 91 metro stations serviced by six lines traveling between the terminal stations. The second of the observed metro networks in Oslo, OMN, also known as T-bane or Oslo Tunnelbane (Fig. 3) is composed of 101 metro stations serviced by five lines traveling between the terminal stations. It is interesting to mention that only 17 out of 101 of the stations in the Oslo's metro network are underground or indoors.

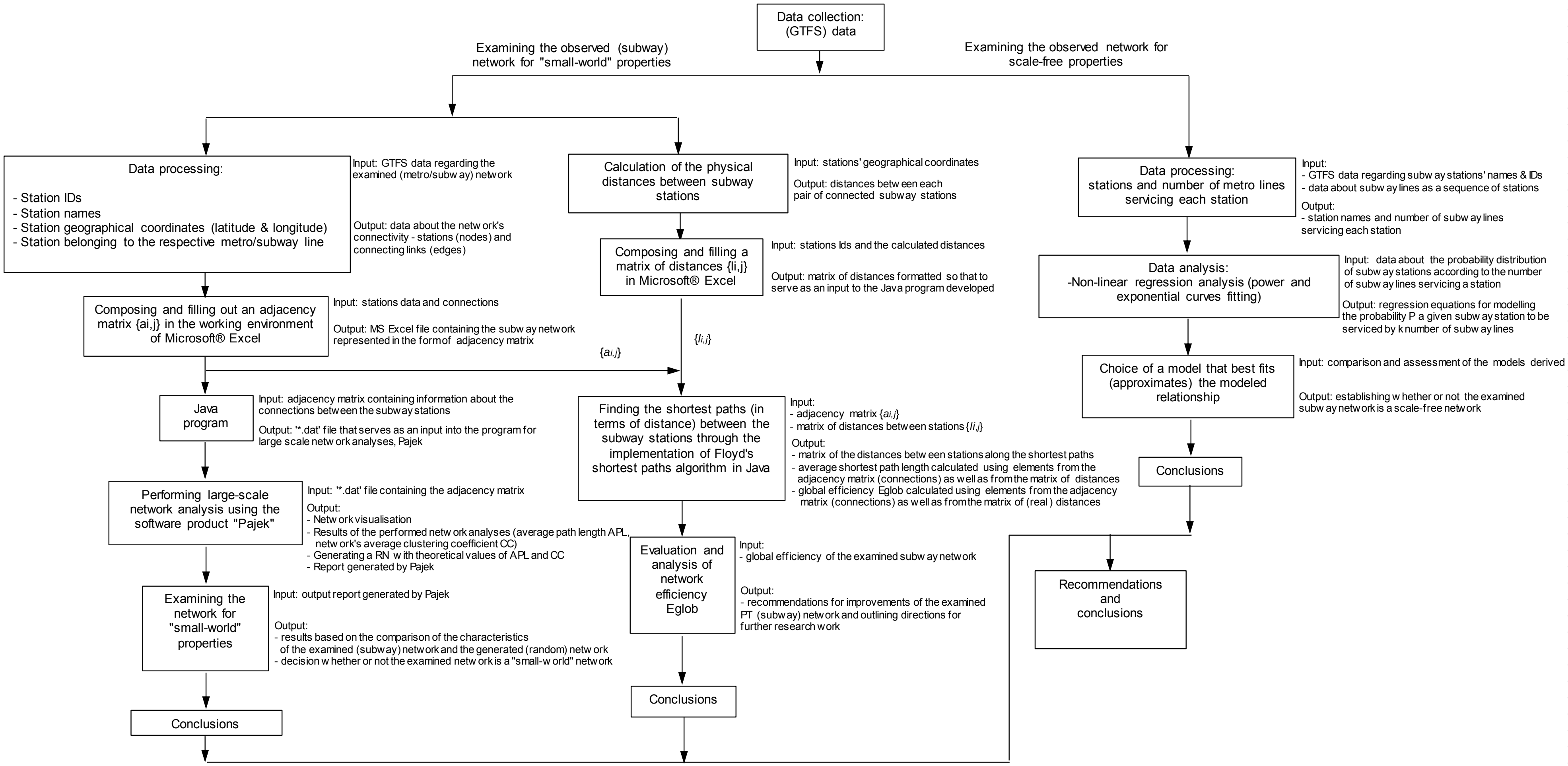

Fig. 1. *Flowchart representing the sequence of steps outlining the logic of the proposed tool for examinining and analysing PT (subway/metro) networks*

