UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Public Transport Networks: Topological Features and Stability Analysis

Author:
Yaryna KORDUBA

Supervisors:
Prof. Yurij HOLOVATCH
Dr. Robin DE REGT

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2019

Declaration of Authorship

I, Yaryna KORDUBA, declare that this thesis titled, "Public Transport Networks: Topological Features and Stability Analysis" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"A developed country isn't a place where the poor have cars. It's where the rich use public transportation."

Gustavo Petro

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Public Transport Networks: Topological Features and Stability Analysis

by Yaryna KORDUBA

Abstract

In this study, we present a quantitative analysis of the Lviv and Bristol public transportation networks (PTN) viewed as complex systems. We integrate methods in statistical physics to investigate the correlation between PTN topological features and their operational stability.

Initially, to present a PTN in the form of a complex network (i.e. a graph consisting of vertices-nodes and edges-links), we perform a coarse-graining procedure. We merge stations considered to be within a reasonable pedestrian walking distance (e.g. stops across the street) by implementing a DBSCAN clustering algorithm to the transport dataset.

Subsequently, we analyse the topological features of the resulting complex networks in various network representations reflecting PTN operational features. In the second part of our analysis we assess the vulnerability of PTNs by removing network constituents according to different protocols (attack scenarios). We observe correlations between network topological features and its stability with respect to random failures and targeted attacks.

Acknowledgements

Firstly, I would like to thank my supervisors: Prof. Yurij Holovatch and Dr. Robin de Regt. Thank you for interesting scientific discussions, sharing the insights and unfailing support throughout the research. I would like to say a special thank you to the members of Laboratory for Statistical Physics of Complex Systems for all of the conversations over a cup of tea from which I also learnt a lot.

I am deeply grateful to Mykhailo Ivankiv. Thank you for always being there to help with no matter the question and for keeping me motivated.

I am very thankful to my grandmothers who hosted me and my groupmates and cared for us during the most intense periods of work.

My deepest gratitude goes to my parents and my brother. Thank you for always being a shoulder to lean and for cheering me up during the whole period of studying.

I also want to thank Volodymyr Spodaryk for providing Lviv public transport datasets.

Special thanks to Oles Kozak and Olena Skibinska for listening to the stories about transport networks and for helping with the illustrations.

Finally, I am very grateful to Adrian Slywotzky and Ukrainian Catholic University for granting me scholarship to cover my tuition fees.

Contents

List of Figures

List of Tables

List of Abbreviations

PTN	**P**ublic **T**ransport **N**etwork
GCC	**G**iant **C**onnected **C**omponent
RA	**R**andom **A**ttack

List of Symbols

N	number of nodes of a network
M	number if edges of a network
k	node degree
l	shortest path length
l_{rand}	random graph shortest path length
l_{η}	path length efficiency
D	diameter (maximal shortest path length)
C	clustering coefficient
C_{rand}	random graph clustering coefficient
C_{β}	betweenness centrality
r	assortativity

Dedicated to public transport users.

Chapter 1

Introduction

An efficient public transport network (PTN) is an essential factor in city planning. The priority of public transport does not only concern itself with traffic congestion but stretches into the realms of even urban economics affecting the ratio between the rich and poor within a city. For example, in London, the priority of public transportation in urban planning and the inconvenience of using personal cars makes accommodation in the city attractive for the rich (Glaeser, Kahn, and Rappaport, 2008). In contrast, in many cities, e.g., Detroit, the inefficiency of the public transport together with authomobilization policy was one of the contributing factors that led to an increase in poverty in the city (Freeman, 2011). A sound transport system must be able to provide efficient movement of passengers and quickly recover in case of unpredictable, disruptive events. Therefore, the efficiency and resiliency of public transport systems are an essential topic for analysis.

In many works considering public transport analysis PTNs are analyzed using complex network science. With a recently growing interest in the features of natural and man-made systems, complex network science has acquired a broad range of applications in different fields. The examples include not only transportation networks, but also social collaboration networks (Wasserman and Faust, 1994, Granovetter, 1977), power grid networks (Amaral et al., 2000, Crucitti, Latora, and Marchiori, 2004, Albert, Albert, and Nakarado, 2004), Internet (Dorogovtsev and Mendes, 2013, Pastor-Satorras and Vespignani, 2007), ecological networks (Sole and Montoya, 2001) etc.

The growth of interest in public transport systems started at the beginning of the 21st century. Figure 1.1 (de Regt et al., 2018) indicates locations where scientists have analyzed public transport systems using a complex network science approach.

A transport system can be analyzed considering different features. Firstly, the transport system is a network. Thus, one can explore the network topology features. Secondly, the routes are embedded in 2D space. Therefore, one can consider their spatial coordinates and Euclidean distances in the analysis. Finally, a transport system contains the processes - transport and passenger movement over time. Thus, one can study the system in terms of its dynamic changes.

The direction of this study is the exploration of the system topology. In this work, we do not consider the spatial coordinates of the routes as well as dynamic processes, e.g. load distributions over time. To this end, we have chosen for the analysis the PTNs of Lviv and Bristol. The motivation behind such a choice is that Lviv, to the best of our knowledge, has never been analyzed using a complex network approach. Bristol was chosen because of the similarity of its size to Lviv.

The rest of this paper is organized as follows: in chapter 2 we describe the previous explorations related to our topic; in chapter 3 the structure of the datasets, data processing flow and network simplification process are explained; chapter 4

FIGURE 1.1: Locations where city PTN have been analyzed within the complex network approach (de Regt et al., 2018)

describes main local and global indicators of Lviv and Bristol PTNs; chapter 5 covers the analysis of the PTNs resilience to random and targeted failures of the nodes; and finally, in chapter 6 the concluding remarks are given. Before proceeding to the transport networks analysis, we would like to define the most important terms related to our topic.

1.1 Main definitions

Graph $G\langle V, E\rangle$ is a mathematical object represented by a set of nodes $V = \{v_i | i = 1, 2, 3, ...N\}$ and a set of edges $E = \{e_{ij} | v_i, v_j \in V\}$. Each edge connects a pair of nodes. If an edge connects two nodes, these nodes are called adjacent, or neighbors. A graph is called complete if all of the nodes are pairwise adjacent. A graph is called connected if a path exists between any pair of nodes. Otherwise, it is disconnected.
A random graph is a collection of nodes and edges that connect pairs of nodes at random (Newman, Strogatz, and Watts, 2001). In this work by the term "random graph" we refer to a random graph generated by Erdös-Rényi model $G(N, M, p)$. In this model a fixed set of nodes N and number of edges M are predefined. The procedure of graph generation is as following:

1. Select two nodes at random and connect them with an edge.

2. Iterate through the previous step until a total of M edges are generated (de Regt, 2018).

Giant Connected Component GCC is a connected subnetwork which in the limit of an infinite network contains a finite fraction of the network (Berche et al., 2009). In the particular case of finite networks, we indicate the GCC as the largest connected component of the network.
Node degree k_i of the node v_i is a number of edges linked to this node. k_{max} stands for the maximal node degree. $\langle k\rangle$ stands for the average node degree and is defined as

$$\langle k\rangle = \frac{2M}{N}, \tag{1.1}$$

where M and N are the number of edges and the number of nodes of the network, respectively.

Shortest path l_{ij} between the nodes v_i and v_j is the path with the smallest number of the edges between them. The average shortest path length $\langle l \rangle$ of the whole network is defined as

$$\langle l \rangle = \frac{\sum_{i \neq j} l_{ij}}{N(N-1)}, \quad (1.2)$$

where l_{ij} is the shortest path length between nodes v_i and v_j and N is the number of the nodes in the graph.
Diameter D is the maximal shortest path length between any two nodes in the graph.
Path length efficiency l_η is a relative measure that compares the mean shortest path of a particular graph with the mean shortest path of a random graph of the same size. It is defined as

$$l_\eta = \frac{\langle l \rangle}{\langle l_{rand} \rangle}, \quad (1.3)$$

where $\langle l \rangle$ and $\langle l_{rand} \rangle$ are the mean shortest paths of a graph and of a random graph of the same size, correspondingly.

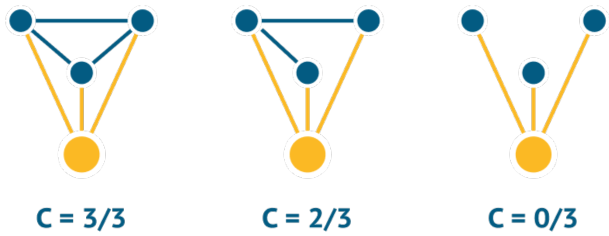


FIGURE 1.2: Clustering coefficient of the node depending on the number of links between its neighbors

Clustering coefficient C_i of the node v_i provides information on how the neighbours of this node are joined (von Ferber et al., 2009b). The clustering coefficient value is defined as

$$C_i = \frac{2y_i}{k_i(k_i-1)}, \quad (1.4)$$

where k_i is the degree of node i and y_i is the number of links between its neighbors (von Ferber et al., 2007). Mean clustering coefficient C is defined as

$$C = \frac{1}{N}\sum_i C_i, \quad (1.5)$$

where N is the number of nodes in the network. The clustering coefficient of random graph is defined as

$$C_{rand} = \frac{\langle k \rangle}{N}, \quad (1.6)$$

where $\langle k \rangle$ is the mean node degree.

Betweenness centrality $C_\beta(i)$ of a node v_i is a measure that estimates the importance of a node in the network. It depends on the number of shortest paths which go through a given node. Betweenness centrality is defined as

$$C_\beta(i) = \sum_{x \neq y} \frac{\sigma_{xy}(v_i)}{\sigma_{xy}}, \quad (1.7)$$

where $\sigma_{xy}(v_i)$ is the number of shortest paths between the nodes x and y that go through the node v_i and σ_{xy} is the number of all shortest paths between the nodes x and y.
Assortativity r indicates whether the nodes of the same degree tend to be connected. For any edge i, let X_i and Y_i be the node degrees of the two nodes connected by this edge. Then assortativity is defined as

$$r = \frac{M^{-1}\sum_i X_i Y_i - (M^{-1}\sum_i \frac{1}{2}(X_i + Y_i))^2}{M^{-1}\sum_i \frac{1}{2}(X_i^2 + Y_i^2) - (M^{-1}\sum_i \frac{1}{2}(X_i + Y_i))^2}, \quad (1.8)$$

where M is the number of links in the network (von Ferber et al., 2009b). If $r > 0$, the nodes of the similar degree are likely connected and the network is called assortative. Otherwise, if $r < 0$, the network is called dissortative. It means that nodes with a high degree tend to link to nodes with a low degree. If $r \approx 0$, no tendencies are observed.
Molloy-Reed criterion is the criterion that can be used to determine the stability of large scale transportation networks. The stability of a network depends on the presence of a GCC. Molloy-Reed criterion states that the GCC is present in any uncorrelated network if

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2, \quad (1.9)$$

where $\langle k \rangle$ stands for mean node degree and $\langle k^2 \rangle$ is the mean square node degree (Molloy and Reed, 1995).
Scale-free network is a network whose node degree distribution obeys a power law. Whereby the name "Scale-free" was first coined by Barabási and Albert, 1999 to describe network exhibiting this property.
Small-world network is a network in which most of the nodes can be reached from any other node with a small number of steps whilst still being highly clustered (Watts and Strogatz, 1998). The explorations of transport networks have mainly concentrated on the question of whether these systems possess the features of a small-world model.

