Statistical analysis of public transport networks of 27 cities in the different network representation

Maryam Kiashemshaki

September 2, 2020

Abstract

In this study, we used the network science approach to analyze the urban public transportation systems of 27 different cities across the world. We performed a comprehensive network analysis with the main goal to identify similarities and differences in the transportation networks of these cities in the different network representations. In particular, we calculated some network measures to examine the network characteristic and categorize the public transport networks with known networks such as small-world and scale-free in three different network representations. By comparing the results, we get a picture of the differences in public transport network, which may due to historical and geographical factors.

1 INTRODUCTION

Recently, network science became a useful tool to analyze and visualize huge amounts of data, and effective approach to understand the underlying structure of real-world networks.

Public transportation system is an important factor in urban sustainability. Most municipalities are focusing to provide more incentives for citizens to leave the car at home and use public transport instead. In order to achieve this goal, public transport should be reliable and efficient. A comprehensive understanding of the structure and properties of the public transport network (PTN) will benefit urban planning, traffic operation, and infectious disease control. Since public transports have a great number of stops and routes in different transportation means, network science becomes a useful approach to study PTN properties, efficiency, and robustness [1, 2, 3, 4, 5, 6].

For applying complex network analysis into PTNs, there are several questions in terms of what kind of graph representations to use, what kind of network measures to investigate, whether randomness dominates public transport networks.

The topological properties of such networks have been subject to comprehensive attention in recent literature [6, 7, 8, 9, 10, 11]. Most of all these studies have been used a quite simple, intuitive representation of networks by modeling stops as nodes and routs as edges. However, there are many other ways to represent PTNs. We will explain them in detail in section 3.

While a PTN of cities is closely related to geographical, historical, and social cultures, some studies showed PTNs of different cities have shared common statistical properties [6, 7, 8, 9, 10, 11]. By using random graph models one can capture common features of PTNs and also find out whether randomness dominates PTNs. To reveal network characteristics, one should compare real-world networks with two common class of random networks: (1) small-world, meaning that in such network the average shortest path length is small compared to the size of the network and also the clustering coefficient is high with comparison to random graph values, (2) scale-freeness are characterized by the presence of power-law behavior in the degree distribution (most of the nodes have a small number of links, while a few have an extremely large number of connections).

For this purpose, we construct PTN of 27 different cities using three different network representations and analyze statistical properties of PTN including degree distribution, average shortest path length, average clustering coefficient, average degree, and assortativity.

This report is organized as follows: In Section 2, we go through the background studies. In section 3 we define different PTN representations. In section 4, we introduce the database and set up the network using three different network representations. In Section 5 various statistical properties are studied. Section 6 gives conclusion.

2 BACKGROUND STUDIES

Several studies have analyzed specific means of public transport, which are the sub-networks of PTNs [9, 12, 13, 14, 15, 16]. Small-world characteristics on Boston subway networks in Lspace were studied in [12]. Structural properties of the Indian railway in Pspace were studied by [9] . Topological and geographic structure of the national road networks of the United States, England, and Denmark in Cspace were studied by [16]. The Bus network was investigated by [15] in Lspace and Pspace.

However, each particular network (the bus network, or the subway network, etc) is not a closed system [1] and should be considered as a subgraph of a wider PTN, to describe the public transport system of a city, the whole system should be analyzed, without restriction to specific means of transport. In this regards, Latora and Marchiori in [1] showed by extending the subway network to the network of subway and bus, the local efficiency of the network will increase. Analysing PTNs in the general type have been done in a small number of studies. Sienkiewicz and Holyst in [7] analyzed statistical properties of PTN (bus and tram) in 22 Polish cities in Lspace and Pspace. They described degree distribution based on network representation, which may follow a power law or an exponential function. Ferber and his coworkers in [6] proposed a growth model of PTNs and analyzed the statistical properties of PTN in fourteen cities in the different network representations.

However,one should notice these studies on PTN usually have dealt with a smaller number of cities or a smaller variety of nodes range. Sienkiewicz and Holyst in [7] studied PTN properties on 22 Polish cities, while the number of nodes varies from 152 to 2811. Ferber and his coworkers in [6] studied 14 different cities with nodes range from 1494 to 44629. In this study, we analyzed PTNs in 27 different cities with nodes range from 549 in Kuopio to 24063 in Sydney in three network representations (Lspace, Pspace, and Cspace), and then we did comparative studies to find out network characteristics. Table 1 shows detail of the cities. As we can see in all cities bus is the main way of the public transportation system.

Table 1: Cities analyzed in this study

City	**Number of stations**	**Number of routes**	**Transportation type**					
			Bus	**Tram**	**Subway**	**Rail**	**Ferry**	**Cable-car**
Adelaide	7548	9234	8950	54	-	230	-	-
Antofagasta	650	963	963	-	-	-	-	-
Athens	6768	7978	7866	50	62	-	-	-
Belfast	1917	2180	2180	-	-	-	-	-
Berlin	4601	6600	5861	358	127	250	4	-
Bordeaux	3435	4026	3798	222	-	-	6	-
Brisbane	9645	11681	11279	-	367	-	35	-
Canberra	2764	3206	3204	-	-	2	-	-
Detroit	5683	5946	5964	-	-	-	-	-
Dublin	4571	5537	5274	106	-	157	-	-
Grenoble	1547	1679	1511168	-	-	-	-	-
Helsinki	6986	9022	8598	301	20	101	2	-
Kuopio	549	699	699	-	-	-	-	-
Lisbon	7073	8817	8691	-	52	68	6	-
Luxembourg	1367	1903	1851	-	-	52	-	-
Melbourne	19493	21434	19406	1737	-	291	-	-
Nantes	2353	2743	2581	162	-	-	-	-
Palermo	2176	2559	2559	-	-	-	-	-
Paris	11950	13726	12522	345	370	489	-	-
Prague	5147	6714	5894	695	119	-	2	4
Rennes	1407	1670	1642	-	28	-	-	-
Rome	7869	10068	9658	133	142	135	-	-
Sydney	24063	28695	27545	44	-	960	46	-
Toulouse	3329	3734	3643	53	38	-	-	-
Turku	1850	2335	2335	-	-	-	-	-
Venice	1874	2647	2391	37	-	-	219	-
Winnipeg	5079	5846	5846	-	-	-	-	-