3.3. *Examining the subway network for small-world properties*

Because of differences in data formats and descriptions used by the public transport operators within the GTFS feeds [17], in contrast to Dimitrov and Ceder (2016) who extracted and processed the required GTFS data by means of a Java program, in this work the data feeds in the form of text ('*.txt') files. The GTFS data [18,19], provided by the Washington Metropolitan Area Transit Authority as well as by Ruter – the management company that plans, coordinates, orders and markets public transport in Oslo, was directly imported and further processed within the Microsoft Excel® working environment. As a result, useful information regarding the subway stations has been extracted, such as station names, codes (an identifying number), geographical coordinates (latitude and longitude) and the sequence of stations along each of the metro lines in the examined networks. Then, the data about the connections between each pair of stations have been further used to assign values in the Excel adjacency matrix $\{a_{i,j}\}$ showing the connectivity between each two nodes within the network, i.e., between each two subway stations. For example, if station i is directly connected by link (railway section) to station j, then the matrix element $a_{i,j} = 1$, otherwise $a_{i,j} = 0$. As a result, the subway network examined is represented as an undirected, unweighted graph within which each node represents a subway station (in this network the number of stations in both directions is the same and these stations are located opposite one another) and the edges (having length 1) represent the connection between two stations. The presence of a link between two stations means that there is at least one subway line servicing these stations.

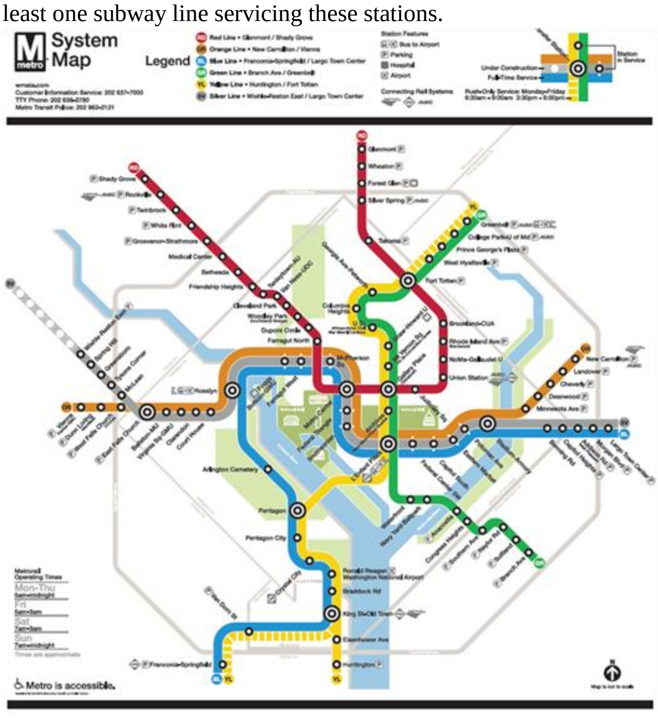

Fig. 2a. *Washington DC Metro system map, adapted from Washington Metropolitan Area Transit Authority [20]*

3.3.1. *Evaluation of the average network's shortest path length and the average network's clustering coefficient*

The information extracted from GTFS files and saved in a MS Excel file is thereafter imported into a developed Java software program automating and accelerating the processing of the data. Within the Java program, the imported adjacency matrix data is then processed, transformed, and as a result exported and saved into a ('*.dat') file that served as an input into the software program for large scale network analyses Pajek [23,7,8] which was used to visualize the network and perform further analyses. By using Pajek, topological properties of the examined network, such as average path length (APL), clustering coefficient (CC) and average network's node degree (ANND), have been computed. Alternatively, apart from Pajek, the average shortest path length for undirected graphs was calculated within the developed Java software program by using the following formulae [24]:

$$(1) \quad L = \frac{2}{N\cdot(N-1)} \cdot \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{i,j} \ ,$$

where:

L – *APL*;

N – number of nodes (stations) in the network;

$d_{i,j}$ – the shortest path (in terms of number of sections) between any two pairs of nodes i and j, and $d_{i,j} = d_{j,i}$.