1.2 Transport graph representations

With the development of transport network studies, various types of topology representations have been developed. One can present a transport network as an undirected graph where the nodes are the stations and links between are undirected edges. Such a type of representation is called $\mathbb{L}$-space. From another perspective, one can display the stations as nodes and link any two stations if they serve the same route. Such a type of representation is called $\mathbb{P}$-space. The general ideas of $\mathbb{L}$- and $\mathbb{P}$-space first appeared in the work of Sen et al., 2003. One can also present the network routes as the nodes and link any two routes if they have at least one common stop. Such representation type is called $\mathbb{C}$-space (von Ferber et al., 2009a). In another representation, $\mathbb{B}$-space (von Ferber et al., 2007, Chang et al., 2007, Zhen-Tao et al., 2008), a public transport network is represented as a bipartite graph with the two types of nodes: the stations and the routes. If a station belongs to the route, the route and the station are linked. One can also use $\mathbb{L}'$- and $\mathbb{P}'$-space. The only difference from the classical representations ($\mathbb{L}$- and $\mathbb{P}$-space) is that in the primed versions multiple links between the nodes can exist (von Ferber et al., 2007). The examples of the spaces mentioned above are presented in Figure 1.3.

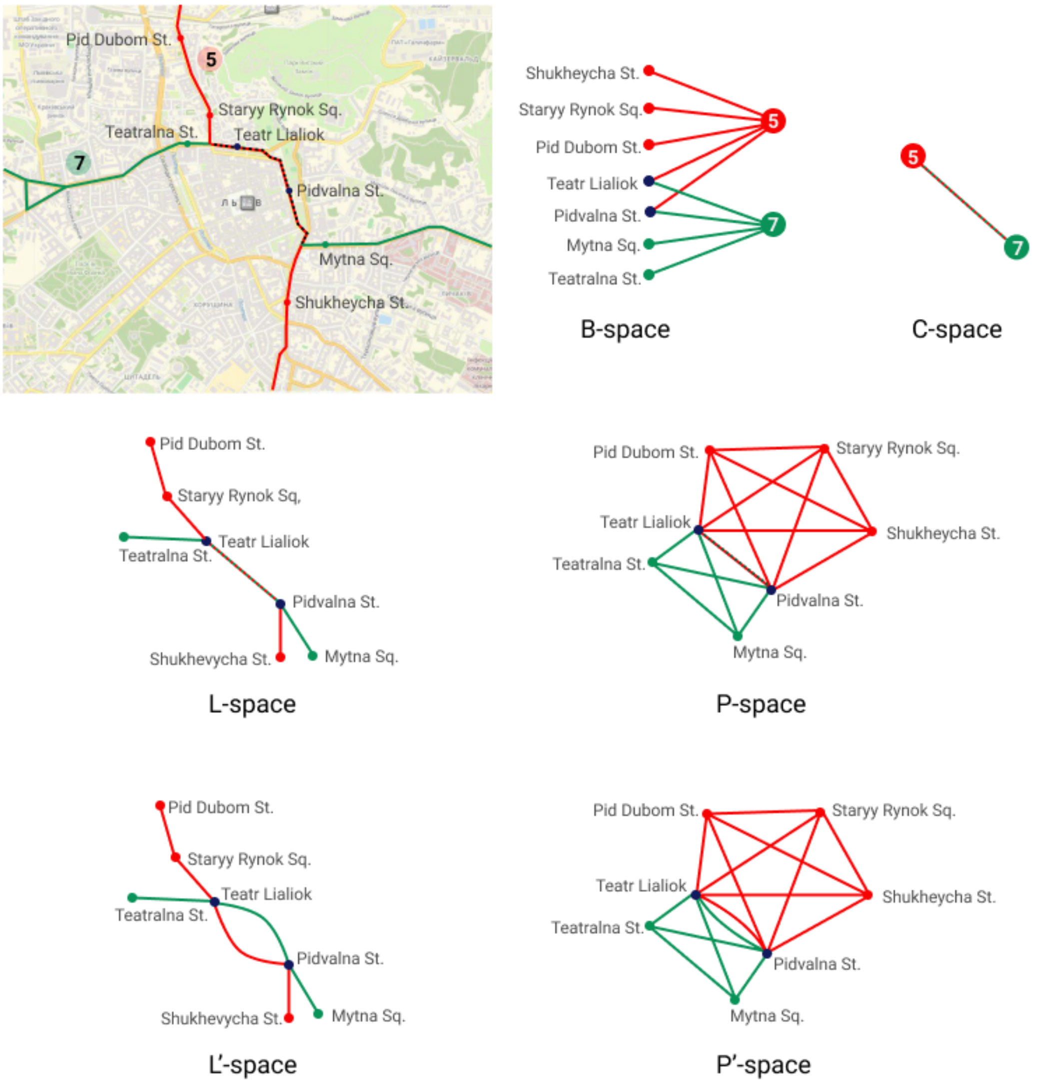


FIGURE 1.3: A piece of Lviv PTN and its representations in different graph "spaces"

Chapter 2

Related works

In recent years, many scientists have used complex network theory in studies of transport systems. The systems studied include not only urban public transport, but also national railways, airlines, etc. It appeared that in spite of the differences in size, transport networks possess certain similar features, for example, scaling behavior.

One of the first studies of public transport topology using complex network theory was presented by Latora and Marchiori in the article "Is the Boston subway a small-world network?" (Latora and Marchiori, 2002). The authors represented the subway system with N stations and M tunnels as a graph with N nodes and M edges. In the analysis, the authors considered such characteristics as the mean shortest path and the clustering coefficient. Later Latora and Marchiori characterized the network in terms of how efficiently it propagated the information. They assumed that the efficiency in the communication between two nodes is inversely proportional to the shortest path length between them. They also distinguished between global efficiency - "the efficiency of the whole network" and local efficiency - "the average efficiency of the subgraph of a generic node i". Besides efficiency, the authors defined the measure of the cost of a network - "the price to pay for the number and length (weight) of edges". The cost is a relative measure, and $Cost = 1$ if the graph is complete. In the case of an unweighted graph $Cost = \frac{2M}{N(N-1)}$, where M is the number of edges and N is the number of nodes.

Another exploration of network efficiency was considered by Gastner and Newman (Gastner and Newman, 2004). In their work, the authors followed the idea that in a "good" network the lengths of the paths between every two vertices should be relatively short and the sum of lengths of the network edges should be low. The efficiency of transport networks was also analyzed by Barthélemy and Flammini (Barthélemy and Flammini, 2006). They presented a study where they described the optimality of traffic networks.

Sen et al., 2003 analyzed the Indian railway system. As it was stated before, in this work, the authors first presented a general idea of $\mathbb{L}$-space. They investigated the properties of the network to see if it possesses some general scaling behavior. The analyzed topological features included degree distribution, clustering coefficient, and assortativity. The authors found that the Indian Railway Network has a disassortative structure. They also concluded that this network possesses small-world properties.

The explorations of airline networks were presented in the articles "Modeling the world-wide airport network" (Guimera and Amaral, 2004) and "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles" (Guimera et al., 2005). Here, the authors concluded that the world-wide airport network also shares small-world properties. They analyzed the correlation between node degree and betweenness centrality. Surprisingly, the explorations showed that in the airport networks the nodes with the highest degree are not always the nodes with the highest betweenness centrality. It meant that the cities with the most significant number of connections are not necessarily those that belong to the largest number of shortest paths. The authors concluded that the nodes with higher betweenness centrality tend to be more important for the connectivity of the network than those with a higher degree. Explorations of the airport networks are also presented in the works of Li et al., 2006 and Guida and Maria, 2007.

The first exploration of the overall city transport system was presented by von Ferber at al. in the article "Scaling in public transport networks" (von Ferber, Holovatch, and Palchykov, 2005). In this study, they analyzed the public transport system (trams, buses, and subways) of three big cities: Düsseldorf, Berlin, and Paris. The authors found the node degree distribution of the networks. It appeared that the node degree distribution obeys Zipf's (power) law. Therefore, the authors concluded that the explored transport networks were scale-free.

In contrast, the assortativity values in $\mathbb{P}$-space were different: negative for small towns and positive for big cities. As the authors explained, small towns usually have a star transport network structure and only a few doubled routes, so there exist a lot of connections between the nodes of high and low degree. Additionally, the authors derived the plots of the dependence of graph nodes on the size and the population of the cities. To end with, the authors concluded that despite the difference in the sizes, considered networks share some common features. These features include the behavior of degree and path length distributions. Another common property is logarithmic dependence of distances on node degrees. All the explored networks appeared to be hierarchically organized and exhibit small-world behavior.

von Ferber et al., 2007 analyzed the public transport systems of 14 major cities of the world in $\mathbb{P}$-, $\mathbb{L}$- and $\mathbb{L}'$-space. The exploration of $\mathbb{P}$- and $\mathbb{L}$-space included such characteristics as mean and maximal shortest path lengths, mean clustering coefficient, and the percolation value. One of the features found in $\mathbb{L}'$-space is that several routes can proceed in parallel on the same road for a sequence of stations. To describe such behavior, the researchers presented a new characteristic - the harness distribution $P(r,s)$. The harness distribution shows the number of sequences of s consecutive stations that are serviced by r parallel routes. The harness distribution for most of the studied cities appeared to be scale-free. Knowing that PTNs possess scale-free properties, the authors proposed a growth model for these networks. They represented a grid of streets by a quadratic 2D lattice and modeled the routes as self-avoiding walks on a lattice. Such an approach aims to find a balance between the area coverage and traveling time.

von Ferber et al., 2009b analyzed the characteristics of PTNs of 14 cities in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space. The characteristics included the node degree distribution in $\mathbb{L}$-space, cumulative node degree distribution in $\mathbb{P}$- and $\mathbb{C}$-space and assortativity. The results showed that for about half of the cities node degree distribution in $\mathbb{L}$-space possessed an exponential tail while another part of the cities possessed power law decay. In $\mathbb{P}$-space power-law behavior, in general, is not observed. In $\mathbb{C}$-space node degree for the explored cities decays exponentially or even faster. Assortativity r of nodes was analyzed considering their nearest neighbors ($r^{(1)}$) and second nearest neighbors ($r^{(2)}$). The authors concluded that there was no linkage between the network size and degree assortativity in $\mathbb{L}$-space. In $\mathbb{P}$-space most of the PTNs were characterized by minimal negative or positive assortativity values. In $\mathbb{C}$-space nearly all the explored PTNs demonstrated clear assortative mixing ($r = 0.1 \div 0.5$). $r^{(2)}$ for the nodes with high values of $r^{(1)}$, in general, appeared to be even stronger. The authors further analyzed the betweenness-degree correlation of Paris in $\mathbb{L}$-, $\mathbb{P}$-, $\mathbb{C}$-

and $\mathbb{B}$-space. The correlation in $\mathbb{L}$- and $\mathbb{C}$-space tends to obey a power law. The correlations in $\mathbb{B}$- and $\mathbb{P}$-space show different behavior depending on the node degree values.

