HAL
open science

Complex transportation networks: resilience, modelling and optimization

Taras Holovatch

To cite this version:

Taras Holovatch. Complex transportation networks: resilience, modelling and optimization. Physics and Society [physics.soc-ph]. Université Henri Poincaré - Nancy 1, 2011. English. NNT : 2011NAN10090 . tel-01746249v2

HAL Id: tel-01746249

https://theses.hal.science/tel-01746249v2

Submitted on 16 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nancy-Université

Faculté des Sciences & Techniques
École Doctorale Energie Mécanique et Matériaux (EMMA)
Formation Doctorale Physique et Chimie de la Matière et des Matériaux (PCMM)

Coventry University

Faculty of Engineering and Computing
Applied Mathematics Research Centre

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy I

en Sciences Physiques

par **Taras Holovatch**

Réseaux de Transport Complexes: Résilience, Modélisation et Optimisation

Complex Transportation Networks: Resilience, Modelling and Optimisation

Soutenance publique effectuée le 19 octobre 2011

Membres du Jury:

Rapporteurs :	Serge Galam	Directeur de Recherche CNRS, Ecole Polytechnique
	Colm Connaughton	Professeur, Université de Warwick, Angleterre
Examinateurs :	Simon Bell	Professeur, Université de Coventry, Angleterre
	Christophe Chatelain	Maître de conférences, UHP, Nancy I
Co-directeurs :	Bertrand Berche	Professeur, UHP, Nancy I
	Christian von Ferber	Professeur, Université de Coventry, Angleterre
Invité :	Robert Low	Professeur, Université de Coventry, Angleterre
	Ralph Kenna	Professeur, Université de Coventry, Angleterre

Institut Jean Lamour - Statistical Physics Group
Faculté des Sciences & Techniques - 54500 Vandœuvre-lès-Nancy

Complex Transportation Networks: Resilience, Modelling and Optimisation

by

Taras Holovatch

A thesis submitted for the award of Doctor of Philosophy

Applied Mathematics Research Centre, Coventry University (UK)
Faculty of Engineering and Computing

Institut Jean Lamour, UMR 7198, CNRS - Nancy-Université (France)
Ecole Doctorale Energie Mécanique et Matériaux

June 2011

Acknowledgements

I want to say "Thank you" to all those people who happened to be in any connection to me and my work during the time of my PhD course in the Applied Mathematics Research Centre, Coventry University and in the Jean Lamour Institute, Nancy University. They were always ready to understand and to help, to give an attention or an advice. And surely I wouldn't manage to solve all of those amazing bureaucratic questions without their help. Not less important was warm personal treatment, and friendly atmosphere in any situation.

And of course there are two people without whom this work would not appear at all. I am grateful for the invaluable guidance and encouragement throughout the period of this work offered by my scientific supervisors Bertrand Berche and Christian von Ferber. I was very lucky to get an opportunity to work with them.

I also want to thank to all people who showed any interest to my work and come with their questions, remarks and suggestions, which very often were useful.

The financial support by the Ecole Doctorale EMMA and AMR Centre of Coventry University which enabled me to study for the Degree of PhD is especially acknowledged.

Last but not least I am sincerely thankful to my family and friends for their support throughout my studies. And simply for being such important people for me.

Contents

Introduction **7**

1 Review of previous work **11**
1.1 Emergence of complex network science 11
1.1.1 How all began 11
1.1.2 Network characteristics 14
1.2 Complex network models 19
1.2.1 Erdös-Rényi random graph 19
1.2.2 Watts-Strogatz small-world model 20
1.2.3 Barabási-Albert scale-free network 21
1.3 Previous studies of public transport networks 22
1.4 Network attack vulnerability and percolation phenomenon 23
1.5 Optimisation 25

2 Empirical analysis **27**
2.1 Description of the database 27
2.2 PT network topology 28
2.3 Local network properties 32
2.3.1 Neighborhood size (node degree) 32
2.3.2 Clustering 37
2.3.3 Generalized assortativities 38
2.4 Global characteristics 40
2.4.1 Shortest paths 40
2.4.2 Betweenness centrality 43
2.4.3 Harness 46
2.4.4 Geographical embedding 49
2.5 Conclusions 51

3 Public transport network models **53**
3.1 Mutually interacting self-avoiding walks in 2d 53
3.1.1 Motivation and description of the model 53
3.1.2 Global topology of model PTN 55
3.1.3 Statistical characteristics of model PTN 57
3.2 Modelling in 1d: analytic results and simulation 61
3.3 Non-interacting walks in 2d 64
3.4 Conclusions 66

4 Public transport network vulnerability and resilience **69**
4.1 Observables and attack strategies 69
4.2 Results in $\mathbb{L}$-space 73
4.2.1 Choice of the 'order-parameter' variable 74
4.2.2 Segmentation concentration 75
4.2.3 Numerical estimates 76
4.2.4 Correlations 77
4.2.5 Comparison of node- and link-targeted attacks 82
4.3 Results in $\mathbb{P}$-space 87
4.3.1 Numerical estimates 88
4.3.2 Correlations 91
4.4 Conclusions 92

5 Optimisation of public transport networks **95**
5.1 Hail and Ride 96
5.2 Hail and Ride with complications 101
5.3 Routes with regular stations 104
5.4 Conclusions 109

Conclusions **111**

Bibliography **113**

Introduction

The present thesis is devoted to an application of the ideas of complex networks theory for analysing, modelling, and, finally, optimising different processes that occur in transportation networks. In particular we will be primarily concerned with the public transport networks, understanding them as the assembly of all means of public transport in a given city. Although studies of such networks have a long standing tradition and are the subject of different applied mathematical and technological disciplines, the novelty of our approach is that for the first time a complex network theory is used for such analysis.

A network is a set of items which we will call nodes with connections between them, which we will call links. Network is a central notion of our time and the explosion of interest to networks is a social and cultural phenomenon which arrived at the end of last century [3, 39, 92, 40, 116, 21, 10, 117]. In mathematical literature the term 'graph' is used and the graph theory constitutes a part of discrete mathematics [17]. A typical example of graph is given in Fig. 1.

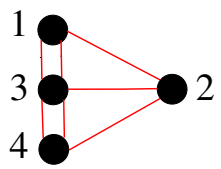


Figure 1: A graph of $N = 4$ *vertices* that are connected by $M = 7$ *edges*. Vertices 1, 2, and 4 have degree three ($k_1 = k_2 = k_4 = 3$), vertex 3 has degree five ($k_3 = 5$).

Systems, which have the form of network are numerous: these are the internet, www, neural networks, metabolic networks, transportation networks, wood webs, distribution networks (such as blood vessels or postal delivery routes), social networks of communication between people, networks of citation between papers

and many, many others. Physicists started the empirical and theoretical analysis of networks only very recently, seminal papers are dated by late 1990th. From an analysis of single small graphs and properties of individual vertices or edges (see Fig. 1.1) the task of the research shifted to consideration of statistical properties of graphs (networks). This change in the task caused also a change in the way of analysis. The breakthrough and the ”birth of network science” occurred due to new technologies, both are due to computers: www allowed for comparatively easy access to databases on different networks whereas computer power allowed for their detailed statistical analysis (which would have been simply impossible without computers for majority of networks of interest). To denote such objects and the type of question one is interested in the term *complex networks* often is used.

It appeared [45, 2, 70, 80], that the most important natural and man-made networks have a special structure, which is characterized by a fat-tail distribution of the number of node links and drastically differs from the classical random graphs, studied before by mathematicians [43, 16]. As a rule, these networks are not static, but they evolve and one cannot understand their structure without understanding the principles of their evolution.

The main motivation of our research was an expectation that applying concepts and ideas of complex network science to the public transport networks will result in a better understanding of their structure, their functioning, in particular their robustness to targeted attacks and random failures. In turn this will allow for an effective modelling of this networks and their optimisations. A certain novelty of our studies is that for the first time we have analysed a public transportation system of a city as a whole (previously only certain sub-networks of public transport were considered [76, 73, 100]); another particular feature of an empirical part of our analysis is considerably large database (before much smaller cities were considered [103, 104]; last but not least one should mention the specific features of public transport networks that were for the first time analysed in our studies (in particular, we were interested in their vulnerability and resilience under attacks, in their specific features such as generalized assortativities, centralities, harness - see below for more details and definitions). Such a comprehensive analysis allowed us to offer public transport networks models: such networks were never modeled before our study. Moreover, the majority of current models of complex networks consider network growth in terms of nodes, the novelty of one of our models consists in its growth in terms of chains. The main results of the thesis are published in: [67, 48, 49, 50, 13, 51, 52, 12, 14]. They were reported at the following meetings: COST ACTION P-10 Physics of Risk (Vilnius, Lithuania,

13-16 May 2006); MECO32 Conference of the Middle European Cooperation in Statistical Physics, (Ladek Zdroj, Poland, 16-18 April 2007); ANet07 Conference on Applications of Networks (Krakow, Poland, 1-5 Nov 2007); MECO33 Conference of the Middle European Cooperation in Statistical Physics(Wels, Austria, 13-17 Apr 2008); AGSOE, DY, DPG Meeting (Dresden, Germany, 23-27 Mar 2009); Statistical Physics and Low Dimensional Systems (Nancy, France, 13-15 May 2009); Statistical Physics: Modern Trends and Applications (Lviv, Ukraine, 23-25 June 2009); MECO35 Conference of the Middle European Cooperation in Statistical Physics(Pont-´-Mousson, France, 15-19 Apr 2010); AGSOE, DY, DPG Meeting (Regensburg, Germany, 21 - 26 Mar 2010); MECO36 Conference of the Middle European Cooperation in Statistical Physics(Lviv, Ukraine, 5-7 Apr 2011)

The set-up of the thesis is the following. In the chapter 1 we give a sketch of the evolution of the networks science, introduce some complex network models, review some of the previous studies about network vulnerability, about public transport networks and their optimisation. Chapter 2 is devoted to an empirical analysis of public transport networks. There, we will introduce different representations for the networks and define different observables in terms of which an analysis will be performed. Different public transport network models will be introduced and analysed in Chapter 3. Some of them will allow for analytic solutions, the other will be considered by numerical simulations. We will show that such models correctly describe essential features of the public transport networks. In Chapter 4 we present results about public transport network vulnerability and resilience. In particular, this will allow to elaborate criteria to determine network stability prior to an attack as well as will allow us to find certain correlation between the theoretical predictions for idealized networks with data for the real-world networks. Chapter 5 deals with network optimisation, conclusions and outlook are collected in the last chapter.

Chapter 1

Review of previous work

In this chapter we will give a brief sketch of the previous work relevant for our studies. In particular we will shortly describe how the science of complex networks emerged, introduce some complex network models, review studies of public transport networks by means of complex networks theory and some of the previous studies about network vulnerability and their optimisation. Some of the features presented below were the subject of review papers [67, 49].

1.1 Emergence of complex network science

1.1.1 How all began

Very often the starting point of the graph theory is attributed to Leonard Euler because of his famous solution of the Seven bridges problem in Königsberg (or rather by proving it to be unsolvable). The graph shown in Fig. 1 is the oldest one in graph theory: its edges correspond to bridges which joined Kneipenhoff island in Königsberg with the mainland. Leonard Euler translated the problem of finding a continuous nonintersecing path along all bridges onto the graph in Fig. 1 and gave rise to the graph theory. Sometimes the beginning of the theory is attributed to Francis Guthrie, who, colouring a map of the counties of England, posed the four color problem which asks if it is possible to color, using only four colors, any map of countries in such a way as to prevent two bordering countries from having the same color. Trying to solve this problem, mathematicians invented many fundamental graph theoretic terms and concepts. Many outstanding mathematicians contributed to the field which resulted in the creation of the graph theory, which

Figure 1.1: Leonard Euler (1707 – 1783) [122]. In 1736 he published a paper *Solutio problematis ad geometriam situs pertinentis* which was the earliest application of graph theory in topology.

is one of the pillars of discrete mathematics [17].

Perhaps the first physical application of graph ideas was discovered by Gustav Kirchhoff with the circuit laws for calculating voltage and current in electric circuits. In XXth century, graph theory became widely applied in many different fields. An obvious example being sociology, where empirical studies belong to the basic tools to establish interrelations between members of a society. The representation of these relations in the form of graphs enables to quantify them. Two spectacular examples of such studies have been performed in the late 60ies and early 70ies of the last century. In 1967 S. Milgram published the results of an experiment in which he sent small packets to supposedly randomly selected individuals in the USA with the task of passing it on by mail to an aquaintance thought to be nearer to a given target person such that the packet will finally reach that target through a chain of aquaintances. From the history of those few packets that reached their target Milgram concluded that on average there is a distance of six steps (six intermediate acquaintances) between any two members of society. This phenomenon known also as ”six degrees of separation” was a precursor of the small world effect discovered for many networks afterwards [117] (see section 1.2.2). The second example is given by a study performed by M. Granovetter, who introduced the concept of ’strong and weak ties’ in social networks. His claim of the importance of weak ties is based on a study in which he inquired ’white collar’ workers about how and through whom they found their job. The paper in which he published the results of his thesis that mainly the weak ties (side contacts in the

Figure 1.2: Francis Guthrie (1831 – 1899) [123]. In 1852 he formulated the Four Colour Problem, which remained one of the most famous unsolved problems in mathematics for more than a century, until it was eventually proven in 1976 using a controversial computer-aided proof.

network of acquaintances) allow to find a new job, has become one of the most cited papers in economic sociology [59]. Being applied by sociologists, the graph theory also evolved due to their contributions and many terms in the theory stem from their sociological applications (see e.g. 'betweenness', explained below).

The object which was intensively studied in graph theory and which is directly related to complex networks is the classical random graph, called also the Erdös-Rényi random graph. It was suggested and studied at the end of 1950ies by Paul Erdös and Alfréd Rényi [43] (see also [16]). We will give the definition of this graph and discuss some of its properties in section 1.2.1. Here we just mention, that it consists of N vertices which are randomly connected by M edges and that random graph theory studies properties of such graphs which arise in the limit $N \to \infty$.

Subsequently, graph theory was used and evolved in the frames of informatics, cybernetics and biology. Physicists appeared on the stage only at the end of 1990, when many real world and man-made networks were analysed [45, 2, 70, 80] and it appeared that their properties have nothing to do with the properties of classical random graphs, which was almost the only object of random graph theory for almost 40 years! The complex networks were found to be small worlds with short distance between nodes, high level of correlations and self-organization. They demonstrated extremely high robustness if their nodes were removed at random [30, 95], however one observed their vulnerability to targeted attacks. Certain of

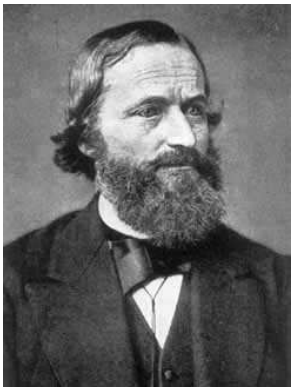


Figure 1.3: Gustav Robert Kirchhoff (1824 – 1887) [124]. In 1845 he published circuit laws (Kirchhoff's laws) starting application of graphs in physics.

their properties were governed by scaling laws, which would signal non-trivial correlations present in their structure [7]. In the rest of this section we present the main observables which are typically used to quantify network behaviour.

1.1.2 Network characteristics

We will use in the rest of this thesis the terms *node* and *link* both when speaking about simple graphs and their complex ensembles, networks. The node degree tells how many links are attached to the node (see Fig. 1). Links may be undirected, as in Fig. 1 or directed (coming in or out of the node, then one speaks about in-degree and out-degree, correspondingly). The complete information about a network is contained in its *adjacency matrix* $\hat{\mathrm{A}}$. For a network of N nodes (here, we will be mainly speaking about a network which does not contain multiple links or loops), $\hat{\mathrm{A}}$ is a square $N \times N$ matrix, with elements a_{ij} equal 1 if there is a link from node i to node j or zero, otherwise. For undirected networks $a_{ij} = a_{ji}$ and $a_{ii} = 0$. Then, for the degree k_i of the node i one gets:

$$k_i = \sum_j a_{ij}. \tag{1.1}$$

Here, and below the sum spans all N nodes of the network.

To give a measure of the 'linear size' of a network, the notions of the mean, $\langle \ell \rangle$ and maximal, $\ell_{\max}$, shortest path are useful. For a connected network of N

Figure 1.4: Paul Erdös (1913 – 1996) [125]. With about 500 coauthors he provides a perfect example of a collaboration network. It is estimated that 90% of the world's active mathematicians have an Erdös number (distance from him in coauthorship) smaller than 8.

nodes, the mean shortest path is defined by

$$\langle \ell \rangle = \frac{2}{N(N-1)} \sum_{i>j} \ell_{i,j}, \tag{1.2}$$

where $\ell_{i,j}$ is the length of the shortest path between nodes $\ell_{i,j}$. Here and below, we refer to Fig. 1.6, where examples are given. Correspondingly, the maximal shortest path is the largest one of all $\ell_{i,j}$ for a given network. Note that the length of the shortest path ℓ between nodes i and j equals to the minimal power of the adjacency matrix with a non-zero $\{ij\}$ element [17, 40]:

$$for\ all\ m < l \qquad (\hat{\mathrm{A}}^m)_{ij} = 0\,, \quad (\hat{\mathrm{A}}^\ell)_{ij} \neq 0\,, \tag{1.3}$$

and non-zero element $(\hat{\mathrm{A}}^\ell)_{ij}$ is the number of paths of length ℓ between i and j.

Whereas the mean shortest path length is a parameter characterising the network as a whole and is a global characteristic, the clustering coefficient defined below is a local value, it characterises a given node. For a node m, the clustering coefficient C_m is a relation between the actual number of links between adjacent nodes, E_m, and the maximal possible number of such links (see Fig. 1.6):

$$C_m = \frac{2E_m}{k_m(k_m-1)}, \qquad C_m \leq 1. \tag{1.4}$$

To derive (1.4), note that for the node of degree k_m the maximal number of links between its nearest neighbours is $k_m(k_m-1)/2$. The clustering coefficient of a

Figure 1.5: Alfréd Rényi ((1921 – 1970) [126]. Together with Paul Erdös he suggested in 1959 a classical random graph.

network C is defined as an average of all C_m of constituting nodes. Again, C can be calculated via the adjacency matrix as [40]:

$$C = \frac{1}{9}\frac{\sum_i(\hat{\mathrm{A}}^3)_{ii}}{\sum_{i\neq j}(\hat{\mathrm{A}}^2)_{ij}}. \tag{1.5}$$

The clustering coefficient indicates, how many of the nearest neighbors of a given node are also nearest neighbours of each other. It quantifies a tendency of cliques (a groups of interconnected nodes) to be formed. From the definition (1.5) it follows, that the clustering coefficient of any node of a tree, that is of a graph without loops of any length, is zero. On the other hand, the clustering coefficient of each node of a fully connected network (a complete graph) is equal to one. C gives the probability that there is a link between two randomly chosen nearest neighbours of a given node, therefore it contains information about loops of length three present in a network. Presence of loops is a specific form of correlations in networks. Typically, real-world networks are highly correlated and often possess values of the mean clustering coefficient close to 1.

One more characteristic of a node is its betweenness centrality. It shows how important is the node to maintain connections in the network and tells how many shortest paths go through a given node. This notion was first introduced in sociology, where individuals (nodes) with higher betweenness centrality possess a central role in the communication between the other nodes of the graph. Between-

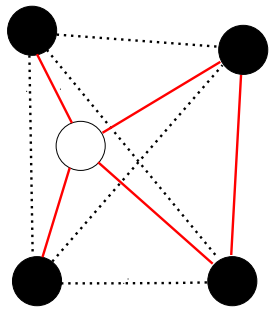


Figure 1.6: A graph of $N = 5$ nodes and $M = 5$ links (shown by red solid lines). It is characterized by the maximal shortest path length $l_{\max} = 2$ and the mean shortest path length $\langle \ell \rangle = 3/2$. To calculate the clustering coefficient of the node m (an open circle, $k_m = 4$) we divide the actual number of links between its nearest neighbours, $E_m = 1$, by the maximal possible number of links between them $k_m(k_m - 1)/2$ (other possible links are shown by dashed lines) and get $C_m = 1/6$. Applying equation 1.6 it is straightforward to show that the betweenness centrality of the node m equals $\mathcal{C}_B(m) = 5$

ness centrality $\mathcal{C}_B(m)$ of a node m is defined as

$$\mathcal{C}_B(m) = \sum_{i \neq m \neq j} \frac{\sigma(i,m,j)}{\sigma(i,j)}, \tag{1.6}$$

where $\sigma(i,j)$ is a the number of shortest paths between nodes i and j, and $\sigma(i,m,j)$ is the number of shortest paths between i and j that go through node m. $\mathcal{C}_B(m)$ is also called 'load' or 'betweenness' of a node.

A central notion in network theory is the node degree distribution $P(k)$. It defines the probability that a given node i has degree $k_i = k$. As has been worked out recently, networks which are characterized by different $P(k)$ demonstrate very different behaviour, a situation which might resemble to the different universality classes in the theory of critical phenomena [108, 37]. Several examples of node degree distributions, most frequently accounted, are shown in Fig. 1.7. These are: the Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \tag{1.7}$$

an exponential distribution:

$$P(k) \sim e^{-k/\langle k \rangle}, \tag{1.8}$$

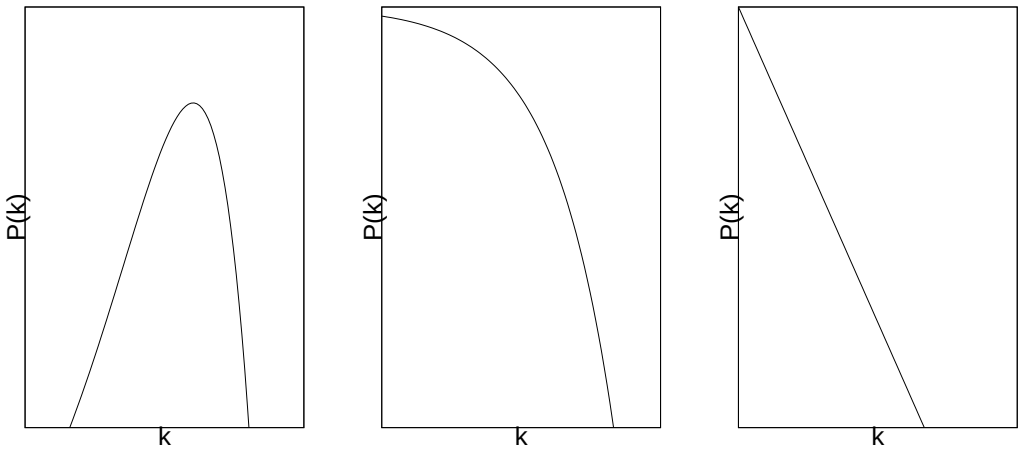


Figure 1.7: Node degree distributions $P(k)$ in log-log scale. From left to right: Poisson distribution $P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$, exponential distribution $P(k) \sim e^{-k/\langle k \rangle}$, and a power law distribution $P(k) \sim 1/k^{\gamma}$.

and a power law distribution:

$$P(k) \sim 1/k^{\gamma},\ k \neq 0. \tag{1.9}$$

Although all functions $P(k)$ above decay for large k, a special feature of distributions (1.7) and (1.8) is that they contain a typical scale. It is either the location of the maximum for the Poisson distribution, or the characteristic decay length for the exponential one. On contrary, a power-law distribution (1.9) does not contain such a scale. Networks with a node degree distribution, such as (1.9), are called *scale-free* networks. And it is these scale free distributions which are often encountered in complex real world networks.

One more principle difference between the above distributions is that whereas all moments of $P(k)$ exist for (1.7), (1.8), it is not the case for the scale-free distribution. Indeed for $P(k)$ in the form of equation (1.9), the moments

$$M_n = \sum_{k=0}^{\infty} k^n P(k) \text{ with } m \geq \gamma - 1 \tag{1.10}$$

diverge. The scale-free distribution allows for nodes with very high degree (hubs), which are ruled out in practice in finite networks with an exponential or Poissonian decay of the distribution (1.7), (1.8). It is their presence which manifests in the behaviour of moments (1.10) and leads to many other specific features of scale-free networks, as we will see below.

1.2 Complex network models

There exists a number of models that help us to explain different phenomena using complex network theory. In this section we will describe three of these models, those that mostly make up our understanding of complex networks. Those are the classical Erdös-Rényi random graph (as was mentioned above it can not describe real-world networks), the Watts-Strogatz small-world network and the Barabási-Albert scale-free network. The last one is an example of a growing network model, and it turns out that some exactly solved models of growing networks have power-law node degree distributions.

1.2.1 Erdös-Rényi random graph

This type of networks is a classical example of an equilibrium network with a fixed number of nodes N. There exist two models of the Erdös-Rényi random graph. In the first model M links are randomly and independently distributed between pairs of nodes from N nodes; in the second one there is a fixed probability to exist for a link between any two nodes. Further it has been shown that for both models when $M \to 0$, $N \to \infty$, the distribution of the node degree k is governed by a Poisson distribution (1.7), with mean degree value: $\langle k \rangle = 2M/N$ for the first model and $\langle k \rangle = mN$ for the second one. A lot of characteristics can be obtained analytically for $N \to \infty$ [43, 16]. The mean shortest path length (1.2) and the clustering coefficient (1.4) values are:

$$\langle \ell \rangle \sim \ln(N)/ln(k), \quad C \sim k/N. \tag{1.11}$$

To generate an Erdös-Rényi random graph the following algorithm is usually used. Let us take N isolated nodes and then we consecutively add links that are connecting random pairs of the nodes. Within this process at the beginning the graph is a set of small disconnected components. They are growing and at some point we will get "giant" cluster of connected nodes and their number will be a finite fraction of N even in the limit $N \to \infty$. It is important to mention that a giant cluster emerges only when the generation probability m of having a link between two nodes reaches a critical treshhold m_c. The giant cluster will appear in a similar way as drops of water are condensed in an oversaturated steam. As a result of this phase transition process the fraction of connected nodes that belongs to the giant cluster can be given as [36]

$$G = 1 - \frac{1}{\langle k \rangle} \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!} (\langle k \rangle e^{-\langle k \rangle})^n, \quad \langle k \rangle = mN. \tag{1.12}$$

As we see, the fraction of connected nodes monotonously grows when increasing the mean degree $\langle k \rangle$.

1.2.2 Watts-Strogatz small-world model

Usually complex network models are generated on computers - as numeric realizations of graphs. However the idea of small-world network model appeared far before it became possible to make such calculations on computers. It is easy to get an idea of small-world just checking your acquaintances, and then acquaintances of your acquaintances (those that do not know you, personally) and so on. It is enough to observe a short chain of such acquaintances to understand that anyone of us can build quite short chains to connect some randomly chosen person (prime-minister for example). In that meaning our world is small, and that fact gave name for the model - small-world.

A computer model of small-world networks was introduced by Watts and Strogatz [115]. It can be described as follows. Let us consider one-dimensional closed chain of N nodes as on Fig. (1.8). At the beginning each node is connected with

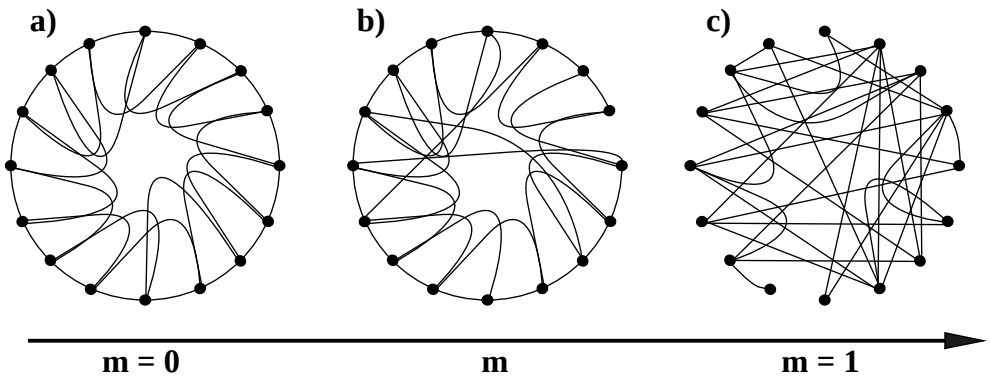


Figure 1.8: Watts-Strogatz rewiring algorithm [115], shows transformation of the regular chain a) into small-world graph b) and then into random graph c). Parameter m define the probability of rewiring links.

k neighbours, where k is an even positive number. Then with some probability m each link randomly changes one of its connecting nodes. This procedure is called rewiring. From Fig. 1.8 one can see that small-world graph generates with small values of probability of rewiring m, when node degree distribution follows Poisson law (1.7).

The Watts-Strogatz model provides an intuitive image for real-world networks, which have neither completely regular topology, nor are completely random. Back to Fig. 1.8: for the case of a regular lattice, when $m = 0$ for large values of N (in particular when $N \gg k \gg \ln k \gg 1$, to consider a connected graph) the mean shortest path length behaves as $\langle \ell \rangle \sim N/2k$ and the clustering coefficient takes the value $C = 3/4$. However from another point of view, for a completely irregular lattice with $m = 1$ (random graph in other words, see section 1.2.1) those characteristics follow equations 1.11. Comparing the characteristics for $m = 0$ and $m = 1$ one can see that the regular lattice is a strongly correlated (clustered) "big-world", where the mean shortest path length grows with N linearly, while the irregular network is the weak correlated small-world ($C \sim k/N \ll 1$), where $\langle \ell \rangle$ grows with N as logarithm. Watts and Strogatz find that the small-world effect (sudden decrease of the mean shortest path length) is observed already for small values of m, while on a local level this transition is nearly imperceptible (the clustering coefficient value C stay large, as for a regular lattice). In current literature the term small-world is used for networks where the mean shortest path length increases slower than any positive power of N [40]. Note that the linear size of the d-dimensional regular lattice grows as $N^{1/d}$.

1.2.3 Barabási-Albert scale-free network

It is observed that many important real-world networks have node degree distributions which exhibit exponential decay (1.8) or power law behaviour (1.9). Because of strong correlations systems governed by a power law do not have any scale of characteristic changes. The main difference between these and random graphs is that with its distribution (1.7), the tail of node degree distribution of the random graph decreases much more faster than for a scale-free network.

Both exponential and power law decay of the degree distribution can be modeled by assuming a non equilibrium growth process of the network by which in consecutive time steps nodes and links are added to an existing network [40]. If the added nodes are arbitrarily linked to any of the existing nodes an exponential tail results, however, if the probability to connect to a given existing node is a linear function of its degree one can show that the resulting degree distribution develops a power law tail.

An algorithm of generation of scale-free networks was proposed by Albert and Barabási [7, 8, 9]. It is based on two concepts that are common for most networks. The first is growth and the second one is preferential attachment. It is important to mention that similar ideas were known earlier in the context of power-law di-

stributions. For example the idea of cumulative advantage [97] was proposed by Price in 1976, and in the Simon model even earlier [105]. The Barabási-Albert model uses the following algorithm: let us have some small number of connected nodes (n_0) and at each time step we add a new node with $n \leq n_0$ links connected to the already existing nodes. Following the idea of preferential attachment the probability of connecting a new node to an already existing node i depends on the degree k_i of the node i:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{1.13}$$

In (1.13) we are summing over all nodes. As Π is proportional to the first power of k_i, this procedure is sometimes called linear preferential attachment. Both numerical simulations and analytical solutions of Barabási-Albert model [7, 9, 74, 38] result in power asymptotic of node degree distribution (1.9) with the exponent value $\gamma = 3$.