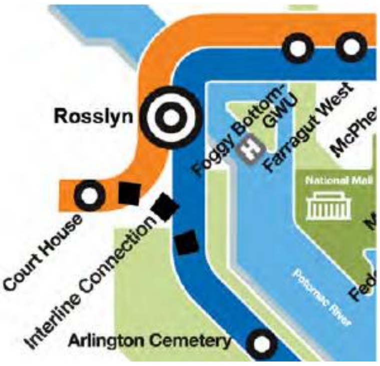

Fig. 2b. *Washington DC Metro system map, Planned improvements, adapted from Momentum [21]*

Fig. 3. *Oslo Metro system map [22]*

Formula (1) was incorporated within the Java program not only to calculate the average network shortest path length as a number of consecutive route sections, but also in kilometers by substituting in $d_{i,j}$ the distances between stations along the shortest paths computed through the Java implementation [25] of Floyd's shortest path algorithm [26]. To verify the output results of the software program developed, the value of the APL (in number of sections) calculated with the Java program was compared with the APL computed by using Pajek.

In order to discover whether the examined public-transport metro networks exhibit small-world characteristics, their properties, as shown in Table 1 and Table 2, have been compared to those of random Bernoulli/Poisson [8] networks that were generated with Pajek using the same number of nodes and average network node degree as an input, as did Watts and Strogatz [27].

Table 1. *Washington DC's metro network compared with a random network generated using the software product Pajek*

Public transport mode	Networks	Network properties				
		Nodes	Edges	ANND	CC	APL
Subway/ Metro	Examined subway/ metro network	91	93	2.044	0.0000	11.51
	Generated random (Bernoulli/Poisson) network		95	2.088	0.0156	5.06

Table 2. *Oslo's metro network compared with a random network generated using the software product Pajek*

Public transport mode	Networks	Network properties				
		Nodes	Edges	ANND	CC	APL
Subway/ Metro	Examined subway/ metro network	101	101	2.000	0.0000	14.68
	Generated random (Bernoulli/Poisson) network		105	2.079	0.0718	5.80

Based on the comparison made between the empirical (calculated) and the theoretical (generated) values of the topological properties, it can be concluded that examined metro networks (Washington DC's and Oslo's), characterized with average path lengths of 11.51 and 14.68 sections, respectively, which exceed more than twice the APL for the generated random networks, and both having clustering coefficient equal to 0, cannot be considered small-world networks. This is because the derived values differ from what is given as a definition for small world graphs [28]: "*A small-world graph is a large-n, sparsely connected, decentralized graph (n >> k_{max} >> 1) that exhibits a characteristic path length close to that of an equivalent random graph (L ≈ L_{random}), yet with a clustering coefficient much greater (C >> C_{random}).*"

3.3.2. Evaluation of network efficiency

In contrast to Watts and Strogatz [27], examining real life networks by using average path length (as a global property) and clustering coefficient (as a local property), Latora and Marchiori [29] introduced the concept of network efficiency – global and local – in their work, measuring the communication between the nodes within a network. In this way, they examined networks' local and global behavior with local efficiency E_{loc} and global efficiency E_{glob} instead of using CC and APL. In other words, E_{loc} and E_{glob} replace CC and APL. It is known [29] that small-world networks have both high E_{loc} and E_{glob} .

As the efficiency $\varepsilon_{i,j}$ between nodes i and j is inversely proportional to the shortest path $d_{i,j}$ between these two nodes [30]:

$$(2) \quad \varepsilon_{i,j} = \frac{1}{d_{i,j}},$$

the global efficiency E_{glob} was calculated via the following formulae implemented in the Java program developed:

$$(3) \quad E_{glob} = \frac{2}{N\cdot(N-1)}\cdot\sum_{i=1}^{N}\sum_{j=i+1}^{N}\varepsilon_{i,j} = \frac{2}{N\cdot(N-1)}\cdot\sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{1}{d_{i,j}},$$

where E_{glob} takes values within the range $0 \le E_{glob} \le 1$ [29], i.e., $E_{glob} \epsilon [0;1]$. Ideally, the global efficiency, considered [30] as the efficiency of the whole network, would have the value of $E_{glob} = 1$, which could be valid in complete graphs in which there is an edge between each pair of nodes connecting them.