One of the recent analyses of transport networks topology was presented in the article "Public transportation in Great Britain viewed as a complex network" (de Regt et al., 2018). To gain information about the robustness and efficiency of the networks, the authors considered topological and spatial features of the PTNs of Greater London, Greater Manchester, West Midlands, and Bristol, as well as the national rail and coach networks of Great Britain. For the analysis, the $\mathbb{L}$-space network representation was chosen. The researchers used the Molloy-Reed criterion to define network stability. In the research, the authors also studied the fractals properties of these systems. The authors completed the study by analyzing the network load and dynamics features. Interestingly, the exploration of network load during a sample day showed that the load distributions of all the analyzed networks follow a similar behavior pattern (Figure 2.1).

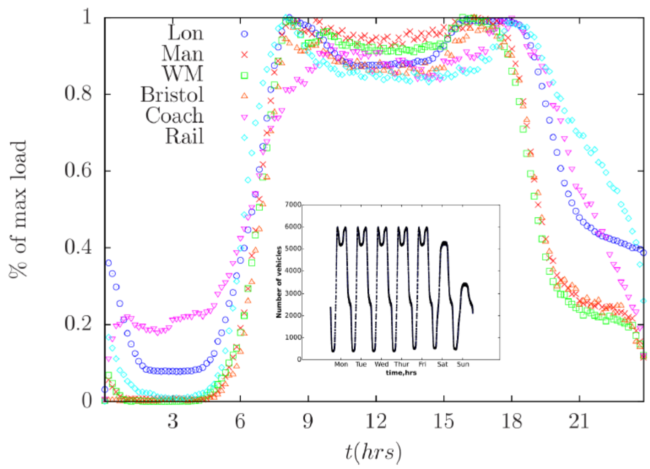


FIGURE 2.1: Distribution of the load in some of the UK transport systems as a function of time over 24 hours (de Regt et al., 2018). The inset shows the London load for the entire week

Another important direction of complex network topology studies is an analysis of the network vulnerability. The vulnerability studies considered different real-world networks: Internet (Cohen et al., 2000, Albert, Jeong, and Barabási, 2000), protein (Jeong et al., 2001), social networks (Yi et al., 2015) etc. Many vulnerability explorations considered the influence of network topology on links and nodes failures. It appeared that complex networks are highly resilient to random failures but vulnerable to targeted removals of the most important nodes or links.

Recently many studies considering public transport network vulnerability have been conducted. The explorations of transport vulnerability include the works of Berche et al. (Berche et al., 2009, Berche et al., 2010), Rodríguez-Núñez and García-Palomares (Rodríguez-Núñez and García-Palomares, 2014), Cats and Jenelius (Jenelius and Cats, 2015) etc.

Berche et al. considered the impact to network topology through simulated attacks in $\mathbb{L}$- and $\mathbb{P}$-space (Berche et al., 2009. The study was based on two types of attack scenarios: random failures and targeted attacks. In targeted attacks, the nodes were removed according to the node lists sorted in order of decreasing node importance. For different scenarios such importance criteria were considered: degree k, closeness C_C , graph C_G, stress C_S, and betweenness C_B centralities, clustering coefficient C, and next nearest neighbors number z_2. After the attack, the properties of the present nodes in the network can change. Therefore, the authors conducted the targeted attacks simulations in two modes: simulations using initial sorted list and simulations with list recalculation after each step. At each step of the simulation, 1% of nodes were removed from the network. After that, the authors analyzed the network state. During the simulations the changes of three characteristics were observed: the normalized size of the GCC ($S = N_{GCC}/N$, where N is the initial number of nodes in the network and N_{GCC} is the number of nodes of the largest connected component), the average shortest path $\langle l \rangle$ and the average inverse shortest path $\langle l^{-1} \rangle$. For the sake of uniqueness, the effectiveness of the attacks was measured with the value of S. To determine the percent of the removed nodes c at which network stops to operate the characteristic concentration of removed nodes c_s was defined. c_s is the percent of nodes at which S decreases to one half of its initial value:

$$S(c_s) = \frac{1}{2} S(c = 0) \tag{2.1}$$

As the simulations showed, the random attacks are the least harmful to the PTNs. The most effective scenarios are usually removals by stress and betweenness centralities, node degree and next nearest neighbors number (Figure 2.2). Also, differences between "initial" and "recalculated" scenarios were observed. These differences are significant for centrality-based attack strategies but smaller for scenarios based on local characteristics (the node degree, the number of second nearest neighbors, etc.).

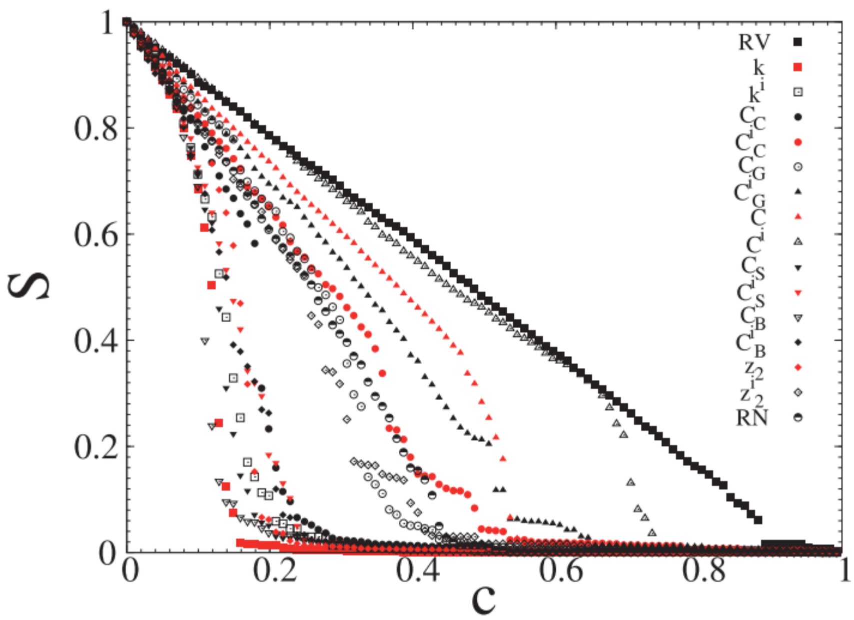


FIGURE 2.2: Typical attack simulation results on example of Paris PTN. S - normalized GCC size, c - fraction of removed nodes. The lists of nodes for removal were prepared according to degree k, closeness centrality C_C, graph centrality C_G, stress centrality C_S, betweenness centrality C_B, - clustering coefficient C, next nearest neighbors number z_2. Subscript i denotes vertices removal from the initial lists. RV and RN indicate removal of random vertex and removal of its randomly chosen neighbor

The authors also observed the correlation in $\mathbb{L}$-space between resilience and the node degree distribution exponent γ. It appeared that in a random attack and node-degree recalculated scenario the PTNs with the smaller value of γ tend to be more resilient.

von Ferber, Holovatch, and Holovatch, 2009 analyzed the resilience of 14 PTNs to the attack in $\mathbb{L}$-space. The random attack scenarios included random vertex removal and random neighbor removal (removal of a randomly chosen neighbor of random vertex). In directed attack scenarios nodes were removed according to the lists sorted by decreasing node degrees k, centralities (closeness centrality, graph centrality, stress centrality, betweenness centrality), the number of second nearest neighbors and increasing clustering coefficient. The directed attacks were implemented in two modes: removal according to the original sorted list and removal with list recalculation after each step. The network degradation under the attacks was judged by the changes in the normalized GCC value S, the inverse shortest path $\langle l^{-1} \rangle$, maximal shortest path l_{max} and the mean shortest path $\langle l \rangle$. In simulations these four attack scenarios appeared to be the most harmful: attacks by degree (with recalculation), attack by the largest number of second neighbors, attacks by betweenness centrality and attacks by stress centrality. The analysis showed that l_{max} as function of the removed node concentration displays a sharp maximum (Figure 2.3). This maximum can indicate the breakup of the network. In contrast, behavior of S and $\langle l^{-1} \rangle$ is smooth. Therefore, l_{max} appeared to be a better criterion to indicate the breakup of a network. Another direction in vulnerability studies considers network capacity limits. In this approach, load changes during the simulations are observed. One of the works in this direction was presented by Cats and Jenelius (Cats and Jenelius, 2018). Here, the authors analyzed the partial capacity reduction in service. They defined capacity as the number of transport units that traverse a particular part of the network under normal conditions. The authors simulated the network disruptions for the morning peak period. Two types of events were considered: planned line-level and unplanned link-level outages. The latter disruptions occurred at the most central links based on the passenger load in the network. The authors assessed the network performance by the changes in passenger waiting time, walking time, in-vehicle time and the number of transfers.

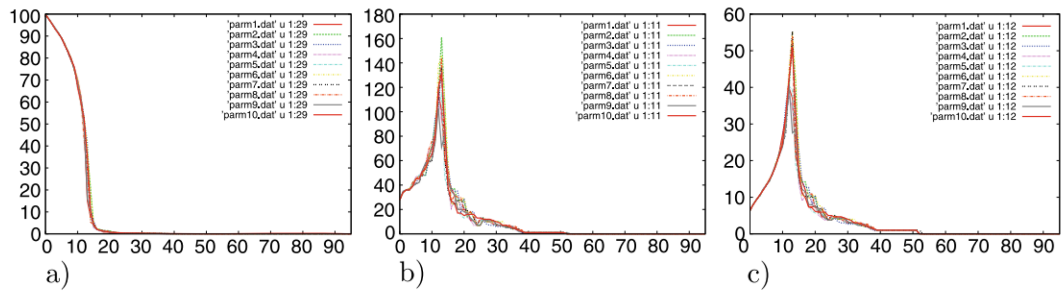


FIGURE 2.3: 10 instances of Paris PTN recalculated highest degree attack scenario (von Ferber, Holovatch, and Holovatch, 2009). Vertical axis: a) the normalized GCC size S, b) the maximal shortest path length l_{max}, c) the mean shortest path length $\langle l \rangle$. Horizontal axis: percent of removed nodes.