3 NETWORK REPRESENTATIONS

To analyze PTN properties, we need to define the network representation. In this study, we go through three most common PTN representations.

In the Lspace (a.k.a space of stations) stops are represented by nodes; two nodes are linked if they are consecutive stops on the route, fig. 1 shows Lspace representation. The average shortest path length in Lspace gives the number of stops one has to pass on average to travel between any two stops; and the degree of a node in this representation gives the number of stops in the neighborhood.

In the Pspace (a.k.a space of changes) same as Lspace stops are represented by nodes; they are linked if they can be reached without changing means of transport as we can see in fig. 2. Therefore, in this representation, the average shortest path length, tells how many changes one has to do to travel between any two stops, while the degree of a node, determines the number of other stations reachable without changing transportation means.

In Cspace (a.k.a dual graph) each route is represented by a node; each link corresponds to a common stop shared by the routes, fig 3 shows Cspace representation. The average shortest path length brings the number of changes one has to do to pass between any two routes. The degree of a node, tells how many routes are directly accessible from the given one.

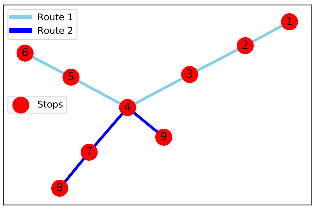

Figure 1: Lspace representation

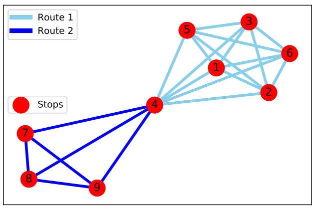

Figure 2: Pspace representation

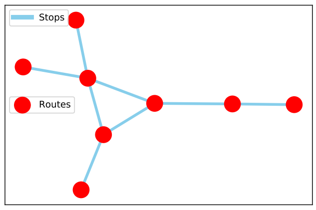

Figure 3: Cspace representation

4 CONSTRUCTION OF THE NETWORK

We analyzed PTN in 27 cities in different continents, these cities are varied regarding the size of the city and population and also the variety of public transport alternatives as shown in table 1. The process of data collection is explained in [17], in this data set number of stops varies from 549 in Kuopio to 24063 in Sydney.

For network construction, since under normal circumstances, passengers can travel from station A to bus station B along a certain route and also travel from station B to station A by using the same route, PTNs are usually treated as undirected graphs. In this study, we created an undirected unweighted graph in the different network representations. Fig.4 is an intuitive example of the different representations of giant connected component (GCC)of Berlin .

5 NETWORK MEASURES

5.1 DEGREE DISTRIBUTION

First we examine the degree distributions in three spaces. Fig.5 shows plots for degree distribution of different cities in different network representations in the log-log plot.

By using curve fitting we can find out the degree distributions can be approximately described by a power-law distribution $p(k) \approx k^{-\gamma}$ where the γ is between 2.63 and 4.14. Almost all cities except Paris have $2 < \gamma < 3$, which characteristics correspond with the preferential attachment model [18]. Braunstein and his colleagues in [19] explained networks with a characteristic exponent greater than 4 are considered close to random graphs. Fig. 6 shows fitted power law for four chosen cities.

Degree distribution in Pspace as shown in fig.5 follows an exponential distribution $P(k) = Ae^{-\alpha k}$ with the exponent $0.01 < \alpha < 0.05$. As Barabási and his coworkers [18] explained the exponential character of the distribution shows nodes are attached generally randomly. Fig. 7 shows fitting exponential distribution on Pspace degree distribution for four chosen cities.