The results showing that the global efficiencies for both the metro networks having values $E_{glob} = 0.14$ (Washington DC) and $E_{glob} = 0.12$ (Oslo) means that these networks have only 14% and 12% of the efficiency of the ideal one ($E_{glob} = 1$), the latter having a direct line from each station to the others. The obtained low values are an indication of poor global efficiency of the networks, lacking direct connections between their stations. The low global efficiency values only show that possible network improvements could be achieved in network's topology (and not by the level of service it provides) through the construction of new links (sections), as long as it is proven that the cost incurred for this would be justified.

Since the network analyses, performed with the software program Pajek for the metro networks studied, established that they are not clustered at all (*CC = 0*), the alternative topological property of the clustering coefficient (which is the local network's efficiency E_{loc}) was not evaluated. Again, the value 0 of CC means that there are no metro stations within the networks having neighbouring stations connected to one another. Therefore, it can be concluded that the networks examined are not fault tolerant in the context that each disruption (disconnected stations) would significantly affect the transport service as there are no alternative connections between the stations. This leads to the conclusion that the higher the number of connections in the networks, the greater the fault tolerance of the networks. This can be achieved by increasing connectivity between stations in the network through constructing more sections linking important subway (transfer) stations – hubs.

The Washington Metropolitan Area Transit Authority's (WMATA) strategic plan [21] for 2013-2025 discusses constructing a rail track (section) that would serve as a connection between the Blue and Orange/Silver metro lines (see the Blue line's Arlington Cemetery station and the Silver/Orange Court House station as shown in Fig. 2b), as one of two potential alternatives. This would be a new edge within the existing subway network enabling passengers to make one seat (non-transfer) trips between Dulles Airport, Tysons Corner, Ballston, The Pentagon, National Airport, and the Alexandria metro stations, thus avoiding the need to travel through the core of the network. Both alternatives would result in an increase of train frequency with five more trains per hour, thus leading to an increase in the vehicle carrying capacity provided for up to 4000 passengers per direction per hour. People from WMATA expect the above measures to reduce the average passenger waiting time by an average of three minutes for around 16000 trips.

To explore how the above constructive changes planned for the subway network would affect the APL, CC, and the network's (global) efficiency, the newly created link was added as an edge into the adjacency matrix $\{a_{i,j}\}$, then the approach described in sections 3.3.1 and 3.3.2 was applied. As a result, the following values of the evaluated topological properties of the network examined were identified: APL = 11.32 sections in the proposed network as opposed to 11.51 for the existing network, and clustering coefficients, 0.0252 as opposed to 0, respectively. The network efficiency remains unchanged (0.14). Even after the expected changes, the network is still not a Small-world network (SWN) as it has a clustering coefficient (CC_{ESP} = *0.0252*) commensurate to that of the Random network (RN) (CC_{RN} = *0.0156*) and still quite long APL (11.32 sections) compared to that value for RN (5.06 sections). Thus, as expected, the structural change above (only one new link added) does not globally affect network connectivity. The positive effect on the network can be felt locally as the increased value of CC from 0 to 0.0252 means that the network would become more faults tolerant. This means that in an event of a breakdown (a disconnected link in the triangle Arlington Cemetery, Court House and Rosslyn), the network would still be functioning because of the presence of an alternative (a newly constructed) connection.