More general γ can be produced if we imply initial preferential parameter a

$$\Pi(k_i) = \frac{k_i + a}{\sum_j (k_j + a)}. \tag{1.14}$$

1.3 Previous studies of public transport networks

While general interest in complex networks was growing very fast, one particular type of networks, which are used everyday, public transport networks (PTN) was analysed in details only recently [47, 103]. PTN are an example of transportation networks, and have general common characteristics of those: evolutionary dynamics, optimisation, two-dimensionality. However topological characteristics of the PTN were less known than for example such characteristics of airports networks that also belong to transportation networks [5, 62, 63, 11, 77, 78, 60]. Railway networks can be mentioned in the same context [101] and networks of electricity [5, 34, 4] as well.

In some studies specific subsets of PTN were analysed, for example the Boston subway network [82, 75, 76, 100], the Vienna subway network [100] or bus networks of three cities in China [119]. However each separate type of public transport (bus network, subway or trams network) is not a closed system: these are only subnetworks of a wider city transport system, or as we called it PTN. To understand and describe public transport characteristics one need to analyse complete PTN, not separate specific parts. And really it turns out that network characteris-

tics obtained when analysing separately subway network and network ”subway + buses” differ a lot. That was shown for Boston in studies [75, 76].

Complete PTN were analysed in studies [47, 103]. In the first one [47] PTN of Berlin, Paris and Düesseldorf were considered, in the second one [103], – public transport systems of 22 polish cities. Study [47] is concerned on scale-free characteristics of PTN. It is shown that for mentioned cities power-law node degree distribution is common. Also power-law was observed for some other characteristics that describe the intensity of movement in PTN. However there were not enough statistical data to make clear conclusions. In study [103] a conclusion was made that node degree distributions for PTN can have both exponential and power-law behaviour, depending on the topology of the network representation chosen. At the same time were analysed other PTN characteristics (clustering coefficient, betweenness centrality, assortativity).

1.4 Network attack vulnerability and percolation phenomenon

The question of resilience or vulnerability of a complex network [44, 109] against failure of its parts has, beside purely academic interest, a whole range of important practical implications. In what follows such failure will be called an *attack*. In practice, the origin of the attack and its scenario may differ to large extent, ranging from random failure, when a node or a link in a network is removed at random to a targeted destruction, when the most influential network constituents are removed according to their operating characteristics. The notion of attack vulnerability of complex networks originates from studies of computer networks and was coined to denote the decrease of network performance as caused by the removal of either nodes or links. The behavior of a complex network under attack has been observed to drastically differ from that of regular lattices. Early evidence of this fact was found in particular for real world networks that show scale-free behavior: the world wide web and the internet [2, 110], as well as metabolic [70], food web [107], and protein [71] networks. It appeared that these networks display a high degree of robustness against random failure. However, if the scenario is changed towards targeted attacks, the same networks may appear to be especially vulnerable [30, 23].

Essential progress towards a theoretical description of the attack vulnerability of complex networks is due to the application of the tools and concepts of

percolation phenomena. On a lattice percolation occurs e.g. when at a given concentration of bonds a spanning cluster appears. This concentration c_{perc} which is determined by an appropriate ensemble average in the thermodynamic limit is the so-called percolation threshold and is in general lattice dependent. On a general network the corresponding phenomenon is the emergence of a giant connected component (GCC) i.e. a connected subnetwork which in the limit of an infinite network contains a finite fraction of the network. For a random graph where given vertices are linked at random this threshold can be shown to be reached at one bond per vertex [43]. However the distribution $p(k)$ of the degrees k of vertices in a random graph is Poissonian. A more general criterion applicable to networks with given degree distribution $p(k)$ but otherwise random linking between vertices has been proposed by Molloy and Reed [30, 23, 84]. For such equilibrium networks a GCC can be shown to be present if

$$\langle k(k-2)\rangle \geq 0 \tag{1.15}$$

with the appropriate ensemble average $\langle \ldots \rangle$ over networks with given degree distribution. Defining the Molloy-Reed parameter as the ratio of the first two moments of the degree distribution

$$\kappa^{(k)} \equiv \langle k^2\rangle / \langle k\rangle \tag{1.16}$$

the percolation threshold can then be determined as

$$\kappa^{(k)} = 2 \quad \text{at} \quad c_{\text{perc}}. \tag{1.17}$$

Taken that for scale-free networks the degree distribution obeys power law scaling

$$p(k) \sim k^{-\gamma} \tag{1.18}$$

one finds that the second moment $\langle k^2\rangle$ diverges for $\gamma < 3$. Thus, the value $\gamma = 3$ separates two different regimes for the percolation on equilibrium scale free networks [30]. Moreover the values $\gamma = 2$ and $\gamma = 4$ are further boundaries [32]. Indeed, for infinite equilibrium scale-free networks if $\gamma < 2$ the distribution has no finite average $\langle k\rangle$, for $\gamma < 3$ a GCC is found to exist at any concentration of removed sites (the network appears to be extremely robust to random removal of nodes and has no percolation threshold with respect to a dilution of its nodes), $\kappa^{(k)}$ (1.16) remains finite for $\gamma > 3$ and finally when $\gamma > 4$ network percolation and other properties are expected to be similar to those of exponentially decaying networks. Therefore, the observed transitions for real-world systems [2, 110, 70, 107, 71]

from the theoretical standpoint may be seen as finite-size effects or resulting from essential degree-degree correlations. The tolerance of scale-free networks to intentional attacks (when the highest degree nodes are removed) was studied in [31]. It was shown that even networks with $\gamma < 3$ may be sensitive to intentional attacks.

Obviously, the above theoretical results apply to ideal complex networks and for ensemble averages and may be confirmed within certain accuracy when applied to different individual real-world networks. Not only finite-size effects are the origin of this discrepancy [72]. Furthermore, even networks of similar type (e.g. of similar node degree distribution and size) may be characterized by a large variety of other characteristics. While some of them may have no impact on the percolation properties [120], others do modify their behavior under attack, as empirically revealed in study [66] for two different real-world scale-free networks (computer and collaboration networks). Therefore, an empirical analysis of the behavior of different real-world networks under attack appears timely and will allow not only to elaborate scenarios for possible defence mechanisms of operating networks but also to create strategies of network constructions, that are robust to attacks of various types.

1.5 Optimisation

As far as public transport networks plays a very important part in the economy of a city and the everyday lives of many of its inhabitants a number of studies have been done on their optimisation [28]. Here we are interested in a mathematical approach. The usual approach is to optimise either the average overall travelling time for customers, or the average travel cost, which can be defined in various ways. To simplify the problem most of studies are concerned with the bus networks, however these represent general characteristics of PTN of other kinds. Another simplification approach is to use simple geometric configurations to model the PTN. Most described are systems with rectangular routeing (along a rectangular grid) [42, 112, 46, 87] and polar routeing (along ring and radial roads) [88, 113, 22, 26] which was introduced by Smeed and Haight [106, 65].

Analytical mathematical approaches on PTN optimisation in general are attempts to simplify urban areas and their networks by using simple geometric bus configurations and travel demand functions as to facilitate the use of simple formulae to determine optimum routes, headways, fleet sizes, shapes, number of units and so on. Of course a real PTN system is more complex, and such simplification of a complex problem usually results in a model that is so far from reality

that their use at operational level is doubtful [1, 87]. However such models are useful in understanding the various relationships within the geographical layout of a PTN at the conceptual level. In our approach we focus on finding the optimal setup of routes in a radial system for a given investment level.

Chapter 2

Empirical analysis

In this chapter we will present results of the empirical analysis of the PTN of 14 major cities of the world. To this end, we will introduce different network representations (different "spaces"), introduce more observables in terms of which network properties will be analysed, and find the values of these observables for each PTN under consideration. Some of the results presented here were published in [48, 50].

2.1 Description of the database

The choice for the selection of fourteen major cities (see Table 2.1) [127, 128] is motivated by the idea to collect network samples from cities of different geographical, cultural, and economical background. Apart from the systematic analysis explained above this choice also extends to PTN of much larger size as compared to previous work [47, 103] which considered PTN of typically hundreds of stations.

All PTN analysed within this study are either operated by a single operator or by a small number of operators with a coordinated schedule, as e.g. expressed by a central website from which our data were obtained. Rather than artificially dividing these centrally organized networks into subnetworks of different means of transport like bus and metro or in an 'urban' and a 'sub-urban' part we treat each full PTN as an entity.

City	N	R	S	Type
Berlin	2992	211	29.4	BSTU
Dallas	5366	117	59.9	B
Düsseldorf	1494	124	28.5	BST
Hamburg	8084	708	25.5	BFSTU
Hong Kong	2024	321	39.6	B
Istanbul	4043	414	31.7	BST
London	10937	922	34.2	BST
Los Angeles	44629	1881	52.9	B
Moscow	3569	679	22.2	BEST
Paris	3728	251	38.2	BS
Rome	3961	681	26.8	BT
Saõ Paolo	7215	997	58.3	B
Sydney	1978	596	16.3	B
Taipei	5311	389	70.5	B

Table 2.1: Cities analysed in this study. N: number of PTN stations; R: number of PTN routes; S: mean route length (mean number of stations per route). Types of transport taken into account: Bus, Electric trolleybus, Ferry, Subway, Tram, Urban train.

2.2 PT network topology

A straightforward representation of a PT map in the form of a graph represents every station by a node while the edges correspond to the links that exist between stations due to the PT routes servicing them (see e.g. Figs. 2.1, 2.2a).

Let us first introduce a simple graph to represent this situation, see Fig. 2.2b. In the following we will refer to this graph as the $\mathbb{L}$-space graph [103] or simply as $\mathbb{L}$-space. This graph represents each station by a node, a link between nodes indicates that there is at least one route that services the two corresponding stations consecutively. No multiple links are allowed. In the analysis of PTN, this $\mathbb{L}$-space representation has been used in studies of Refs. [75, 47, 103, 6, 119].

A somewhat different concept is that of a bipartite graph which was proven useful in the analysis of cooperation networks [92, 61]. In this representation called $\mathbb{B}$-space both routes and stations are represented by nodes [121, 48, 27]. Each route node is linked to all station nodes that it services. No direct links between nodes of same type occur (see Fig. 2.2c). Obviously, in $\mathbb{B}$-space the

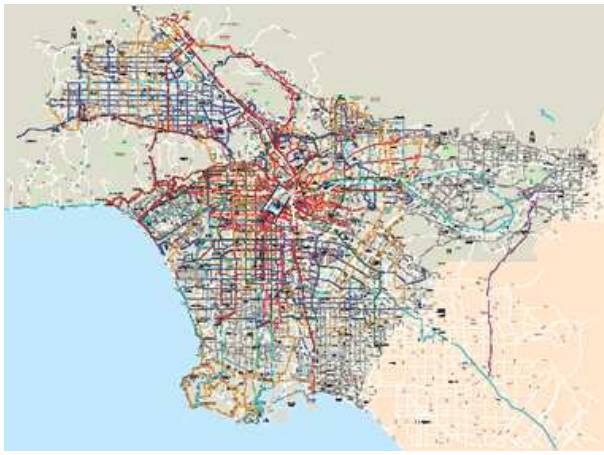


Figure 2.1: One of the networks we analyse in this study. The Los Angeles PTN consists of $R = 1881$ routes and $N = 44629$ stations, some of them are shown in this map.

neighbors of a given route node are all stations that it services while the neighbors of a given station node are all routes that service it.

There are two one-mode projections of the bipartite graph of $\mathbb{B}$-space. The projection to the set of station nodes is the so-called $\mathbb{P}$-space graph, Fig. 2.2d. The complementary projection to route nodes leads to the $\mathbb{C}$-space graph, Fig. 2.2e, of route nodes where any two route nodes are neighbors if they share a common station.

The $\mathbb{P}$-space graph representation [101, 103] has proven particularly useful in the analysis of PTN [101, 100, 103, 48, 119]. The nodes of this graph are stations and they are linked if they are serviced by at least one common route. In this way the neighbors of a $\mathbb{P}$-space node are all stations that can be reached without changing means of transport and each route gives rise to a complete $\mathbb{P}$-subgraph, see Fig. 2.2d.

It is worthwhile to note the real world significance of these seemingly abstract 'spaces'. To give an example, the average length of a shortest path $\langle \ell_{\mathbb{L}} \rangle$ in an $\mathbb{L}$-space graph is the average number of stops one has to pass to travel between any two stations. When represented in $\mathbb{P}$- space, the mean shortest path $\langle \ell_{\mathbb{P}} \rangle$ counts the average number of changes one has to do to travel between two stations while the corresponding mean $\mathbb{C}$- space path length $\langle \ell_{\mathbb{C}} \rangle$ counts the average number of changes needed to pass between any two routes. As another example let us

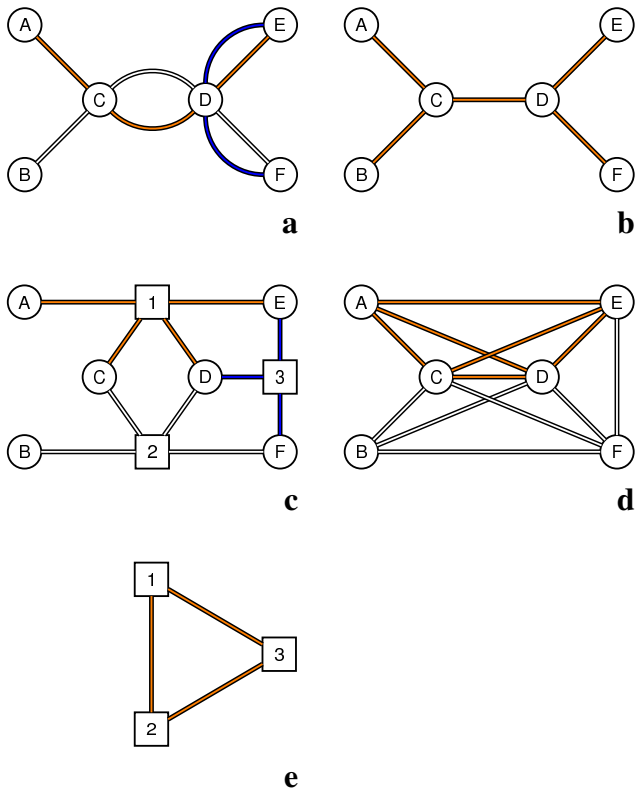


Figure 2.2: **a**: a simple public transport map. Stations A-F are serviced by routes No 1 (shaded orange), No 2 (white), and No 3 (dark blue). **b**: $\mathbb{L}$-space graph. **c**: $\mathbb{B}$-space bipartite graph. Route nodes are shown as squares. **d**: $\mathbb{P}$-space graph, the complete sub-graph corresponding to route No 1 is highlighted (shaded orange). **e**: $\mathbb{C}$-space graph of routes.

note the node degree k: for the $\mathbb{L}$-space graph the node degree of a station is the number of other stations within one stop distance; in the bipartite $\mathbb{B}$-space graph the degree of a station is the number of routes servicing it, while the degree of a route is the number of its stations; in the $\mathbb{P}$-space graph the degree of $k_{\mathbb{P}}$ of a station is the number of stations reachable without changing the route; whereas in the $\mathbb{C}$-space graph the degree $k_{\mathbb{C}}$ of a route is the number of other routes one can transfer to.

In summary, the explicit definitions of our 'spaces' are as follows. In $\mathbb{L}$-space each node represents station, and a link between any two nodes indicates that there is at least one route that services the two corresponding stations consecutively. In

City	$\langle k_{\mathbb{L}} \rangle$	$\kappa_{\mathbb{L}}$	$\ell_{\mathbb{L}}^{\max}$	$\langle \ell_{\mathbb{L}} \rangle$	$\langle \mathcal{C}_{\mathbb{L}B} \rangle$	$c_{\mathbb{L}}$	$\langle k_{\mathbb{P}} \rangle$	$\kappa_{\mathbb{P}}$	$\ell_{\mathbb{P}}^{\max}$	$\langle \ell_{\mathbb{P}} \rangle$	$\langle \mathcal{C}_{\mathbb{P}B} \rangle$	$c_{\mathbb{P}}$	$\langle k_{\mathbb{C}} \rangle$	$\kappa_{\mathbb{C}}$	$\ell_{\mathbb{C}}^{\max}$	$\langle \ell_{\mathbb{C}} \rangle$	$\langle \mathcal{C}_{\mathbb{C}B} \rangle$	$c_{\mathbb{C}}$
Berlin	2.58	1.96	68	18.5	$2.6 \cdot 10^4$	52.8	56.61	11.47	5	2.9	$2.9 \cdot 10^3$	41.9	27.56	4.43	5	2.2	$1.2 \cdot 10^2$	4.75
Dallas	2.18	1.28	156	52.0	$1.4 \cdot 10^5$	55.0	100.58	11.23	8	3.2	$5.9 \cdot 10^3$	48.6	11.09	3.45	7	2.7	$9.2 \cdot 10^1$	5.34
Düsseldorf	2.57	1.96	48	12.5	$8.6 \cdot 10^3$	24.4	59.01	10.56	5	2.6	$1.2 \cdot 10^3$	19.7	32.18	2.47	4	1.8	$4.9 \cdot 10^1$	2.23
Hamburg	2.65	1.85	156	39.7	$1.4 \cdot 10^5$	254.7	50.38	7.96	11	4.7	$1.4 \cdot 10^4$	132.2	17.51	4.49	10	4.0	$9.9 \cdot 10^2$	28.3
Hong Kong	3.59	3.24	60	11.0	$1.0 \cdot 10^4$	60.3	125.67	10.20	4	2.2	$1.3 \cdot 10^3$	11.7	98.98	2.12	3	1.7	$1.2 \cdot 10^2$	2.14
Istanbul	2.30	1.54	131	29.7	$5.7 \cdot 10^4$	41.0	76.88	10.59	6	3.1	$4.2 \cdot 10^3$	41.5	52.81	3.86	5	2.3	$2.6 \cdot 10^2$	5.00
London	2.60	1.87	107	26.5	$1.4 \cdot 10^5$	320.6	90.60	16.97	6	3.3	$1.2 \cdot 10^4$	90.0	49.91	6.80	6	2.6	$7.4 \cdot 10^2$	11.1
Los Angeles	2.37	1.59	210	37.1	$7.9 \cdot 10^5$	645.3	97.99	17.21	11	4.4	$7.4 \cdot 10^4$	399.6	40.11	8.42	10	3.6	$2.3 \cdot 10^3$	22.1
Moscow	3.32	6.25	27	7.0	$1.1 \cdot 10^4$	127.4	65.47	26.48	5	2.5	$2.7 \cdot 10^3$	38.0	109.37	4.57	4	1.9	$3.2 \cdot 10^2$	3.59
Paris	3.73	5.32	28	6.4	$1.0 \cdot 10^4$	78.5	50.92	24.06	5	2.7	$3.1 \cdot 10^3$	59.6	39.95	4.67	4	1.9	$1.1 \cdot 10^2$	2.72
Rome	2.95	2.02	87	26.4	$5.0 \cdot 10^4$	163.4	69.05	11.34	6	3.1	$4.2 \cdot 10^3$	41.4	59.40	4.86	5	2.5	$5.1 \cdot 10^2$	7.04
Saõ Paolo	3.21	4.17	33	10.3	$3.4 \cdot 10^4$	268.0	137.46	19.61	5	2.7	$6.0 \cdot 10^3$	38.2	151.72	4.25	4	2.0	$5.2 \cdot 10^2$	4.27
Sydney	3.33	2.54	34	12.3	$7.3 \cdot 10^3$	82.9	42.88	7.79	7	3.0	$1.3 \cdot 10^3$	33.6	65.02	2.92	6	2.4	$3.5 \cdot 10^2$	6.30
Taipei	3.12	2.42	74	20.9	$5.3 \cdot 10^4$	186.2	236.65	12.96	6	2.4	$3.6 \cdot 10^3$	15.4	93.33	2.95	5	1.8	$1.6 \cdot 10^2$	2.44

Table 2.2: PTN characteristics in different spaces (subscripts refer to $\mathbb{L}$-, $\mathbb{P}$-, and $\mathbb{C}$-spaces, correspondingly). k: node degree; $\kappa = \langle z \rangle / \langle k \rangle$ where z is the number of next nearest neighbors; $\ell^{\max}$, $\langle \ell \rangle$: maximal and mean shortest path length (1.2); $\mathcal{C}_B$: betweenness centrality (1.6); c: relation of the mean clustering coefficient to that of the classical random graph of equal size (1.11). Averaging has been performed with respect to corresponding network, only the mean shortest path $\langle \ell \rangle$ is calculated with respect to the largest connected component.

$\mathbb{P}$-space each node represents station, they are linked if they are serviced by at least one common route. In $\mathbb{C}$-space each node represents a route, and two routes are linked if they share a common station. In $\mathbb{B}$-space both routes and stations each represented by nodes: route nodes and station nodes. The $\mathbb{B}$-space graph is a bipartite graph where each edge links a station node with a route node. Each route node is linked to all station nodes that it services. No direct links between nodes of same type occur. In all 'spaces' neither multiple links nor loops are allowed. The $\mathbb{P}$-space graph results from a one-mode projection of the bipartite graph to the station nodes while the $\mathbb{C}$-space graph results from one mode projection of the bipartite graph to the route nodes.

Table 2.2 lists some of the PTN characteristics we have obtained for the cities under consideration using publicly available data from the web pages of local transport organizations [127, 128]. To limit the data presented, this and further tables are restricted to the basic results discussed within this thesis. The interested reader may find supplementary material in [49].

2.3 Local network properties

Let us first examine the properties of the PTN determined by the immediate neighborhood of the nodes as measured by its size, its interconnectedness and the correlations within this neighborhood. To this end we will recall some network characteristics that were already mentioned in section 1.1.2.

2.3.1 Neighborhood size (node degree)

The size of the neighborhood of a node as given by its degree often indicates its importance e.g. as a hub within the network. In large networks created by randomly connecting nodes, hubs are rare while in real networks they are often found with much higher probability. Formally this is measured by the behavior of the tail of the node degree distribution. Denoting by $p(k)$ the normalized node degree distribution, the mean node degree k is given by the average

$$\langle k \rangle = \sum_{k=1}^{\mathrm{k}^{\max}} p(k)k = \frac{2M}{N} \,. \tag{2.1}$$

Here, M is the number of links and N the number of edges of the graph while $k^{\max}$ stands for the maximal node degree. For the finite size Erdös-Rényi [43, 16]

random graph the node degree distribution $p(k)$ is binomial, which in the infinite case becomes a Poisson distribution (1.7).

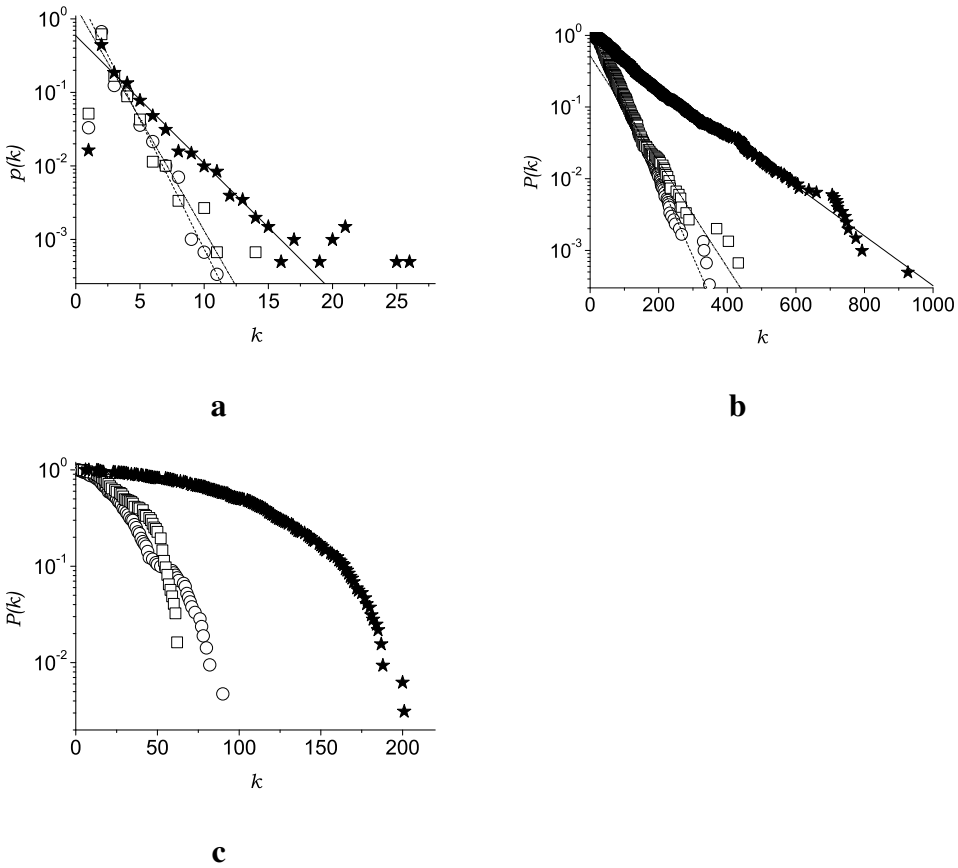


Figure 2.3: **a**: Node degree distributions of PTN of several cities in $\mathbb{L}$-space. **b**: Cumulative node degree distribution in $\mathbb{P}$-space. **c**: Cumulative node degree distribution in $\mathbb{C}$-space. Berlin (circles, $\hat{k}_{\mathbb{L}} = 1.24$, $\hat{k}_{\mathbb{P}} = 39.7$), Düsseldorf (squares, $\hat{k}_{\mathbb{L}} = 1.43$, $\hat{k}_{\mathbb{P}} = 58.8$), Hong Kong (stars, $\hat{k}_{\mathbb{L}} = 2.50$, $\hat{k}_{\mathbb{P}} = 125.1$).

As was mentioned in section 1.2.3 the higher organization of real world networks usually leads to slower decaying distributions. Typical classes of such networks have either exponential or power law tails. Both exponential and power law decay of the degree distribution can be modeled by assuming a non equilibrium growth process of the network [40]. As far as PTN obviously are evolving networks, their evolution may be expected to follow similar mechanisms. However,

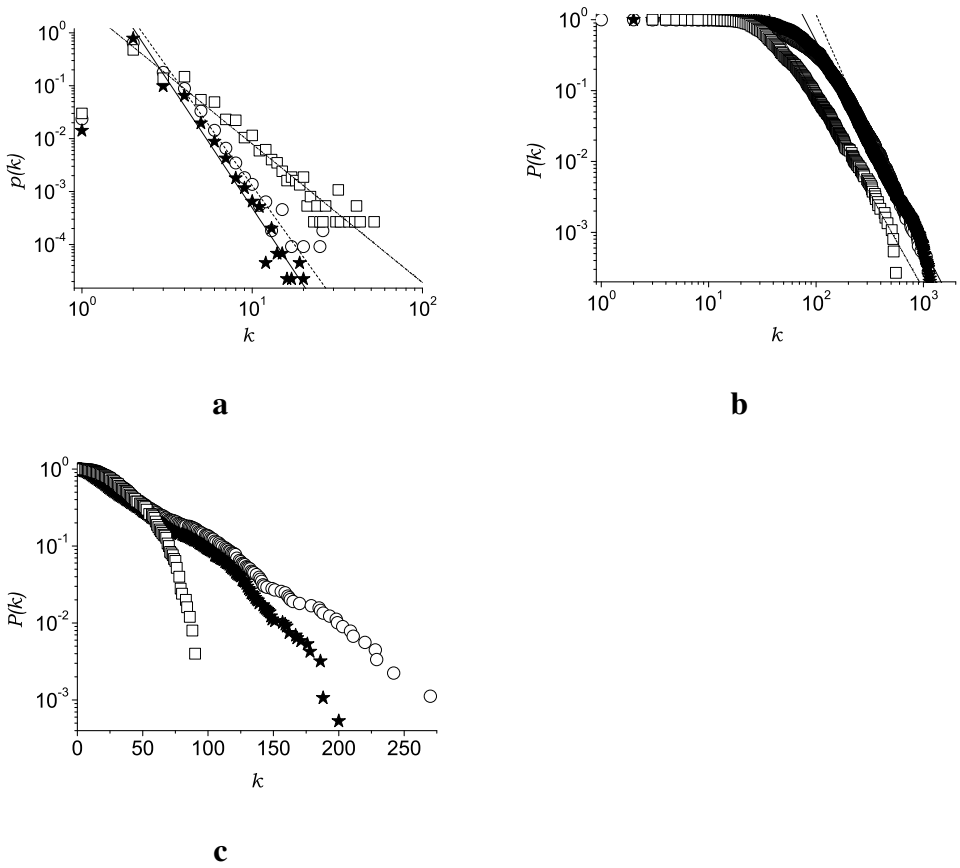


Figure 2.4: **a**: Node degree distributions of the PTN of several cities in $\mathbb{L}$-space. **b**: Cumulative node degree distributions in $\mathbb{P}$-space. **c**: Cumulative node degree distribution in $\mathbb{C}$-space. London (circles, $\gamma_{\mathbb{L}} = 4.48$, $\gamma_{\mathbb{P}} = 3.89$), Los Angeles (stars, $\gamma_{\mathbb{L}} = 4.85$, $\gamma_{\mathbb{P}} = 3.92$), Paris (squares, $\gamma_{\mathbb{L}} = 2.62$, $\gamma_{\mathbb{P}} = 3.70$).

scale-free networks have also been shown to arise when minimizing both the effort for communication and the cost for maintaining connections [53, 57]. Moreover, this kind of optimisation was shown to lead to small world properties [83] and to explain the appearance of power laws in a general context [54]. Therefore, scale-free behavior in PTN could also be related to obvious objectives to optimise their operation.

Figs. 2.3 and 2.4 show the node degree distributions for PTN of several cities in $\mathbb{L}$-, $\mathbb{P}$-, and $\mathbb{C}$-spaces. Note, that the monotonously decreasing curves displayed

for the $\mathbb{P}$- and $\mathbb{C}$-spaces are cumulative distributions defined as:

$$P(k) = \sum_{q=k}^{k^{\max}} p(q). \tag{2.2}$$

The data for $\mathbb{L}$- and $\mathbb{P}$-spaces in Fig. 2.3a,b is shown in log-linear plots together with fits to an exponential decay (1.8). The latter distributions are nicely described by an exponential decay. As far as the $\mathbb{L}$-space data is concerned, we find evidence for an exponential decay for about half of the cities analysed, while the other part rather follow a power law decay (1.9), see Table 2.3.

City	$\gamma_{\mathbb{L}}$	$\hat{k}_{\mathbb{L}}$	$\gamma_{\mathbb{P}}$	$\hat{k}_{\mathbb{P}}$
Berlin	(4.30)	1.24	(5.85)	39.7
Dallas	5.49	(0.78)	(4.67)	64.2
Düsseldorf	3.76	(1.43)	(4.62)	(58.8)
Hamburg	(4.74)	1.46	4.38	(60.7)
Hong Kong	(2.99)	2.50	(4.40)	125.1
Istanbul	4.04	(1.13)	(2.70)	86.7
London	4.48	(1.44)	3.89	(143.3)
Los Angeles	4.85	(1.52)	3.92	(201.0)
Moscow	(3.22)	(2.15)	(2.91)	50.0
Paris	2.62	(3.30)	3.70	(100.0)
Rome	(3.95)	1.71	(5.02)	54.8
Saõ Paolo	2.72	(4.20)	(4.06)	225.0
Sydney	(4.03)	1.88	(5.66)	38.7
Taipei	(3.74)	1.75	(5.16)	201.0

Table 2.3: Parameters of the PTN node degree distributions fit to an exponential (1.8) and power law (1.9) behavior. Bracketed values indicate less reliable fits. Subscripts refer to $\mathbb{L}$- and $\mathbb{P}$-spaces [128].