Zhang, Fu, and Li, 2016 and Zhang et al., 2018 presented studies of cascading node failures. In the research, they modeled a weighted PTN in a modified $\mathbb{L}$-space, which describes passenger flow. The edge weights depend on the bus route density and departure frequency of the two stations adjacent to the edge. They considered the network capacity changes during the attack. Due to the attack, the load distribution from the removed node redistributed to its neighbors. Sometimes it exceeded the capacity limit of these neighbors and caused their failure. Therefore, after the removal of a node from the network, the consequently cascading failure of other nodes occured.

Recently Candelieri et al., 2019 presented a vulnerability study of the PTNs of two Italian cities. The PTN networks were modeled as directed multi-graphs. Firstly, the authors simulated targeted attacks at the nodes with the highest degree and betweenness centrality. Secondly, they conducted cascading failure simulations for the network. The cascading failures started from the node with the highest betweenness centrality. In the simulations, authors used the strategy of maintaining the routes passing through an unavailable station. Thus, a real transport situation was modeled. In such a situation it is possible to bypass the removed station and move to other stations on the route.

Chapter 3

Data processing

The dataset for Bristol urban and suburban public transport is taken from National Transport Data Repository (*National Transport Data Repository*). The original dataset is presented in ATCO-CIF timetable format. The data represent the logs of all journeys with their stops during one observed week in 2011. For each stop, the geospatial coordinates, name, and departure and arrival times are specified. Figure 3.1 shows an example of ATCO-CIF structure.

```
ATCO-CIF0510              Bristol - BUS              ATCOPT20111206170024
QSNFB001B41892010040599999999111100 XX10      0              I
QO0100053214  0820   T1F0
QI5330AWB3517708450845B   T1F0
QI5330AWB3517208450845B   T0F0
QT5330WDB224350850   T1F0
QLN5330AWB35177Thornwell Schools (NW)                            E0039362
QBN5330AWB35177353721  191951
QLN5330AWB35172Warren Slade (NW)                                 E0039362
QBN5330AWB35172353615  192192
QLN5330WDB22435Chepstow Bus Station (4)                          E0054379
QBN5330WDB22435353177  193748
```

FIGURE 3.1: One journey snapshot in ATCO-CIF timetable file. A line starting with QS represents route information, the last character in the route information row defines whether the route is incoming (I) or outcoming (O). Lines starting with QO, QI, QT define the information about start, intermediate and terminal stations. The following 12 characters are unique station identifiers. The next 4 characters indicate arrival and departure time at the station. QL and QB denote lines with the information about station coordinatesand name.

The Lviv static dataset was taken from UA-Gis Track system. The original dataset has GTFS (General Transit Feed Specification) format (*GTFS Static Overview*). The Lviv original dataset includes files with the following information:

- agency.txt contains information about transport agencies.
- stops.txt contains information about stop IDs, names, spatial coordinates, etc.
- routes.txt contains information about route IDs, names, route vehicle types, etc.
- trips.txt contains information about trip IDs, directions, the route IDs that they belong to, etc.
- stop_times.txt contains information about stop IDs, trip IDs, arrival and departure times, etc.
- calendar.txt indicates weekdays of transport service availability.

- calendar_dates.txt indicates dates of transport service availability.
- feed_info.txt contains general information about the dataset.

For purposes of reproducibility, the original Lviv and Bristol datasets are presented in public repository (*Original Datasets, GitHub*).

Before the analysis, we process the datasets into one universal form. Firstly, we convert the datasets from their original formats into JSON form. To process the Bristol dataset, we use ATCO-CIF parser (*ATCO-CIF parser*). Secondly, we convert the datasets into the form of two files:

- stops.json. Contains information about stop id, label, spatial coordinates, stop vehicle type, ids of the routes that the stop serves, ids of adjacent stops (neighboring stops on the routes).
- journeys.json. Contains information about route id, name, stops, direction (incoming or outcoming), and route vehicle type.

In the Bristol dataset, there are logs of the city as well as outskirt routes. Therefore, to compare only the city networks, we consider the city boundaries [1] (Figure A.1) and filter the datasets according to these boundaries. The information about the explored cities is indicated in Table 3.1.

City	Population	Area	N	R	Vehicle types
Lviv	721 301	$182km^2$	768	77	BET
Bristol	535 907	$110km^2$	1474	143	BF

TABLE 3.1: General information about the cities. N - considered stops number, R - considered routes number, B - bus, E - electric trolley, T - tram, F - ferry.

A lot of the stops in the city networks were close to each other (e.g., stops across the street). To traverse between two close stops a passenger often needs less than a minute. In that case, there is no need to consider such stops separately, and one can simplify the network. In the previous exploration of Bristol PTN de Regt et al., 2018, used reducing of the routes of the network. In most of the Bristol routes, outgoing routes approximately duplicate incoming. Therefore, the authors rejected the outcoming routes. In Lviv, more than 50% of the routes have differences in incoming and outcoming routes. Thus, we cannot use such an approach. Gallotti and Barthelemy, 2015 connected different layers of the multilayer network by aggregating the stops associated with different modes of public transport into a single network of nodes with the help of a coarse-graining procedure. In our work, we also apply coarse-graining to merge the stops that fall within a small distance of each other. For this purpose, we use a density-based clustering algorithm DBSCAN (Ester et al., 1996). DBSCAN algorithm considers clustering radius R and a minimal number of points $MinPts$ for creating the clusters of the points that are tightly packed together. The algorithm divides all points, in a dataset, into a few categories:

- Core points. A point p is a core point if at least $MinPts$ points are within radius R from it.

[1] To filter Bristol dataset, we used Bristol county boundaries which consist of Bristol city area and a part of the river. However, it does not affect the results of the PTN data filtering.

- Directly density reachable points. A point p is directly reachable if it is within radius R from a core point c.
- Density reachable points. A point p is reachable from a core point q if there exists a chain of points $p_1, p_2, ...p, p_1 = q$ and p_{i+1} is directly density-reachable from p_i
- Noise points. A point is a noise point if it does not belong to any cluster.

Core points together with their reachable points form clusters.

To apply the algorithm to the transport datasets, one should choose the reasonable R. In the selection of R such factors are important:

- R should be at a reasonable pedestrian walking distance
- if R is too small no stops in the network will be clustered (Figure 3.2)
- if R is too large all stops in the network will be grouped in one cluster (Figure 3.2)

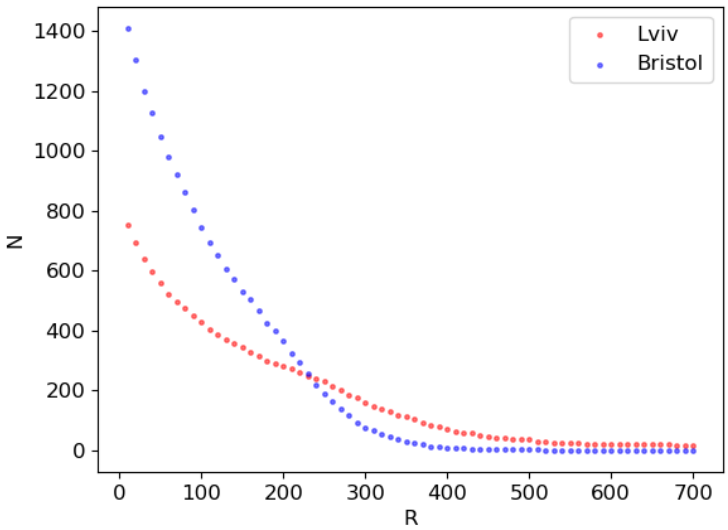
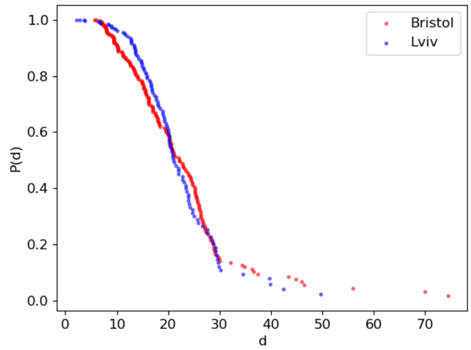


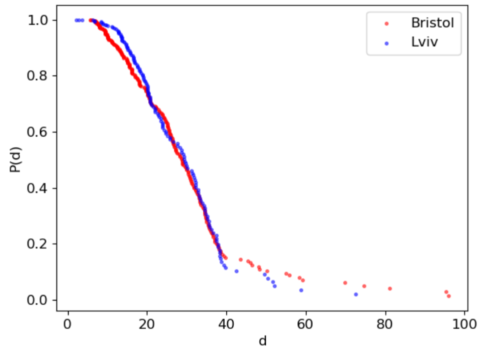
FIGURE 3.2: Dependence of the stops number N on the clustering radius R (in meters)

We assume that a reasonable pedestrian walking distance should not exceed 100m. One spends around one minute to pass such a distance on foot. Figure 3.3 indicates cumulative distributions of maximal distances in the clusters for Lviv and Bristol PTN. With $R = 40$m the maximal cluster distances for both cities are high, but do not exceed 100m. Therefore, $R = 40$m is considered as an optimal clustering radius for these particular datasets. The heatmaps of the clustered Lviv and Bristol PTNs are shown on the Figure A.2.

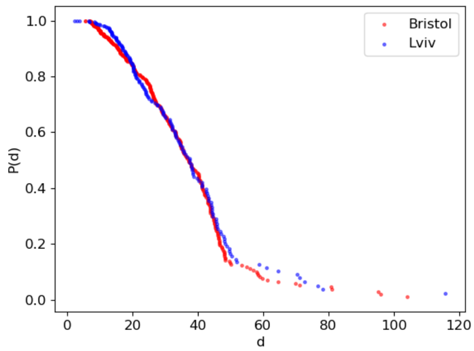
To simplify the network, we also ignore the route directions. Thus, we can construct an undirected graph from our transport network. If a few edges are present between two vertices, we leave only one of them. Therefore, we can build a simple graph. Figure 3.4 shows the whole process of network simplification.