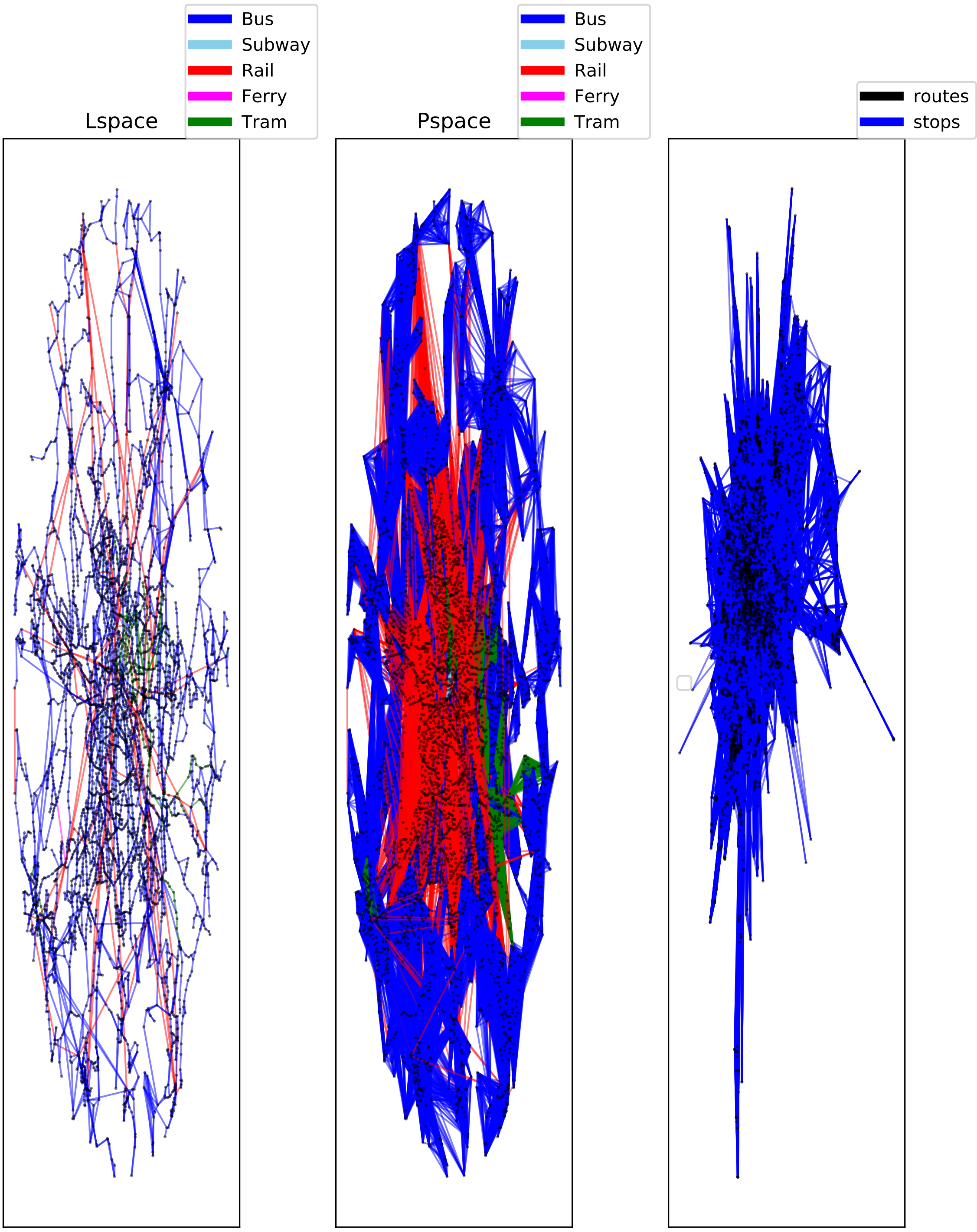

Figure 4: Different network representation of GCC of Berlin

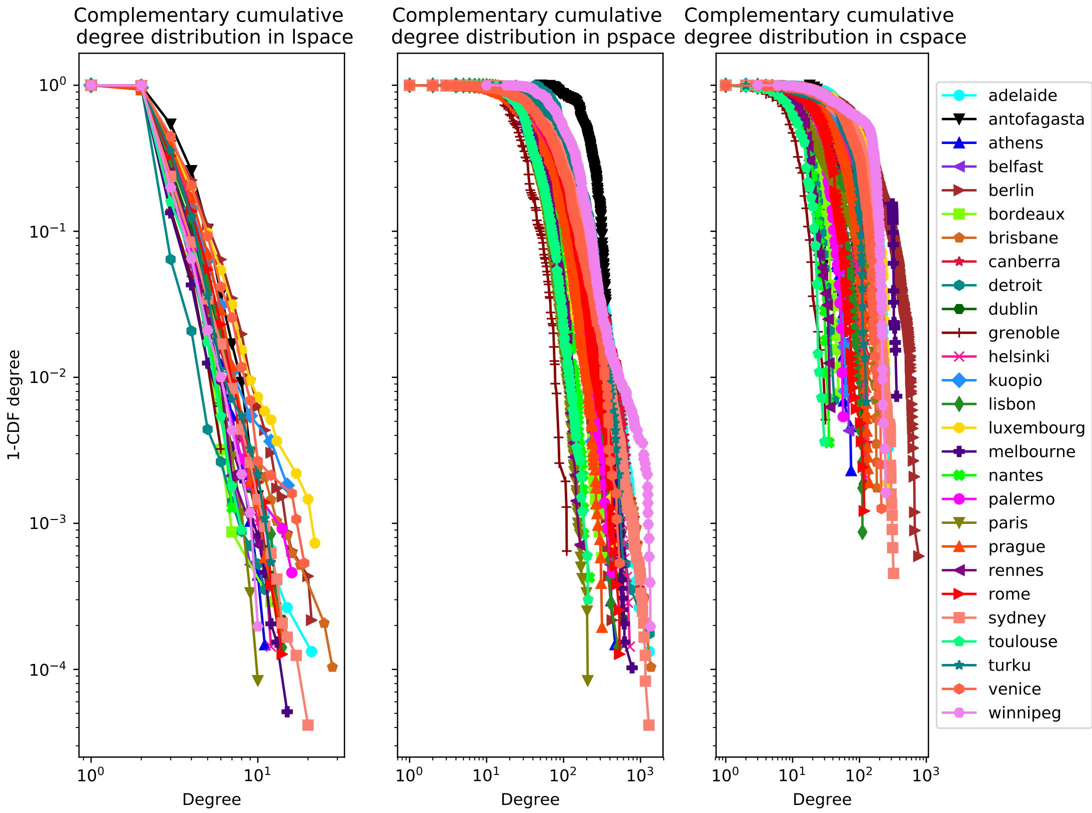

Figure 5: Degree distribution in different network representation

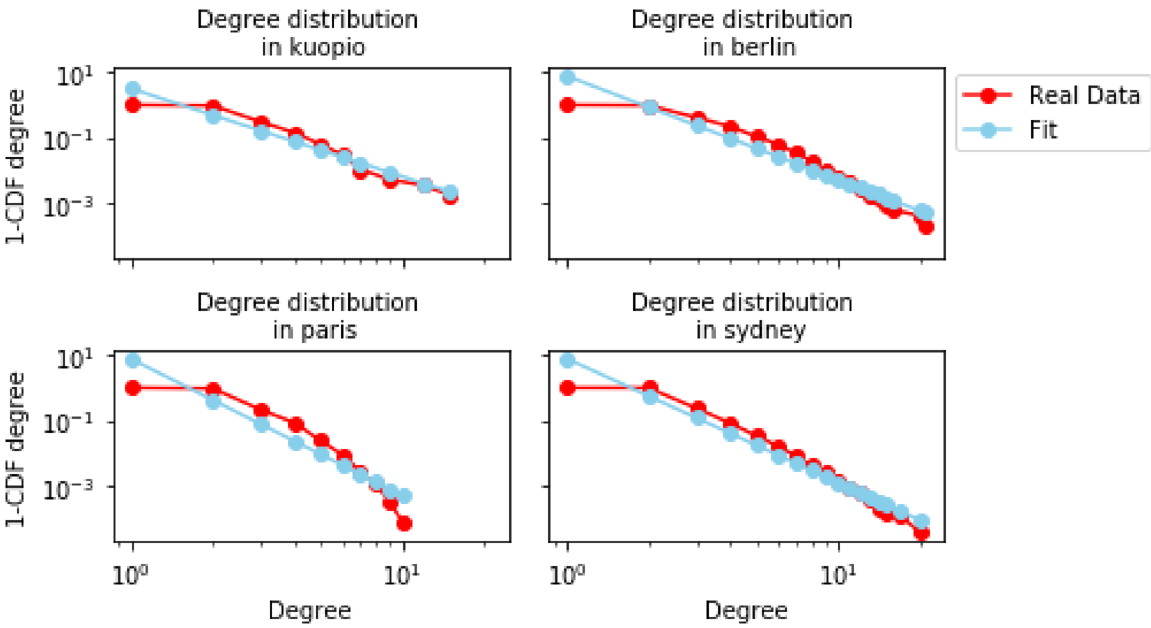

Figure 6: Degree distributions and fitted power-law distribution in the Lspace for four chosen cities.