Figs. 2.4a,b show the corresponding plots for three other cities on a log-log scale. Here, these plots are shown together with fits to a power law (1.9). Numerical values of the fit parameters $\hat{k}$ and γ for different cities are given in Table 2.3. Here, values in parentheses indicate a less reliable fit. In the case when none of equations (1.8), (1.9) lead to reliable data, both fit parameters are given in parentheses in the table. The typical range of data points which could be fitted was

of the order of 90 % or more for $\mathbb{L}$- and $\mathbb{P}$-spaces. The value of the fit parameters was considered to be reliable if the absolute value of the Pearson correlation coefficient exceeded $R_{\mathbb{L}} = 0.984$ and $R_{\mathbb{P}} = 0.990$ in $\mathbb{L}$ and $\mathbb{P}$-spaces, correspondingly. Exceptions from this rule are the $\mathbb{L}$-space fits for the PTN of Paris, Rome ($R_{\mathbb{L}} \simeq 0.97$, but 97 % of data points are covered), and London (with 72 % of data points covered and $R_{\mathbb{L}} \simeq 0.985$). For $\mathbb{P}$-space, exceptions are the PTN of Paris ($R_{\mathbb{P}} = 0.993$) and Saõ Paolo ($R_{\mathbb{P}} = 0.999$), where the fit covered only ~ 60 % of data points. Note, that for $\mathbb{L}$-space the fit was done for the plain node degree distribution $p(k)$, whereas for $\mathbb{P}$-space the parameters $\gamma_{\mathbb{P}}$ or $\hat{k}_{\mathbb{P}}$ were determined by fitting the cumulative distribution (2.2).

While the node degree distribution of almost half of the cities in the $\mathbb{L}$-space representation display a power law decay (1.9), this is in general not the case for the $\mathbb{P}$-space. However, the data for the PTN of Hamburg, London, Los Angeles, and Paris (see Fig. 2.4b) give first evidence of power law behavior of $P(k)$ even in the $\mathbb{P}$-space representation. Previous results concerning node-degree distributions of PTN in $\mathbb{L}$- and $\mathbb{P}$-spaces [103, 119] seemed to indicate that in general the degree distribution may be power-law like in $\mathbb{L}$-space but never in $\mathbb{P}$-space. This was interpreted [103] as being due to strongly correlated connections between stations in $\mathbb{L}$-space and nearly randomly linked routes, as also expressed by a low clustering coefficient in $\mathbb{C}$-space, see below. Our present study, which includes a much less homogeneous selection of cities (study [103] was exclusively based on Polish cities) shows that almost any combination of different distributions in $\mathbb{L}$- and $\mathbb{P}$-spaces may occur. We note that even within the small sub-group formed by Hamburg, Los Angeles, London and Paris there is no alignment to 'typical behaviour'.

In $\mathbb{C}$-space the decay of the node degree distribution is exponential or faster, as one can see from the plots in Figs. 2.3c and 2.4c. From the cities presented there, only the PTN of Berlin, London, and Los Angeles are governed by an exponential decay.

For most cities that show a power law degree distribution in $\mathbb{L}$-space the corresponding exponent $\gamma_{\mathbb{L}}$ is $\gamma_{\mathbb{L}} \sim 4$. Also the exponents found for the PTN of Polish cities of similar size N also lie in this region: $\gamma_{\mathbb{L}} = 3.77$ for Kraków (with number of stations $N = 940$), $\gamma_{\mathbb{L}} = 3.9$ for Łódź ($N = 1023$), $\gamma_{\mathbb{L}} = 3.44$ for Warsaw ($N = 1530$) [103]. According to the general classification of scale-free networks [39] this indicates that in many respect these networks are expected to behave similar to those with exponential node degree distribution. Prominent exceptions to this rule are the PTN of Paris ($\gamma_{\mathbb{L}} = 2.62$) and Saõ-Paolo ($\gamma_{\mathbb{L}} = 2.72$). Note, that values of $\gamma_{\mathbb{L}}$ in the range $2.5 - 3.0$ were recently reported for the bus networks

of three cities in China: Beijing ($N = 3938$), Shanghai ($N = 2063$), and Nanjing ($N = 1150$) [119].

A conclusion from our survey of the various degree distributions is that they appear much more diverse than expected and that with respect to these there is no simple division of the PTN at hand into two or even three classes.

2.3.2 Clustering

While the node degree counts the neighbors of a node, the connectivity within its neighborhood may be quantified in terms of the so called clustering coefficient. The latter is defined in section 1.2.1 1.4 ($C_i = \frac{2y_i}{k_i(k_i-1)}$ for $k_i \geq 2$, where y_i is the number of links between the k_i nearest neighbors of the node i). $C_i \equiv 0$ for $k_i = 0, 1$. The clustering coefficient of a node may also be defined as the probability of any two of its randomly chosen neighbors to be connected. For the mean value of the clustering coefficient of an Erdös-Rényi random graph one finds 1.11 ($\langle C \rangle_{\mathrm{ER}} = \frac{\langle k \rangle}{N} = \frac{2M}{N^2}$).

In Table 2.2 we give the values of the mean clustering coefficient in $\mathbb{L}$-, $\mathbb{P}$-, and $\mathbb{C}$-spaces. The highest absolute values of the clustering coefficient are found in $\mathbb{P}$-space, where their range is given by $\langle C_{\mathbb{P}} \rangle = 0.7 - 0.9$ (c.f. with $\langle C_{\mathbb{L}} \rangle = 0.02 - 0.1$). This is not surprising since in $\mathbb{P}$-space each route gives rise to a fully connected (complete) subgraph between all of its stations. In order to make numbers comparable we normalize the mean clustering coefficient by that of a random graph (1.11) of the same size:

$$c = N^2 \langle C \rangle / (2M). \tag{2.3}$$

In $\mathbb{L}$- and $\mathbb{P}$-representations we find the mean clustering coefficient to be larger by orders of magnitude relative to the random graph. This difference is less pronounced in $\mathbb{C}$-space indicating a lower degree of organization in these graphs. Most prominently, we find the values to vary strongly within the sample of the 14 cities.

In $\mathbb{P}$-space the clustering coefficient of a node is strongly correlated with the node degree. All stations i belonging to the complete subgraph of a single route have $C_i = 1$, while C_i generally decreases if i belongs to more than one route. Averaging the $\mathbb{P}$-space clustering coefficient over all nodes with given degree k we confirm that it decays as function of k according to a power law

$$\langle C_{\mathbb{P}}(k) \rangle \sim k^{-\beta}. \tag{2.4}$$

Within a simple model of networks with star-like topology this exponent is found to be $\beta = 1$ [103]. In transport networks, this behavior has been observed before for the Indian railway network [101] as well as for Polish PTN [103]. In our case, the values of the exponent β for the networks studied range from 0.65 (Saõ Paolo) to 0.96 (Los Angeles) again showing significant diversity within our sample.

These obvious differences in the locally observable structure may be assumed to reflect a strong diversity within the concepts according to which various PTN are structured. Comparing the division between weak and strongly clustered PTN we find no alignment with the different classes of degree distributions adding to the idea of an individual profile of each city's PTN with respect to the various network characteristics.

2.3.3 Generalized assortativities

To describe correlations between the properties of neighboring nodes in a network the notion of assortativity was introduced. This quantity measures the correlation between the node degrees of neighboring nodes in terms of the mean Pearson correlation coefficient [91, 93]. Here, we propose to generalize this concept to measure also the correlations between the values of other node characteristics (other observables). For any link i let X_i and Y_i be the values of the observable at the two nodes connected by this link. Then the correlation coefficient is given by:

$$r = \frac{M^{-1}\sum_i X_i Y_i - [M^{-1}\sum_i \frac{1}{2}(X_i + Y_i)]^2}{M^{-1}\sum_i \frac{1}{2}(X_i^2 + Y_i^2) - [M^{-1}\sum_i \frac{1}{2}(X_i + Y_i)]^2} \tag{2.5}$$

where summation is performed with respect to the M links of the network. Taking X_i and Y_i to be the node degrees equation (2.5) is equivalent to the usual formula for the assortativity of a network [91]. Here, we will call this special case the degree assortativity $r^{(1)}$. It is although possible to investigate generalized assortativities for a number of other network characteristics, however obtained results not always are interesting to discuss. Here, besides the assortativity $r^{(1)}$, we discuss the behavior of the generalized assortativity $r^{(2)}$ for the number z of next nearest neighbors. The numerical values of the assortativities $r^{(1)}$ and $r^{(2)}$ of all PTN are listed in Table 2.4 for the $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-spaces. With respect to the values of the standard node degree assortativity $r^{(1)}_{\mathbb{L}}$ in $\mathbb{L}$-space, we find two groups of cities. The first one is characterized by values $r^{(1)}_{\mathbb{L}} = 0.1 - 0.3$. Although these values are still small they signal a finite preference for assortative mixing. That is, links tend to connect nodes of similar degree. In the second group of cities these values are

very small $r_{\mathbb{L}}^{(1)} = -0.02 - 0.08$ showing no preference in linkage between nodes with respect to node degrees. PTN of both large and medium sizes are present in each of the groups. This indicates the absence of correlations between network size and degree assortativity $r_{\mathbb{L}}^{(1)}$ in $\mathbb{L}$-space. Measuring the same quantity in the $\mathbb{P}$- and $\mathbb{C}$-spaces, we observe different behavior. In $\mathbb{P}$-space almost all cities are characterized by very small (positive or negative) values of $r_{\mathbb{P}}^{(1)}$ with the exception of the PTN of Istanbul ($r_{\mathbb{P}}^{(1)} = -0.12$) and Los Angeles ($r_{\mathbb{P}}^{(1)} = 0.12$). On the contrary, in $\mathbb{C}$-space PTN demonstrate clear assortative mixing with $r_{\mathbb{C}}^{(1)} = 0.1 - 0.5$. An exception is the PTN of Paris with $r_{\mathbb{C}}^{(1)} = 0.06$.

City	$r_{\mathbb{L}}^{(1)}$	$r_{\mathbb{L}}^{(2)}$	$r_{\mathbb{P}}^{(1)}$	$r_{\mathbb{P}}^{(2)}$	$r_{\mathbb{C}}^{(1)}$	$r_{\mathbb{C}}^{(2)}$
Berlin	0.158	0.616	0.065	0.441	0.086	0.318
Dallas	0.150	0.712	0.154	0.728	0.290	0.550
Düsseldorf	0.083	0.650	0.041	0.494	0.244	0.180
Hamburg	0.297	0.697	0.087	0.551	0.246	0.605
Hong Kong	0.205	0.632	-0.067	0.238	0.131	0.087
Istanbul	0.176	0.726	-0.124	0.378	0.282	0.505
London	0.221	0.589	0.090	0.470	0.395	0.620
Los Angeles	0.240	0.728	0.124	0.500	0.465	0.753
Moscow	0.002	0.312	-0.041	0.296	0.208	0.011
Paris	0.064	0.344	-0.010	0.258	0.060	-0.008
Rome	0.237	0.719	0.044	0.525	0.384	0.619
Saõ Paolo	-0.018	0.437	-0.047	0.266	0.211	0.418
Sydney	0.154	0.642	0.077	0.608	0.458	0.424
Taipei	0.270	0.721	0.009	0.328	0.100	0.041

Table 2.4: Nearest neighbor and next nearest neighbor assortativities $r^{(1)}$ and $r^{(2)}$ in different spaces for the whole PTN.

As we have seen above, all PTN demonstrate assortative ($r^{(1)} > 0$) or neutral ($r^{(1)} \sim 0$) mixing with respect to the node degree (first nearest neighbors number) k. Defining an assortativity $r^{(2)}$ with respect to the number z of second next nearest neighbors we explore the correlation of a wider environment of adjacent nodes. Due to the fact that in this case the two connected nodes share at least part of this environment (the first nearest neighbors of a node form part of the second nearest neighbors of the adjacent node) one may expect the assortativity

$r^{(2)}$ to be non-negative. The results for $r^{(2)}$ shown in Table 2.4 appear to confirm this assumption. In all the spaces considered, we find that all PTN that belong to the group of neutral mixing with respect to k also belong to the same group with respect to the second nearest neighbors. For those PTN that display significant nearest neighbors assortativity $r^{(1)}$ we find that the second nearest neighbor assortativity $r^{(2)}$ is in general even stronger in line with the above reasoning.

From the above observations on assortativity within our sample of PTN we note further evidence for diversity ranging from indefinite to clearly pronounced assortativities $r_{\mathbb{L}}^{(1)}$ and $r_{\mathbb{C}}^{(1)}$ which appear uncorrelated with other properties of the network such as the size or the specific behavior of e.g. the degree distribution.

2.4 Global characteristics

2.4.1 Shortest paths

As was mentioned in section 1.1.2 let $\ell_{i,j}$ be the length of a shortest path between sites i and j in a given graph. Note, that $\ell_{i,j}$ is well-defined only if the nodes i and j belong to the same connected component of the graph. In the following we will restrict considerations to the largest (so-called giant) connected component, GCC. Denoting the path length distribution within the GCC as $\Pi(\ell)$, the mean shortest path is

$$\langle \ell \rangle = \sum_{\ell=1}^{\ell^{\max}} \Pi(\ell)\ell, \tag{2.6}$$

where $\ell^{\max}$ is the maximal shortest path length found within the GCC. In general, the shortest path length distributions obtained in $\mathbb{L}$-, $\mathbb{P}$-, and $\mathbb{C}$-spaces that we have analysed [49] are nicely described by an asymmetric unimodal distribution [103]:

$$\Pi(\ell) = A\ell \exp\left(-B\ell^2 + C\ell\right), \tag{2.7}$$

where A, B, and C are parameters. However, additional structures may lead to deviations from this behavior as can be seen from Fig. 2.5, which shows the mean shortest path length distribution in $\mathbb{L}$-space $P_{\mathbb{L}}(\ell)$ for Los Angeles. One observes a second local maximum on the right shoulder of the distribution. Qualitatively this behavior may be explained by assuming that the PTN consists of more than one community. For the simple case of one large community and a second smaller one at some distance this situation will result in short intra-community paths which will give rise to a global maximum and a set of longer paths that connect the larger

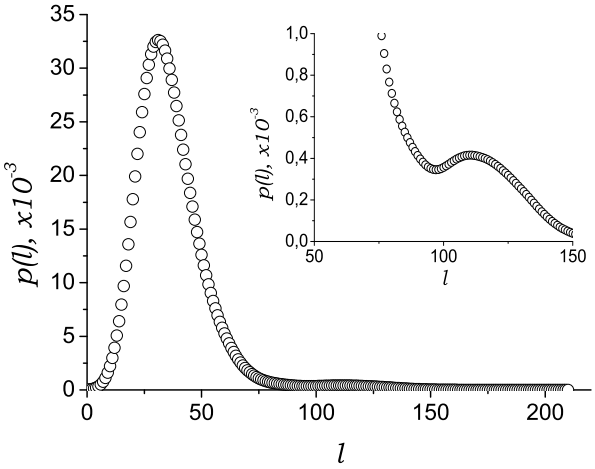


Figure 2.5: Shortest path length distribution in $\mathbb{L}$-space, $P_{\mathbb{L}}(\ell)$, for the PTN of Los Angeles.

to the smaller community resulting in additional local maxima. Such a situation definitely appears to be present in the case of the Los Angeles PTN, see Fig. 2.1.

Of particular interest is the mean shortest path length between nodes of given degrees k and q, $\ell(k,q)$. As has been shown in [68], this relation can be approximated by

$$\ell(k,q) = A - B\log(kq). \tag{2.8}$$

For random networks the coefficients A and B can be calculated exactly [56]. A rather good agreement with equation (2.8) was found for the majority of the $\mathbb{L}$-space graphs of Polish PTN analysed in [103]. Within our study which includes PTN of much larger size, we do not observe a similar alignment for all cities. The suggested logarithmic dependence (2.8) does occur also for the $\mathbb{L}$-space graphs of larger cities, however, with a much more pronounced scattering of data for large values of the product kq. In Fig. 2.6 we plot the mean path $\ell_{\mathbb{L}}(k,q)$ for the $\mathbb{L}$-space graphs of the PTN of Berlin, Hong Kong, Rome, and Taipei, where the relation (2.8) is observed with better accuracy. Note, however, that due to the scatter of data a logarithmic dependence frequently is indistinguishable from a power law with a small exponent.

The dependency of the average path length on the degrees of both end nodes

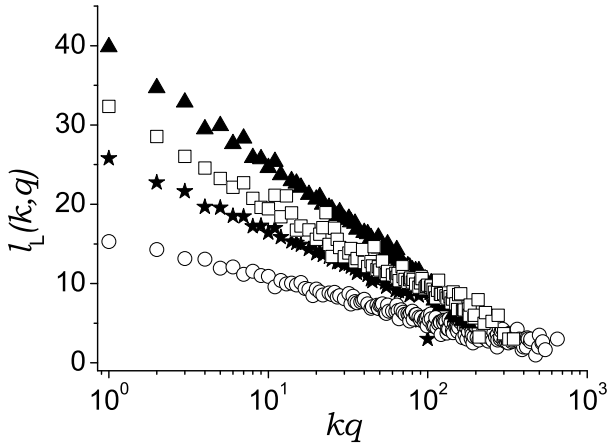


Figure 2.6: Mean $\mathbb{L}$-space paths $\ell_{\mathbb{L}}(k,q)$ as function of kq for the PTN of Berlin (stars), Hong Kong (circles), Rome (triangles), and Taipei (squares).

of the path may be reduced to a dependency on the degree of a single end node. We define $\ell(k)$, the mean shortest path between any node of degree k and other nodes of the network. For the majority of the cities analysed the dependence of the mean path $\ell_{\mathbb{L}}(k)$ on the node degree k in $\mathbb{L}$-space can be approximated by a power law

$$\ell_{\mathbb{L}}(k) \sim k^{-\alpha_{\mathbb{L}}}. \tag{2.9}$$

We find that the value of the exponent varies in the range $\alpha_{\mathbb{L}} = 0.17 - 0.27$. It is instructive to compare this result with results obtained in [41] for the same characteristics calculated for correlated growing networks. For deterministic scale-free networks $\ell(k)$ was found to be characterized by a logarithmic law with power-law corrections, whereas for stochastic scale-free networks $\ell(k)$ was shown to follow a logarithmic behaviour. Furthermore, networks with an exponential node-degree distribution displayed a linear law $\ell(k) \sim a - bk$. Obviously, the small values of the exponent $\alpha_{\mathbb{L}}$ found for the PTN in our study do not exclude a logarithmic law, however the linear dependence can be ruled out. Note, that within our sample of PTN one finds both scale-free and exponential node degree distributions. However, an essential difference between the construction principles of PTN and of the graphs of [41] is that the latter are so-called 'citation graphs' (where new connections do not emerge between already existing nodes), whereas there is no

such restriction for PTN.

In $\mathbb{P}$-space, the shortest path length $\ell_{i,j}$ gives the minimal number of routes required to be used in order to reach site j starting from the site i. The higher the node degree, the easier it is to access other routes in the network. Therefore, also in $\mathbb{P}$-space one expects a decrease of $\ell_{\mathbb{P}}(k)$ when k increases. Apart from an expected decrease we find a tendency to a power-law decay with small powers, sometimes almost indistinguishable from a logarithmic behavior. The value of the exponent $\alpha_{\mathbb{P}}$ varies in the interval $\alpha_{\mathbb{P}} = 0.09$ (for Sydney) to $\alpha_{\mathbb{P}} = 0.17$ (for Dallas) and is centered around $\alpha_{\mathbb{P}} = 0.12 - 0.13$. The mean path $\ell_{\mathbb{P}}(k,q)$ is found to decrease as a function of kq also in $\mathbb{P}$-space, but with much more pronounced scattering than in $\mathbb{L}$-space.

Concluding we note that the mean lengths of the shortest paths as function of the end node degrees show no special structure within the sample of PTN studied. In general the observed behavior does not significantly deviate from the logarithmic behavior that is expected for random graphs.

2.4.2 Betweenness centrality

To measure the importance of a given node with respect to different properties of a graph a number of so-called 'centrality measures' have been introduced quite long time ago in social sciences. In particular, the closeness centrality $\mathcal{C}_C(j)$ (Sabidussi 1966, [99])

$$\mathcal{C}_C(j) = \frac{1}{\sum_{t\in\mathcal{N}} \ell(j,t)}, \tag{2.10}$$

the graph centrality $\mathcal{C}_G(j)$ (Hage and Halary 1995, [64])

$$\mathcal{C}_G(j) = \frac{1}{\max_{t\in\mathcal{N}} \ell(j,t)}, \tag{2.11}$$

the stress centrality $\mathcal{C}_S(j)$ (Shimbel 1953, [102])

$$\mathcal{C}_S(j) = \sum_{s\neq j\neq t\in\mathcal{N}} \sigma_{st}(j), \tag{2.12}$$

and finally the most important betweenness centrality $\mathcal{C}_B(j)$ which measures the importance of a node with respect to the connectivity between other nodes of the network and was already mentioned in section 1.1.2 1.6

$$\mathcal{C}_B(j) = \sum_{s\neq j\neq t\in\mathcal{N}} \frac{\sigma_{st}(j)}{\sigma_{st}}.$$

It was introduced by Freeman in 1977 [55]. In equations (2.10-2.12, 1.6), $\ell(j,t)$ is the length of a shortest path between the nodes j,t that belong to the network $\mathcal{N}$, σ_{st} is the number of shortest paths between the two nodes $s,t \in \mathcal{N}$, and $\sigma_{st}(j)$ is the number of shortest paths between nodes s and t that go through the node j. A reliable algorithm to calculate betweenness centrality was proposed by Brandes [18] (all other centralities can be calculated within this algorithm). Numerical values of the mean betweenness centrality (1.6) are given in Table 2.2 for the $\mathbb{L}$-, $\mathbb{P}$- and $\mathbb{C}$-space graphs.

The betweenness centrality (1.6) of a given node measures the share of the shortest paths between nodes that this node mediates. It is obvious that a node with a high degree has a higher probability to be part of any path connecting other nodes. This relation between $\mathcal{C}_B$ and the node degree may be quantified by plotting the mean betweenness centrality $\langle \mathcal{C}_B(k) \rangle$ averaged among nodes with degree k as function of k. In Fig. 2.7 we present the corresponding results for the PTN of Paris in $\mathbb{L}$-, $\mathbb{C}$- and $\mathbb{P}$-, and $\mathbb{B}$-spaces. Especially well expressed is the betweenness-degree correlation in $\mathbb{L}$-space (Fig. 2.7a) and with somewhat less precision in $\mathbb{C}$-space (Fig. 2.7b). In both cases there is a clear tendency to a power law $\langle \mathcal{C}_B(k) \rangle \sim k^{\eta}$ with an exponent $\eta = 2-3$.

In the plots for both $\mathbb{B}$- and $\mathbb{P}$-spaces we observe the occurrence of two regimes which correspond to small and large degrees k. This separation however has a different origin in each of these cases. In the $\mathbb{B}$-space representation, the network consists of nodes of two types, route nodes and station nodes. Typically, station nodes are connected only to a low number of routes while there is a minimal number of stations per route. One may thus identify the low degree behavior as describing the betweenness of station nodes, while the high degree behavior corresponds to that of route nodes. In the overlap region of the two regimes one may observe that when having the same degree station nodes have a higher betweenness than route nodes.

In the $\mathbb{P}$-space representation on the other hand, the occurrence of two regimes is a feature of this representation. Stations that are part of only a single route and thus within the $\mathbb{P}$-graph belong only to the complete subgraph corresponding to this route (recall Fig. 2.2d) are not part of any shortest $\mathbb{P}$-space path between other nodes and have a betweenness centrality of $\mathcal{C}_B = 0$. The decreasing contribution of these stations to the average $\langle \mathcal{C}_B(k) \rangle$ leads a steep slope in the low degree regime. For degrees higher than the maximal route length these stations do no longer contribute and the slope rather describes the correlation between the degree and finite mean betweenness values. Instead of a steep slope in the low degree regime the study [103] observes a saturation; this may be due to the exclu-

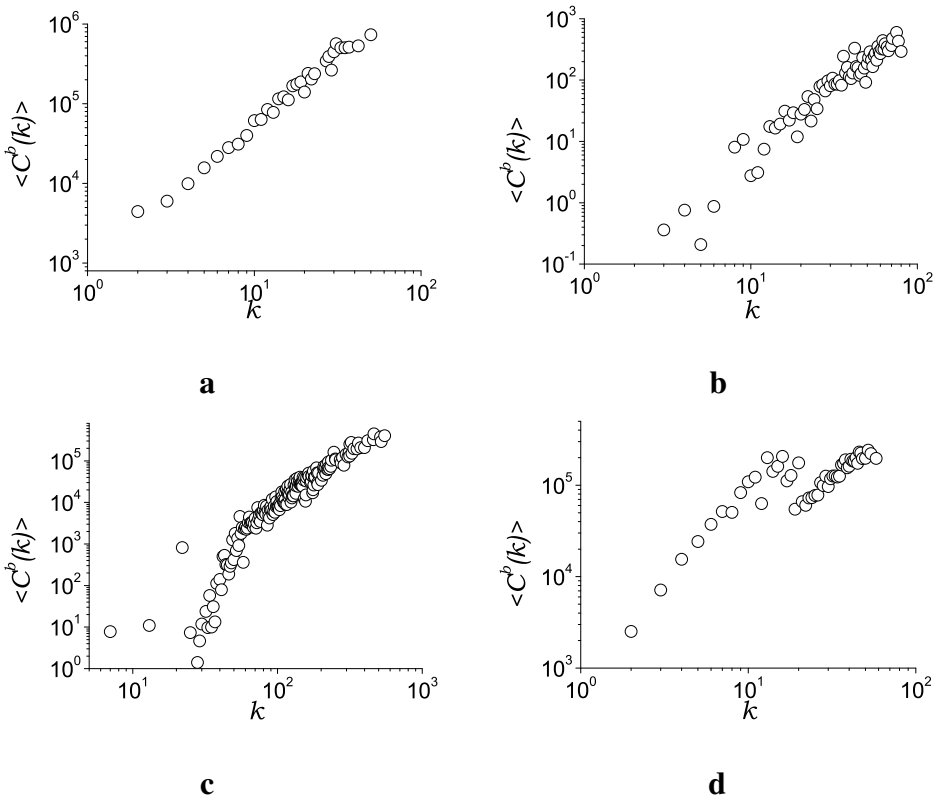


Figure 2.7: Mean betweenness centrality $\langle C_B(k) \rangle$ - degree k correlations for the PTN of Paris in (**a**) $\mathbb{L}$-, (**b**) $\mathbb{C}$-, (**c**) $\mathbb{P}$-, and (**d**) $\mathbb{B}$-spaces.

sion of the zero-betweenness nodes from the average. Very similar betweenness – degree relations as shown in Fig. 2.7 are found for most of the other cities in our sample with slightly varying quality of expression. We emphasize however, that this uniformity of the correlation between the degrees of the nodes and their respective betweenness is strictly speaking valid only for the average value $\langle C_B(k) \rangle$. When analysing the importance of individual nodes e.g. with respect to the vulnerability of the network against failure or attack the betweenness centrality turns out to be a much more sensitive measure than the node degree, see Ref. [52] and chapter 4.

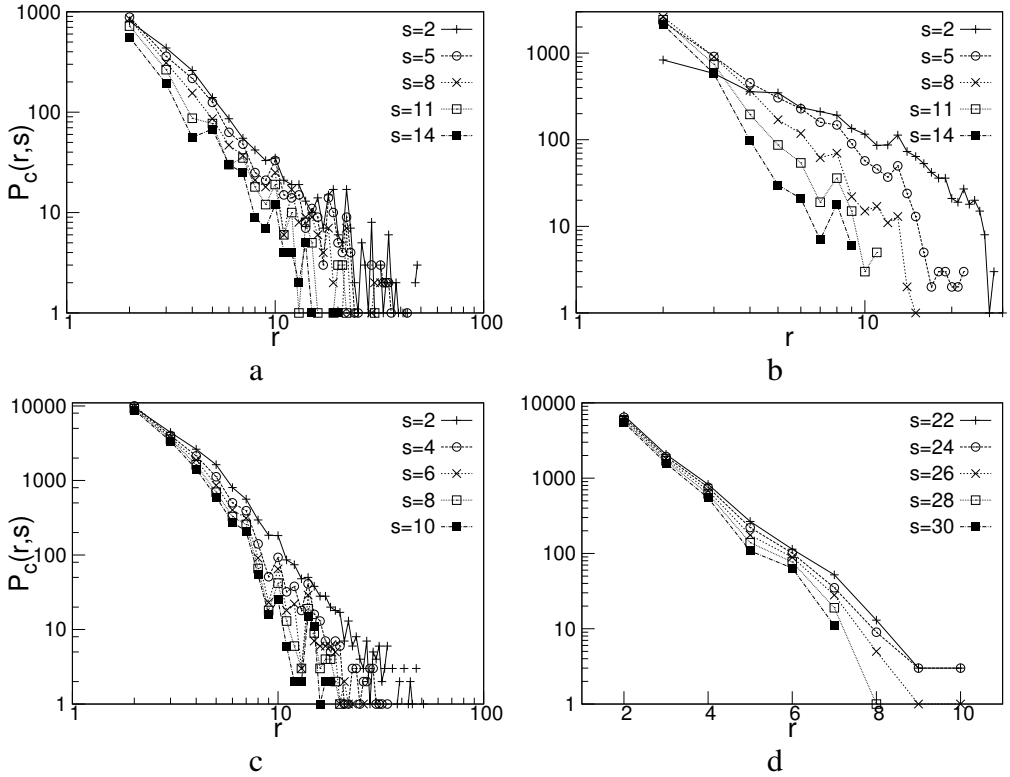


Figure 2.8: s-cumulative harness distribution $P_c(r,\hat{s})$ as function of r for fixed $\hat{s}$. Log-log for a) Istanbul and b) Taipei $\hat{s} = 2,5,8,11,14$. c) Log-log for Los Angeles $\hat{s} = 2,4,6,8,10$ and d) log-lin for Los Angeles $\hat{s} = 22,24,26,28,30$.

2.4.3 Harness

Besides the local and global properties of networks described above which can be defined in any type of network, there are some characteristics that are unique for PTN and networks with similar construction principles. Specific effects may be observed for networks on which a set of walks or paths is defined. A particularly striking example is the fact that as far as the routes share the same grid of streets and tracks often a number of routes will proceed in parallel along shorter or longer sequences of stations. Similar phenomena are observed in networks built with space consuming links such as cables [35], vessels [25], pipes [69], neurons [118], etc. In the present case this behavior may be easily worked out on the basis of sequences of stations serviced by each route. To quantify this behavior we use the recently introduced notion of network harness [48]. It is described by the harness

distribution $P(r,s)$: the number of sequences of s consecutive stations that are serviced by r parallel routes.

Similar to the node-degree distributions, we observe non-vanishing harness distributions $P(r,s) > 0$ even for long sequences s and high numbers of routes r. This is what we call a "strong" harness effect (examples are Sao Paolo, Hong Kong, Istanbul, Los Angeles, Rome, Sydney, Taipei, Moscow, London; some of them are shown in Fig. 2.8). For other PTN the maximal values of s and r with $P(r,s) > 0$ were found to be smaller than 10 (Berlin, Paris, Dallas, Duesseldorf, Hamburg) - this we call a "weak" harness effect (Fig. 2.9). It is important to note that the division into these two classes does not correlate with neither the average number of routes R in the PTN nor their average length S (Table 2.1).