3.4. Examining the subway network for scale-free properties.

3.4.1. Data analysis

Based on the GTFS data processed regarding the metro stations and the metro lines in the two networks (see Figures 4 and 5) it turned out that the majority of the stations (76 stations) in the first (Washington DC's) network are serviced by only one metro line (48 stations) or two metro lines (28 stations). There are also 13 more stations serviced by three lines. On the other hand, there are only two stations serviced by four and five subway lines, respectively, that can be considered hubs. Thus, the network topology of the metro network examined resembles the topology of scale-free networks in which most of the nodes have a small number of links, with only a small number of nodes (called hubs) having multiple connections [31], which is in contrast with the second examined metro network in Oslo. Similar to the Washington's metro network, the majority of the stations in the Oslo's metro network (89 stations) are also serviced by one metro line (61 stations) or two lines (28 stations). However, Oslo's network does not have small number highly connected stations (hubs). Instead, it is characterized with 3 stations serviced by 3 lines, another 3 stations serviced by 4 lines, and 6 more stations serviced by 6 metro lines. For the sake of clarity, the 5th line of the Oslo's metro network (in green) can be considered as composed of two separated lines as it runs twice through nine stations and follows two separate routes. That is, when a vehicle runs in the direction from Vestli to Sognsvann (see Fig. 3), it can either pass through the station "Ullevål stadion" and continues to station Nydalen (the first route), or continues in the direction station Berg.

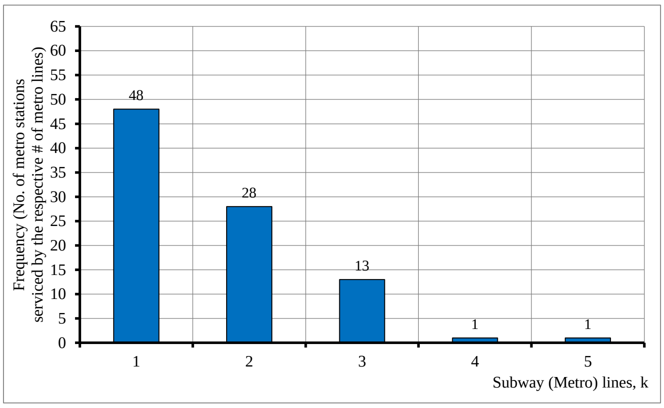

Fig. 4. *Distribution of the Washington DC metro stations according to the number of lines k servicing a station*

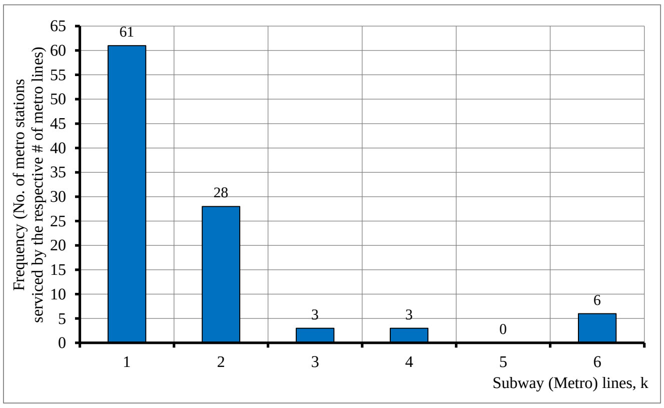

Fig. 5. *Distribution of the Oslo's metro stations according to the number of lines k servicing a station*

With the intention of confirming or falsifying the above supposition, in order to estimate the probability *P(k)* with which a randomly selected station is being serviced by *k* lines, two types of functions – a power-law [32,33] and exponential – were approximated through performing a statistical data analysis [34] as shown in Table 3 (for Washington DC's network) and Table 4 (for Oslo's network) below. It is obvious (from Table 3) that the coefficient *A* in the power function has value which is close to 1, i.e.: $A = 0.948 \approx 1$. Therefore, it can be accepted that the formula $P(k) \sim k^{-B}$ satisfactorily represents the established statistical relationship between *P* (the dependent variable) and *k* (the independent variable), i.e., $P(k) \sim k^{-2.61}$ in which the scaling factor *B* falls within the range $2 < B < 3$ [31]. Based on this result, it can be accepted that the network examined exhibits characteristics of scale-free networks. On the other hand, as shown in Table 3, the results of the least squares fitting procedure performed for the empirical values also revealed that the observed nonlinear relationship between *P* and *k* is better fitted by means of an exponential ($R^2 = 0.902$) rather than a power-law function ($R^2 = 0.813$). In contrast to the Washington DC's metro network the low R-square value ($R^2 = 0.69$) in Table 4 clearly shows for the Oslo's network that the probability *P(k)* is not satisfactorily fitted by a power-law function $P(k) \sim k^{-B}$ describing scale-free networks, which means that this network does not exhibit the specific characteristics for scale-free networks. Therefore, it does not appear that the studied metro network of Oslo is a scale-free network.