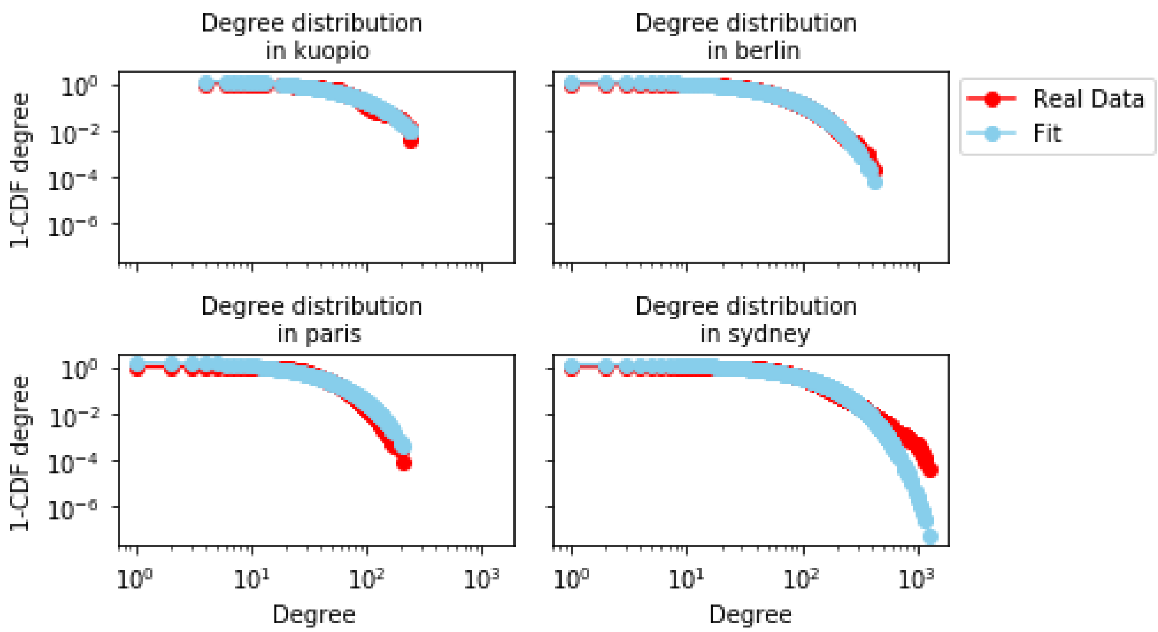

Figure 7: Degree distributions and fitted exponential distribution in the Pspace for four chosen cities.

Degree distribution in Cspace follows an exponential distribution too with the exponent $0.01 < \alpha < 0.14$. Fig. 8 shows fitted degree distribution for four chosen cities.