Another result is that similar to the node-degree distributions we observe that the harness distribution $P(r,s)$ for some of the cities (Sao Paolo, Hong-Kong, Istanbul, Los Angeles, Rome, Sydney) may be described by a power law (1.9), $P(r,s) \sim r^{-\gamma_s}$, for fixed s, whereas the PTN of other cities (Taipei, Moscow, London) are better described by an exponential decay (1.8), $P(r,s) \sim e^{-r/\hat{r}_s}$, for fixed s. We illustrate this behavior in Figs. 2.8a,b showing the harness distribution for Istanbul and for Taipei. Obviously by such a criterium we can separate into different groups only those PTN that have "strong" harness effect, because for the others there is not enough data to be fitted. In some cases (e.g. for Rome and Los Angeles) there is a crossover between regimes (1.8) and (1.9) at larger s as shown in the case of Los Angeles (Fig. 2.8c). Here, one can see that for small values of s the results are better described by a power law (1.9). With increasing s a tendency to an exponential decay (1.8) appears (Fig. 2.8d). This is less obvious for other cities analysed, however in all cases the harness distribution $P(r,s)$ as function of r decays faster for longer sequence lengths and while also attaining a more pronounced curvature.

Note that in Fig. 2.8 we plot s-cumulative distributions $P_c(r,\hat{s})$ where a sequence with maximal length $s = 9$ will be counted once as a sequence of length $\hat{s} = 9$ and twice as a sequence of length $\hat{s} = 8$ etc:

$$P_c(r,\hat{s}) = \sum_{s=\hat{s}}^{S} (s+1-\hat{s}) P(r,s) \tag{2.13}$$

It may be surprising that these curves e.g. for Taipei (Fig. 2.8b) intersect for low values of r. We will discuss this effect below.

For PTN for which the harness distribution follows a power law (1.9) the corresponding exponents γ_s are found in the range of $\gamma_s = 2-4$. For those distributions with an exponential decay the scale $\hat{r}_s$ (see eq.(1.8)) varies in the range

$\hat{r}_s = 1.5 - 4$. The power laws observed for the behavior of $P(r,s)$ indicate a certain level of organization and planning which may be driven by the need to minimize the costs of infrastructure and secondly by the fact that points of interest tend to be clustered in certain locations of a city. Note that this effect may be seen as a result of the strong interdependence of the evolutions of both the city and its PTN. It may also be seen as a result of geographical or topological circumstances (see next section). We want to emphasize that the harness effect is a feature of the network given in terms of its routes but it is invisible in any of the complex network representations of public transport networks presented so far, such as L-space, P-space or B-space. It is possible, that the notion of harness may be useful also for the description of other networks with similar properties. On the one hand, the harness distribution is closely related to distributions of flow and load on the network. On the other hand, in the situation of space-consuming links (such as tracks, cables, neurons, pipes, vessels) the information about the harness behavior may be important with respect to the spatial optimisation of networks. A generalization may be readily formulated to account for real-world networks in which links (such as cables) are organized in parallel over a certain spatial distance. While for the PTN this distance is simply measured by the length of a sequence of stations, a more general measure would be the length of the contour along which these links proceed in parallel.

For the cities observed no correlation appears to occur between the harness distribution behavior and other well-known network characteristics that were analysed, as for example the node-degree distribution of PTN.

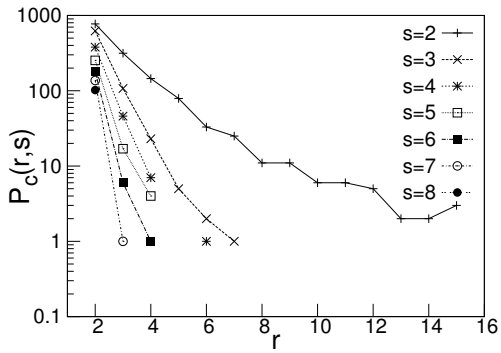


Figure 2.9: s-cumulative harness distribution $P_c(r,\hat{s})$ as function of r for fixed $\hat{s}$ for Paris ($\hat{s} = 2 - 8$). $P_c(r,\hat{s}) = 0$ if $\hat{s} > 2$ for $r > 7$ and if $\hat{s} > 7$ even for $r > 2$.

However, the extent to which harness properties are expressed may obvious-

ly play a role for the attack vulnerability of a PTN. Interestingly, a particular correlation was found between harness effect and the vulnerability, since our investigations of network vulnerability (see chapter 4) show that the Paris PTN is most resilient to any type of random or directed attacks (in terms of percolation concepts) among all analysed PTN. At the same time it exhibits the "weakest" behavior with respect to the harness effect (Fig. 2.9). One may expect such a result: routes that do not share the same streets are more resilient. However, for other cities no apparent correlation between the harness effect behavior and their vulnerability has been found so far.

It should be emphasized that with respect to network optimisation the harness property may at first seem completely counter-intuitive: Why should a route that is e.g. added to the network follow the path of previous, already existing routes, instead of exploring yet unserviced nearby areas? We may name at least two possible reasons for this empirically confirmed harness behavior: The first is the minimization of the cost for infrastructure which is most evident for means of transport that need tracks but relevant also with respect to maintaining e.g. bus stops. Other, more operation related reasons are those of interconnectivity minimizing the effort needed to change from one route to the other and of system redundancy, ensuring a higher transport frequency on important segments of the routes.

As noted above the interesting question to answer is: are there any structural evolution purposes behind this effect, or can it be found just as well within simple random scenarios.

Related unexpected behavior of the routes concerning their geographical embedding is observed and discussed in the following section.

2.4.4 Geographical embedding

So far, we have discussed the properties of PTN without reference to their geographical embedding. The fact that this subject has so far been left aside also by previous studies of PTN with respect to their complex network behavior, is due mainly to the lack of easily accessible data on the locations of stations and routes. Note, however, a study on the fractal dimension of railway networks, [15]. For the present work we have been able to obtain such data for stations of the Berlin PTN as well as for those of the metro subnetwork of Paris. For the Berlin network the positions of the stations were extracted in an automated way from interactive maps provided on the web-pages of the operator [129] which (invisibly) contain the geographical coordinates of the stations. For the Metro network of Paris these

coordinates were retrieved by hand using a free web based map service [130].

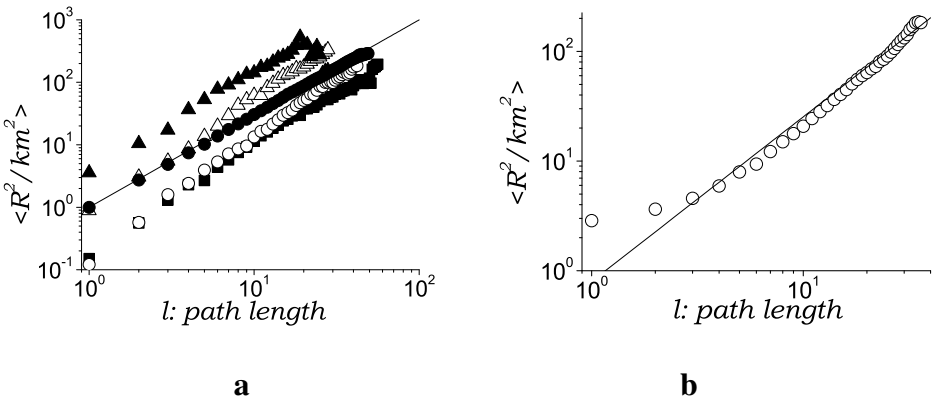


Figure 2.10: Distance – path length – relation $\langle \mathcal{R}^2(\ell) \rangle$ in comparison with that of a two dimensional self-avoiding walk (solid line, $\sim \ell^{3/2}$) for **a**) different means of transport within the Berlin PTN (bus(■), tram (○), u-bahn (△), s-bahn (▲), simulated city (●). and **b**) for the Paris metro network.

The question we pose here is, what is the distance $\mathcal{R}(\updownarrow)$ between initial and final stations of a passenger's journey travelling for ℓ stops on a single route? For routes optimising the time of passenger travel a naive consideration might lead to the expectation of distance growing linearly with path length ℓ at least on larger scales. Surprisingly, the empirical data show quite a different behavior (see Fig. 2.10). For all means of transport analysed within the Berlin PTN as well as for the metro network the dependence of the mean square distance $\langle \mathcal{R}^2(\ell) \rangle$ on ℓ is well described by a power law

$$\langle \mathcal{R}^2(\ell) \rangle \sim \ell^{2\nu} \tag{2.14}$$

with an exponent ν that is significantly smaller than one. For most transport routes this exponent appears to be near to $\nu = 3/4$, which is the well known self-avoiding walk (or Flory-) exponent in two dimensions [94] corresponding to a fractal dimension of $D \sim 1.33$. For the different Berlin subnetworks we find exponents ranging from $\nu = 0.82$ for the bus routes to $\nu = 0.9$ and 0.96 for the subway and tram routes. The s-bahn data is distorted due to a ring structure within this sub-network. The Paris metro data supports an exponent of $\nu = 0.82$ when excluding the short distance contributions. For comparison, the fractal dimensions D of some regional railway networks (not individual routes) reported in study [15] are of the order $D \sim 1.5 - 1.8$.

Self-avoiding walks, apart from observing the constraint of non-self-intersection evolve randomly. The fact that PT routes at least within the present sample appear to display the same scaling symmetry is quite unexpected. In particular, this behavior seems to be at odds with the requirement of minimizing passengers travelling time between origin to destination. The latter argument, however, ignores the time passengers spend walking to the initial and from the final stations. Including these, one understands the need for the routes to cover larger areas by meandering through neighborhoods. Given the requirements for a PTN to cover a metropolitan area with a limited number of routes while simultaneously offering fast transport across the city one may speculate that routes scaling like SAWs may present an optimal solution. Further research on this subject is presented in chapter 3.

2.5 Conclusions

In this chapter we intended to present a systematic survey of statistical properties of PTN based on the data for cities of so far unexplored network size. Especially helpful in our analysis was the use of different network representations (different spaces, introduced in section 2.2). Whereas former PTN studies used some of these, here within a systematic approach we calculate PTN characteristics as they show up in all $\mathbb{L}$-, $\mathbb{P}$-, $\mathbb{C}$-, and $\mathbb{B}$-spaces.

The networks under consideration appear to be strongly correlated small-world structures with high values of clustering coefficients and comparatively low mean shortest path values. Standard network characteristics that we find in these various representations correspond to features a passenger is interested in when using public transport. For example, any two stops in Paris are on the average separated by $\langle \ell_{\mathbb{L}} \rangle - 1 = 5.4$ stations (with a maximal value of 27) and to travel between them one should do $\langle \ell_{\mathbb{P}} \rangle - 1 = 1.7$ changes on average. The power-law node degree distributions observed for many networks in $\mathbb{L}$- and for some in $\mathbb{P}$-space give strong evidence of correlations within these networks. However, for the properties of degree distributions as well as for features of these networks, such as clustering, assortativity and others we find considerable diversity in their expression. Recent work on urban street networks found classifications that discriminate between properties of different classes of city organization. For the present sample of PTN however, we conclude that there is no simple division of the PTN we studied into well defined groups as e.g. seen for street and canal networks [24, 114] where a division into a few groups was found (however analysing only small areas of city

maps). This result is far from obvious: one might have expected that networks all set up in large urban areas and serving an almost identical purpose would turn out to display strongly aligned properties. However, this diversity is an empirical fact and one that would remain hidden if we had restricted our observations to only a handful of measurements.

Beyond traditional network characteristics there are specific features unique for PTN and networks with similar construction principles that we have addressed. In particular, public transport routes are often found to proceed in parallel for a sequence of stations. While the very fact that several routes should follow the same path may seem counter-intuitive (why should a route retrace another's path instead of exploring nearby unserviced areas?), we have quantified this behavior in terms of the harness distribution and given possible explanations noting costs of infrastructure, and operational advantages such as system redundancy. The harness concept may also be useful for a quantitative description of other embedded networks with real space links such as cables, pipes, or neurons etc.

Moreover, our analysis of the geographical data for Berlin and Paris reveals a self-avoiding walk scaling of PTN routes a fact strongly supported by the empirical study which again appears to be counter-intuitive (should a line not be straight to minimize time of travel). We give possible explanation speculating that this shape of the routes may result from an optimisation with respect to total passenger travelling time, area coverage and costs of operation.

It has to be noted that analysing degree distributions for scale free properties only simple least mean square fits were performed. Therefore it would be interesting to apply more elaborated tools to quantify the quality of fits, using more recent methods [29].

With the results of the above empirical analysis at hand, we are in the position to proceed further and analyse PTN properties by introducing different models.

Chapter 3

Public transport network models

In this chapter we will further analyse different properties of PTN introducing models that may reproduce some of their characteristics behaviour. A particular feature of the models considered below is that opposite to the majority of complex network models, where a network grows by adding individual nodes, in our case the growth is in terms of PT lines - which are sequences (chains) of nodes. Being motivated by the results of the empirical analysis of several real-world PTN presented in chapter 2, we will mainly use random walks to describe such chains. In the most general case (section 3.1) we will consider mutually interacting self-avoiding walks in 2d. However, being the most realistic, such a model does not allow for an analytic solution. Therefore, we will first consider and solve analytically a 1d model (section 3.2). In this case we will be primarily interested in description of the PTN harness effect. The latter property will be discussed within the model of non-interacting walks in 2d again in section 3.3. Some of the results presented in this chapter were published in papers Refs. [50, 51, 12].

3.1 Mutually interacting self-avoiding walks in 2d

3.1.1 Motivation and description of the model

Having at hand the above described wealth of empirical data and analysis with respect to typical scenarios found in a variety of real-world PTN we feel in the position to propose a model that may capture the characteristic features of these networks. In view of the diversity found in our sample, it would be in vein to try to construct a model that quantitatively reproduces the data of a given city. The

aim of the present model is to show that a few simple rules and a low number of parameters suffice to generate PTN that display profiles which with respect to most observables that are within the range of those found in real world PTN. Nonetheless it should be capable of discriminating between some of the various scenarios observed.

Essential basic properties of PTN that we intend to implement or reproduce within our model are the following: a) the model is to be based on routes and stations and allows for $\mathbb{L}$- $\mathbb{P}$- $\mathbb{C}$- and $\mathbb{B}$-space representations; b) the model should be embedded in two dimensions and reproduce the SAW (self-avoiding walk) scaling behavior of the routes; c) the model should be able to generate realistic degree distributions; d) the model must generate realistic harness distributions.

If we were only to reproduce the degree distribution of the network, standard models such as random networks [40, 89] or preferential attachment type models [7, 8, 5, 81, 90, 79, 98] would suffice. The evolution of such networks however is based on the attachment of nodes. For the description of PTN the concept of routes as finite sequences of stations is essential [67, 6, 48, 51] and allows for the representation with respect to the spaces defined above. Moreover, taking a route as the essential element of PTN growth allows to account for the bipartite structure of this network [100, 121, 27, 61]. Therefore, the growth dynamics in terms of routes will be a central ingredient of our model. Another obvious requirement is the embedding of this model in two-dimensional space. To simplify matters we will restrict the model to a two-dimensional grid, in particular to square lattice. Both the observations of power law degree distributions as well as the occurrence of the corresponding harness distributions described above indicate a preference of routes to service common stations (i.e. an attraction between routes).

Let us describe our model in more detail. As noticed above, a route will be modeled as a sequence of stations that are adjacent nodes on a two-dimensional square lattice. Following the observation of SAW scaling symmetry for the geographical embedding we choose each PTN route to be a self-avoiding walk. To incorporate all the above features the model is set up as follows. A model PTN consists of R routes each with S stations constructed on a possibly periodic $X \times X$ square lattice. The dynamics of the route generation adheres to the following rules:

- 1. Construct the first route as a SAW of S lattice sites.

- 2. Construct the $R-1$ subsequent routes as SAWs with the following preferential attachment rules:

a) choose a terminal station at $\vec{x}_0$ with probability

$$p \sim k_{\vec{x}_0} + a/X^2; \tag{3.1}$$

b) choose any subsequent station $\vec{x}$ of the route with probability

$$p \sim k_{\vec{x}} + b. \tag{3.2}$$

In (3.1), (3.2) $k_{\vec{x}}$ is the number of times the lattice site $\vec{x}$ has been visited before (the number of routes that pass through $\vec{x}$). Note, that to ensure the SAW property any route that intersects itself is discarded and its construction is restarted with step 2a).

3.1.2 Global topology of model PTN

Let us first investigate the global topology of this model as function of its parameters. We first fix both the number of routes R and the number of stations S per route as well as the size of the lattice X. This leaves us with essentially two parameters a and b, from equations (3.1), (3.2). Dependencies on R and S will be studied below.

For the real-world PTN as studied in the previous sections, almost all stations belong to a single component, GCC, with the possible exception of a very small number of routes. Within the network however we often observe the harness effect of several routes proceeding in parallel for a sequence of stations. Let us first investigate from a global point of view which parameters a and b reproduce realistic maps of PTN. In Fig. 3.1 we show simulated PTN on lattices 300×300 for $R = 1024$, $S = 64$ and different values of the parameters a and b. Each route is represented by a continuous line tracing the path along its sequence of stations. For representation purposes, parallel routes are shown slightly shifted. Thus, the line thickness and intensity of colors indicate the density of the routes.

The parameter a quantifies the possibility to start a new route outside the existing network. For vanishing $a = 0$ the resulting network always consists of a single connected component, while for finite values of a a few or many disconnected components may occur. The results for $a = 0$ and varying b parameters are independent of the lattice size X provided X is sufficiently large to accommodate the network without boundary effects. Parameter b governs the evolution of each single subsequent route. If $a = 0$ and $b = 0$ the only allowed sites according to equations (3.1), (3.2) are those of the first SAW route as far as the choice is restricted to sites x with a finite number k_x of previous visits. The shape variation of

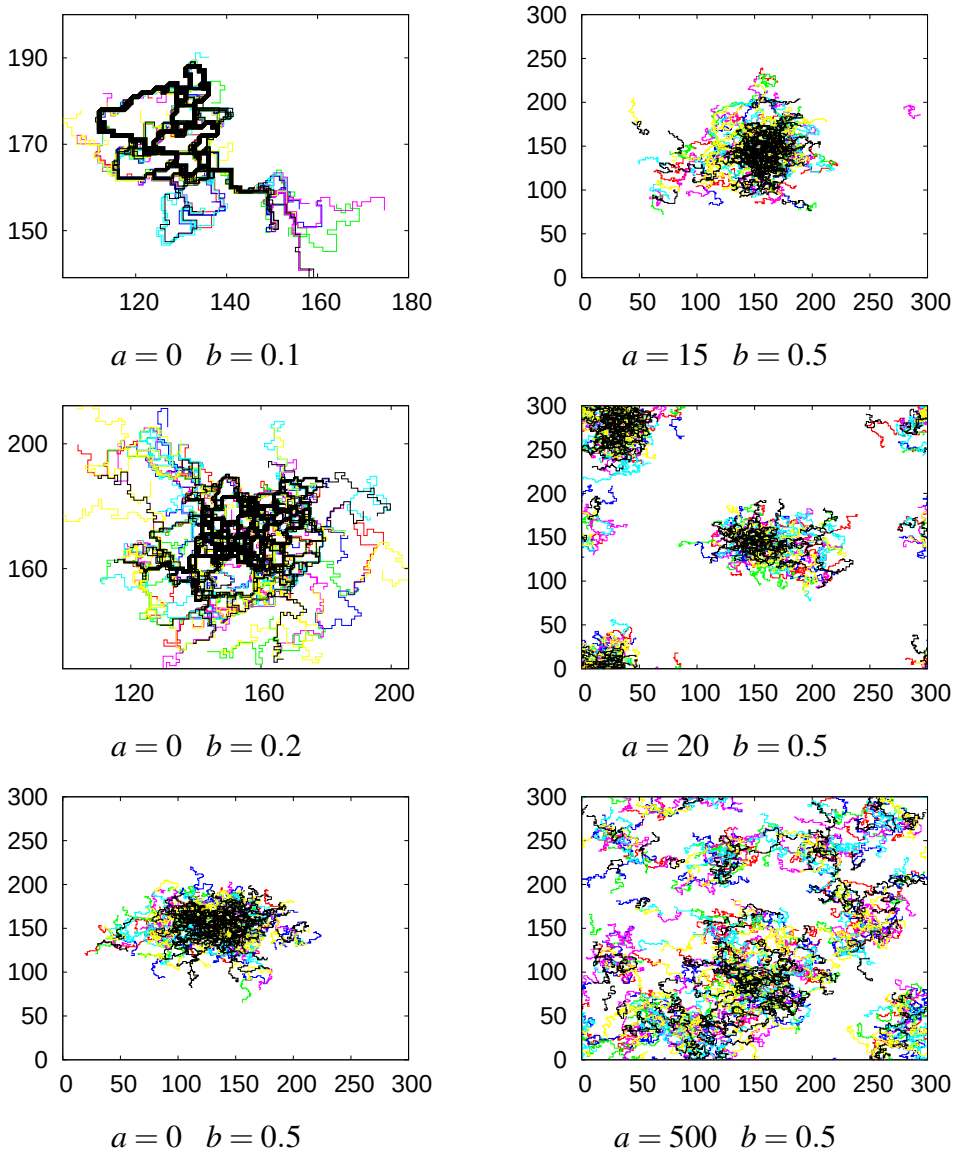


Figure 3.1: PTN maps of different simulated cities of size 300×300 with $R = 1024$ routes of $S = 64$ stations each (color online). First column: $a = 0$, $b = 0.1 - 0.5$. Second column: $b = 0.5$, $a = 15 - 500$. Increasing b the routes cover more and more area. Increasing a leads to a breakup of the network.

the simulated PTN as b is increased for fixed $a = 0$ is shown in the first column of Fig. 3.1. For small values of $b = 0 - 0.1$ almost all routes of the simulated PTN follow the same path with only a few deviations. Shifting b to $b = 0.2$ the area covered by the routes increases while the majority of the routes are concentrated on a small number of paths. Further shifting b to $b = 0.5$ and beyond we find a wider distributed coverage with the central part of the network remaining the most densely covered area. This is due to the non-equilibrium growth process described by equations (3.1), (3.2).

When introducing a finite a parameter, new routes may be started anywhere on the lattice which results in a lattice size dependency. To partly compensate for this, the impact of a is normalized by X^2 in (3.1). The variation of the simulated PTN maps for increasing a and fixed $b = 0.5$ is shown in the second column of Fig. 3.1. For $a < 15$ one observes the formation of a single large cluster with only a few individual routes occurring outside this cluster. Slightly increasing a beyond $a = 15$ one finds a sharp transition to a situation with several (two or more) clusters. For much larger values of a the number of clusters further increases and the situation becomes more and more homogeneous: the routes tend to cover all available lattice space area.

3.1.3 Statistical characteristics of model PTN

From the above qualitative investigation we conclude that realistic PTN maps are obtained for small or vanishing a and $b \geq 0.5$. In the following we will fix $a = 0$ and X large enough as discussed above.

To quantitatively investigate the behavior of the simulated networks as function of the remaining parameters including R and S let us now compare their statistical characteristics with those we have empirically obtained for real-world networks. In Table 3.1 we have chosen to list the same characteristics of the simulated PTN as selected for the real-world networks in Table 2.2. To provide for additional checks of the correlations between simulated and real-world networks, we present the characteristics in all $\mathbb{L}$-, $\mathbb{P}$-, and $\mathbb{C}$-spaces. Let us note that our choice of the underlying grid to be a square lattice limits the number of nearest neighbors of a given station in $\mathbb{L}$-space to $k_{\mathbb{L}} \leq 4$. Moreover, as far as no direct links between these neighbors occur, the clustering coefficient in $\mathbb{L}$-space vanishes, $c_{\mathbb{L}} = 0$. Nonetheless, as we discuss below, both characteristics display nontrivial behavior similar to real-world networks when measured for $\mathbb{P}$- and $\mathbb{C}$-spaces.

As noted above we choose a vanishing parameter $a = 0$ and $b = 0.5$ and

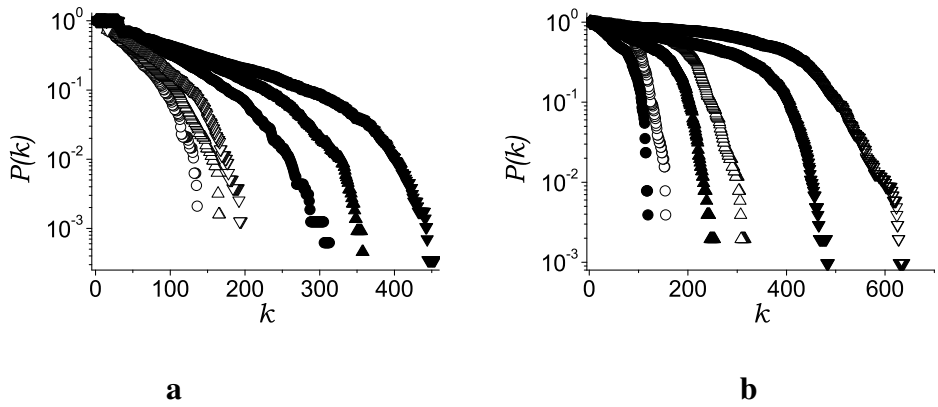


Figure 3.2: Cumulative node degree distributions $P(k)$ (2.1) for several simulated PTN in (**a**) $\mathbb{P}$- and (**b**) $\mathbb{C}$-spaces. $R = 256, S = 16$ (○), $R = 256, S = 32$ (●), $R = 512, S = 16$ (△), $R = 512, S = 32$ (▲), $R = 1024, S = 16$ (▽), $R = 1024, S = 32$ (▼).

for comparison $b = 5.0$. The data shown in the Table was obtained for simulated PTN of different numbers of routes, $R = 256, 512, 1024$ and route lengths $S = 16, 32, 64$. In the range of parameters covered in the Table we observe only weak changes of the various characteristics. Natural trends are that with the increase of the number of routes R the maximal and mean shortest path length increases in all spaces. This is most pronounced in $\mathbb{L}$-space, while it is weakest in $\mathbb{C}$-space. A similar increase is observed in $\mathbb{L}$-space when increasing the number of stations S per route. Choosing the values of R in the range $R = 256 - 1024$ and $S = 16$, $S = 32$ the average and maximal values of the characteristics studied here are found within the ranges seen for real-world PTN, see Table 2.2. More detailed information is contained in the distributions of these characteristics and their correlations.

Let us examine the node degree distributions of some selected PTN. As explained above, the $\mathbb{L}$-space degrees are restricted by the geometry of the underlying square lattice. Thus we may observe non-trivial distributions only in $\mathbb{P}$-, $\mathbb{C}$-, and $\mathbb{B}$-spaces. The cumulative node degree distributions in $\mathbb{P}$-space are shown in Fig. 3.2a. All these distributions display two regions each governed by an exponential decay with a separate scale. Note, that increasing both S and R leads to an increase of the ranges over which these regions extend. This is in line with the results for real world PTN found in previous studies [103, 119] as well as in

R	S	b	$\langle k_{\mathbb{L}} \rangle$	$\kappa_{\mathbb{L}}$	$\ell_{\mathbb{L}}^{\max}$	$\langle \ell_{\mathbb{L}} \rangle$	$\langle \mathcal{C}_{\mathbb{L}B} \rangle$	$\langle k_{\mathbb{P}} \rangle$	$\kappa_{\mathbb{P}}$	$\ell_{\mathbb{P}}^{\max}$	$\langle \ell_{\mathbb{P}} \rangle$	$\langle \mathcal{C}_{\mathbb{P}B} \rangle$	$c_{\mathbb{P}}$	$\langle k_{\mathbb{C}} \rangle$	$\kappa_{\mathbb{C}}$	$\ell_{\mathbb{C}}^{\max}$	$\langle \ell_{\mathbb{C}} \rangle$	$\langle \mathcal{C}_{\mathbb{C}B} \rangle$	$c_{\mathbb{C}}$
256	16	0.5	2.92	1.66	61	20.8	4.7×10^3	44.15	3.18	7	3.0	4.7×10^2	7.98	86.39	1.36	6	1.9	1.2×10^2	2.22
256	16	5.0	2.99	1.74	80	21.7	7.5×10^3	42.95	3.76	9	3.4	8.8×10^2	11.7	59.96	1.99	8	2.2	1.5×10^2	2.79
256	32	0.5	2.76	1.60	127	38.1	3.0×10^4	84.45	4.32	8	3.3	1.9×10^3	13.6	60.51	1.75	7	2.2	1.6×10^2	2.90
256	32	5.0	2.90	1.72	177	43.1	5.3×10^4	74.24	5.22	10	4.0	3.8×10^3	23.7	33.06	2.69	9	2.8	2.3×10^2	4.55
512	16	0.5	2.95	1.68	73	22.5	6.7×10^3	50.07	3.39	7	3.1	6.5×10^2	9.14	169.7	1.44	6	1.9	2.3×10^2	2.25
512	16	5.0	3.12	1.78	80	23.3	1.0×10^4	51.56	3.79	10	3.5	1.2×10^3	12.3	115.3	2.24	9	2.1	2.9×10^2	2.88
512	32	0.5	2.83	1.63	166	44.2	4.7×10^4	99.53	4.56	10	3.6	2.8×10^3	15.7	118.4	2.03	9	2.2	3.0×10^2	2.92
512	32	5.0	3.12	1.79	175	44.6	7.2×10^4	97.05	5.37	9	3.9	4.7×10^3	22.2	60.36	3.08	8	2.7	4.4×10^2	5.04
1024	64	0.5	2.86	1.66	325	80.7	3.3×10^5	242.2	6.32	9	3.7	1.1×10^4	23.4	213.3	2.42	8	2.2	6.1×10^2	3.10
1024	64	1.0	2.97	1.72	355	88.5	4.8×10^5	222.2	6.74	12	4.2	1.7×10^4	32.4	143.9	2.97	11	2.5	7.9×10^2	4.39

Table 3.1: Characteristics of the simulated PTN with $X = 300$, $a = 0$ for different parameters R, S, and b. The rest of notations as in Table 2.2.

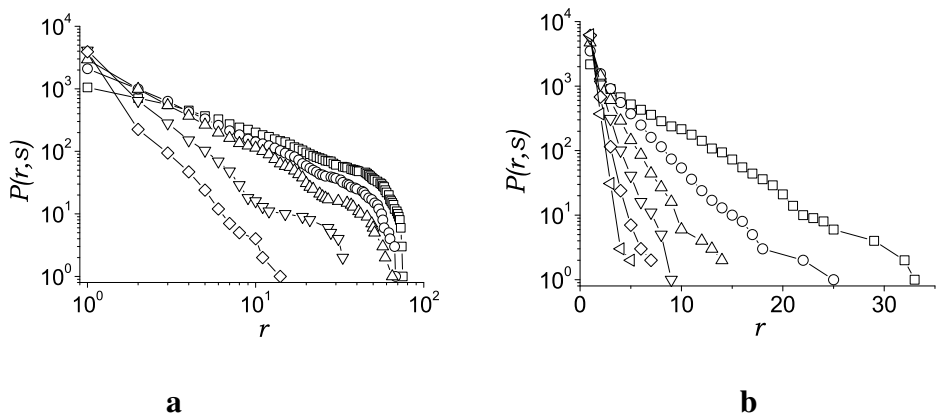


Figure 3.3: s-cumulative harness distributions $P_c(r,\hat{s})$ for the simulated PTN with $R = 256$, $S = 32$. **a**: $a = 0$, $b = 0.2$, $s = 2(\square)$, $4(\circ)$, $6(\triangle)$, $11(\triangledown)$, $16(\lozenge)$. **b**: $a = 0$, $b = 1.0$, $s = 2(\square)$, $3(\circ)$, $4(\triangle)$, $5(\triangledown)$, $6(\lozenge)$, $7(\triangleleft)$. Compare with plots in Fig. 2.8 for the real-world networks.

section 2.3.1. Within the parameter ranges chosen here the current model does not seem to attain a power law node degree distribution in $\mathbb{P}$-space.