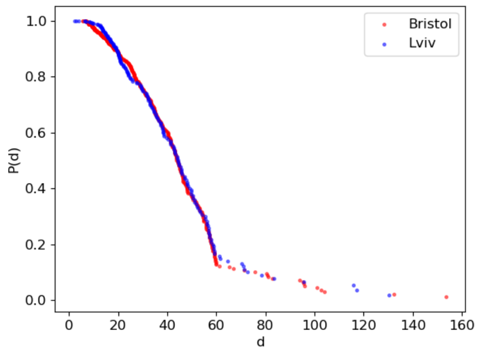
(A) R=30m



(B) R=40m



(C) R=50m



(D) R=60m

FIGURE 3.3: Cumulative distribution of maximal distances (in meters) between the stops in each cluster. R is the clustering radius, d is the maximal distance in the cluster.

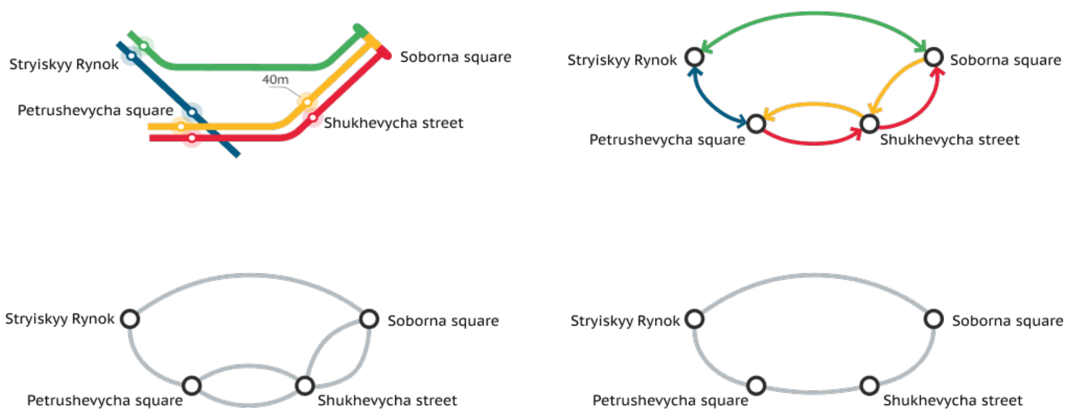


FIGURE 3.4: Network simplification process. Firstly, we do coarse-graining procedure for the stops that lay in radius R=40m from each other. Secondly, we reject directions. Thirdly, we reject parallel edges in the graph.

Chapter 4

Main network indicators of several PTNs

In this study, we use $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space topology to represent Lviv and Bristol PTNs as a complex network. In such way we can make conclusions about networks from different aspects: analyze connections between the stops ($\mathbb{L}$- and $\mathbb{P}$-space) and connections between the routes ($\mathbb{C}$-space). To explore the topology of a network we define its local and global characteristics (Eqs. 1.1 - 1.8).

4.1 Local network characteristics

Local network characteristics are determined by the immediate neighborhood of the nodes (von Ferber et al., 2009b). The local characteristics include the node connectivity, the number of neighbors of the nodes and the tendencies in building the connections between the nodes. The numerical values of local characteristics for PTNs in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space are listed in Table 4.1

City	$\langle k_{\mathbb{L}} \rangle$	$k_{max\mathbb{L}}$	$C_{\mathbb{L}}$	$C_{rand\mathbb{L}}$	$r_{\mathbb{L}}$	$\langle k_{\mathbb{P}} \rangle$	$k_{max\mathbb{P}}$	$C_{\mathbb{P}}$	$C_{rand\mathbb{P}}$	$r_{\mathbb{P}}$	$\langle k_{\mathbb{C}} \rangle$	$k_{max\mathbb{C}}$	$C_{\mathbb{C}}$	$C_{rand\mathbb{C}}$	$r_{\mathbb{C}}$
Lviv	2.558	10	0.047	0.004	-0.03	90.605	411	0.637	0.152	-0.06	38.805	65	0.745	0.511	-0.05
Bristol	3.372	25	0.104	0.003	0.31	100.483	620	0.622	0.089	-0.03	34.014	85	0.623	0.24	0.06

TABLE 4.1: PTNs local characteristics in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space. $\langle k \rangle$ and k_{max} - mean and maximal node degrees of a PTN, C and C_{rand} - clustering coefficient of a PTN and of a random graph of the same size, r - assortativity. The subscripts $\mathbb{L}$, $\mathbb{P}$, and $\mathbb{C}$ indicate the spaces.

4.1.1 Node degree

Node degree k is one of the characteristics that indicate the importance of the node in the network. The nodes with high node degree are called hubs. To analyze the overall network we use mean node degree $\langle k \rangle$ (Eq. 1.1). The analysis of PTNs showed that in $\mathbb{P}$- and $\mathbb{C}$-space mean node degrees are relatively high. In these representations the networks usually possess high connectivity. The values of $\langle k \rangle$ in $\mathbb{L}$-space are much lower: $\langle k \rangle = 2.558$ for Lviv PTN and $\langle k \rangle = 3.372$ for Bristol PTN. Mean node degrees in $\mathbb{L}$-space are usually close to 2. Such tendency can be explained by the common structure of the PTNs. The number of terminal stops with one connection as well as hubs is usually small. Most of the stations are intermediate stops with degree $k = 2$.

Additionally, we determine k_{max}, the maximal node degree in the network. Interestingly, the maximal node degrees of Bristol in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space are much higher than those of Lviv. The strongest difference occurs in $\mathbb{L}$-space: $k_{max} = 10$ for Lviv PTN and $k_{max} = 25$ for Bristol PTN. It means that the largest transport hub of Bristol has 2.5 times more connections than the one in Lviv.

Typical real-world networks usually have slow decaying node degree distributions $p(k)$. As von Ferber et al., 2009b state, these distributions obey power-law or exponential behavior. Power law decay is described by

$$p(k) \sim k^{-\gamma} \tag{4.1}$$

Exponentially decaying distributions can be defined as:

$$p(k) \sim \exp\left(-k/\hat{k}\right), \tag{4.2}$$

where $\hat{k}$ is the scale of the order of mean node degree. Such decays are also observed in the particular case of Lviv and Bristol. Figure 4.1 shows the node degree distributions for Bristol and Lviv PTNs in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space. Nota that the continuously decreasing curves represent cumulative distributions defined as

$$P(k) = \sum_{c=k}^{k_{max}} p(c) \tag{4.3}$$

As the plots show, in $\mathbb{L}$-space and $\mathbb{P}$-space in both PTNs the distributions possess exponentially decaying tails. The distribution for Lviv PTN in $\mathbb{C}$-space possesses an exponential decay as well. The distribution of Bristol PTN in $\mathbb{C}$-space (Figure 4.1d) possesses a large gradient, which could indicate the truncated power law or exponential decay.

4.1.2 Clustering coefficient

Clustering coefficient C of a network (Eq. 1.5) defines the average connectivity within the neighborhood of a graph node. Interestingly, the correlation between Lviv and Bristol clustering coefficients differs in each of the spaces. The value of C for Bristol PTN in $\mathbb{L}$-space is more than twice higher than that of Lviv PTN. In $\mathbb{P}$-space the values of C are almost equal, while in $\mathbb{C}$-space Lviv PTN possesses a higher clustering coefficient.

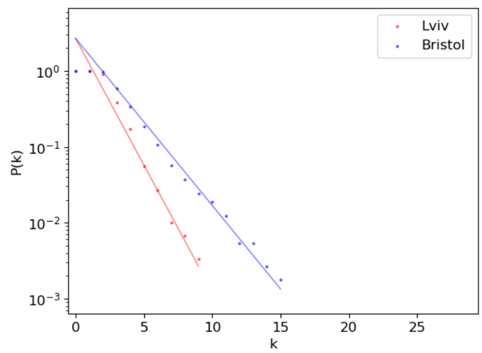
A useful characteristic is the relation c between the clustering coefficient of a graph and the clustering coefficient of a random graph of the same size:

$$c = \frac{C}{C_{rand}} \tag{4.4}$$

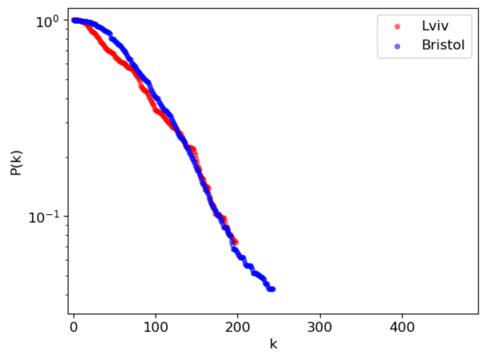
The values of c (Table 4.4) are derived considering the clustering coefficients presented in Table 4.1. Although the ratios for Bristol PTN representations are considerably higher than those for Lviv, both PTNs possess large values of c. These correlations can indicate the small-world structure of the PTNs.

City	$c_{\mathbb{L}}$	$c_{\mathbb{P}}$	$c_{\mathbb{C}}$
Bristol	34.67	6.99	2.6
Lviv	11.75	4.19	1.46

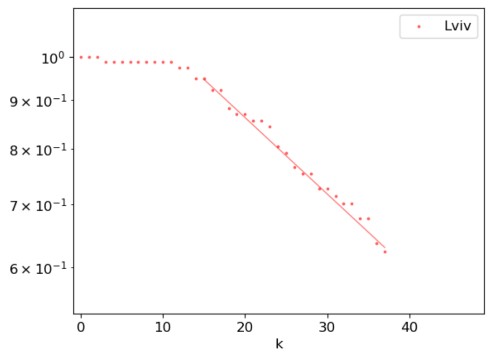
TABLE 4.2: The ratio of the mean clustering coefficient of the graphs to the clustering coefficient of the random graphs of the same size



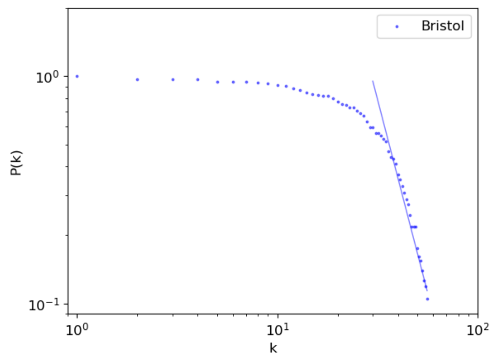
(A) Bristol and Lviv, $\mathbb{L}$-space. The lines show the fits to exponential function.



(B) Bristol and Lviv, $\mathbb{P}$-space.



(C) Lviv, $\mathbb{C}$-space. The line shows the fit to exponential function.



(D) Bristol, $\mathbb{C}$-space. The line shows the fit to power-law.