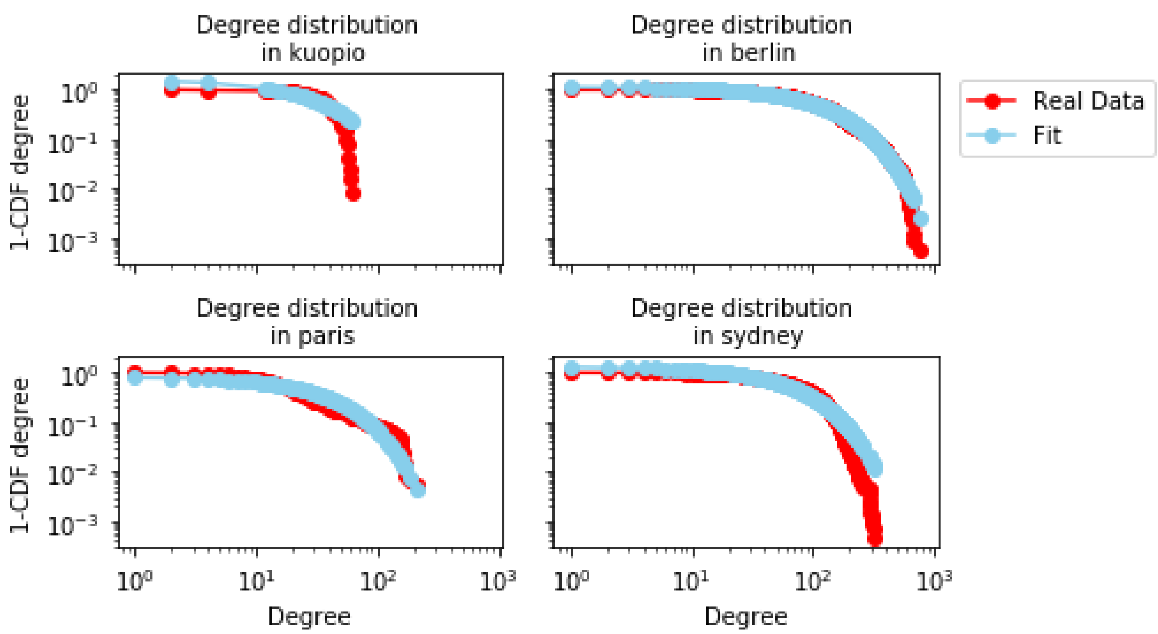

Figure 8: Degree distributions and fitted exponential distribution in the Cspace for four chosen cities.

5.2 AVERAGE DEGREE

Fig.9 shows average degree in different representation. As we can see the average degree in all spaces for cities are not dependent on the network size (number of stops). In Lspace average degree is between 2.09 and 2.96, this means almost in all scale cities the number of directions a passenger can travel from a random station is 2.45. In Pspace we have $28.09 < \langle k \rangle < 199.75$, which means in all cities on average 77.18 stations reachable without changing transportation mean. In Cspace average degree for different cities varies from 9.81 to 143.35, we can say approximately 50.25 routes are directly accessible from the given route.

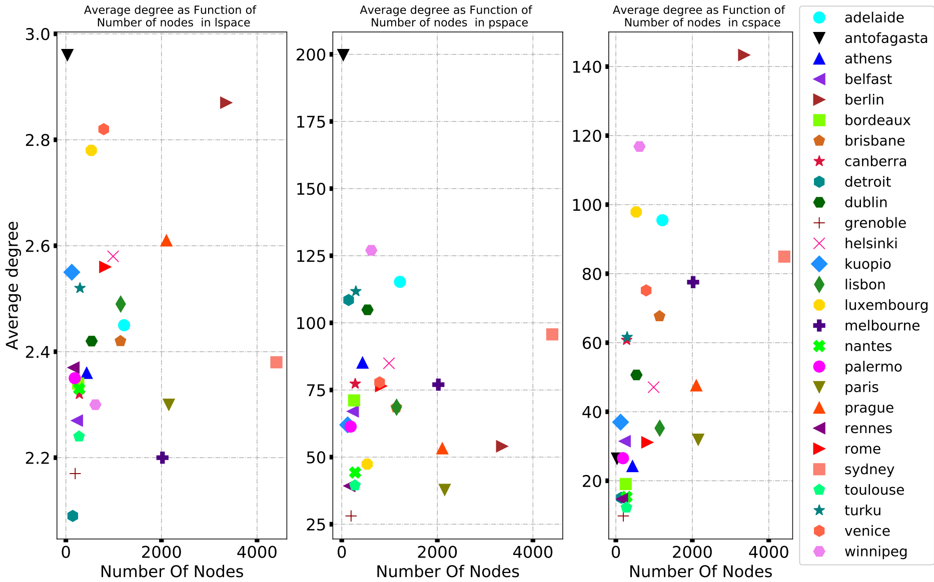

Figure 9: Average degree in different network representation

5.3 DEGREE-DEGREE CORRELATIONS

To analyze degree-degree correlations in PTN we have used the assortativity coefficient r. A large positive value of r means that connected nodes very much tend to connect to similar degree nodes, a value close to 0 means no strong association between connected nodes and a large negative value means that connected nodes tend to possess very different degrees. Fig. 10 shows assortativity coefficient in different representation. Values of the assortativity coefficient are independent from the network sizes in all spaces. These values in the Lspace are always positive, while some cities in Pspace have negative assortativity coefficient. The assortativity coefficients in Cspace have almost the same trend as this values in Pspace.