Comparing the $\mathbb{C}$-space node degree distributions for real-world and simulated PTN (Figs. 2.3 and 3.2, correspondingly) one again finds a definite tendency to an exponential behavior with two different scales in both cases. As can be expected we observe that the scale of the exponential decay increases with the number of routes R while it decreases with the number of stations per route S.

s-cumulative harness distributions $P_c(r,\hat{s})$ for two simulated networks with different values of the parameter b ($b = 0.2$, $b = 1.0$) are shown in Fig. 3.3. These appear to reproduce the harness behavior of real world networks as given in Fig. 2.8. Both exponential and scale-free behavior as observed for the real-world PTN is found. A prominent feature demonstrated by Fig. 3.3 is that one can tune the decay behavior by changing the parameter b. For small values of b the probability of a route to proceed in parallel with other routes is high. Thus for small b the $P(r,s)$ distribution shows a high probability for the formation of 'hubs' of parallel routes as reflected by its power-law decay distribution. For larger b such hubs are suppressed as shown by the exponential decay of their distribution.

Summarizing, the comparison of the statistical characteristics of real world networks with those of simulated ones one can definitely state that the model proposed above captures many essential features of real world PTN. This is especially

evident if one includes into the the comparison different network representations (different spaces) as performed above.

However it is interesting to investigate if some of characteristics have such behaviour results mainly from the chain structure and thus can be represented in a much simpler model. In particular further we will try to represent the harness effect in one-dimensional space.

3.2 Modelling in 1d: analytic results and simulation

Let us first investigate a network model with routes placed randomly in one dimensional space. Although being very simple this model can mimic a harness effect and as we will see below, it allows an analytical solution. The model is formulated in the following way:

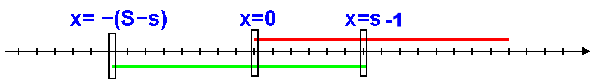


Figure 3.4: $R = 2$ routes given as simple sequences of $S = 15$ consecutive sites are placed at random on a line with N sites.

The left terminals of R routes of length S are placed at random on a line with N sites, with periodic boundary conditions. E.g. in Fig. 3.4 we show two routes of length $S = 15$ with left terminals at $x = 0$ and $x = -8$. We define the route density as $\rho = R/N$ routes per site.

The distribution of left-terminals on a given site, e.g. site x will be $P_x(r)$: the probability that r routes have their left-terminal on site x

$$P_x(r) = \binom{R}{r} \cdot (\frac{1}{N})^r \cdot (1 - \frac{1}{N})^{R-r}, \tag{3.3}$$

where the first term counts the number of ways to select r of R routes, the second term is the probability that the r left terminals lie on site x and the third term is the probability that no left terminal of an unselected route lies on site x. In other words, by definition $P_x(r)$ is a binomial distribution. For $N \to \infty$, but fixed $\rho = R/N$ this distribution has the limiting behavior of a Poisson distribution:

$$P_x(r) \approx e^{-\rho} \cdot \frac{\rho^r}{r!}. \tag{3.4}$$

Let us now calculate the probability that there is a sequence of maximal length s and maximal width r of r routes in parallel between sites $x = 0$ and $x = s$. This implies that at least one of the r routes starts at $x = 0$ and at least one of the routes ends at $x = s - 1$. The latter route then starts at $x = -(S - s) \equiv -\bar{s}$ (Fig. 3.4). The other $r - 2$ routes may start anywhere in between $-\bar{s} \leq x \leq 0$. We denote the number of routes starting at $x = 0$ as r_0, those starting at $x < 0$ as r_{-x}.

In the limit $R \gg r$ and $N \gg \bar{s}$ we may consider $P_0(r_0)$, $P_1(r_1)$, ... ,$P_{\bar{s}}(r_{\bar{s}})$ as independent probabilities (this is not true for small systems, however, as we see later the correlation between these probabilities is negligible for the cases studied here. The overall probability to find a sequence of length s and width r starting at $x = 0$ is then the sum over all combinations leading to the result:

$$\begin{aligned} P_0(r,s) &= \sum_{\{r_i\}, r_0 \geq 1, r_{\bar{s}} \geq 1}^{(r)} P(r_0) \cdot P(r_1) \cdot \ldots \cdot P(r_{\bar{s}}) = \\ &= \sum_{\{r_i\}, r_0 \geq 1, r_{\bar{s}} \geq 1}^{(r)} e^{-\bar{s}\rho} \cdot \frac{\rho^{r_0}}{r_0!} \cdot \frac{\rho^{r_1}}{r_1!} \cdot \ldots \cdot \frac{\rho^{r_{\bar{s}}}}{r_{\bar{s}}!} = \\ &= e^{-\bar{s}\rho} \cdot \rho^r \cdot \sum_{\{r_i\}, r_0 \geq 1, r_{\bar{s}} \geq 1}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!}, \end{aligned} \tag{3.5}$$

where $\sum_{\{r_i\}}^{(r)}$ denotes a sum over $\{r_i\}$ with $r = r_0 + r_1 + \ldots + r_{\bar{s}}$.

Now, without the conditions $r_0 \geq 1$ and $r_{\bar{s}} \geq 1$ this sum can be derived from:

$$(\bar{s}+1)^r = (1+1+1+\ldots+1)^r = \sum_{\{r_i\}}^{(r)} \frac{r!}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!}. \tag{3.6}$$

The sum with these conditions however can be written as:

$$\begin{aligned} &\sum_{\{r_i\}, r_0 \geq 1, r_{\bar{s}} \geq 1}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!} = \sum_{\{r_i\}}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!} - \\ &- \sum_{\{r_i\}, r_0 = 0}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!} - \sum_{\{r_i\}, r_{\bar{s}} = 0}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!} + \\ &+ \sum_{\{r_i\}, r_0 = 0, r_{\bar{s}} = 0}^{(r)} \frac{1}{r_0! \cdot r_1! \cdot \ldots \cdot r_{\bar{s}}!} = \frac{(\bar{s}+1)^r}{r!} - 2\frac{(\bar{s})^r}{r!} + \frac{(\bar{s})^r}{r!}. \end{aligned} \tag{3.7}$$

Thus:

$$P_0(r,s) = e^{(-\bar{s}\cdot\rho)} \cdot \rho^r \cdot [(\bar{s}+1)^r - 2(\bar{s})^r + (\bar{s}-1)^r]/r!. \tag{3.8}$$

With this formula we count sequences that start at $x = 0$. To receive the overall probability this is to be multiplied by N.

$$P(r,s) = N \cdot e^{(-\bar{s}\cdot\rho)} \cdot \rho^r \cdot [(\bar{s}+1)^r - 2(\bar{s})^r + (\bar{s}-1)^r]/r!. \tag{3.9}$$

Simple arithmetic allows us to calculate the s-cumulative distributions $P_c(r,\hat{s})$, see equation(2.13).

As mentioned above, the probabilities we use are appropriate for infinite systems. To test their validity for finite cases we performed some simple simulations. It turns out that the average results fit this formula with very good accuracy even for small N, for example $N = 10$ and of course for any larger values of N (Fig. 3.5a). For a large range of parameters R, S, N the behavior of $P(r,s)$ looks similar to what is shown in Fig. 3.5b. For high overall density $\rho \cdot S$ we observe that curves for different s intersect at small r as for some real-world PTN. In one dimension, this effect has the following explanation: when the space is overcrowded one will in general find more than two routes to overlap for small sequences of stations.

This model has three parameters: the number of routes R, the route length S and the number of sites N. As is obvious from equation 3.9 the harness distribution $P(r,s)$ for all r and s is close to a Poisson decay (1.7) for any set of parameters. Therefore PTN with an exponential behavior of the harness distribution $P(r,s)$ may be compared with the results of the one-dimensional approach. As an example we compare the normalized harness distribution for Moscow and the one-dimensional set of lines (Fig. 3.5c,d), where the number of routes R and the route length S were chosen to match those of the Moscow PTN (S set to the average route length).

However, the quantitative results for all observed PTN are several orders of magnitude higher than the result obtained with equation 3.9 in one dimension for the same R and S (for any N). Furthermore, PTN that show a power law behavior (1.9) are even qualitatively different from the random 1D approach. Another difference is that the harness distribution curves for different s are very similar in shape and both slope and curvature vary much less than for the PTN harness distributions. We will propose possible explanation for this difference further.

In the following we test a two-dimensional model with the simple simulations.

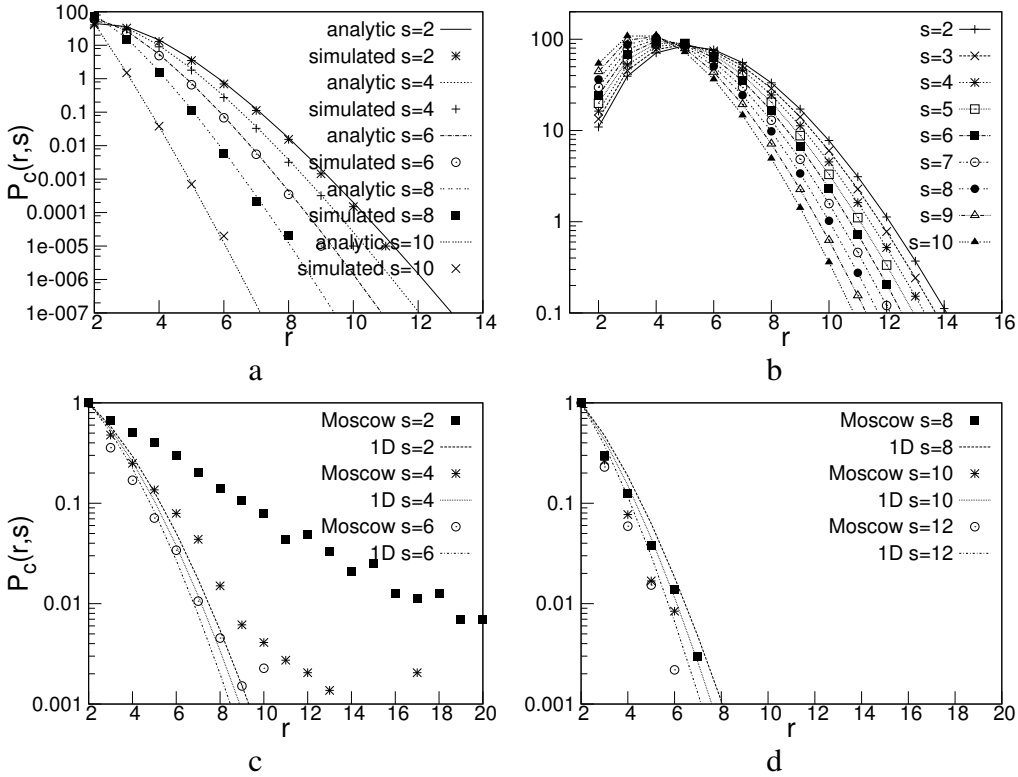


Figure 3.5: s-cumulative harness distribution $P_c(r,\hat{s})$ as function of r for fixed $\hat{s}$. Log-lin scale. a) Comparing the analytical solution and numerical simulations for $N = 10000$, $R = 1000$, S=10; b) for analytical solution for $N = 10000$, $R = 2000$, S=20; c,d) comparing the analytical solution with empirical results for the Moscow PTN ($R_{an} = R_{Moscow} = 679$, $S_{an} = \bar{S}_{Moscow} = 22$, $N = 5250$) for different s normalized by $P_c(2,s)$.

3.3 Non-interacting walks in 2d

It is obvious that simply throwing random lines parallel to the axis' of a 2d square lattice with periodic boundary conditions will lead to the original 1d problem: If the lattice has $X \times X$ sites one would get $2X$ independent one-dimensional systems. However, it is not a priory clear what results one will find for more general sets of walks on a 2d square lattice.

To work this out, we implemented the following simulations. We work on a 2d $X \times X$ square lattice with periodic boundary conditions. On this lattice we chose a

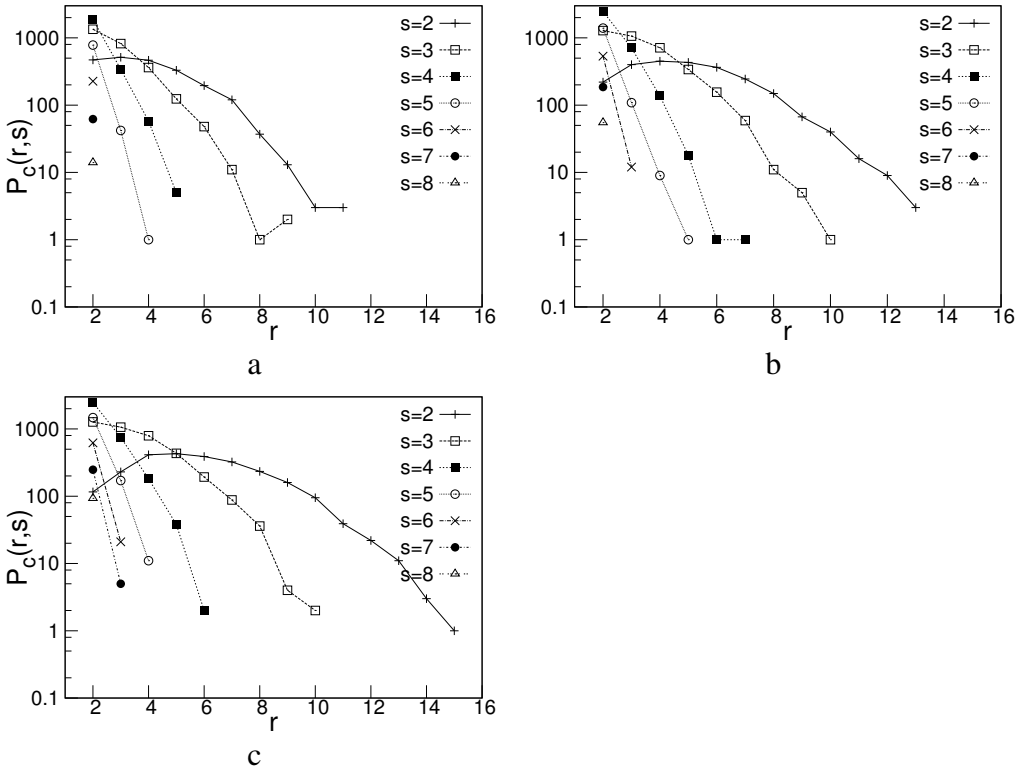


Figure 3.6: s-cumulative harness distribution $P_c(r,\hat{s})$ as function of r for fixed $\hat{s}$ ($\hat{s} = 2-8$), for $R = 500$, $S = 30$, $X = 50$. Routes are generated as: a)RW b)NRRW c)SAW. $P(r,s) = 0$ if $s > 2$ for $r > 11$ and if $s > 7$ even for $r > 2$.

set of R walks each of length S (number of steps plus 1). The routes are built either as random walks (RW), non-reversal random walks (NRRW) that cannot reverse the previous step, or self-avoiding random walks (SAW), that may not intersect themselves.

These models have three parameters: the number of routes R, the route length S and the lattice size X. We choose the first two parameters to match those of different real PTN.

Let us summarize here some of the main features of the harness distribution $P_c(r,\hat{s})$ of these models. Besides the finding that the harness effect is "weak", some similarity between the the harness effects seen in the three models is observed (Fig 3.6a,b,c). Curvature and slope evolve in a similar way. Also intersections between the curves for different s are found to occur at lower values of r in

all cases. Differences are that the RW-generated networks demonstrate a "weaker" harness effect, while NRRW- and SAW-generated networks result in harness distributions $P_c(r,\hat{s})$ of similar order of magnitude.

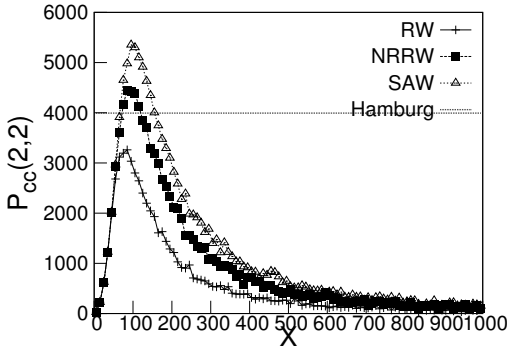


Figure 3.7: Situations observed for the r- and s-cumulative harness distribution $P_{cc}(\hat{r},\hat{s})$ at $\hat{r}=2$, $\hat{s}=2$ as function of lattice size X. For all three of models with $R = R_{Hamburg}$, $S = \bar{S}_{Hamburg}$. The empirical value of $P_{cc}(2,2)$ for Hamburg is shown for comparison.

It turns out that for fixed R and S, increasing the lattice size X, the harness distributions $P(r,s)$, for all fixed $r < R$ and $s < S$ show non-monotonous behavior (Fig. 3.7). As function of X it first increases, and then after reaching a maximum it starts to decrease. Comparing with the empirical values found for real PTN we observe, that for some PTN the empirical values even for small r and s are significantly larger than the maximum that could be obtained with RW. For NRRW or SAW the empirical values are within the observed range, however only for a small interval of X.

This proves numerically the not surprising observation, that with any of the proposed random or quasi-random walks only a very "weak" harness effect may be obtained. In turn, this strongly indicates that for most of the observed cities the harness effect must have a structural background, that is not to be modeled by any of the random approaches taken here.

3.4 Conclusions

The main purpose of this chapter was to introduce some models that allow further analysis of PTN. This was done both in 2d and 1d embedding spaces. A particular feature of the models considered is that opposite to the majority of complex

network models, where a network grows due to adding separate nodes, in our case the grows is in terms of PT lines - which are sequences (chains) of nodes.

First we have considered a model of mutually interacting self-avoiding walks in 2d. The network growth model that we developed captures both special features of PTN as well as generating profiles of network characteristics in the various representations which are in line with those found for real world PTN. By varying only a single parameter one may e.g. discriminate between scale-free and exponential harness distributions, both of which are observed in real cities. The method used, a non equilibrium growth model in terms of attractive self-avoiding walks on a square lattice may further be extended to study the effects of geographical constraints e.g. coast-lines, rivers and bridges or disorder.

However, being the most realistic, the above model does not allow for an analytic solution. Therefore, we have further considered and solved analytically a 1d model. In this case we were interested in description of the PTN harness. The latter property was discussed within the model of non-interacting walks in 2d as well. While in one dimension an analytic treatment was successful, the two dimensional case was studied by simulations showing that the empirical results for real PTN deviate significantly from those expected for randomly placed routes. Here, the main conclusions may be summarized as follows:

- A one dimensional model for harness distributions was solved analytically.
- Exponentially decaying harness distributions may be reproduced by the 1d approach.
- Simple random placement of RW, SAW or NRRW on a two dimensional square lattice results in weak harness distributions; in the RW case, much weaker than for real PTN.
- The s-cumulative distributions for different s intersect at low values of r for all models for combinatorial reasons.
- A model of mutually interacting SAWs reproduces many of the empirically observed features of harness distributions in spite of the morphology of the networks which is visibly different from that of real PTN.

Chapter 4

Public transport network vulnerability and resilience

This chapter presents on about public transport network vulnerability and resilience. In particular, we elaborate several criteria to determine network stability and test the theoretical predictions derived for different idealized networks using data for the real-world networks. First we will define ways in which PTN constituents are removed (we will call these attack strategies) and the observables that will be used to monitor the network properties. Then we perform attack simulations in different PTN representations (different spaces introduced in chapter 2) and analyse correlations between PTN properties prior to the attack and its robustness. Some of the results presented in this chapter were published in [13, 52, 14].

4.1 Observables and attack strategies

In the following we mostly consider the removal of nodes. Along the lines of the lattice site percolation problem the removal of a node implies the removal of all links that this node contributed to the network. Thus, in terms of the PTN this interrupts all routes that pass through the corresponding station splitting any such route into two independent parts. No 'detour' links will be inserted to reconnect these routes. This may reflect e.g. an instance where a tram or subway station becomes blocked. When the links are removed nonetheless the corresponding neighbouring nodes survive.

Let us take the $\mathbb{L}$-space representation to introduce the observables we will use to quantify the PTN behavior under attack. Keep in mind however, that in our

analysis presented in section 4.3 we will also deal with the $\mathbb{P}$-space. There are two intrinsically connected questions that naturally arise when one wants to describe quantitatively how a certain network changes when its nodes are removed.

The first is how to choose the 'order-parameter' variable that signals the quantitative change in the network behavior (i.e. the break down of the network), the second is how to locate the value of concentration of removed nodes at which this change occurs. As we have mentioned in the chapter 1, in a theoretical description a useful quantity is the GCC: its disappearance can be associated with a network breakdown. Strictly speaking, the GCC is well-defined only in the $N \to \infty$ limit, therefore in practice dealing with a network of finite size N it is substituted by the size of the largest connected component. We will use in the following its normalized value defined by:

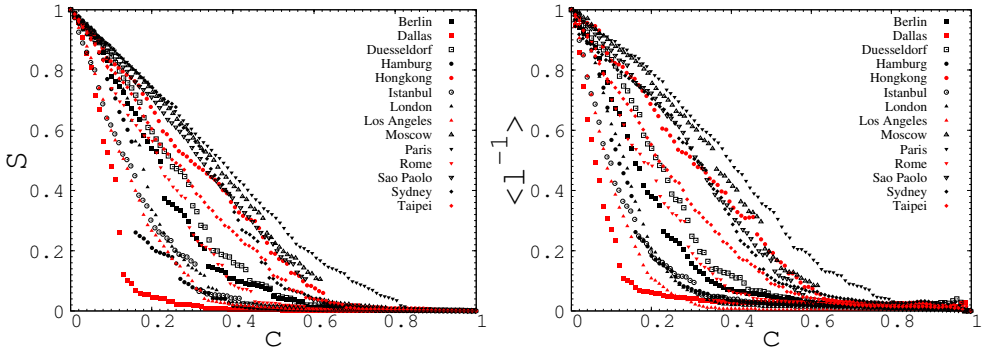
$$S = N_1/N, \tag{4.1}$$

with N and N_1 respectively being number of nodes of the network and of its largest component correspondingly. By definition (4.1), a largest component is always present in a network of non-zero size. A useful quantity to measure network connectivity is the mean shortest path defined in chapter 1, equation 1.2 ($\langle \ell \rangle = \frac{2}{N(N-1)} \sum_{i>j} \ell(i,j)$, where $\ell(i,j)$ is the length of a shortest path from node i to j and the sum spans all pairs i, j of sites of the network). However, $\langle \ell \rangle$ is ill-defined for a disconnected network. Alternatively, one can suitably define the mean inverse shortest path length [66] by:

$$\langle \ell^{-1} \rangle = \frac{2}{N(N-1)} \sum_{i>j} \ell^{-1}(i,j), \tag{4.2}$$

with $\ell^{-1}(i,j) = 0$ if nodes i, j are disconnected. As one can see, equation (4.2) is well-defined even for a disconnected network and as such can be used to trace changes of network behavior under attack. To give an example, we show in Fig. 4.1 how the largest component fraction S, (equation 4.1) and the mean inverse shortest path length $\langle \ell^{-1} \rangle$, (equation 4.2), change upon random removal of nodes in each of fourteen PTN selected for our study. More precisely, we measure these quantities as functions of the fraction of removed nodes c starting from the unperturbed network ($c = 0$) and eliminating at random step-by-step 1 % of the nodes up to $c = 1$. In what follows below we will call this scenario a *random scenario*.

Note, that in Fig. 4.1 we display the result of a specific random attack. We have however verified that random permutations do not influence the results to extents that were visible on the scale of Fig. 4.1. This question will be further investigated in more detail within the discussion of Fig. 4.7 below.



a **b**

Figure 4.1: $\mathbb{L}$-space. Random scenario. Size of the largest cluster S **a**) and an average inverse mean shortest path length $\langle \ell^{-1} \rangle$ **b**) as functions of a fraction of removed nodes c normalized by their values at $c = 0$.

Already this first attack attempt brings about interesting (and in part unexpected) PTN features. Namely:

(i) different PTN react on random removal of their nodes in different ways, that range from rapid abrupt breakdown (Dallas) to a slow almost linear decrease (Paris);

(ii) although qualitatively similar, the observed impact of the attack differs depending on which variable is used as indicator, either S or $\langle \ell^{-1} \rangle$. Ordering the PTN by their vulnerability, this order may thus differ depending on the applied indicator;

(iii) up to $c = 1$, there is no general 'percolation threshold' concentration of removed nodes c at which S (or $\langle \ell^{-1} \rangle$) vanishes that would hold for all PTN. Rather for some individual PTN one observes various values of c at which these PTN show abrupt changes of their properties.

Figs. 4.1**a**,**b** display how the different PTN react on a *random* removal of their nodes. Obviously, the question immediately arises how this behavior changes if one removes the nodes not at random, but following a given order or scheme (we call this the scenario of the attack). As we have mentioned in chapter 1, a number of different attack scenarios have been proposed [2, 23, 66, 7, 8, 20, 58, 62, 63, 13]. These are generally based on the intuitive assumption that the largest impact on a network is caused by the removal of its most 'important' nodes. A number of indicators have been developed in particular in applications of graph theory for social

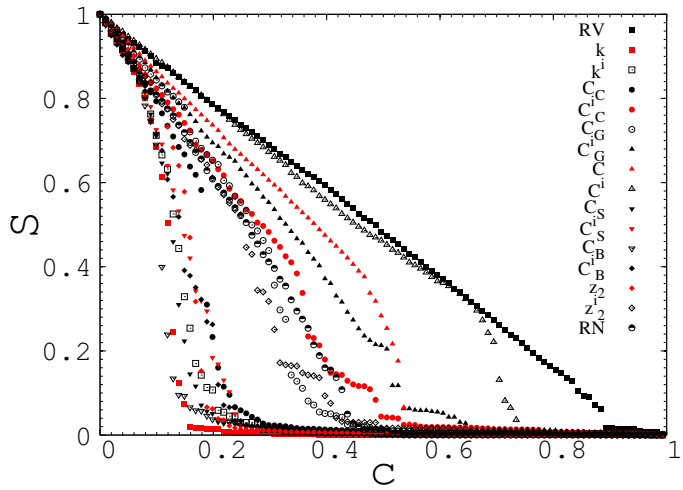


Figure 4.2: Largest component size of the PTN of Paris as function of the fraction of removed nodes for different attack scenarios. Each curve corresponds to a different scenario as indicated in the legend. Lists of removed nodes were prepared according to their degree k, closeness $\mathcal{C}_C$, graph $\mathcal{C}_G$, stress $\mathcal{C}_S$, and betweenness $\mathcal{C}_B$ centralities, clustering coefficient C, and next nearest neighbors number z_2. A superscript i refers to lists prepared for the initial PTN before the attack. RV and RN denote the removal of a random vertex (RV) or of its randomly chosen neighbor (RN), respectively.

science to measure the importance of a node. Besides the node degree k_j, which is equivalent to the number of nearest neighbors $z_1(j)$ of a given node j, different centralities have been introduced for this purpose. In particular, the closeness $\mathcal{C}_C(j)$, graph $\mathcal{C}_G(j)$, stress $\mathcal{C}_S(j)$, and betweenness centralities $\mathcal{C}_B(j)$ of a node j that were mentioned in chapter 2 (equations 2.10-2.12, 1.6). Alternatively, one may measure the importance of a given node j by the number of its second nearest neighbors $z_2(j)$ or its clustering coefficient $C(j)$ mentioned in chapter 1 (equation 1.4 - $C(j) = \frac{2E_j}{k_j(k_j-1)}$ - the ratio of the number of links E_j between the k_j nearest neighbors of j and the maximal possible number of mutual links between them).

Removing important nodes according to lists prepared in the order of decreasing node degrees k, centralities (equations 2.10-2.12, 1.6), number of their second nearest neighbors z_2, and increasing clustering coefficient C defines seven different attack scenarios. Those scenarios can be either implemented according to lists prepared for the *initial* PTN before the attacks (we will indicate the corresponding scenario by a superscript i, e.g. $\mathcal{C}_B^i$) or by lists rebuilt by recalculating the order

of the remaining nodes after each step. Together, this leads to fourteen different attack scenarios. In addition, we will keep the above described random scenario (denoted further as RV) and add one scenario more, removing a randomly chosen neighbor of a randomly chosen node (RN). The latter scenario appears to be effective for immunization problems [33] and it is based on the fact, that in this way nodes with a high number of neighbors will be selected with higher probability. Note that in this scenario only a neighbor node is removed and not the initially chosen one.

All together, this defines sixteen different scenarios to attack a network and we apply these to all fourteen PTN that form our database. A typical result for a single PTN is displayed in Fig. 4.2. Here, we show how the largest connected component size S of the Paris PTN changes under the influence of the above described attack scenarios. Already from this plot one may discriminate between the most effective scenarios that result in a fast decrease of the largest component size (those governed by betweenness and stress centralities, node degree, and next nearest neighbors number – see the Figure) and the least harmful ones (those governed by clustering coefficient, graph and closeness centralities and random offer scenario). In the following, instead of displaying the results of all attacks for all different PTN we will focus on the results of the most effective scenarios comparing them with those of random failure as introduced by the random scenario. As outlined in the introduction, we make use of different PTN representations (different 'spaces' of Fig. 2.2). In the following section, we present the analysis of PTN resilience in the $\mathbb{L}$-space representation.

4.2 Results in $\mathbb{L}$-space

The $\mathbb{L}$-space representation of a PTN is a graph that represents each station by a node, a link between nodes indicates that there is at least one route that services the two corresponding stations consecutively. No multiple links are allowed (see Fig. 2.2**b**). Therefore, attacks in the $\mathbb{L}$-space correspond to situations, in which given public transport stations cease to operate for all means of traffic that go through them. Note however, that in this representation, the removal of a station node does not otherwise interfere with the operation of a route that includes this station. It rather splits this route into two (operating) pieces. An alternative situation will be considered in the forthcoming section.

4.2.1 Choice of the 'order-parameter' variable

In order to answer some of the questions raised in section 4.1, let us return to Fig. 4.2, where the impact on the largest component size S of the PTN of Paris is shown for sixteen different attack scenarios as function of the fraction of removed nodes. As we have already remarked, for this PTN the most influential are the scenarios where nodes are removed according to lists ordered by $\mathcal{C}_B$, k, $\mathcal{C}_S$, k^i, $\mathcal{C}_B^i$, $\mathcal{C}_S^i$ (we list the characteristics in a decreasing order of effectiveness of the corresponding scenario). For a small value of c ($c < 0.07$) these scenarios cause practically indistinguishable impact on S with a linear behavior $S \sim (1-c)$. As c increases, deviations from the linear behavior arise and the impact of different scenarios start to vary. In particular, there appear differences between the role played by the nodes with highest value of k and highest betweenness centrality $\mathcal{C}_B$. Whereas the first quantity is a local one, i.e. it is calculated from properties of the immediate environment of each node, the second one is global. Moreover, the k-based strategy aims to remove a maximal number of edges whereas the $\mathcal{C}_B$-based strategy aims to cut as many shortest paths as possible. In addition, there arise differences between the 'initial' and 'recalculated' scenarios, suggesting that the network structure changes as important nodes are removed. Similar behavior of $S(c)$ is observed for all PTN included in this study, with certain peculiarities in the order of effectiveness of different attack scenarios. Note however, that the difference between 'initial' and 'recalculated' scenarios is less evident for strategies based on local characteristics, as e.g. the node degree or the number of second nearest neighbors (c.f. curves for k, k^i and z_2, z_2^i, respectively). This difference between initial and recalculated characteristics is however more pronounced for the centrality-based scenarios.