FIGURE 4.1: Cumulative node degree distributions

4.1.3 Assortativity

Assortativity r (Eq. 1.8) indicates the correlations between the node degrees of the neighboring nodes. The observed assortativity values show that only Bristol PTN in $\mathbb{L}$-space has a clear preference for assortative mixing ($r = 0.31$) meaning that links tend to connect nodes of similar degree. Lviv PTN in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space and Bristol PTN in $\mathbb{P}$- and $\mathbb{C}$-space possess very small negative or positive assortativity values. Thus, in these cases, there are no clear preferences in node linkages.

4.2 Global network characteristics

Global characteristics characterize a network as a whole. They include the shortest path length and betweenness centrality. The numerical values of global characteristics for PTNs in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space are listed in Table 4.3.

City	$\langle l_{\mathbb{L}} \rangle$	$l_{rand\mathbb{L}}$	$D_{\mathbb{L}}$	$\langle l_{\mathbb{P}} \rangle$	$l_{rand\mathbb{P}}$	$D_{\mathbb{P}}$	$\langle l_{\mathbb{C}} \rangle$	$l_{rand\mathbb{C}}$	$D_{\mathbb{C}}$
Bristol	12.281	5.186	37	2.088	1.736	5	1.854	1.854	4
Lviv	14.287	5.849	40	1.966	1.624	5	1.498	1.514	3

TABLE 4.3: PTNs global characteristics. $\langle l \rangle$ and $\langle l_{rand} \rangle$ - mean shortest path length of a PTN and of a random graph of equal size. $\langle D \rangle$ - diameter. The subscripts $\mathbb{L}$, $\mathbb{P}$, and $\mathbb{C}$ indicate the spaces.

4.2.1 Shortest paths

Shortest path l_{ij} measures the smallest number of edges between nodes i and j. For the whole network we calculate mean shortest path length $\langle l \rangle$ (Eq. 1.2). Shortest path length can be well defined only for the nodes that belong to the same connected component (von Ferber et al., 2009b). Thus, the further calculations related to shortest path length (mean shortest path, diameter, path length efficiency, betweenness centrality) will be calculated for GCC.

In $\mathbb{L}$-space shortest path length defines the smallest number of stops to pass from one station to another. The shortest path lengths of Lviv and Bristol are almost similar in each of the spaces.

Of particular interest are the shortest path values in $\mathbb{P}$- and $\mathbb{C}$-space. In $\mathbb{P}$-space shortest path is related to the number of transfers T between the stops. $l_{ij} = 1$ if stop v_i has at least one common route with stop v_j. In $\mathbb{C}$-space shortest path is related to the number of changes between the routes. $l_{ij} = 1$ if route v_i has at least one common stop with route v_j. Therefore, in $\mathbb{P}$- and $\mathbb{C}$-space the number of transfers between two nodes can be described as $T_{ij} = l_{ij} - 1$. Therefore, mean number of transfers can be defined as $\langle T \rangle = \langle l \rangle - 1$. In $\mathbb{P}$-space $\langle l \rangle \approx 2$ for both Lviv and Bristol PTNs. Considering this value, in both PTNs the average number of changes between any two stops is $\langle T \rangle \approx 1$. In $\mathbb{C}$-space $\langle l \rangle = 1.498$ and $\langle l \rangle = 1.854$ for Lviv and Bristol PTN, accordingly. Therefore, the number of changes between any two routes is $\langle T \rangle \approx 0.5$ in Lviv PTN and $\langle T \rangle \approx 0.9$ in Bristol PTN.

One can also compare two PTNs by their diameters D. The diameters of Lviv and Bristol are approximately the same in all three spaces.

To conclude, we have estimated path length efficiency l_η (Eq. 1.3). Path length efficiency is a useful criterion that indicates the ratio of the mean shortest path of a network to the mean shortest path of a random graph of the same size. The smaller l_η, the more efficient a network is. The results show that in terms of path lengths Bristol PTN is more efficient than Lviv PTN in $\mathbb{L}$- and $\mathbb{P}$-space, but less efficient in $\mathbb{C}$-space. All the l_η values are relatively small. Together with the high clustering coefficient values, it indicates that the observed PTNs are small-world networks.

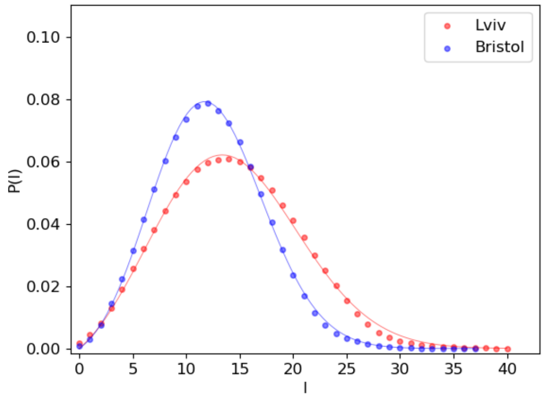
City	$l_{\eta\mathbb{L}}$	$l_{\eta\mathbb{P}}$	$l_{\eta\mathbb{C}}$
Bristol	2.368	1.203	1.224
Lviv	2.443	1.21	1.131

TABLE 4.4: The ratio of the mean clustering coefficient of a graph to the clustering coefficient of a random graph of the same size

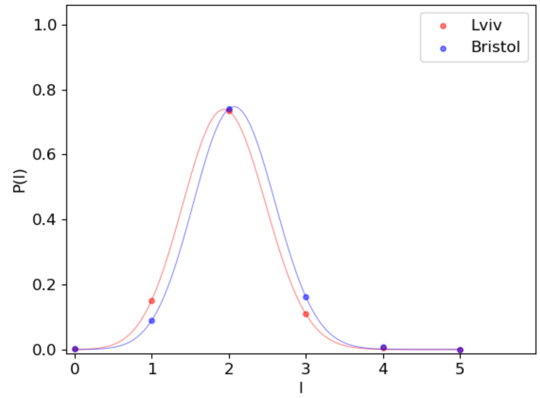
As in von Ferber et al., 2009b, the shortest path length distributions $\Pi(l)$ of the explored cities in $\mathbb{L}$- (Figure 4.2a), $\mathbb{P}$- (Figure 4.2b) and $\mathbb{C}$-space (Figure 4.2c) can be described by an asymmetric unimodal distribution:

$$\Pi(l) = Al \exp(-Bl^2 + Cl), \tag{4.5}$$

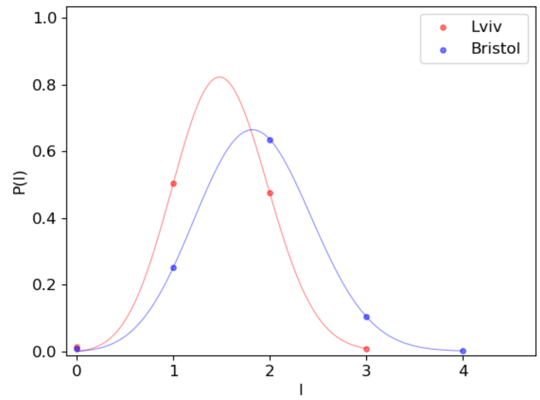
where A, B, and C are parameters.



(A) 𝕃-space



(B) ℙ-space



(C) ℂ-space

FIGURE 4.2: Shortest path lengths distributions. Solid lines represent the fits to Equation (4.5).

Chapter 5

Analysis of PTNs resilience

To be stable, a network must maintain its overall connectivity: there should exist a path between any two nodes. Generally, a network is considered functional, if it has a significantly large component that remains connected (Ferber et al., 2012).

Unpredictable events in some parts of a network can influence its operational properties and consequently harm connectivity. In critical cases, they can cause the overall network collapse. Such events include random failures in the network and targeted attacks. Random failures might be caused by car accidents, weather conditions, substantial traffic jams, etc. Targeted attacks include terrorist acts, strikes, etc. Targeted attacks usually occur at the most important parts of the network.

In our work we analyze resilience of the network in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-spaces. One of the measures of a network resilience is Molloy-Reed criterion (Eq. 1.9). It states that a network is stable if the value κ related to mean node degree and the average variance of node degree is $\kappa \geq 2$. The higher κ is, the more stable the network. This criterion is normally used to define stability of infinite uncorrelated networks under random failures. Although in our exploration the PTNs are finite and correlated networks, Molloy-Reed criterion can be nicely applied to determine their stability to random failures as well. The value of κ is the most useful for network comparison in $\mathbb{L}$-space, as the networks in $\mathbb{P}$- and $\mathbb{C}$-space are strongly connected. Their κ values are high by default and do not differ sharply. By Molloy-Reed criterion both Lviv and Bristol PTNs in $\mathbb{L}$-space are resilient to random failures, and the Bristol PTN seems to be more stable (Table 5.1).

City	$\kappa_{\mathbb{L}}$	$\kappa_{\mathbb{P}}$	$\kappa_{\mathbb{C}}$
Bristol	4.493	145.456	43.104
Lviv	3.099	138.413	44.245

TABLE 5.1: Molloy-Reed criterion k for PTNs in $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space

For assessing the network vulnerability under different scenarios, the attack simulations are often used. In such simulations, network behavior is analyzed under successive removal of its constituents. In our work, we consider attacks with node removal.

Attack simulations have different interpretations in each of the spaces. In $\mathbb{L}$-space deletion of a node usually causes disconnection of the routes that pass through it. In the real world it corresponds to the situation when the stop and the road near it are unavailable (e.g., due to the traffic jam) and one should choose another way to reach the destination. In $\mathbb{P}$-space, if a node is removed the other nodes on the same route are still connected. It can correspond to the situation when the transport still uses the same road, but one of the stops was canceled (Berche et al., 2009). In $\mathbb{C}$-space removal of a node corresponds to the removal of the whole route.

5.1 Attack simulation scenarios

The influence of the attack on the network depends on initial network characteristics and the attack scenario. There are many scenarios of nodes removal in the simulations. While in random attacks the nodes are removed at random, in targeted attacks they are removed according to the node lists sorted by decreasing node importance. The authors of other PTNs explorations defined the node importance in terms of node degrees, closeness, graph, stress and betweenness centralities, increasing clustering coefficient etc. (von Ferber, Holovatch, and Holovatch, 2009, Berche et al., 2010, Berche et al., 2012). In our work we examine nodes removal considering the three different scenarios:

- random node removal
- removal by decreasing node degree values
- removal by decreasing betweenness centrality values

Removal by degrees and removal by betweenness centrality have different meanings. The former type of attack aims to remove a maximal number of edges, while the latter aims to cut a maximal number of shortest paths (Berche et al., 2009).

On each step of a simulation, we delete 1% from the initial number of the nodes in the network. We repeat the procedure until network destruction. To assess the changes in the size of the network we observe the decay $S(c)$, where c is the ratio of the number of removed nodes to the overall number of nodes, and S (von Ferber, Holovatch, and Holovatch, 2009) is the normalized giant connected component (GCC) size. S is defined as follows

$$S = \frac{N_{GCC}}{N_{init}}, \tag{5.1}$$

where N_{GCC} is number of the nodes of GCC and N_{init} is initial number of the nodes in the network.