Table 3. *Derived regression equations with their estimated parameters, metro network Washington DC*

PT network	Method	Function	Type	Parameter	Value	R-square
Subway lines	Least squares fitting	*P(k) = A.k* $^{-B}$	Power	*A*	*0.948*	*0.813*
				B	*2.610*	
		P(k) = A.e $^{-B.K}$	Exponential	*A*	*2.149*	*0.902*
				B	*1.100*	

Table 4. *Derived regression equations with their estimated parameters, metro network in Oslo*

PT network	Method	Function	Type	Parameter	Value	R-square
Subway lines	Least squares fitting	*P(k) = A.k* $^{-B}$	Power	*A*	*0.505*	*0,69*
				B	*1.656*	

Regardless of the fact that for the Washington DC's metro network examined the presence of a small number of highly connected nodes (considered hubs) was established, and that a power-law function satisfactorily describes the node degree distribution of that network, it still cannot be considered a purely scale-free network. The fact that the examined network is better described by an exponential rather than a power-law function supports this statement. As Barabási and Albert [35] wrote, "*a common feature of the ER and WS models is that the probability of finding a highly connected vertex (that is, a large k) decreases exponentially with k; thus, vertices with large connectivity are practically absent*." Accordingly, "*such networks are called the exponential networks*" [36]. It is known [37] that in RNs, each node has approximately the same number of links, which means that the node degree k_i of each node *i* is close to the average network's node degree $\langle k \rangle$, i.e. $k_i = \langle k \rangle$. In the case study, the network was presented as of bus-station network which "*retains the basic topological features of public transportation networks*" [38]. In this network the average node degree having a value of 1.67 lines servicing each node is on the average 2.4 and 3.0 times less than the degree of the two highly connected nodes in the network (serviced by four and five subway lines, respectively). Therefore, due to the presence of hubs, the network also cannot be considered as an entirely random network. According to Albert and Barabási [39] who wrote that "*if all processes shaping the topology of a certain network are properly incorporated, the resulting P(k) often has a rather complex form, described by a combination of power laws and exponentials*", it can be expected that the network examined should have a complex form, which shows that the PTN examined can be considered a complex network. The latter is not valid for the metro network in Oslo as the *P(k)* is neither good fitted by an exponential nor by a power-law function. In accordance with Albert and Barabási who wrote that the "*evolving networks can develop both power-law and exponential degree distributions*" [39] as the network examined can satisfactorily be described by a power-law as well as by exponential functions, it can be expected for it to be an evolving network.

4. Conclusion

The results of the examination of two real-life metro networks and the analyses performed as part of this case study, by using the proposed network science tool, have led to a few general conclusions that can be summarized as follows:

The network analysis performed in exploring the topological properties of the Washington DC's and the Oslo's metro networks showed that when represented in an L-space network topology [40], the networks examined do not exhibit small-world properties, and hence, they are not small-world networks.

The examination of the Washington DC's metro network and its analyses also showed that the network is neither a scale-free nor random network; this is based on the consideration of network's node degree distribution, the number of the metro lines servicing each station and representing the network as a bus station network. In contrast to the Oslo's metro network, the metro network in Washington appears to be a complex network.

The analysis considering the networks' global efficiency, performed by using network science concepts and findings, showed that both the metro networks examined appear not to be faults tolerant. In case of a railway section's breakdown, this would significantly affect the quality of the transport service provided, because of the lack of alternative connections linking the disconnected metro stations. These results illustrate connectivity weaknesses, thus indicating that there is a room for improving the connectivity between the stations within the networks.

This case study does not only examine real-world metro networks with their topological properties, but also performs network's reconstruction analysis for the Washington DC's network. That is, it considers a hypothetical construction of a section connecting two stations within the Washington DC's network to show that a new link would not sensibly improve the global network's efficiency; this is explicated by measuring the presence of direct connections between the metro stations of the network.