Now let us return to some of the observations of section 4.1. Namely, we noted that the observed impact of an attack may differ depending on which observable is used as the 'order-parameter' variable (c.f. Fig. 4.1 where this is shown for the RV attack scenario taking either S or $\langle \ell^{-1} \rangle$ as 'order-parameter'). Similar differences we observe also in the case of the other scenarios. For the sake of uniqueness in the following we will use the value of S to measure the effectiveness of a given attack. This choice is motivated by several reasons: (i) in an infinite network limit S defines an order parameter of the classical percolation problem [44, 109]; (ii) differences between network resilience as judged e.g. by the behavior of S or by that of $\langle \ell^{-1} \rangle$ are not significant enough to be a subject of special analysis (at least not for the PTN we consider); (iii) considering S naturally leads to other useful characteristics that allow to estimate the PTN operating ability and its

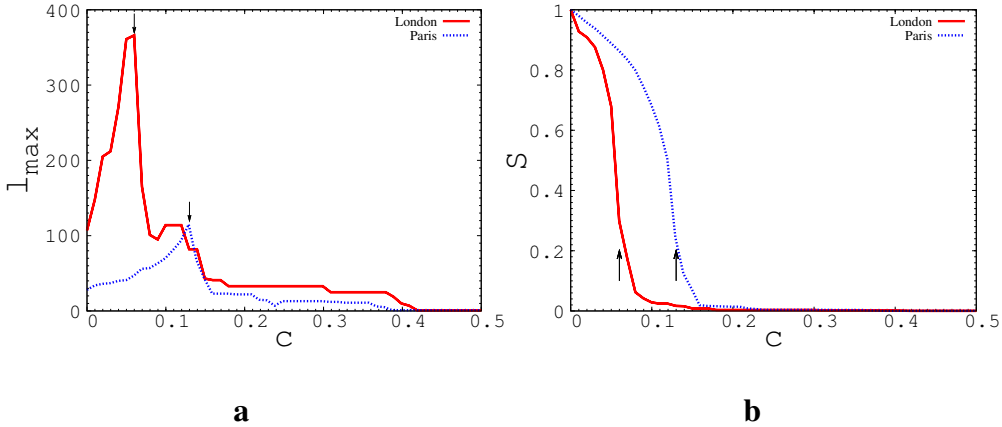


Figure 4.3: $\mathbb{L}$-space. Recalculated highest degree scenario. **a**) behavior of the maximal shortest path $\ell_{\max}$ for the PTN of Paris and London. Note the characteristic peaks that occur at $c = 0.13$ (Paris) and $c = 0.06$ (London). **b**) Size of largest connected cluster S as function of a fraction of removed nodes for the same networks. The arrows indicate the values of c at which the peak for $\ell_{\max}$ appears.

segmentation. Let us stop to elaborate the latter point in more detail.

4.2.2 Segmentation concentration

As we have already emphasized, there is no well defined 'percolation threshold' concentration of removed nodes c_{perc} at which S (or $\langle \ell^{-1} \rangle$) vanishes (see Figs. 4.1, 4.2) which could serve as evidence of a break down of the largest PTN component and hence of the loss of operating ability (another obvious difference between the largest cluster S of a network as considered here and the spanning cluster on a lattice is that below the percolation threshold a spanning cluster is absent while there is always some largest cluster, however its relative size may vanish in the infinite network limit.). It is possible to use the behavior of maximal shortest path length $\ell_{\max}$ as a possible indicator of the network break down. This is based on the observation, that as the concentration of removed nodes c increases, the value of $\ell_{\max}$ for different PTN displays similar typical behavior: initial growth and then an abrupt decrease when a certain threshold is reached (see e.g. Fig. 4.3 **a** where this value is shown for the recalculated highest degree attack scenario of the PTN of Paris and London). Obviously, removing the nodes initially increases the path lengths as deviations from the original shortest paths need to be taken

into account. Further removing nodes then at some point leads to the breakup of the network into smaller components on which the paths are naturally limited by the size of these components which explains the sudden decrease of their lengths. For comparison, in Fig. 4.3 **b** we show how the value of S changes under the recalculated highest degree scenario for the above PTN.

However, being certainly useful for many instances of the PTN analysed, the above $\ell_{\max}$-based criterion cannot serve as an universal tool to determine the region of c, where the network stops to operate. One of the reasons is that for certain PTN (as well as for certain attack scenarios) we have found that $\ell_{\max}$ does not show a pronounced maximum, but rather shows several maxima at different values of c. Therefore, to devise a criterion which may be equally well used for any of the networks we decided to define characteristic concentration of removed nodes c_{s} at which the size of the largest component S decreases to one half of its initial value. This characteristic concentration allows us to compare the effective robustness of different PTN or of the same PTN when different attack scenarios are applied. In what follows below, we will call this concentration the *segmentation* concentration c_{s}, with the obvious condition:

$$S(c_{\mathrm{s}}) = \frac{1}{2}S(c=0). \tag{4.3}$$

In Fig. 4.4 we plot the size of the largest connected component S for different PTN as function of the fraction of removed nodes c for the random vertex scenario (RV) in $\mathbb{L}$-space. The choice of the lowest S value $S = 1/2$ in this figure enables one to find the value c_{s} as the crossing point of $S(c)$ with the horizontal axis. The values of c_{s} obtained for this scenario are given in the last column of Table 4.1. Note that the PTN under consideration react on random attack in many different ways: some of them slowly decrease without any abrupt changes in S (like PTN of Paris, Moscow, Sydney) while others are characterized by rather fast decay of S (Dallas, Los Angeles, Istanbul).

4.2.3 Numerical estimates

Now, applying these attacks according to the sixteen scenarios described above we are in the position to discriminate them by their degree of destruction and to single out those with the highest impact on each of the PTN considered. To this end, for each PTN we give in Table 4.1 the segmentation concentration c_{s} for the five most harmful attack scenarios. The obtained values of c_{s} are given in increasing order. Near each value we denote the scenario that was implemented.

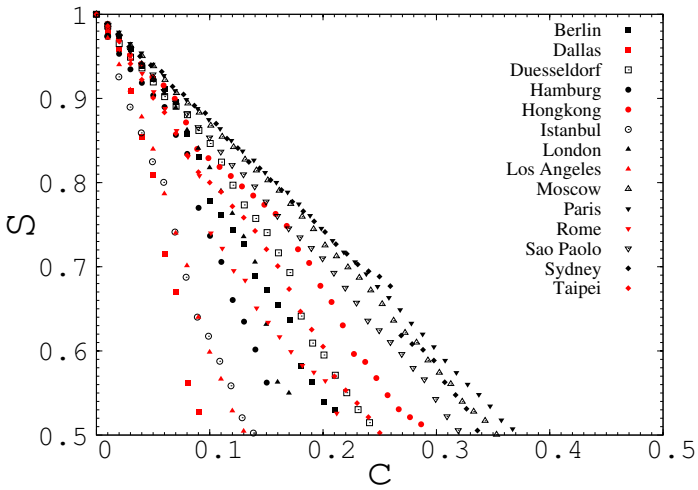


Figure 4.4: $\mathbb{L}$-space. Random scenario. Size of the largest cluster S normalized by its value at $c = 0$ as function of a fraction of removed nodes. From this figure it is easy to define the fraction of nodes c_s which satisfies equation 4.3.

Our analysis reveals the most harmful scenarios as those targeted at nodes with the highest values of either the node degree k, the betweenness centrality $\mathcal{C}_B$, the next nearest neighbor number z_2, or the stress centrality $\mathcal{C}_S$ recalculated after each step of the attack. While the least harmful are those targeted at nodes with the highest values of either the initial graph centrality $\mathcal{C}_G^i$, or the clustering coefficient C^i, C (either initial or recalculated after each step), or random vertex scenario (RV).

4.2.4 Correlations

Molloy-Reed parameter κ

It is instructive to observe correlations between the characteristics of unperturbed PTN (see Table 2.2) and their robustness to attacks. Such correlations may allow for an a priory estimate of the resilience of a network with respect to attacks. As discussed in the chapter 1, percolation theory for uncorrelated networks predicts that the value of the Molloy-Reed parameter $\kappa^{(k)}$, (equation 1.16), can be used to measure the distance to the percolation point $\kappa^{(k)} = 2$. We may therefore expect that networks with a higher value of $\kappa^{(k)}$ show higher resilience. To this end let us first compare the values of c_s for certain scenarios with the value of $\kappa^{(k)}$ for the unperturbed PTN. Before doing this let us note that for an uncorrelated network

City	c_s		c_s		c_s		c_s		c_s		c_s	
Berlin	.060	C_B	.065	k^i	.065	C_S	.070	k	.075	z_2	.220	RV
Dallas	.025	k^i	.030	k	.030	C_B	.045	z_2	.055	z_2^i	.090	RV
Düsseldorf	.075	C_B	.080	k	.080	k^i	.095	C_S	.105	z_2	.240	RV
Hamburg	.040	C_B	.040	C_C	.045	C_S	.045	k^i	.060	z_2	.150	RV
Hong Kong	.030	C_B	.040	C_C	.050	z_2^i	.060	C_S	.090	k^i	.300	RV
Istanbul	.025	C_S	.030	C_C	.030	C_B	.035	k^i	.035	k	.140	RV
London	.055	k	.060	k^i	.065	C_B	.075	C_C	.085	z_2	.175	RV
Los Angeles	.040	k	.060	k^i	.065	z_2	.075	C_B	.100	z_2^i	.130	RV
Moscow	.070	C_B	.085	C_S	.085	k	.085	k^i	.100	C_C	.350	RV
Paris	.105	C_B	.120	k	.125	C_S	.130	k^i	.140	C_B^i	.375	RV
Rome	.050	C_B	.060	C_C	.065	k	.065	k^i	.085	C_S	.215	RV
Saõ Paolo	.040	k	.040	k^i	.045	C_B	.060	C_S	.060	C_S^i	.320	RV
Sydney	.040	C_B	.040	C_C	.065	C_S	.075	k^i	.085	C_G, k	.350	RV
Taipei	.105	C_B	.105	C_G	.115	k	.120	k^i	.120	C_C	.240	RV

Table 4.1: Segmentation concentration c_s for different attack scenarios applied to different PTN. For each city, the Table displays the results of the five most destructive attack scenarios ordered by increasing values of c_s. The scenario is indicated after corresponding value of c_s. The scenarios are abbreviated by the name of the characteristics used to prepare the lists of removed nodes (see section 4.1 for detailed explanation). In the last column the value of c_s for the random scenario (RV) is shown.

the value of $\kappa^{(k)}$ can be equally represented by the ratio between the mean next neighbour number z_1 of a node (which is by definition equal to the mean node degree $\langle k \rangle$) and the mean second nearest neighbour number z_2:

$$\kappa^{(z)} = z_2/z_1. \tag{4.4}$$

Indeed, given that for a network (see e.g. [3, 39, 92, 40])

$$z_2 = \langle k^2 \rangle - \langle k \rangle, \tag{4.5}$$

one may rewrite equation 1.16 ($\kappa^{(k)} \equiv \langle k^2 \rangle / \langle k \rangle$) as:

$$\kappa^{(z)} = 1 \quad \text{at} \quad c_{\text{perc}}. \tag{4.6}$$

The relation $\kappa^{(k)} = \kappa^{(z)} + 1$ holds only approximately for the real-world networks we consider in our study, as one can see, e.g., from the Table 2.2. In Fig. 4.5**a**

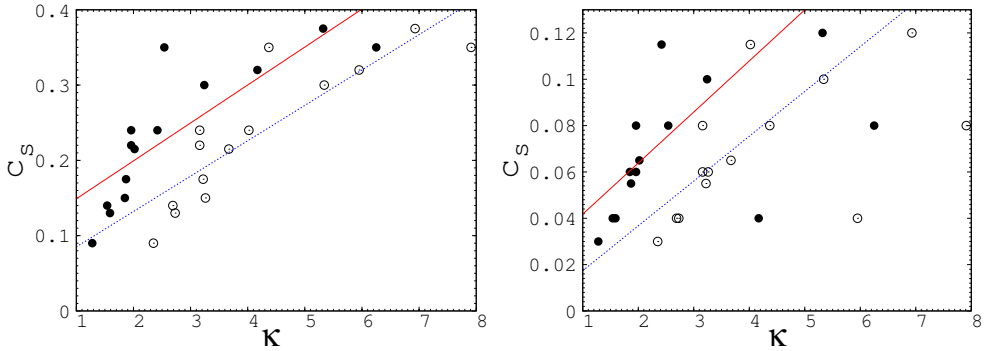


Figure 4.5: $\mathbb{L}$-space. Correlations between the ratio κ, (equations 1.16, 4.4) and segmentation concentration c_s. Open circles: $\kappa^{(k)} = \langle k^2 \rangle / \langle k \rangle$, filled circles: $\kappa^{(z)} = z_2/z_1$. The lines serve as guides to observe the tendency of c_s to increase for higher values of κ. **a.** *Random scenario.* Most out-of-range are the points $c_s = 0.35$, $\kappa^{(z)} = 2.54$, $\kappa^{(k)} = 4.37$ (Sydney) and $c_s = 0.35$, $\kappa^{(z)} = 6.25$, $\kappa^{(k)} = 7.91$ (Moscow). **b.** *Recalculated node-degree scenario.* Two PTN are out of range: $c_s = 0.04$, $\kappa^{(z)} = 4.17$, $\kappa^{(k)} = 5.95$ (Saõ Paolo) and $c_s = 0.08$, $\kappa^{(z)} = 6.25$, $\kappa^{(k)} = 7.91$ (Moscow).

we compare both quantities $\kappa^{(k)}$, $\kappa^{(z)}$ for unperturbed PTN with the corresponding segmentation concentration c_s for the random attack scenario. Within the expected scatter of data one can definitely observe a general tendency of c_s to increase with both $\kappa^{(k)}$ and $\kappa^{(z)}$: the higher the value of κ for an unperturbed network, the more robust it is to random removal of its vertices. This conclusion, however with a more pronounced scatter of data even holds if one repeats the same analysis for the case of the scenario based on recalculated node degrees, as shown in Fig. 4.5**b**. Again, one observes c_s to increase with increasing κ. For the betweenness-based attack scenarios the data is however more scattered and a prediction based on the a priori calculated ratios is unreliable.

Node-degree distribution decay exponent γ

Another useful observation concerns the correlation between the PTN attack resilience and the node-degree distribution exponent γ (1.9). As we have shown in chapter 2 some of the PTN under consideration are scale-free: their node-degree distributions have been fitted to a power-law decay (1.9) with the exponents shown in Table 2.3. Others are characterized rather by an exponential decay, but up to

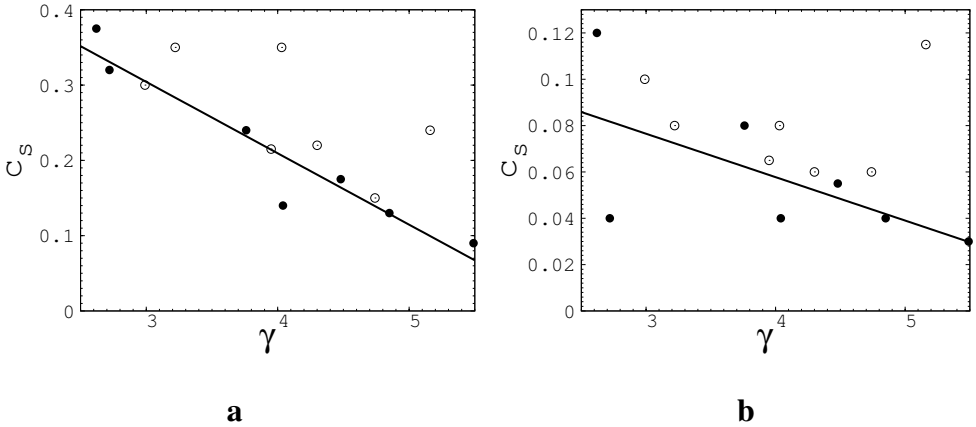


Figure 4.6: $\mathbb{L}$-space. Correlations between the node-degree distribution exponent γ and segmentation concentration c_s. Filled circles: scale-free PTN, open circles: PTN with less pronounced power-law decay. Solid lines serve as guides to observe the tendency of c_s to decay with an increase of γ. **a**. *Random scenario*. Most out of range are the points at $c_s = 0.24$, $\gamma = (5.16)$ (Taipei) and at $c_s = 0.35$, $\gamma = 4.03$ (Sydney). **b**. *Recalculated node-degree scenario*. Most out of range are the points at $c_s = 0.04$, $\gamma = 2.72$ (Sao Paolo) and at $c_s = 0.115$, $\gamma = (5.16)$ (Taipei).

a certain accuracy they can also be approximated by a power-law behavior (then, the corresponding exponent is shown in Table 2.3 in brackets). In Fig. 4.6**a** we show the correlation between the fitted node-degree distribution exponent γ and c_s for the random attack scenario. Filled circles correspond to scale-free PTN, open circles correspond to the PTN where the scale-free behavior is less pronounced. It is interesting to observe, that even if we include the PTN which are better described by the exponential decay of the node-degree distributions, there is a notable tendency to find PTN with smaller values of γ to be more resilient as indicated by larger values of c_s. This tendency is again confirmed if one considers the recalculated node degree attack scenario, as shown in Fig. 4.6**b**.

The above observed correlation between the exponent γ that characterizes the unperturbed network (i.e. a PTN at $c = 0$) and the segmentation concentration c_s at which however the PTN is still to a large part unperturbed indicates that some global properties of the node-degree distribution may remain essentially unchanged when the nodes are removed (i.e. a scale-free distribution remains scale-free as c increases, $0 < c < c_s$). To check that assumption for the RV scenario, we analysed the averaged cumulative node degree distributions for each of the PTN

with 3 %, 5 %, and 10 % of removed nodes. The cumulative distribution $P(k)$ is defined in terms of the node-degree distribution $p(q)$ (1.9) as:

$$P(k) = \sum_{q=k}^{k^{\max}} p(q), \tag{4.7}$$

with $k^{\max}$ the maximal node degree in the given PTN. Typical results of this analysis are shown in Fig. 4.7, for the PTN of Paris. We compare the cumulative node degree distribution $P(k)$ of the unperturbed PTN with that of the PTN where a given fraction c part of the nodes ($c = 0.03$, 0.05, and 0.1, correspondingly) was removed according to the random attack scenario (RV). For each of the concentrations of the removed nodes, $P(k)$ was averaged over 2000 repeated attacks.

In the first plot, Fig. 4.7**a**, we compare the three resulting average distributions (for $c = 0.03$, 0.05, and 0.1) with the original one ($c = 0$). One clearly sees that there is no qualitative or even quantitative (change of exponent) change of the distributions for any of the three cases. Indeed, if one has a large set of nodes with a given node-degree distribution any sufficiently large random subset of these nodes should have the same distribution; in particular this holds if one averages these subset distributions over many instances. The above argument seems to ignore the change of degrees in the subset due to cutting off those vertices not remaining in the set. However, due to the random choice of the removed nodes the share of lost degree will on the average be proportional to the degree of each vertex: the higher its degree the more probable it is that one of its neighbors is chosen to be removed and this probability is proportional to its degree. Thus, the sum of degrees in the remaining subset is lower; but the degree distribution $P(k)$ is effectively transformed to $P'(ck) = nP(k)$ where c is the probability of any node being removed and $P'(k)$ is the distribution in the remaining subset of nodes, n a normalization. For an exponential distribution this transformation shifts the scale. However, a scale free distribution keeps its exponent under such a transformation.

In the other three plots, Figs. 4.7**b**-**d** we show for each amount of removed nodes the average cumulative distribution together with statistical errors calculated as the standard deviation within the ensemble of the 2000 instances generated in the sample. Even on the logarithmic scale these are very small for all but the very high degrees where on the average fluctuations of small numbers of often less than one node for a given degree occur.

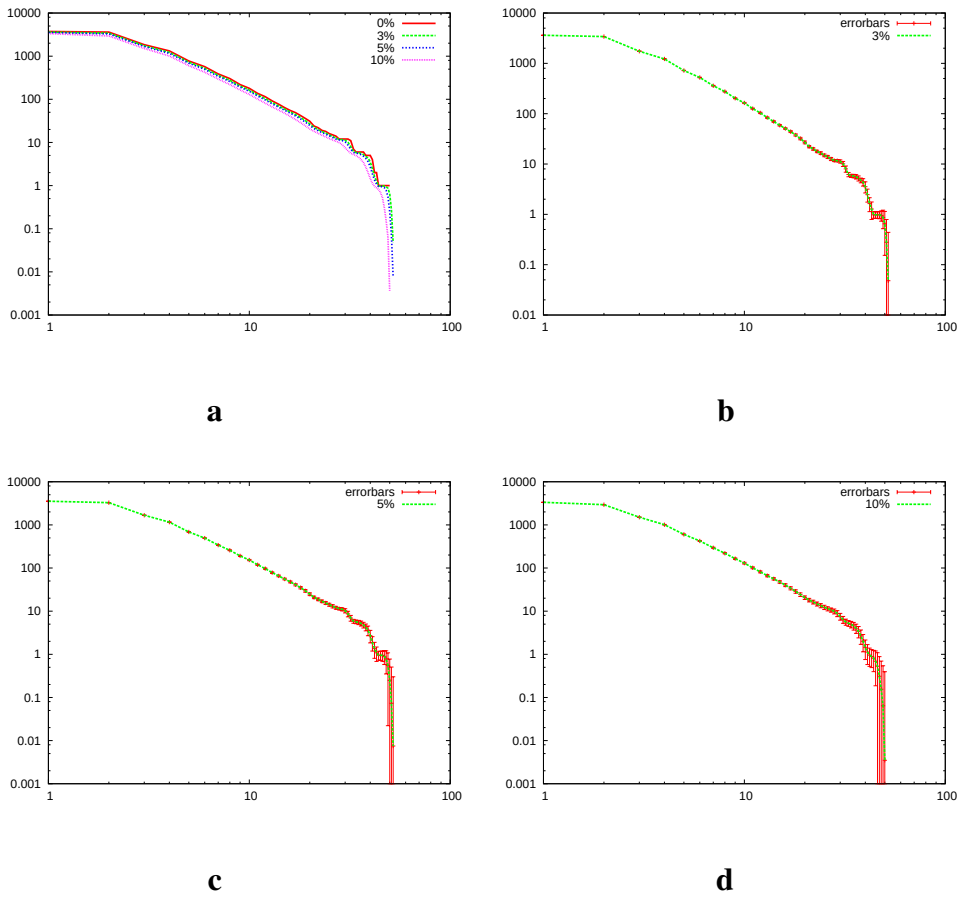
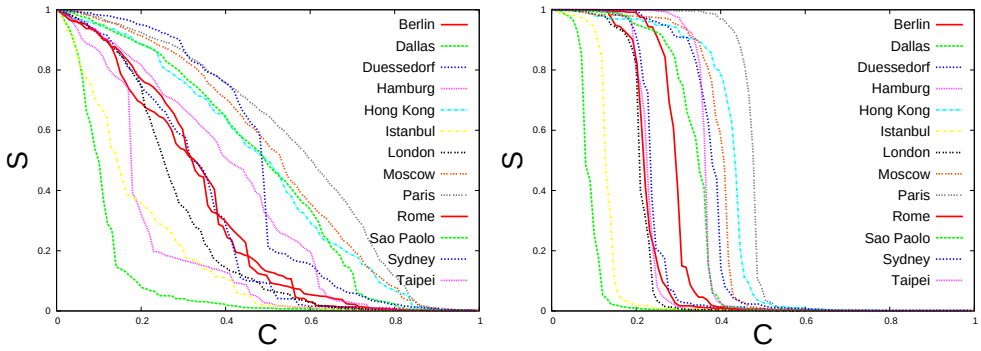


Figure 4.7: $\mathbb{L}$-space. Average cumulative node degree distributions for Paris PTN for the random attack scenario. Comparison of the initial distribution (red curve, $c = 0$) with those of the PTN with $c = 0.03$, $c = 0.05$, $c = 0.1$ (**a**). Average cumulative node degree distribution together with statistical errors for $c = 0.03$ (**b**), $c = 0.05$ (**c**), $c = 0.1$ (**d**).

4.2.5 Comparison of node- and link-targeted attacks

In this subsection we observe link-targeted attacks in $\mathbb{L}$-space and compare the results with those for node-targeted attacks.

As mentioned before, when a link is removed the neighbouring node survives. So in link-targeted attacks all nodes survive to the end, however the giant connected component (GCC) decays in a similar way as for node-targeted scenarios. As in the latter case we study the resilience of the $\mathbb{L}$-space PTN graphs to attacks



a **b**

Figure 4.8: $\mathbb{L}$-space. Size of the largest cluster S as functions of a fraction of removed nodes c normalized by their values at $c = 0$. **a**. For random link-targeted scenario. **b**. For recalculated link-degree attack scenario.

performed following five different scenarios. Scenarios are designed analog to the most important scenarios for the node-targeted case. We try to remove the most 'important' links. Taking into account results of node-targeted attacks two main indicators were chosen for these attack scenarios: the link degree $k(l)$ and the link betweenness centrality $\mathcal{C}_B(l)$. The link degree $k(l)$ of the link between nodes i and j equal sum of neighbouring nodes degrees k_i and k_j minus two (corresponding input of ij link into this sum).

$$k(l)_{ij} = k_i + k_j - 2. \tag{4.8}$$

So for a link in a simple graph with two vertices (smallest possible component containing a link) the link degree will be zero, $k(l) = 0$, while for any link in the connected graph with more than two vertices the link degree will be at least one, $k(l) \geqslant 1$. The link betweenness centrality $\mathcal{C}_B(l)_j$ measures the importance of a link j with respect to the connectivity between the nodes of the network. The link betweenness centrality is defined as

$$\mathcal{C}_B(l)_j = \sum_{s \neq t \in \mathcal{N}} \frac{\sigma_{st}(j)}{\sigma_{st}}, \tag{4.9}$$

where σ_{st} is the number of shortest paths between the two nodes $s,t \in \mathcal{N}$, that belong to the network $\mathcal{N}$, and $\sigma_{st}(j)$ is the number of shortest paths between

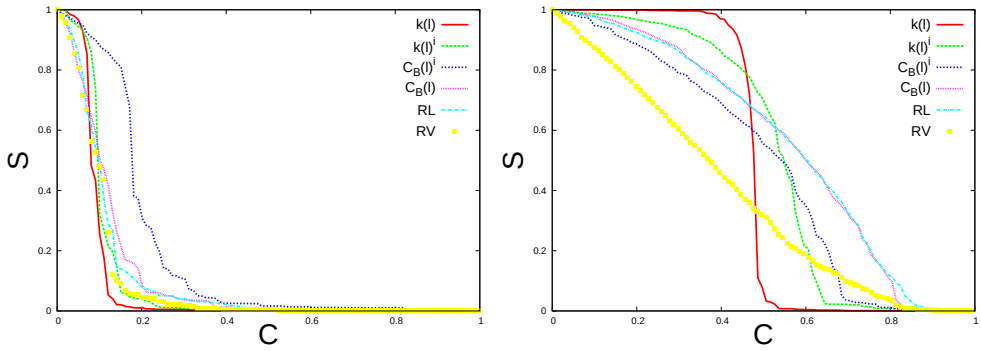


Figure 4.9: $\mathbb{L}$-space. Largest component size of the PTN as function of the fraction of removed links for different attack scenarios. Each curve corresponds to a different scenario as indicated in the legend. Lists of removed links were prepared according to their degree $k(l)$ and betweenness $\mathcal{C}_B(l)$ centrality. A superscript i refers to lists prepared for the initial PTN before the attack; RL and RV denote the removal of a random link and removal of random node respectively. **a**. For PTN of Dallas. **b**. For PTN of Paris.

nodes s and t that go through the link j. The most reliable algorithm to calculate the link (edge) betweenness centralities was proposed by Brandes [19].

Removing important links according to lists prepared in the order of decreasing link degrees $k(l)$ and link betweenness centralities (equations 4.8-4.9) defines two different attack scenarios. Those scenarios can be either implemented according to lists prepared for the *initial* PTN before the attacks (we will indicate the corresponding scenario by a superscript i, e.g. $\mathcal{C}_B(l)^i$) or by lists rebuilt by recalculating the order of the remaining links after each step. Together, this leads to four different attack scenarios. In addition, we will use the random link removal scenario (denoted further as RL). All together, this defines five different scenarios to attack network links and we apply these to thirteen PTN that form our database (all except Los Angeles).

In Fig. 4.8 **a** we show the change of the size of the largest cluster S under random link-targeted attacks (RL). If one compares this behavior with that observed for the random node removal scenario (RV) (see Fig. 4.1) one sees, that for most PTN that have strong resilience to random node-targeted attacks random link removal is even less effective. On the other hand, for PTN with weak resi-

lience there seems to be no significant difference. And in the same way as for random node attacks (RV) random link attacks (RL) lead to changes of the largest connected component S that range from an abrupt breakdown (Dallas) to a slow smooth decrease (Paris). Even slower than for random node removal - removing a link does not necessary lead to removing a node from the largest cluster, while removing a node from the completely connected network decreases it at least by one node. The value of $S(c_s)$ defined by the condition (4.3) is given in the Table 4.2.

City	c_s	c_s	c_s	c_s	c_s	c_s
Berlin	0.22 $k(l)$	0.26 $k(l)^i$	0.30 $C_B(l)$	0.33 $C_B(l)^i$	0.34 RL	0.22 RV
Dallas	0.08 $k(l)$	0.10 $k(l)^i$	0.10 RL	0.11 $C_B(l)$	0.18 $C_B(l)^i$	0.09 RV
Düsseldorf	0.24 $k(l)$	0.30 $k(l)^i$	0.32 $C_B(l)^i$	0.33 $C_B(l)$	0.33 RL	0.24 RV
Hamburg	0.14 $C_B(l)^i$	0.18 RL	0.20 $C_B(l)$	0.22 $k(l)^i$	0.23 $k(l)$	0.15 RV
Hong Kong	0.36 $C_B(l)^i$	0.42 $k(l)^i$	0.44 $k(l)$	0.48 $C_B(l)$	0.51 RL	0.30 RV
Istanbul	0.13 $k(l)$	0.14 $C_B(l)^i$	0.15 $C_B(l)$	0.15 $k(l)^i$	0.15 RL	0.14 RV
London	0.21 $k(l)$	0.23 $k(l)^i$	0.24 $C_B(l)$	0.26 RL	0.27 $C_B(l)^i$	0.18 RV
Moscow	0.41 $k(l)$	0.43 $C_B(l)^i$	0.45 $k(l)^i$	0.53 $C_B(l)$	0.54 RL	0.35 RV
Paris	0.49 $k(l)$	0.55 $C_B(l)^i$	0.56 $k(l)^i$	0.61 $C_B(l)$	0.61 RL	0.38 RV
Rome	0.29 $k(l)^i$	0.30 $k(l)$	0.30 $C_B(l)^i$	0.33 RL	0.36 $C_B(l)$	0.22 RV
Saõ Paolo	0.35 $k(l)$	0.35 $k(l)^i$	0.35 $C_B(l)^i$	0.50 $C_B(l)$	0.50 RL	0.32 RV
Sydney	0.19 $C_B(l)^i$	0.35 $k(l)^i$	0.38 $k(l)$	0.49 RL	0.53 $C_B(l)$	0.35 RV
Taipei	0.37 $k(l)$	0.37 $C_B(l)^i$	0.38 $k(l)^i$	0.39 $C_B(l)$	0.41 RL	0.24 RV

Table 4.2: $\mathbb{L}$-space. Segmentation concentration c_s for different link-target attack scenarios applied to different PTN. For each city, the Table displays the results of the five attack scenarios ordered by increasing values of c_s. The scenario is indicated after the corresponding value of c_s. The scenarios are abbreviated by the name of the characteristics used to prepare the lists of removed nodes (as explained above). In the last column the value of c_s for the random vertex scenario (RV) is shown.