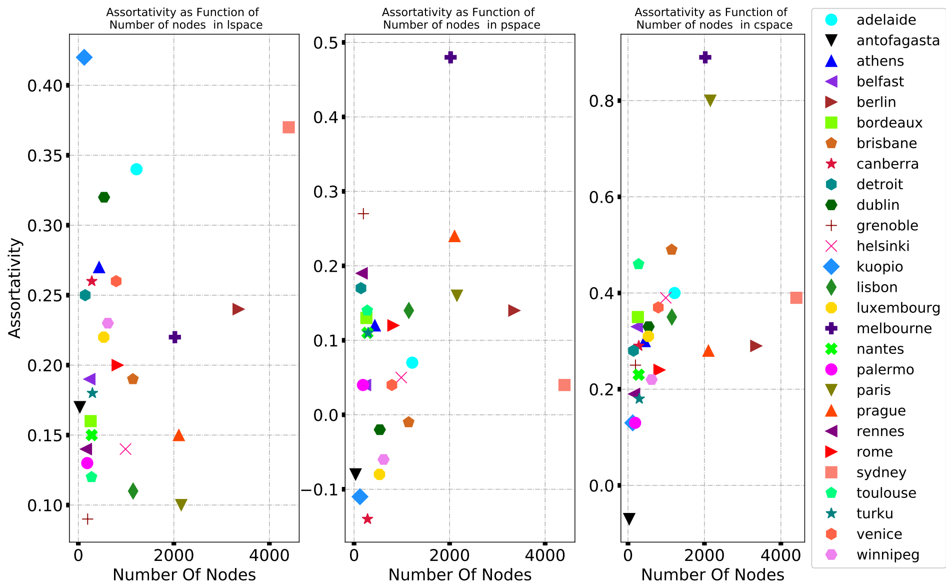

Figure 10: Assortativity coefficient in different network representation

Fig. 11 shows assortativity coefficient of mean value of assortativity coefficient of 100 random graphs generating by configuration model. From the figure we can see the assortativity in graphs generated by configuration model is different from the value from real world data.

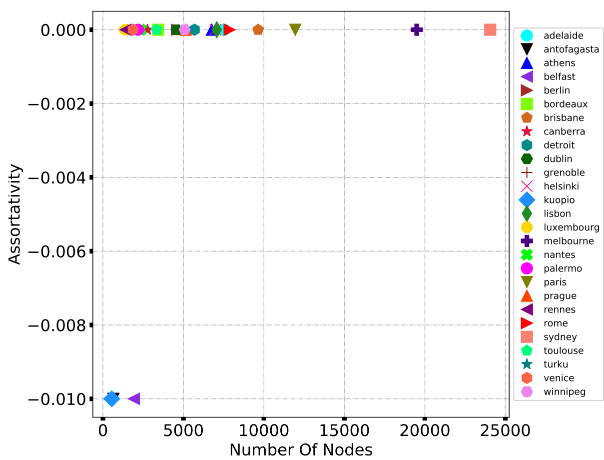

Figure 11: Mean of assortativity coefficient in 100 random graphs

5.4 AVERAGE SHORTEST PATH LENGTH

Fig. 12 shows the average shortest path length in the different representations. Since some cities have more than one component, the average shortest path length was measured in the giant connected component(GCC). We can see in the Lspace average shortest path length, which gives the number of stops one has to pass on average to travel between any two stations is between 10.58 and 75.29. In the Pspace number of changes one has to do to travel between any two stations varies from 1.7 to 5.01. In Cspace the number of changes one has to do to pass between any two routes or average shortest path length in Cspace for different cities is between 1.15 and 4.74.

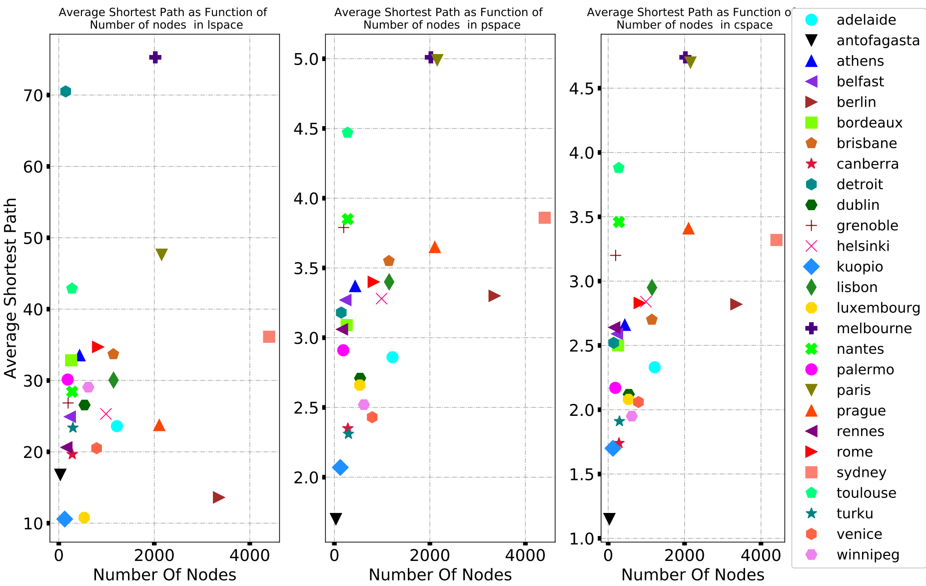

Figure 12: Average shortest path length in different network representation

The average shortest path length from random graphs is different from the real value of the average shortest path length. Fig. 13 shows a comparison between the real value of the average shortest path and distribution of it from 100 graphs construction by configuration model for four chosen cities.