The failure of one node can cause changes in the properties of another node. The order of the nodes in the node degrees and betweenness centrality lists can change. Thus, for targeted attacks we consider two simulation modes:

- removal of the nodes from original sorted list
- removal of the nodes with recalculation of the sorted list after each simulation step

5.2 Numerical results

The results of simulations (Figure 5.1) showed that Bristol and Lviv PTNs' behavior under different attack scenarios is close to similar. One can see that the networks in $\mathbb{L}$-space (Figures 5.1a, 5.1b) deteriorate rapidly, while in $\mathbb{P}$-space (Figures 5.1c, 5.1d) and $\mathbb{C}$-space (Figures 5.1e, 5.1f) the size S of GCC under removal of fraction c of nodes decays slower. It can be explained by the higher connectivity of the networks in the latter two spaces.

To numerically compare the stability of PTNs under different attack scenarios, we have used the value of the area A under different $S(c)$ curves:

$$A = \int_0^1 S(c)dc \tag{5.2}$$

The area A captures the network reaction over the whole attack sequence (Berche et al., 2012). The higher is the value of A, the more robust is the network under a particular type of attack. The choice of such parameter to monitor network robustness under attack sequences was suggested by Schneider et al., 2011. The values of A are given in Table 5.2.

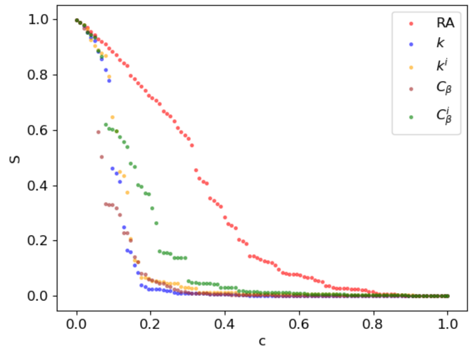
Space	City	RA	$\langle k^i \rangle$	k	C^i_β	C_β
$\mathbb{L}$-space	Bristol	0.304	0.125	0.109	0.159	0.095
	Lviv	0.234	0.087	0.075	0.159	0.059
$\mathbb{P}$-space	Bristol	0.498	0.438	0.439	0.416	0.31
	Lviv	0.497	0.423	0.403	0.4	0.321
$\mathbb{C}$-space	Bristol	0.481	0.432	0.404	0.395	0.343
	Lviv	0.498	0.47	0.464	0.465	0.426

TABLE 5.2: The area A (Eq. 5.2) under the $S(c)$ curves for different attack scenarios. $S(c)$ - the function of the the size S of GCC under removal of fraction c of nodes. RA - random attack, k - recalculated node degrees, k^i - initial node degrees, C_β - recalculated betweenness centrality, C^i_β - initial betweenness centrality.

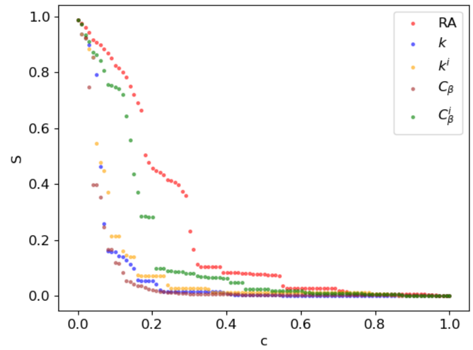
In all the explored representations the most inefficient for both PTNs is random attack scenario. It possesses the slowest decrease of S and, accordingly, the highest values of A. S in $\mathbb{P}$- and $\mathbb{C}$-space under random attack decreases linearly for both PTNs. The results of the simulations in $\mathbb{L}$-space confirmed the analysis of κ values presented earlier: Bristol PTN appeared to be a bit more stable to random attacks comparing to Lviv PTN. Note that the results of random attacks are represented by single random sequences. However, as in Ferber et al., 2012, we observed a "self-averaging" effect. Due to the large size of the PTNs, the averaging over many sequences of random attacks gives results almost identical to those presented on the plot.

The most harmful for PTNs in all three spaces is node removal by recalculated betweenness centrality.

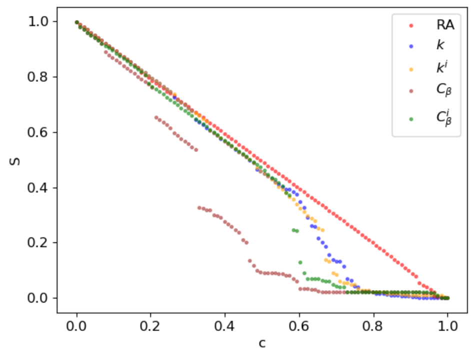
The attack simulation results also indicated that scenarios of targeted attacks based on recalculated lists are more harmful than the same scenarios based on initial lists.



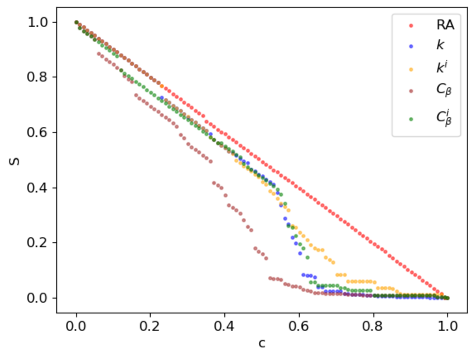
(A) Bristol, 𝕃-space



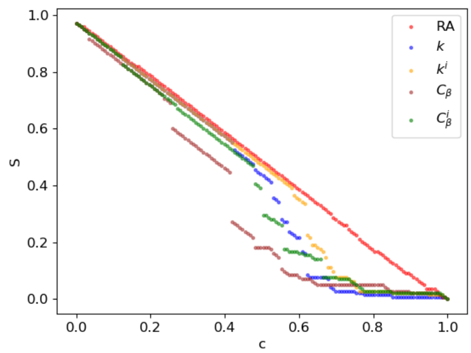
(B) Lviv, 𝕃-space



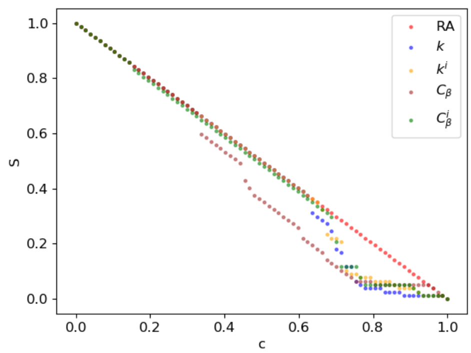
(C) Bristol, ℙ-space



(D) Lviv, ℙ-space



(E) Bristol, ℂ-space



(F) Lviv, ℂ-space

FIGURE 5.1: Attacks on PTNs in 𝕃-, ℙ- and ℂ-space. *RA* - random attack, k - recalculated node degrees, k^i - initial node degrees, C_β - recalculated betweenness centrality, C_β^i - initial betweenness centrality.

Chapter 6

Conclusions

This analysis of public transport networks was driven by two main goals. Firstly, we wanted to define the main statistical properties of Lviv and Bristol PTNs and compare them. Secondly, we aimed to assess the vulnerability of both PTNs to unpredictable events, namely, to random and targeted attacks. To achieve this goal we conducted a set of simulations of various attack scenarios.

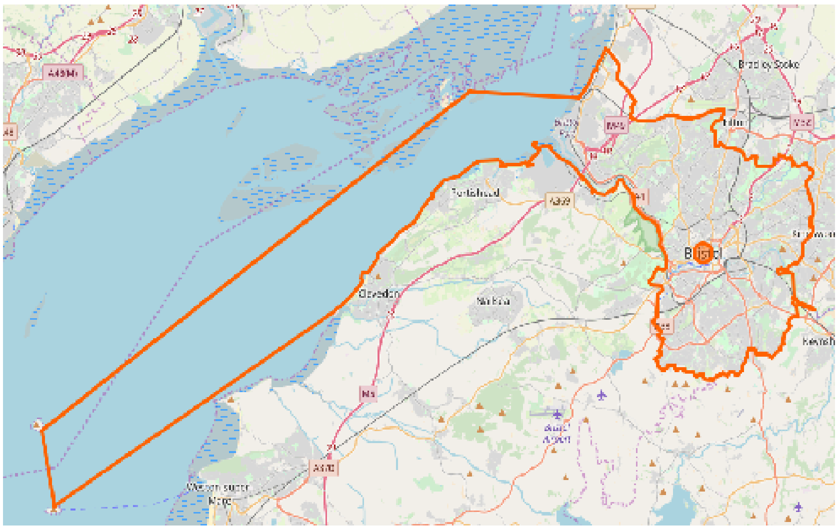
The results of the analysis showed that Lviv and Bristol PTNs share a lot of common features. The observed patterns of the PTNs behavior under attack simulations give strong evidence of the similarity of the networks in terms of stability. Both Lviv and Bristol PTNs are highly resilient to random failures of the nodes. The most dangerous for the networks are attacks (with recalculation) at the nodes with the highest betweenness centrality. Most of the global characteristics of the two PTNs, such as path length efficiency, diameters, and mean shortest paths in some of the spaces have no sharp differences. Furthermore, Lviv and Bristol PTNs appear to be small-world networks as they possess relatively low path lengths values and large clustering coefficients. Also, all the path length distributions can be described by an asymmetric unimodal function. Most of the node degree distributions possess exponential tail.

However, there are considerable diversities in the local network characteristics. Namely, Bristol PTN possesses higher values of mean and maximal node degrees. The most substantial difference is observed in k_{max} values in $\mathbb{L}$-space: k_{max} of Bristol PTN is 2.5 times higher than the of Lviv PTN.