Typical results for a single PTN under different types of link-targeted attacks are displayed in Fig. 4.9. Here, we show how the largest connected component size S of the Dallas (**a**) and Paris (**b**) PTN change under the influence of the above described attack scenarios. As one can see, for the PTN of Dallas there is no significant difference between the effectiveness of most scenarios including the random one. That vulnerable behavior of the Dallas PTN under link-targeted at-

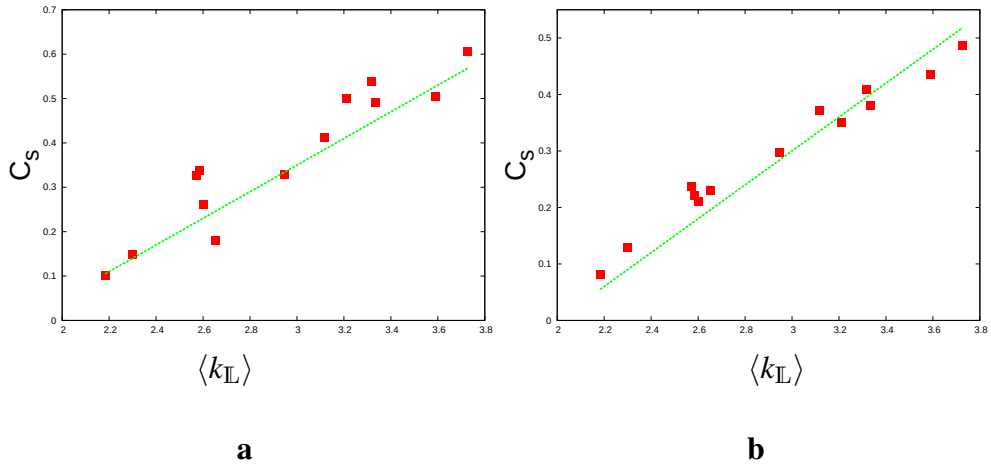


Figure 4.10: $\mathbb{L}$-space. Correlations between the mean degree value $\langle k_{\mathbb{L}} \rangle$ and segmentation concentration c_s for link-targeted attacks. Solid lines serve as guides to observe the tendency of c_s to increase for higher values of $\langle k_{\mathbb{L}} \rangle$. **a**. *Random link-targeted scenario*. No points definitely out of range. **b**. *Recalculated link-degree scenario*. Again no points definitely out of range.

tacks is the same as under random vertex removal approach. For Paris the situation is different. The main observation is that the random vertex attack is more effective than any link-targeted attack from the beginning, until breakdown and further, and only after the maximal cluster attains less than 40% of its initial size the recalculated link degree ($k(l)$) targeted scenario starts to be more harmful. Comparing just link-targeted scenarios one can see that they mostly have similar behavior, only the recalculated degree scenario line initially decays slower, however become more effective near to the breakdown.

To further detail the situation, similar as in subsection 4.2.3, we summarize in Table 4.2 the outcome of five attack scenarios for all cities and compare those with the random vertex removal scenario. Several conclusions can be made from those results. We can see that the segmentation concentration c_S values are of the same order for all types of link-targeted attacks. The most effective attack scenario for the majority of the observed PTN is the recalculated link degree ($k(l)$) scenario. In Fig. 4.8 b we plot its behavior for all analysed PTN. However the difference between the most effective and the less harmful scenarios usually is insignificant. 'Initial' and 'recalculated' scenarios behavior is very similar either for degree-targeted or betweenness centrality-targeted attacks. However often the 'initial'

approach occurs to be more effective. It is interesting to mention that for three PTN (Hamburg, Istanbul, Sydney) which are not very resilient against any kind of attacks (however not for PTN of Dallas, which is least), most efficient is the scenario of removing links with initial highest values of the betweenness centrality $C_B(l)^i$. The last observation from Table 4.2 is that all link-targeted scenarios are of the same order, or less effective than the scenario of random vertex removal (RV), which is given in the last column for comparison. Only the scenario of removing links with initial highest values of the betweenness centrality $C_B(l)^i$ for PTN of Sydney occurs to be significantly more effective than the RV scenario.

To conclude this subsection, we ask the question if a simple criterion can be found that allows to predict a priori the PTN vulnerability under the link-targeted attacks. Namely, given the general PTN characteristics (see Table 2.2) can one forecast resilience against such attacks? It seems that such a criterion does exist and is even more simple than for node-targeted attacks. As follows from Table 4.2 all links occur to be of the same importance in the majority of the observed PTN. If 'quality' is unimportant then the only difference between PTN is in 'quantity'. The normalized quantity of links is their density and it is represented by the mean node degree value $\langle k_{\mathbb{L}} \rangle$. In support of the above reasoning, in Fig. 4.10 we plot c_s as function of $\langle k_{\mathbb{L}} \rangle$ for attacks based on the random removal of links (**a**) and highest recalculated degree of the link scenario (**b**). There, for both cases, within the expected scatter of data one observes a clear evidence of the decrease of c_s with $\langle k_{\mathbb{L}} \rangle$, i.e. networks with smaller mean node degree $\langle k_{\mathbb{L}} \rangle$ break down at smaller values of c and are thus more vulnerable to link-targeted attacks.

It is worthwhile to note here, that the order of the PTN according to their vulnerability under link-targeted attacks is similar to that for the node-targeted scenarios, there are just few light shifts.

4.3 Results in $\mathbb{P}$-space

Let us complement the $\mathbb{L}$-space analysis performed above by observing the reaction of PTN graphs under attack when one observes them in another representation. In particular, we will investigate $\mathbb{P}$-space graphs.

First let us recall that in this representation each node corresponds to a PTN station, i.e. it has the same interpretation as in the $\mathbb{L}$-space. However, the interpretation of a link differs from that in the $\mathbb{L}$-space: now all station-nodes that belong to the same route are connected and thus each route enters the $\mathbb{P}$-space network as a complete subgraph. This results in the main peculiarity of the interpretation of

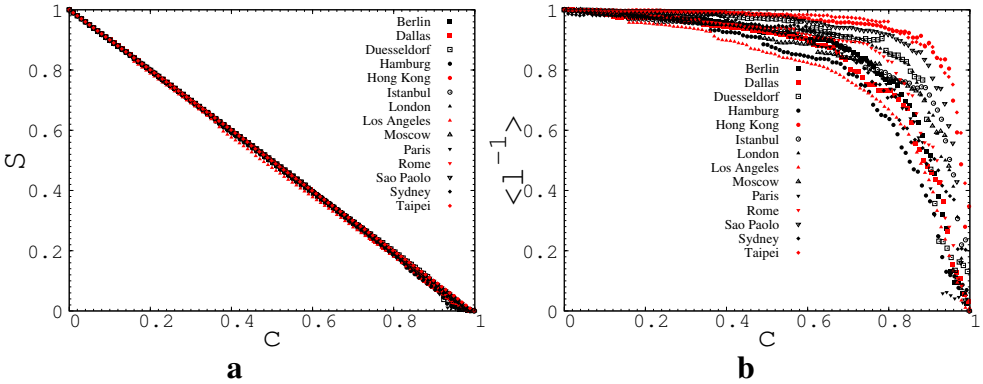


Figure 4.11: $\mathbb{P}$-space. Random scenario. (**a**) size of the largest cluster S and (**b**) the average inverse mean shortest path length $\langle \ell^{-1} \rangle$ as functions of the fraction of removed nodes c normalized by their values at $c = 0$.

the behavior under attacks of these graphs. Consider as an example the $\mathbb{P}$-space graph of Fig. 2.2**d** and compare it to the original PTN map, Fig. 2.2**a**. Whereas the removal of station node C in the map (Fig. 2.2**a**) disconnects the nodes A and B, the removal of the same node in the $\mathbb{P}$-space (Fig. 2.2**d**) keeps nodes A and B connected, as far as they still belong to the same route. Therefore, the removal of nodes in $\mathbb{P}$-space, performed either in a random way or according to certain lists, has a different interpretation in comparison to that occurring in the $\mathbb{L}$-space. An interpretation of the removal of nodes in $\mathbb{P}$-space is the following: if a node is removed, the corresponding stop of the route is canceled while the route otherwise keeps operating. If in the above example the station-node C is removed, route No 2 still keeps operating and station-node B can be reached from D, only without stopping at C (e.g. the bus takes a shortcut). In this way, as we will see below, the removal of nodes in $\mathbb{P}$-space allows us to gain additional insight into the PTN structure.

4.3.1 Numerical estimates

As in the case of the $\mathbb{L}$-space representation, we study the resilience of the $\mathbb{P}$-space PTN graphs to attacks performed following the sixteen different scenarios defined in section 4.1. In Fig. 4.11 we show the change of the size of the largest cluster S (**a**) and the average inverse mean shortest path length $\langle \ell^{-1} \rangle$ (**b**) under random attacks (RV). If one compares this behavior with that observed for the RV scenario

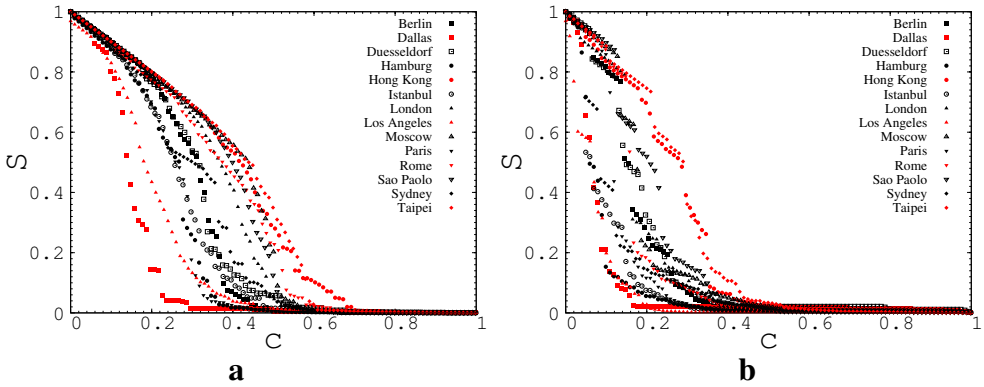


Figure 4.12: $\mathbb{P}$-space, size of the largest cluster S at **a**: highest degree scenario (recalculated), **b**: highest betweenness scenario (recalculated).

City	c_s		c_s		c_s		c_s		c_s		c_s	
Berlin	.155	C_B	.175	C_C	.215	C_S	.285	C^i	.290	C_B^i	.490	RV
Dallas	.065	C_B	.075	C_C	.095	C_S	.115	C	.130	C^i	.490	RV
Düsseldorf	.160	C_B	.185	C_S	.255	C_C	.295	C^i	.300	k^i	.495	RV
Hamburg	.050	C_C	.065	C_B	.145	C_G	.170	C	.175	C_C^i	.490	RV
Hong Kong	.285	C_B	.295	C_S	.335	C_C	.365	C	.380	C^i	.505	RV
Istanbul	.060	C_C	.060	C_B	.060	C_B^i	.115	C_C^i	.175	C	.500	RV
London	.155	C_B	.205	C_C	.305	C_G	.330	C	.350	C^i	.495	RV
Los Angeles	.065	C_B	.095	C_C	.145	C_S	.145	C_B^i	.150	C	.480	RV
Moscow	.175	C_B	.255	C_C	.285	C_S	.345	C	.395	$^i, C_S^i$	.495	RV
Paris	.115	C_B	.165	C_S	.215	C_C	.235	C_B^i	.240	C, C^i	.500	RV
Rome	.135	C_C	.160	C_B	.225	C_G	.285	C_S	.305	C	.495	RV
Saõ Paolo	.205	C_B, C_C	.240	C_S	.355	C_G	.365	C	.390	C^i	.500	RV
Sydney	.075	C_C	.085	C_B	.105	C_S	.225	C	.240	C^i	.510	RV
Taipei	.290	C_B	.320	C_S	.370	C_C	.430	C_G	.440	k, C_S^i	.495	RV

Table 4.3: Segmentation concentration c_s for different attack scenarios applied to different PTN in $\mathbb{P}$-space. For each city, the Table shows the five most effective attack scenarios ordered by increasing values of c_s. The scenario is indicated after corresponding value of c_s. The scenarios are abbreviated by the name of the characteristics used to prepare the lists of removed nodes (see section 4.1 for detailed explanation). In the last column the value of c_s for the random scenario (RV) is shown.

in $\mathbb{L}$-space (see Fig. 4.1) one sees, that all PTN under consideration react in a much more homogeneous way. In $\mathbb{L}$-space random attacks lead to changes of the largest connected component S that range from an abrupt breakdown (Dallas) to a slow smooth decrease (Paris). In $\mathbb{P}$-space one observes for the same scenario only a decrease of S which corresponds to the number of removed nodes. No breakdown of this cluster occurs in this scenario. The value of $S(c_{\mathrm{s}})$ defined by the condition (4.3) is given in the last column of Table 4.3. It is worth to note, that the behavior of the mean inverse shortest path length $\langle \ell^{-1} \rangle$ as function of the fraction c of disabled nodes is also qualitatively different between the two RV scenarios in $\mathbb{L}$- (Fig. 4.1**b**) and $\mathbb{P}$- (Fig. 4.11**b**) spaces. In $\mathbb{L}$-space $\langle \ell^{-1} \rangle$ decreases in general faster than linearly indicating an increase of the path length between the nodes as well as partitioning of the network. In $\mathbb{P}$-space $\langle \ell^{-1} \rangle$ remains for a large part unperturbed as the nodes of the complete subgraph remain essentially connected and the shortest path length remains almost unchanged until only a small fraction of the network remains.

To further detail the situation, similar as in section 4.2, we summarize in Table 4.3 the outcome of the five most harmful attack scenarios and compare those with the random attack scenario. As it follows from the Table and as is further supported by Fig. 4.12, the betweenness-targeted scenarios appear to be the most harmful. Following this observation let us investigate the role of the highest betweenness nodes: above all these are the nodes (and not the highest-k hubs) that control the PTN behavior under attack. The $\mathbb{P}$-space degrees of these high-betweenness nodes do not essentially differ from those of the hubs, therefore they cannot be easily distinguished from the other nodes during attacks according to highest-k scenario. To support this assumption, let us recall that in the $\mathbb{P}$-space representation each route enters the overall network as a complete subgraph, with all nodes interconnected. Removing nodes from a complete graph does not lead to any segmentation. The decrease of the normalized size of this graph will be given by the exact formula $S = 1 - c$ (which is - almost - reproduced by the RV scenario, c.f. Fig. 4.11**a**). Under such circumstances a special role is played by those nodes that join different complete graphs (different routes). The removal of such nodes will separate different complete routes and as a result may lead to network segmentation. Naturally, being between different complete subgraphs such nodes are characterized by high centrality indices, as observed above. Moreover, as far as their direct neighbors belong to different complete graphs, these neighbors are not connected between each other resulting in a lower value of the clustering coefficient C. From Table 4.3 one sees that attacks based on choosing nodes with low-C values are very effective in $\mathbb{P}$-space.

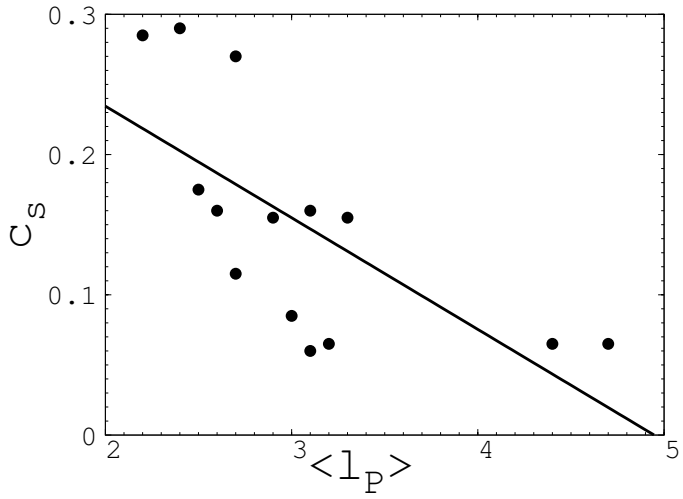


Figure 4.13: $\mathbb{P}$-space. Correlations between the mean shortest path length $\langle \ell_{\mathbb{P}} \rangle$ and segmentation concentration c_s in the highest betweenness centrality scenario. The line serves as a guide to observe the tendency of c_s to decrease with increasing $\langle \ell_{\mathbb{P}} \rangle$.

4.3.2 Correlations

To conclude this section, we ask the question if a simple criterion can be found that allows to predict a priori the $\mathbb{P}$-space PTN vulnerability. Namely, given the general PTN characteristics (see Table 2.2) can one forecast resilience against attacks in $\mathbb{P}$-space? The answer is given by the observation that the networks with low mean shortest path length $\langle \ell_{\mathbb{P}} \rangle$ are the best connected in $\mathbb{P}$-space and hence may be expected to be less vulnerable. Indeed, on the one hand, for the above example of a complete graph (a single PTN route) $\langle \ell_{\mathbb{P}} \rangle = 1$ and it is extremely robust to $\mathbb{P}$-space attacks. On the other hand, a high value of $\langle \ell_{\mathbb{P}} \rangle$ indicates numerous intermediate nodes between different routes. As we have checked above, the targeted removal of such nodes leads to rapid network segmentation. In support of the above reasoning, in Fig. 4.13 we plot c_{s} as function of $\langle \ell_{\mathbb{P}} \rangle$ for attacks based on the highest betweenness centrality scenario. There, within the expected scatter of data one observes a clear evidence of the decrease of c_s with $\langle \ell_{\mathbb{P}} \rangle$, i.e. networks with higher mean path length break down at smaller values of c and are thus more vulnerable.

It is worth to note here, that in $\mathbb{P}$-space it is only the RV attack that has very similar impact on all PTN (see Fig. 4.11). As we have just observed, similar to the

$\mathbb{L}$-space also in $\mathbb{P}$-space the PTN manifest different levels of robustness against attacks targeted at the most important nodes. However, the order of vulnerability changes if one compares the outcome of the $\mathbb{L}$-space and $\mathbb{P}$-space attacks. This means that PTN that were vulnerable in the $\mathbb{L}$-space may appear to be robust against attacks in $\mathbb{P}$-space. From Table 4.3 we see that the PTN that are most stable against highest $\mathcal{C}_B$-targeted attacks in $\mathbb{P}$-space are the PTN of Hong Kong, Saõ Paolo, and Moscow, with $c_s = 0.285$, 0.205, and 0.175, correspondingly. When attacked in $\mathbb{L}$-space, the PTN of Moscow keeps its robustness: $c_s = 0.07$ during $\mathcal{C}_B$-targeted attack, which displays one of highest c_s values for the $\mathbb{L}$-space, see Table 4.1. This is however not the case for the PTN of Hong Kong and Saõ Paolo. In $\mathbb{L}$-space, these belong to the most vulnerable PTN.

4.4 Conclusions

In this chapter, we have studied the behavior of city PTN under attacks. In our analysis we have examined PTN of fourteen major cities of the world. The principal motivation behind this study was to observe the behavior under attack of a sample of networks that were constructed for the same purpose, to compare these with available analytical results for percolation of complex networks, and possibly to derive some conclusions about correlations between PTN characteristics calculated a priory and the resilience to attacks. Furthermore, the resilience behavior of a network against different attack scenarios gives additional insight into the network architecture, discovering structures on different scales. This approach has also been termed the 'tomography' of a network [120].

In our study we have also attempted to compare our results with the predictions of percolation theory on networks. Due to the sizes of these systems which are far from the thermodynamic limit and the rather small sample of networks no quantitative comparison appeared possible. However, qualitative predictions about the location of segmentation thresholds and thus the vulnerability could be verified. Although our study was not primarily motivated by applications, some of the results and methods developed within this study may be useful for planning and risk assessment of PTN. Our analysis has identified PTN structures which are especially vulnerable and others, which are particulary resilient against attacks. Further investigation of other relevant network properties may reveal mechanisms behind this structural resilience. Furthermore we note that the methods developed here also allow to identify minimal strategies to obstruct the operation of the PTN of a city e.g. for the purposes of industrial action and possibly achieve a successful

end of a social conflict.

To analyse PTN resilience we have applied different attack scenarios, either node or link targeted, that range from random failure to targeted destruction, when the most influential network nodes are removed according to their operating characteristics. To choose the most influential nodes, we have used different graph theoretical indicators and determined in such a way the most effective attack scenarios. Our work shows that even within a sample of networks all created for the same purpose one observes essential diversity with respect to their behavior under attacks of various scenarios. Results of our analysis show that PTN demonstrate a rich variety of behavior under attacks, that range from smooth decay to abrupt change.

Concerning random scenarios we have also verified a self-averaging effect that results in a suppression of deviations between different random scenarios and a stability of the network degree distribution against moderate impact of random attacks.

We applied node-targeted attack scenarios to PTN representations both for $\mathbb{L}$-space and $\mathbb{P}$-space graphs. In practice the $\mathbb{L}$-space interpretation of a station failure rather correponds to the failure of transport that is rail-mounted. If a station fails then travel between stations on both parts of the remaining track one needs to be diverted through other routes. The $\mathbb{P}$-space interpretation of this situation on the other hand would correspond to bus transport where the the bus route may be diverted to other roads leaving out the failed station. In both cases the connections provided between different routes at the given station are lost while in the $\mathbb{L}$-space case in addition the routes are cut into disjunct pieces.

As shown in this work, the impact of attacks may be measured by different quantities. As a criterion that is well defined and easily reproducible we choose to define the segmentation concentration c_{s} to correspond to the situation where the largest remaining cluster contains one half of the original nodes of the network. Let us note as well, that definitely not all of the PTN analysed demonstrate scale-free behavior in $\mathbb{P}$-space (and even less in $\mathbb{L}$-space). Nevertheless, in spite of the diversity of behavior we clearly see common tendencies in their reaction to attacks. In particular, this enabled us to propose criteria that allow an a priori estimate of PTN robustness. In $\mathbb{L}$-space resilience to node-targeted attacks is indicated by a high value of the Molloy-Reed parameter κ, (equations 1.9, 4.4) or by a small value of the exponent γ, if a power law is observed for the PTN node degree distribution, resilience to link-targeted attacks in $\mathbb{L}$-space is indicated by a high value of the mean node degree $\langle k_{\mathbb{L}} \rangle$ and in $\mathbb{P}$-space high resilience is indicated by a small mean shortest path length $\langle \ell_{\mathbb{P}} \rangle$.

In this study PTNs were analysed only as unweighted graphs. In particular the resilience study was performed in this way. However, it is obvious that it is important to take into account the limited load that any given link may support. Re-routing the passengers of a broken link as we assume in the present study will in general overloading other links and lead to a breakdown before the giant component splits. As one of the possible continuations of this work this suggests to study e.g. cascade-based attacks [85, 86] on PTN.

Chapter 5

Optimisation of public transport networks

For obvious reasons, optimisation is very important issue, in particular for public transport networks. Quite number of works have been published in this field, however many of these are rather concerned with traffic jams, schedules, logistics, etc. Here we study another interesting question. Given a city that extends over a circular area with an overall constant population, can the topology of the public transport network (including trams, buses and other kinds of public transport operating in the city) be optimised and in what way. This approach is quite different from the usual problem of optimising transport between a set of given sites leading to a combinatorial problem [96, 111, 87]. We will be looking at simple topologies, having more or less only one input parameter - the number of buses N_B

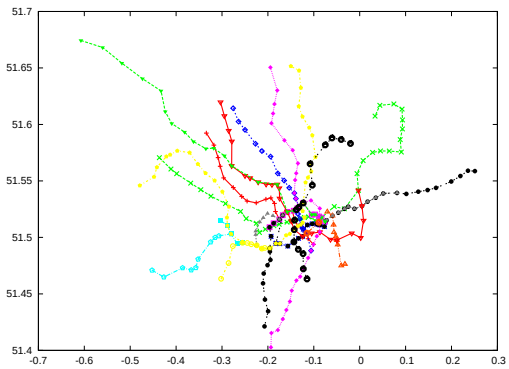


Figure 5.1: London tube network routes representation in GPS coordinates.

running on this network. This number is essentially determined by the amount the 'city' is willing to invest. The aim is to optimise either the average travelling time of all citizens, or their mean velocity, while travelling. We will consider mainly different regular radial models. While it is often possible to find an analytic solution for these and they are quite common in real world scenarios. They are found in particular for metro networks of the large cities e.g. London tube network (see Fig. 5.1).

5.1 Hail and Ride

In the so-called hail-and-ride system buses will stop anywhere along designated segments of their route when indicated by a potential passenger [131].

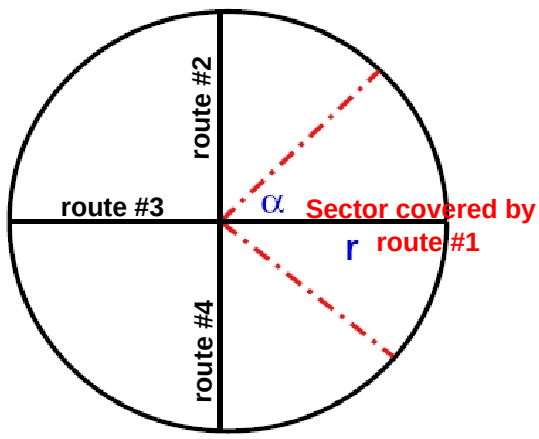


Figure 5.2: City interpretation. Each radius represents route. Sector between red lines is covered with one route. Each point of the circle represents one passenger.

Let us take that the city area covers a circle with the radius $r > 0$ as shown in Fig. 5.2. It will represent the city. Per unit area of the circle area we assume one inhabitant. Thus the density of the population is regular. We further simplify the situation by assuming that each inhabitant needs to travel to the center of the circle only. There is only one way for the habitant to do that: to walk to the nearest bus route and catch the next bus. The bus routes are represented by regular radii (with equal angles between them). Thus, if we have $N_R > 1$ routes, then the angle between them will be $2\alpha = 2\pi/N_R$. There are no stations, and the bus will stop anywhere along the route. Because of this we will further assume that the bus

spends no time when stopping. Therefore the fastest way for the passenger to get to the bus is to walk to the nearest route along a line that is perpendicular to the route. In this very simple case we ignore the fact, that actually it maybe better for the passenger not to take the perpendicular but a more "center-directed" way. However it turns out that in our model this strategy leads to very small improvement, such that our approximation does not influence our qualitative and quantitative results. We will show this in detail in section 5.2.

Let us implement $N_B > 0$ buses running on this system servicing each route on a regular basis. Then at each route point (potential stopping point) the time between the two consecutive buses will be

$$t_{b2b} = \frac{2rN_R}{N_B V_B}, \tag{5.1}$$

where V_B is the bus velocity. Further let us assume that there is no known schedule. Therefore in a first approximation the average waiting time for any passenger will be

$$t_{waiting} = \frac{rN_R}{N_B V_B}. \tag{5.2}$$

Here we will not take into account that actually this is the minimal average waiting time. Which may be increased by any bus delay. However this induces no qualitative changes as shown in section 5.2.

Let us denote by V_w the walking velocity, and by $k = V_B/V_w$ the velocity ratio.

In this simple case the inhabitants are not allowed to walk directly to the center, even if it may be the fastest way to get there.

In such a system with the above limitations the time that passenger i spends travelling to the center of the city is

$$t_i = t_{walking} + t_{waiting} + t_{bus}, \tag{5.3}$$

where

$$t_{walking} = \frac{y_i}{V_w}, \tag{5.4}$$

$$t_{waiting} = \frac{rN_R}{N_B V_B} \tag{5.5}$$

and

$$t_{bus} = \frac{x_i}{V_B}. \tag{5.6}$$

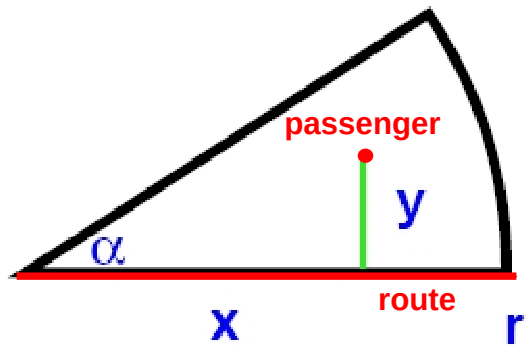


Figure 5.3: Half of the sector between routes bounded by one route, and the corresponding sector boundary. The sector angle is $\alpha = \pi/N_R$
.

With this the average travelling time will be

$$T = \frac{1}{N_{sites}} \sum_{N_{sites}} t_i. \tag{5.7}$$

Because of the regular geometry of the routes the average time T_{circle} over the complete area of the circle is equal to the average time T_α over the sector with angle $\alpha = \frac{\pi}{N_R}$. See Fig. 5.3.

For such a sector we may look for the integral of all travelling time averaged over this sector.

$$\sum t_i = \int\int_{sector} [\frac{y}{V_w} + \frac{rN_R}{N_B V_B} + \frac{x}{V_B}]\, dx\, dy \tag{5.8}$$

However, it is easier to solve this in polar coordinates.

$$\begin{aligned} \sum t_i = \int_0^\alpha \int_0^r [\frac{\rho sin\varphi}{V_w} + \frac{rN_R}{N_B V_B} + \frac{\rho cos\varphi}{V_B}]\rho\, d\rho\, d\varphi = \\ r^3 \int_0^\alpha [\frac{sin\varphi}{3V_w} + \frac{N_R}{2N_B V_B} + \frac{cos\varphi}{3V_B}]\, d\varphi = \\ \frac{r^3}{V_w}(\frac{1}{3} - \frac{cos\alpha}{3} + \frac{\pi}{2N_B k} + \frac{sin\alpha}{3k}). \end{aligned} \tag{5.9}$$

With this, the travelling time averaged over the sector (and therefore over the whole circle) will be

$$T = \frac{2N_R}{\pi r^2} \cdot \frac{r^3}{V_w}\left(\frac{1}{3} - \frac{cos\frac{\pi}{N_R}}{3} + \frac{\pi}{2N_B k} + \frac{sin\frac{\pi}{N_R}}{3k}\right) =$$
$$\frac{2r}{\pi V_w}\left(\underbrace{\frac{N_R}{3} - \frac{N_R cos\frac{\pi}{N_R}}{3}}_{\text{walking}} + \underbrace{\frac{\pi N_R}{2N_B k}}_{\text{waiting}} + \underbrace{\frac{N_R sin\frac{\pi}{N_R}}{3k}}_{\text{bus}}\right). \tag{5.10}$$

From this equation we see that the average travelling time depends linearly on the radius of the city. While the number of routes, the number of buses and the velocity ratio enter in a non-linear form.

Let us make the assumption that the number of routes N_R is much larger than π (much more than three in other words). Then, in good approximation we can assume that $sin\frac{\pi}{N_R} \approx \frac{\pi}{N_R}$ and $cos\frac{\pi}{N_R} \approx 1 - \frac{\pi^2}{2N_R^2}$. Then

$$T = \frac{2r}{V_w}\left(\frac{\pi}{6N_R} + \frac{N_R}{2N_B k} + \frac{1}{3k}\right). \tag{5.11}$$

Our goal is to minimize the average travelling time as far as possible for any given number of buses (level of investment). The velocity ratio is more or less fixed. So we can treat all other variables in this formula as constants. Therefore the only dependency is on the number of routes. For minimal T the first derivative should be equal to zero $T(N_R)' = 0$. Then

$$\frac{1}{2N_B k} = \frac{\pi}{6N_R^2}. \tag{5.12}$$

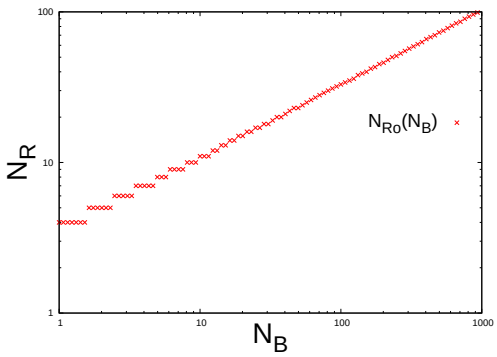


Figure 5.4: The optimal number of routes N_{Ro} as a function of the number of buses N_B. The velocity ratio is fixed as $k = 10$. The plot is shown in a log-log scale.