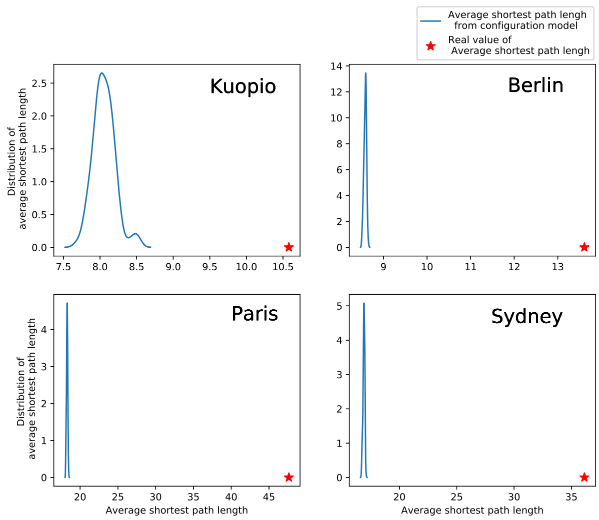

Figure 13: Average shortest path distribution from 100 random graphs for 4 chosen city

The average shortest path for a two-dimensional regular lattice with N nodes is $\langle l \rangle \approx N^{\frac{1}{2}}$, this value for small-world networks is $\langle l \rangle \approx \log N$. In table 2 we can see the value of average shortest path length in Lspace and also $N^{\frac{1}{2}}$ and $\log N$ for cities.

Table 2: Comparison of average shortest path with small-world network and $2d$ lattice

City	Number of nodes	Average shortest path in Lspace	$N^{\frac{1}{2}}$	$\log N$
Adelaide	7548	23.6	86.88	8.93
Antofagasta	650	16.79	25.5	6.48
Athens	6768	33.53	82.27	8.82
Belfast	1917	24.92	43.78	7.56
Berlin	4601	13.61	67.83	8.43
Bordeaux	3435	32.83	58.61	8.14
Brisbane	9645	33.69	98.21	9.17
Canberra	2764	19.65	52.57	7.92
Detroit	5683	70.51	75.39	8.65
Dublin	4571	26.56	67.61	8.43
Grenoble	1547	26.85	39.33	7.34
Helsinki	6986	25.31	83.58	8.85
Kuopio	549	10.58	23.43	6.31
Lisbon	7073	30.05	84.10	8.86
Luxembourg	1367	10.79	36.97	7.22
Melbourne	19493	75.29	139.62	9.88
Nantes	2353	28.42	48.51	7.76
Palermo	2176	13.13	46.65	7.69
Paris	11950	47.63	109.32	9.39
Prague	5147	23.75	71.74	8.55
Rennes	1407	20.62	37.51	7.25
Rome	7869	34.70	88.71	8.97
Sydney	24063	36.13	155.12	10.09
Toulouse	3329	42.88	57.7	8.11
Turku	1850	23.37	43.01	7.52
Venice	1874	20.51	43.29	7.54
Winnipeg	5079	29.06	71.27	8.53

5.5 AVERAGE CLUSTERING COEFFICIENT

Fig.14 shows the average clustering coefficient in the different representations. The average clustering coefficient in Lspace varies in the range $[0, 0.12]$. This value in Pspace is between 0.67 and 0.95, and in Cspace is from 0.51 to 0.87.

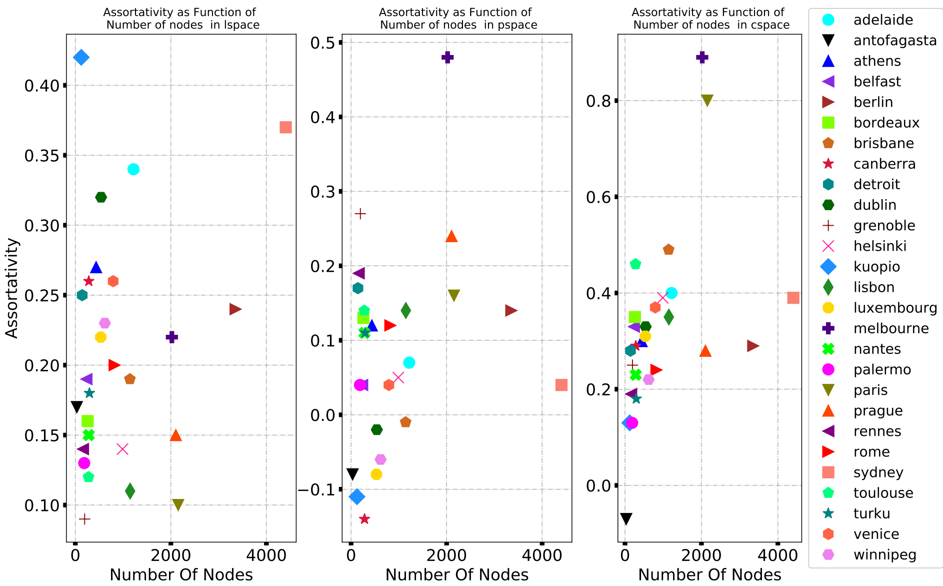

Figure 14: Average clustering coefficient in different network representation

The average clustering coefficients in all spaces are larger than $C_{ER} \approx 1/N \approx 10^{-3} - 10^{-2}$ corresponding to a random ER graph. Fig 15 shows mean value of the average clustering coefficient of 100 random graphs generating by configuration model.