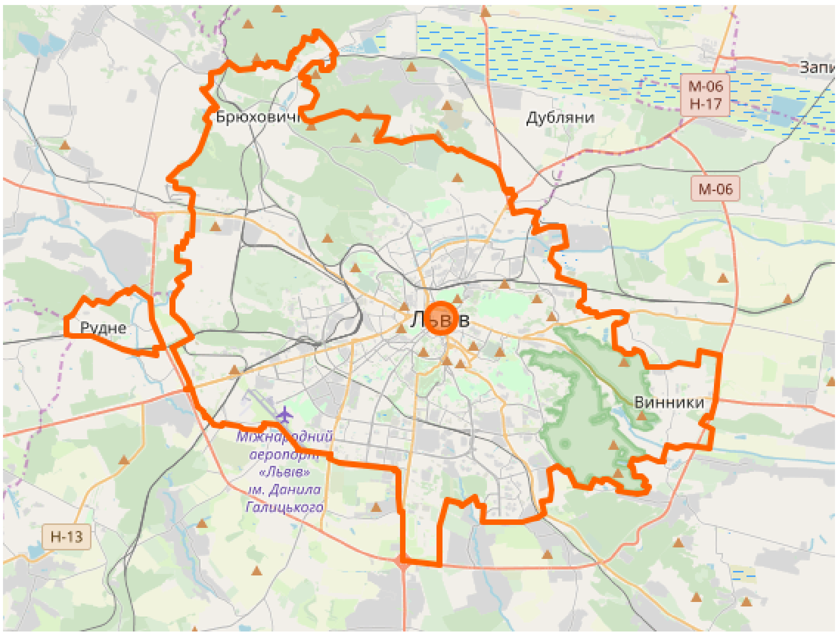
The future research on the properties of PTNs might be extended by investigation of the dynamic features of the networks. These include analysis of passenger and transport load, changes of average commuting times during a sample period, etc. Besides that, the load capacity of the nodes might be considered in further explorations of resilience. After the attack, the load redistributes and exceeds the capacity limits at some of the nodes. Therefore, cascading failure simulations might prove an important area for future research.

Appendix A

Maps

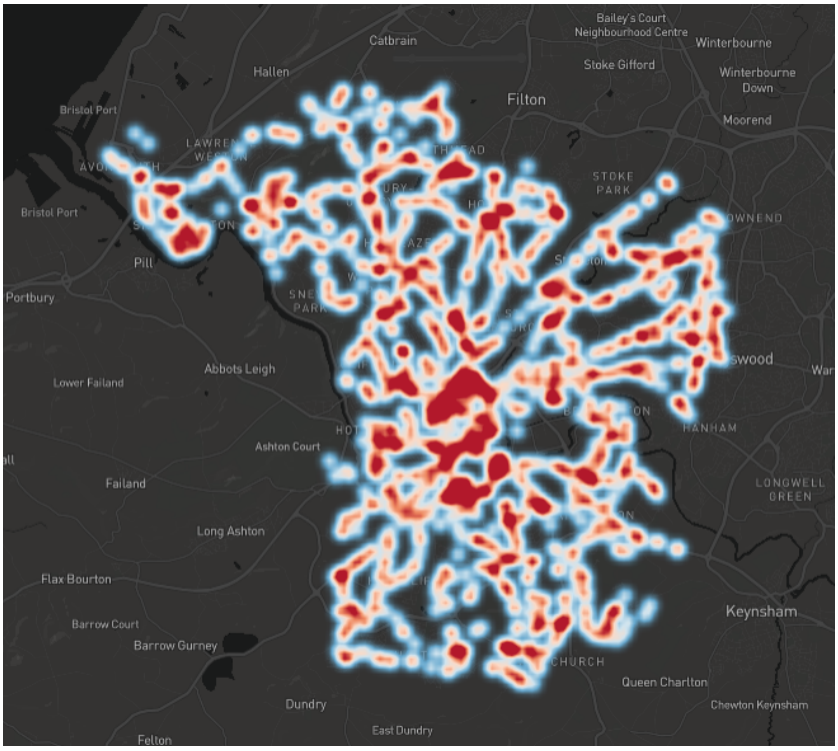


(A) Bristol



(B) Lviv

FIGURE A.1: City boundaries used to crop the datasets



(A) Bristol

(B) Lviv

FIGURE A.2: The Heatmaps of PTNs

Bibliography

Albert, Réka, István Albert, and Gary L Nakarado (2004). "Structural vulnerability of the North American power grid". In: *Physical review E* 69.2, p. 025103.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási (2000). "Error and attack tolerance of complex networks". In: *nature* 406.6794, p. 378.

Amaral, Luıs A Nunes et al. (2000). "Classes of small-world networks". In: *Proceedings of the national academy of sciences* 97.21, pp. 11149–11152.

ATCO-CIF parser. https://github.com/davidjrice/atco. Accessed: 2019-04-30.

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *science* 286.5439, pp. 509–512.

Barthélemy, Marc and Alessandro Flammini (2006). "Optimal traffic networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2006.07, p. L07002.

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

Berche, Bertrand et al. (2010). "Public transport networks under random failure and directed attack". In: *arXiv preprint arXiv:1002.2300*.

Berche, Bertrand et al. (2012). "Transportation network stability: a case study of city transit". In: *Advances in Complex Systems* 15.supp01, p. 1250063.

Candelieri, Antonio et al. (2019). "Vulnerability of public transportation networks against directed attacks and cascading failures". In: *Public Transport*, pp. 1–23.

Cats, Oded and Erik Jenelius (2018). "Beyond a complete failure: the impact of partial capacity degradation on public transport network vulnerability". In: *Transportmetrica B* 6, pp. 77–96. DOI: 10.1080/21680566.2016.1267596.

Chang, Hui et al. (2007). "Assortativity and act degree distribution of some collaboration networks". In: *Physica A: Statistical Mechanics and its Applications* 383.2, pp. 687–702.

Cohen, Reuven et al. (2000). "Resilience of the internet to random breakdowns". In: *Physical review letters* 85.21, p. 4626.

Crucitti, Paolo, Vito Latora, and Massimo Marchiori (2004). "A topological analysis of the Italian electric power grid". In: *Physica A: Statistical mechanics and its applications* 338.1-2, pp. 92–97.

de Regt, Robin (2018). "Complex Networks: topology, shape and spatial embedding". PhD thesis. Coventry University.

de Regt, Robin et al. (2018). "Public transportation in Great Britain viewed as a complex network". In: *Transportmetrica A: Transport Science*, pp. 1–27.

Dorogovtsev, Sergei N and José FF Mendes (2013). *Evolution of networks: From biological nets to the Internet and WWW*. OUP Oxford.

Ester, Martin et al. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *Kdd*. Vol. 96. 34, pp. 226–231.

Ferber, Christian von et al. (2012). "A tale of two cities". In: *Journal of Transportation Security* 5.3, pp. 199–216.

Freeman, Richard (2011). "Death of Detroit: Harbinger of collapse of deindustrialized America". In:

Gallotti, Riccardo and Marc Barthelemy (2015). "The multilayer temporal network of public transport in Great Britain". In: *Scientific data* 2, p. 140056.

Gastner, Michael T and MEJ Newman (2004). "Shape and efficiency in spatial distribution networks". In: *arXiv preprint cond-mat/0409702*.

Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport (2008). "Why do the poor live in cities? The role of public transportation". In: *Journal of urban Economics* 63.1, pp. 1–24.

Granovetter, Mark S (1977). "The strength of weak ties". In: *Social networks*. Elsevier, pp. 347–367.

GTFS Static Overview. https://developers.google.com/transit/gtfs/. Accessed: 2019-04-30.

Guida, Michele and Funaro Maria (2007). "Topology of the Italian airport network: A scale-free small-world network with a fractal structure?" In: *Chaos, Solitons & Fractals* 31.3, pp. 527–536.

Guimera, Roger and Luıs A Nunes Amaral (2004). "Modeling the world-wide airport network". In: *The European Physical Journal B* 38.2, pp. 381–385.

Guimera, Roger et al. (2005). "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles". In: *Proceedings of the National Academy of Sciences* 102.22, pp. 7794–7799.

Jenelius, Erik and Oded Cats (2015). "The value of new public transport links for network robustness and redundancy". In: *Transportmetrica A: Transport Science* 11.9, pp. 819–835.

Jeong, Hawoong et al. (2001). "Lethality and centrality in protein networks". In: *Nature* 411.6833, p. 41.

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

Li, Wei et al. (2006). "How to fit the degree distribution of the air network?" In: *Physica A: Statistical Mechanics and its Applications* 368.1, pp. 262–272.

Molloy, Michael and Bruce Reed (1995). "A critical point for random graphs with a given degree sequence". In: *Random structures & algorithms* 6.2-3, pp. 161–180.

National Transport Data Repository. URL: http://data.gov.uk/dataset/nptdr.

Newman, M. E. J., S. H. Strogatz, and D. J. Watts (2001). "Random graphs with arbitrary degree distributions and their applications". In: *Phys. Rev. E* 64 (2), p. 026118. DOI: 10.1103/PhysRevE.64.026118. URL: https://link.aps.org/doi/10.1103/PhysRevE.64.026118.

Original Datasets, GitHub. URL: https://github.com/yarynakorduba/transport-networks-analysis/tree/master/data/initial_datasets.

Pastor-Satorras, Romualdo and Alessandro Vespignani (2007). *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press.

Rodríguez-Núñez, Eduardo and Juan Carlos García-Palomares (2014). "Measuring the vulnerability of public transport networks". In: *Journal of transport geography* 35, pp. 50–63.

Schneider, Christian M et al. (2011). "Mitigation of malicious attacks on networks". In: *Proceedings of the National Academy of Sciences* 108.10, pp. 3838–3841.

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

Sole, Ricard V and Mª Montoya (2001). "Complexity and fragility in ecological networks". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1480, pp. 2039–2045.

von Ferber, C, Yu Holovatch, and V Palchykov (2005). “Scaling in public transport networks”. In: *arXiv preprint cond-mat/0501296*.

von Ferber, Christian, Taras Holovatch, and Yurij Holovatch (2009). “Attack vulnerability of public transport networks”. In: *Traffic and Granular Flow’07*. Springer, pp. 721–731.

von Ferber, Christian et al. (2007). “Network harness: Metropolis public transport”. In: *Physica A: Statistical Mechanics and its Applications* 380, pp. 585–591.

von Ferber, Christian et al. (2009a). “Modeling metropolis public transport”. In: *Traffic and Granular Flow’07*. Springer, pp. 709–719.

von Ferber, Christian et al. (2009b). “Public transport networks: empirical analysis and modeling”. In: *The European Physical Journal B* 68.2, pp. 261–275.

Wasserman, Stanley and Katherine Faust (1994). *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.

Watts, Duncan J and Steven H Strogatz (1998). “Collective dynamics of ’small-world’ networks”. In: *nature* 393.6684, p. 440.

Yi, Chengqi et al. (2015). “Modeling cascading failures with the crisis of trust in social networks”. In: *Physica A: Statistical Mechanics and its Applications* 436, pp. 256–271.

Zhang, Lin, Bai-bai Fu, and Yun-xuan Li (2016). “Cascading failure of urban weighted public transit network under single station happening emergency”. In: *Procedia engineering* 137, pp. 259–266.

Zhang, Lin et al. (2018). “A cascading failures perspective based mesoscopic reliability model of weighted public transit network considering congestion effect and user equilibrium evacuation”. In: *Mathematical Problems in Engineering* 2018.

Zhen-Tao, Zhu et al. (2008). “An evolutionary model of urban bus transport network based on B-space”. In: *Chinese physics B* 17.8, p. 2874.