It turned out that both the metro networks examined have low valued global network's efficiencies. This global efficiency does not take into account the routes of the operational metro lines and the passenger trips. Thus, the above statement does not necessarily mean that these networks are inefficient in serving passengers and covering the existing PT demand.

References

[1] Börner, K., Sanyal, S., & Vespignani, A. (2007). Network science. *Annual review of information science and technology*, *41*(1), 537-607.

[2] Euler, L. (1953). Leonhard Euler and the Königsberg bridges. *Scientific American*, *189*(1), 66-70.

[3] De Solla Price, D. J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510-515.

[4] Boissevain, J. (1979). Network analysis: a reappraisal. *Current anthropology*, *20*(2), 392-394.

[5] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol.8). Cambridge university press.

[6] Wasserman, S. (1980). Analyzing social networks as stochastic processes. *Journal of the American statistical association*, *75*(370), 280-294.

[7] De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek,* Cambridge University Press, New York.

[8] De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek: Revised and Expanded Second Edition.* Cambridge University Press.

[9] Barabási, A. L. (2013). Network science. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(1987), http://rsta.royalsocietypublishing.org/content/roypta/371/1987/20120375.full.pdf.

[10] Lewis, T. G., (2009). *Network science,* John Wiley & Sons.

[11] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4), 175-308.

[12] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.

[13] Ceder, A. (2016). *Public transit planning and operation: Modeling, practice and behavior*. 2nd ed. Boca Raton, FL: CRC Press.

[14] Ceder, A. (2007). *Public transit planning and operation: Theory, modeling and practice.* London: Elsevier, Butterworth-Heinemann.

[15] Dimitrov S. D., Ceder, A., (2016). Method of examining the structure and topological properties of public-transport networks, *Physica A: Statistical Mechanics and its Applications*, *451*(1), 373–387.

[16] https://developers.google.com/transit/gtfs/

[17] https://developers.google.com/transit/gtfs/#overview-of-a-gtfs-feed

[18] http://transitfeeds.com/p/wmata/75

[19] http://transitfeeds.com/p/ruter/240

[20] http://wmata.com/pdfs/pocket_guides/english.pdf

[21] https://www.wmata.com/momentum/momentum-full.pdf

[22] https://ruter.no/globalassets/rutetabeller/oslo/t-banen-linjekart-03042016.pdf

[23] http://Pajek.imfm.si/doku.php?id=pajek

[24] Prettejohn, B. J., Berryman, M. J., & McDonnell, M. D. (2011). Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists, *Frontiers in computational neuroscience*, *5*,11.

[25] Lau, H. T. (2006). *A Java library of graph algorithms and optimization*. CRC Press.

[26] Floyd, R. W. (1962). Algorithm 97: shortest path. *Communications of the ACM*, *5*(6), 345.

[27] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

[28] Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon 1. *American Journal of sociology*, *105*(2), 493-527.

[29] Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters*, *87*(19), http://link.aps.org/doi/10.1103/PhysRevLett.87.198701.

[30] Latora, V., & Marchiori, M. (2002). Is the Boston subway a small-world network? *Physica A: Statistical Mechanics and its Applications*, *314*(1), 109-113.

[31] Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 50-59.

[32] Adamic, L. A. (2000). Zipf, Power-laws, and Pareto - a ranking tutorial. *Xerox Palo Alto Research Center, Palo Alto, CA.* http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html. Accessed August 2015.

[33] Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, *46*(5), 323-351.

[34] Kirkup, L. (2002). *Data analysis with Excel®: an introduction for physical scientists*. Cambridge University Press.

[35] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.

[36] Chen, Q., & Shi, D. (2004). The modeling of scale-free networks. *Physica A: Statistical Mechanics and its Applications*, *335*(1), 240-248.

[37] Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378-382.

[38] Lu, H., & Shi, Y. (2007). Complexity of public transport networks. *Tsinghua Science & Technology*, *12*(2), 204-213.

[39] Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

[40] Sienkiewicz, J., & Hołyst, J. A. (2005). Statistical analysis of 22 public transport networks in Poland. *Physical Review E*, *72*(4), 046127.