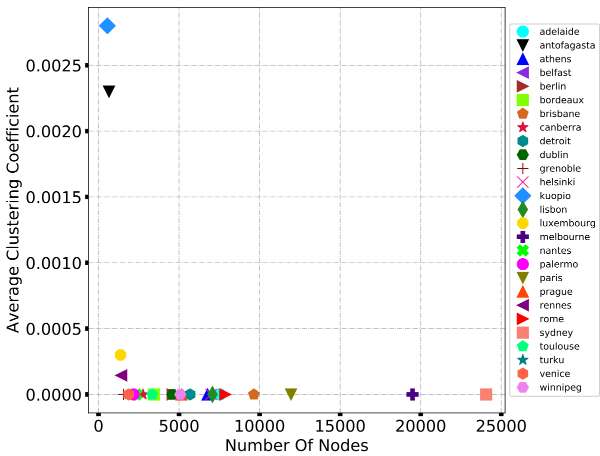

Figure 15: Mean of average Clustering coefficient in 100 random graphs

6 CONCLUSION

In this study, we analyzed PTNs in 27 different cities with nodes range from 549 in Kuopio to 24063 in Sydney in three network representations. We showed in Lspace degree distribution approximately follows a power-law distribution with exponent $2.63 < \gamma < 4.14$, while in Pspace and Cspace the degree distribution is exponential with characteristic exponents $\alpha \in [0.01, 0.05]$ and $\alpha \in [0.01, 0.14]$, respectively. The average degrees in cities were disregard to the network size. The assortativity coefficients measured in the space L were positive for the whole range of N while in the Pspace it changes from negative values for small networks to positive values for large cities. In Cspace, the assortativity coefficients in the most large cities have a high positive value while in small cities were almost close to zero. The value of the average shortest path lengths was much smaller than the network size, and the average clustering coefficients in all cities were higher than these values in random graphs. Therefore, studied networks appear to be correlated small-world structures with the high values of clustering coefficients and low average shortest path lengths compare to the network size .

Since the network properties do not relate to the network size, we can conclude the differences in the public transport networks may due to historical and geographical factors.

ACKNOWLEDGMENT

The calculations presented above were performed using computer resources within the Aalto University School of Science "Science-IT" project.

References

[1] Vito Latora and Massimo Marchiori. "Efficient behavior of small-world networks". In: *Physical review letters* 87.19 (2001), p. 198701.

[2] Bertrand Berche et al. "Resilience of public transport networks against attacks". In: *The European Physical Journal B* 71.1 (2009), pp. 125–137.

[3] Jingyi Lin and Yifang Ban. "Complex network topology of transportation systems". In: *Transport reviews* 33.6 (2013), pp. 658–685.

[4] Zbigniew Tarapata. "Modelling and analysis of transportation networks using complex networks: Poland case study". In: *Archives of Transport* 36 (2015).

[5] A. Háznagy et al. "Complex network analysis of public transportation networks: A comprehensive study". In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. 2015, pp. 371–378.

[6] Christian Von Ferber et al. "Public transport networks: empirical analysis and modeling". In: *The European Physical Journal B* 68.2 (2009), pp. 261–275.

[7] Julian Sienkiewicz and Janusz A Hołyst. "Statistical analysis of 22 public transport networks in Poland". In: *Physical Review E* 72.4 (2005), p. 046127.

[8] Yihong Hu and Daoli Zhu. "Empirical analysis of the worldwide maritime transportation network". In: *Physica A: Statistical Mechanics and its Applications* 388.10 (2009), pp. 2061–2071.

[9] Parongama Sen et al. "Small-world properties of the Indian railway network". In: *Physical Review E* 67.3 (2003), p. 036106.

[10] Huapu Lu and Ye Shi. "Complexity of public transport networks". In: *Tsinghua Science and Technology* 12.2 (2007), pp. 204–213.

[11] Maciej Kurant and Patrick Thiran. "Extraction and analysis of traffic and topologies of transportation networks". In: *Physical Review E* 74.3 (2006), p. 036114.

[12] Vito Latora and Massimo Marchiori. "Is the Boston subway a small-world network?" In: *Physica A: Statistical Mechanics and its Applications* 314.1-4 (2002), pp. 109–113.

[13] Katherine A Seaton and Lisa M Hackett. "Stations, trains and small-world networks". In: *Physica A: Statistical Mechanics and its Applications* 339.3-4 (2004), pp. 635–644.

[14] Panagiotis Angeloudis and David Fisk. "Large subway systems as complex networks". In: *Physica A: Statistical Mechanics and its Applications* 367 (2006), pp. 553–558.

[15] Hui Zhang et al. "The analysis of the properties of bus network topology in Beijing basing on complex networks". In: *Mathematical Problems in Engineering* 2013 (2013).

[16] Vamsi Kalapala et al. "Scale invariance in road networks". In: *Physical Review E* 73.2 (2006), p. 026130.

[17] Rainer Kujala et al. "A collection of public transport network data sets for 25 cities". In: *Scientific data* 5 (2018), p. 180089.

[18] Albert-László Barabási, Réka Albert, and Hawoong Jeong. "Mean-field theory for scale-free random networks". In: *Physica A: Statistical Mechanics and its Applications* 272.1-2 (1999), pp. 173–187.

[19] Lidia A Braunstein et al. "Optimal paths in disordered complex networks". In: *Physical review letters* 91.16 (2003), p. 168701.