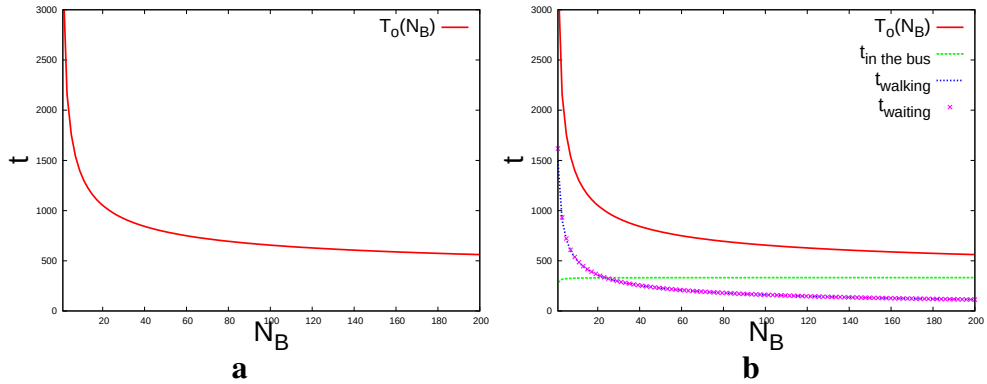


Figure 5.5: Mean travelling time T as a function of the number of buses N_B, while the number of routes N_R is optimal. $V_w = 1$, $V_B = 10$, $k = 10$, and $r = 5000$. **a**) Optimal travelling time $T_o(N_B)$ **b**) Optimal travelling time $T_o(N_B)$ and its components - mean walking time $t_{walking}$, mean waiting time $t_{waiting}$, and mean time spent in the bus t_{bus}

From this condition we find that the minimal average travelling time for fixed velocity ratio and for given number of buses occurs for the following number of routes

$$N_R = \sqrt{\frac{\pi N_B k}{3}}. \tag{5.13}$$

This is the optimal number of routes N_{Ro} for this model. As we can see it does not depend on the city size (radius r) at all. Its dependence on the number of buses shown on Fig. 5.4.

Now, if we insert N_{Ro} into equation (5.10), and assume that $k = 10$, $V_w = 1$, $V_B = 10$ and $r = 5000$ we find the following behavior of the optimal average travelling time as function of the number of buses.

$$T_o(N_B) = \frac{2r}{V_w}\left(\frac{\pi\sqrt{3}}{6\sqrt{\pi N_B k}} + \frac{\sqrt{\pi N_B k}}{2N_B k\sqrt{3}} + \frac{1}{3k}\right) = \frac{2r}{3V_B}\left(\underbrace{\frac{\sqrt{3\pi k}}{2\sqrt{N_B}}}_{\text{walking}} + \underbrace{\frac{\sqrt{3\pi k}}{2\sqrt{N_B}}}_{\text{waiting}} + \underbrace{1}_{\text{bus}}\right). \tag{5.14}$$

$$T_o(N_B) = \frac{2r}{3V_B} + \frac{2r\sqrt{3\pi k}}{3V_B\sqrt{N_B}}. \tag{5.15}$$

As one can see from Fig. 5.5a beyond some point further increasing the number of buses (and therefore number of routes) is not a very efficient strategy. In other words there is some value beyond which it doesn't make sense to invest more into a public transport network of this topology. Fig. 5.5b shows how the individual components of the mean travelling time behave. From that plot and from the equation 5.14 it turns out that at the optimal point the mean walking time is equal to the mean waiting time, while the time spent in the bus doesn't depend on number of buses N_B.

$$t_{bus} = \frac{2r}{3V_B}. \tag{5.16}$$

$$t_{walking} = t_{waiting} = \frac{2r\sqrt{3\pi k}}{3V_B\sqrt{N_B}}. \tag{5.17}$$

$$T_o(N_B) = t_{bus}(r, V_B) + 2t_{waiting}(r, V_B, V_w, N_B). \tag{5.18}$$

Thus we conclude that the optimal mean travelling time $T_o(N_B)$ is reduced by increasing the number N_B of buses, however this reduction follows the inverse square root ofN_B.

$$T_o(N_B) = A + BN_B^{-\frac{1}{2}}. \tag{5.19}$$

5.2 Hail and Ride with complications

As noted above we have neglected so far two potentially perturbing factors: first the possibility of the passenger walking not perpendicular way to the route, but rather center-directed and secondly possible bus delays. Let us implement those.

Let us first treat the case where a passenger takes an alternative center-directed path, increasing the walking distance (see Fig. 5.6). Let us denote the length of the alternative path $qy \geqq y$, where $q \geqq 1$ is the walking coefficient. Then, for this passenger the distance taken by bus decreases by $y\sqrt{q^2-1}$.

Now look at Fig. 5.7. Let us take that the expected average time interval between the buses is $t_{b2b} = 5$ and at each time interval one passenger arrives to the route. If buses arrive on time like on Fig. 5.7a then the average waiting time will be $t_{expected} = (0.5 + 1.5 + ... + 4.5) * 2/10 = 2.5$, exactly one half of

the time interval between the buses, as it should be. However if a bus delays for 2 time intervals (like on Fig. 5.7b) then the average waiting time increases $t_{delays} = (0.5+1.5+...+6.5+0.5+1.5+2.5)/10 = 2.9$. Actually it can increase up to the factor of two, when $t_{delays} = t_{b2b}$. So let us take that $1 \leq z \leq 2$ is the delay factor.

Taking into account those complications the passenger travelling time (equation 5.3) will be

$$t_i = \frac{qy_i}{V_w} + \frac{zrN_R}{N_B V_B} + \frac{x_i - y_i\sqrt{q^2-1}}{V_B} \tag{5.20}$$

and therefore

$$\sum t_i = \int_0^{\alpha}\int_0^{r}[\frac{q\rho sin\varphi}{V_w} + \frac{zrN_R}{N_B V_B} + \frac{\rho cos\varphi - \rho sin\varphi\sqrt{q^2-1}}{V_B}]\rho\, d\rho\, d\varphi = \frac{r^3}{V_w}(\frac{q}{3} - \frac{qcos\alpha}{3} + \frac{z\pi}{2N_B k} + \frac{sin\alpha + \sqrt{q^2-1}(cos\alpha - 1)}{3k}). \tag{5.21}$$

So then taking the same approximations as above the average travelling time will be

$$T = \frac{2r}{V_w}(\underbrace{\frac{\pi q}{6N_R}}_{\text{walking}} + \underbrace{\frac{zN_R}{2N_B k}}_{\text{waiting}} + \underbrace{\frac{1}{3k} - \frac{\pi\sqrt{q^2-1}}{6kN_R}}_{\text{bus}}). \tag{5.22}$$

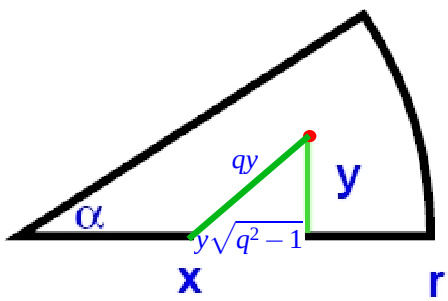


Figure 5.6: Half of the sector between two routes bounded with one route, and its corresponding sector boundary. Sector angle $\alpha = \pi/N_R$. The perpendicular footpath is denoted by y, the alternative footpath by qy and the route interval reduction by $y\sqrt{q^2-1}$.

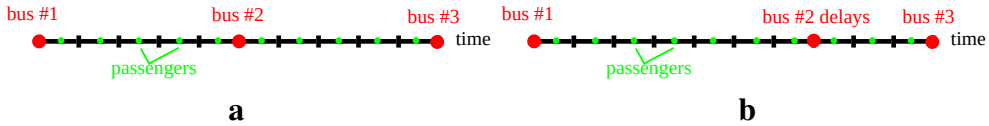


Figure 5.7: Time line. Red points represent instances when three consecutive buses arrive at the same point of the route. At the centre of each time interval one passenger arrives to that point. a) all buses arrive as expected b) bus 2 is delayed by 2 time intervals.

As follows from equation 5.22 t_{bus} in this case also depends in a similar way as before on the number of routes N_R even after simplification.

$$T = \frac{r}{3V_B}\left(\frac{\pi(kq-\sqrt{q^2-1})}{N_R} + \frac{3zN_R}{N_B} + 2\right). \tag{5.23}$$

Taking $T(N_R)' = 0$ for the optimal choise of N_R we find

$$\frac{3z}{N_B} = \frac{\pi(kq-\sqrt{q^2-1})}{N_R^2} \tag{5.24}$$

and therefore the optimal number of routes

$$N_{Ro}(N_B) = \sqrt{\frac{\pi N_B(kq-\sqrt{q^2-1})}{3z}}, \tag{5.25}$$

and average optimal travelling time

$$T_o(N_B) = \frac{2r}{3V_B} + \frac{2r\sqrt{3\pi(kq-\sqrt{q^2-1})z}}{3V_B\sqrt{N_B}}. \tag{5.26}$$

Comparing to previous results we see no qualitative changes. The optimal mean travelling time $T_o(N_B)$ linearly depends on $N_B^{-\frac{1}{2}}$, and the optimal number of routes N_R depends on number of buses N_B as a square root. Still $t_{walking} \approx t_{waiting}$ and $t_{bus} \approx \frac{2r}{3V_B}$. Looking on quantitative outcomes we see that due to delays the optimal waiting and walking time increase with the factor of $1 \leq \sqrt{z} \leq 1.41$. Our walking coefficient q instead should decrease travelling time. Let us look for the minimum of the next expression (the term in brackets in equation 5.25)

$$f(q) = kq - \sqrt{q^2-1} \tag{5.27}$$

$$f'(q) = k - \frac{1}{\sqrt{q^2 - 1}}$$

When $f'(q) = 0$ our walking coefficient is $q = \frac{\sqrt{k^2+1}}{k}$. Then our expression gives $\sqrt{k^2+1} - 1/k$. For the bus to walking speed ratio $k = 10$ this decreases the optimal walking and waiting time by a factor of 1.0025. That is definitely factor which can be ignored.

5.3 Routes with regular stations

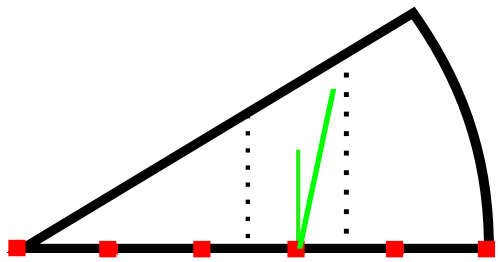


Figure 5.8: Half of the sector between routes with implemented stations.

Now let us to turn to the case of routes with the stations. The distance between stations is regular and equal. The number of stations N_S is an additional variable (Fig. 5.6). In this case integration gives us integrals we cannot solve analytically, therefore we chose to use computer simulations of this model. This allows us to refine the model:
a) if it is overall faster for a passenger to walk not to the nearest station, but to the next one - he takes that way. However it is expected that such situation would occur rarely - due to the results observed in previous section;
b)Some passengers living near to the center of the city will walk directly there;
c) each passenger lives at a given lattice node. For our simulations we used a lattice with a unit grid of $X = \sqrt{\pi r^2/2N_R}/500$;
d) at each station the bus waits for some regular time $t_{boarding}$.

Thus we have four fixed parameters: $t_{boarding} = 30$, $V_w = 1$, $V_B = 10$ and therefore $k = 10$. And we have four parameters to play with: N_B in the range [20,1200], N_R in the range [10,550], N_S in the range [5, 275] and r in the range [5000, 50000].

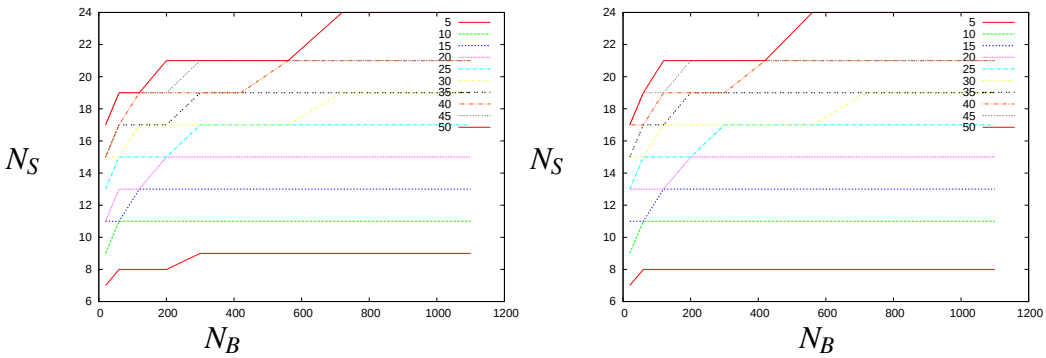


Figure 5.9: The optimal number of stations N_{So} as a function of the number of buses N_B, while the number of routes N_R is optimal. Different lines represent different chosen radii r. On the first plot the optimal values are searched as to minimise the average overall time T, while on the second plot we maximise the average overall velocity V.

For this model we also check how the overall average 'velocity' V behaves. While we define the travelling velocity of a passenger as $v_i = \sqrt{x_i^2 + y_i^2}/t_i$, it is not actually the velocity of passenger movement along his trajectory, but the average velocity of a hypothetical movement along the straight line to the center.

From these simulations we find a set of optimal pairs N_{Ro} and N_{So} for different combinations of N_B and r.

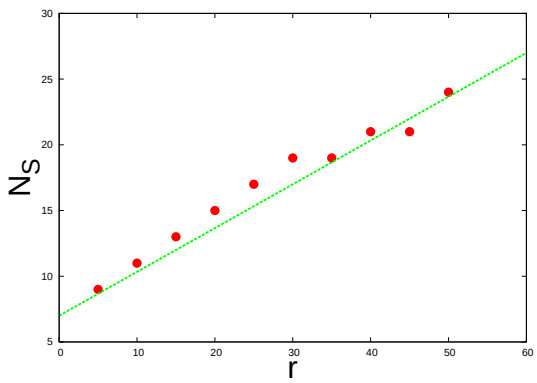


Figure 5.10: The maximal optimal number of stations N_{So} as a function of the radius of the city r, while the number of routes N_R is optimal. Optimal values are searched as to minimise the average overall time T

On Fig. 5.9 we see how the optimal number of stations N_{So} changes with the number of buses, while the number of routes N_R is optimal for the current N_B. Different lines represent different chosen radii r. On the first plot the optimal values are searched as to minimise the average overall time T, while on the second plot we maximise the average overall velocity V. There is no significant difference. Increasing the number of buses N_B the optimal number of stations N_{So} increase. But only up to some value. After this it becomes inefficient to further increase the number of stations. Perhaps, because loosing boarding time on each station has a stronger effect than decreasing slightly the walking time. The maximal optimal number of stations N_{So} increases linearly with the size of the city as shown on Fig. 5.10.

$$N_{So}\ A + Br \tag{5.28}$$

In Fig. 5.11 we show how the optimal number of routes N_{Ro} changes as function of the number of buses, while the number of stations N_S is optimal for the given N_B. Different lines represent different chosen radii r. On the first plot the optimal values are found by minimising the average overall time T, while in the second plot we maximise the average overall velocity V. On the first plot one can see that the original no-station solution fits the simulated results quite well. It again turns out that the optimal number of routes N_{Ro} does not depend on the radius r at all, and increases almost as the square root of the number of buses N_B.

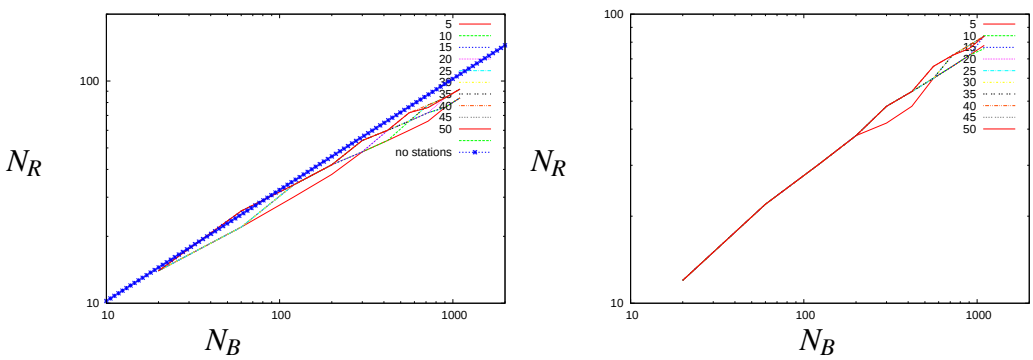


Figure 5.11: The optimal number of routes N_{Ro} as a function of the number of buses N_B, while the number of stations N_S is optimal. Different lines represent different chosen radii r. On the first plot the optimal values are searched as to minimise the average overall time T, while on the second plot we maximise the average overall velocity V.

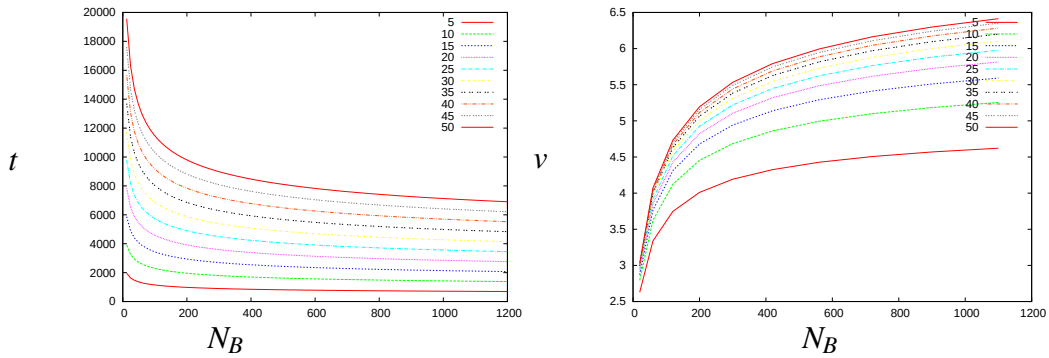


Figure 5.12: Optimal average overall time T_o and optimal average overall velocity V_o as a functions of the number of buses N_B, while the number of routes N_R and the number of stations N_S are optimal. Different lines represent different chosen radii r. On the first plot the optimal values are searched as to minimise the average overall time T, while on the second plot we maximise the average overall velocity V.

In Fig. 5.12 we represent the behavior of the optimised variables, optimal overall mean time T_o, and optimal overall mean velocity V_o, as functions of the number of buses N_B. The number of stations N_S and the number of routes N_R are

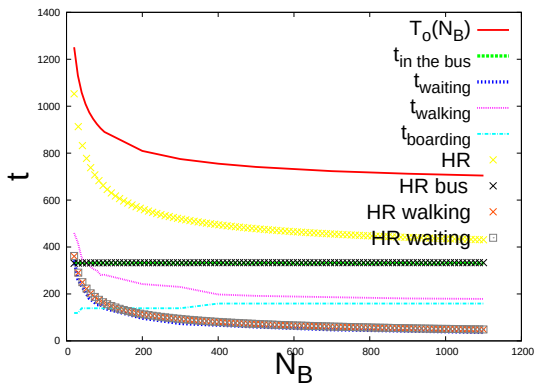


Figure 5.13: Optimal average overall time T_o and its components $t_{walking}$, $t_{waiting}$, t_{bus} and $t_{boarding}$ as a functions of the number of buses N_B, while the number of routes N_R and the number of stations N_S are optimal. Radius r is taken as 5000. Solid and dashed lines represent simulation results for model with stations, while lines of crosses and squares represent results for hail and ride case.

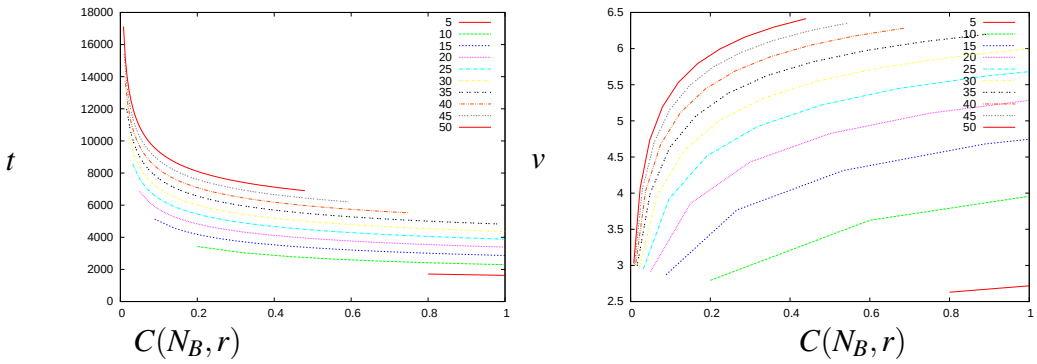


Figure 5.14: The average overall optimal time T_o and the average overall optimal velocity V_o as a functions of the cost function $C(N_B, r)$, while the number of routes N_R and the number of stations N_S are optimal. Different lines represent different chosen radii r. On the first plot the optimal values are searched as to minimise the average overall time T, while on the second plot we maximise the average overall velocity V.

always chosen as the optimal ones for current number of buses N_B. Different lines represent different chosen radii r. As one can see from both plots, the effect of increasing the number of buses becomes weaker for larger N_B. The slope of the lines become smaller and smaller. The smaller the city is, the faster this happens.

Let us look how the time components behave when we implement stations. In Fig. 5.13 the optimal mean travelling time T_o and its components $t_{walking}$, $t_{waiting}$, t_{bus} and $t_{boarding}$ are shown as a functions of the number of buses N_B, while the number of routes N_R and the number of stations N_S are optimal. For comparisonn we also show the corresponding optimal mean travelling time and its components for the hail and ride system. The radius r is fixed to 5000 for both cases. As we can see $t_{boarding}$ increases a bit for small values of N_B exactly at those values where the optimal number of stations N_{So} increases. It linearly depends on the optimal number of stations N_{So}.

$$t_{boarding} \sim A + BN_{So}. \tag{5.29}$$

The time spent in the bus t_{bus} is exactly fitted by the bus time for the hail and ride case. So we can conclude that

$$t_{bus} \sim \frac{2r}{3V_B}. \tag{5.30}$$

The average waiting time $t_{waiting}$ is also well reproduced by the hail and ride result.

$$t_{waiting} \sim \frac{r\sqrt{\pi}}{V_B\sqrt{3N_B}}. \tag{5.31}$$

The average walking time $t_{walking}$ again behaves in the same way as for the hail and ride case, and so in the same way as $t_{waiting}$, however its values increase significantly, due to the fact that most passengers cannot take the shortest footpath to the route, but need to take a footpath that leads directly to the station.

$$t_{walking} \sim A + Bt_{waiting}. \tag{5.32}$$

Therefore the overall mean travelling value increases in comparison to the hail and ride case, however its behavior follows the same dependance as before.

$$T_o(N_B) \sim A + BN_B^{-\frac{1}{2}}. \tag{5.33}$$

Another observation from Fig. 5.12 is that for small cities it is much more expensive to sustain such bus network at all. In our model the number of buses N_B represents the investment into the public transport. While the circle area represents the number of inhabitants. Then we can say that the ratio

$$C = N_B/r^2 \tag{5.34}$$

(with some price coefficient) represents, how much each inhabitant should pay for such kind of bus network. Let us call this the cost function. Fig. 5.14 shows how the average overall time T, and the average overall velocity V depend on the cost function $C(N_B, r)$. The number of stations N_S and the number of routes N_R are optimal for the given N_B. Different lines represent different chosen radii r. On the first plot the optimal values are determined as to minimize the average overall time T, while on the second plot they are determined to maximize the average overall velocity V. These demonstrate that it is very expensive per inhabitant of a small city to have a transport network at all. On the other hand for big city it is comparatively cheap to have a well equipped network and even operate a very efficient number of buses.

5.4 Conclusions

The main conclusions from this part of our work are the following:

- For the described radially structured public transport network there exist an optimal number of routes N_{Ro} as function of the number of buses N_B and the relation is given by $N_{Ro} \sim \sqrt{N_B}$. It can be found analytically for the simpler case of a hail and ride system and the result fits perfectly our simulated results for more complicated cases. This optimal number of routes N_{Ro} does not depend on the size of the city. Perhaps this observation may be related to the radial geometry, or, maybe, just due to two-dimensional space.

- For the described radially structured public transport network there exist an optimal number of stations N_{So} as function of the number of buses N_B. It grows linearly with the radius r of the city and depends only weakly on the number of buses N_B. It seems that for any fixed radius there is always a maximal optimal number of stations, whatever the other parameters are.

- If the mean travelling time $T_o(N_B)$ is optimal (and therefore number of routes N_R is optimal) - then for hail and ride case, the average waiting time is equal to the average walking time $t_{waiting} = t_{walking}$. For the case of fixed stations along routes these times are proportional to each other.

- Optimising either the average travelling time or the mean velocity we find a limit to the effective investment costs (implying more buses in other words). A reason for this is that the optimal travelling time depends on the number of buses N_B as $T_o(N_B) \sim A + BN_B^{-\frac{1}{2}}$, where A and B are constant, for fixed radius r and velocities.

- Finally we observe that for this kind of public network there is an inverse dependency between the size of the city (or number of inhabitants) and the price that should be paid by each inhabitant for an optimal transportation network.

Conclusions

In this study, we have performed a comprehensive analysis of public transport networks (PTN) combining tools of complex network theory, computer modelling, and analytical calculations. We have started from an empirical analysis of the PTN of 14 major cities of the world and have determined their principal characteristics in terms of the complex network theory. With these data at hand, we proceeded with PTN modelling. In both approaches we were interested in PTN characteristics as well as in particular phenomena which might take place on PTN (attack vulnerability being the most prominent example). Especially helpful in our analysis was the use of different network representations (different spaces, introduced in section 2.2). Whereas former PTN studies used some of these, here within a systematic approach we calculate PTN characteristics as they show up in all $\mathbb{L}$-, $\mathbb{P}$-, $\mathbb{C}$-, and $\mathbb{B}$-spaces. Detailed conclusions of our studies are given at the end of each of the thesis chapters. Summarizing them we can make the following statements.

Our empirical analysis gives strong evidence, that the networks under consideration appear to be strongly correlated small-world structures with high values of clustering coefficients and comparatively low mean shortest path values. Standard network characteristics that we find correspond to features a passenger is interested in when using public transport (they are summarized in table 2.2). Beyond traditional network characteristics there are specific features unique for PTN and networks with similar construction principles that we have addressed. In particular, public transport routes are often found to proceed in parallel for a sequence of stations. We have quantified this behavior in terms of the harness distribution. The harness concept may be also be useful for a quantitative description of other embedded networks with real space links such as cables, pipes, or neurons etc. Moreover, our analysis of the geographical data for Berlin and Paris reveals a self-avoiding walk scaling of PTN routes.

We further have introduced several PTN models. This was done both in 2d

and 1d embedding spaces. A particular feature of the models considered is that opposite to the majority of complex network models, where a network grows due to adding separate nodes, in our case the growth is in terms of public transport lines. In particular, our model of mutually interacting self-avoiding walks in 2d captures both special features of PTN as well as generating profiles of network characteristics in the various representations. The method used, a non equilibrium growth model in terms of attractive self-avoiding walks on a square lattice may further be extended to study the effects of geographical constraints e.g. coastlines, rivers and bridges or disorder.

We continued our analysis by studying the behavior of a PTN under attacks. Similar to other real-world and model complex networks, the PTN manifest very different behaviour under attacks of different scenarios. With some notable exceptions they appear to be robust to random attacks but more vulnerable to attacks targeted at nodes with particular importance as measured by the values of certain characteristics (the most significant being the first and second neighbour numbers, as well as the betweenness and stress centralities). The observed difference between attack scenarios based on the initial and the recalculated distributions shows that the network structure changes essentially during the attack sequence. This is necessarily to be taken into account in the construction of efficient strategies for the protection of these network. In our study we have also attempted to compare our results with the predictions of percolation theory on networks. Due to the sizes of these systems which are far from the thermodynamic limit and the rather small sample of networks no quantitative comparison appeared possible. However, qualitative predictions about the location of segmentation thresholds and thus the vulnerability could be verified. In particular, this enabled us to propose criteria that allow an a priori estimate of PTN robustness.

In the concluding part of our analysis we have considered PTN optimisation. The majority of studies of this issue are concerned with dynamical processes that occur on networks (traffic jams, schedules, logistics, etc). The purpose of our study was to analyse the optimisation of topology of PTN for a given simple city model, optimising either the average travelling time of all citizens, or their mean velocity, while travelling. In particular, for the model we considered it follows that optimising either the average travelling time or mean velocity there seems to be a limit on the effective investment costs. Moreover, it has been observed that there is an inverse dependence between the size of the 'city' (or number of inhabitants) and the price that should be paid for the transportation network.

Bibliography

[1] C. Achim, R. Chapleau, C. Chriqui and M. Florian (1976) *Transit Route Development and Evaluation Techniques: a Survey of the State of the Art*, Centre de Recherche sur les Transports, University of Montreal, Canada, 47.

[2] R. Albert, H. Jeong, and A.-L. Barabási (1999) *Nature (London)* **401**, 130.

[3] R. Albert and A.-L. Barabasi (2002) *Rev. Mod. Phys.* **74**, 47.

[4] R. Albert, I. Albert, G. L. Nakarado (2004) *Phys. Rev. E* **69**, 025103.

[5] L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11149.

[6] P. Angeloudis and D. Fisk (2006) *Physica A* **367**, 553 .

[7] A.-L. Barabási, R. Albert (1999) *Science* **286**, 509.

[8] A.-L. Barabási, R. Albert, and H. Jeong (1999) *Physica A* **272**, 173.

[9] A.-L. Barabási, R. Albert, H. Jeong (2000) *Physica A* **281**, 69.

[10] A.-L. Barabási (2002) *Linked: The New Science of Networks*, Perseus Press, New York.

[11] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani (2004) *Proc. Nat. Acad. Sci. USA* **101**, 3747.

[12] B. Berche, C. von Ferber, and T. Holovatch (2009) *AIP Conference Proceedings, Melville, New York* **1198**, 3.

[13] B. Berche, C. von Ferber, T. Holovatch, Yu. Holovatch (2009) *Eur. Phys. J. B* **71**, 125.

[14] B. Berche, C. von Ferber, Yu. Holovatch, and T. Holovatch (2010) *Dynamics of Socio-Economic Systems* **2** 42.

[15] L. Benguigui (1992) *J. Phys. I France* **2**, 385.

[16] B. Bollobás (1985) *Random Graphs*, Academic, London.

[17] S. Bornholdt and H. Schuster eds. (2003) *Handbook of Graphs and Networks*, Wiley-VCH, Weinheim.

[18] U. Brandes (2001) *J. Math. Sociology* **25**, 163.

[19] U. Brandes (2008) *Social Networks* **30(2)** 136.

[20] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener (2000) *Comput. Netw.* **33**, 309.

[21] M. Buchanan (2002) *Small Worlds and the Groundbreaking Science of Networks*, Norton, New York.

[22] B. F. Byrne (1975) *Transportation Res*. **9**, 97.

[23] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, D. J. Watts (2000) *Phys. Rev. Lett.* **85**, 5468.

[24] A. Cardillo, S. Scellato, V. Latora, and S. Porta (2006) *Phys. Rev. E* **73**, 066107.

[25] P. Carmeliet, and M. Tessier-Lavigne (2005) *Nature* **436**, 193.

[26] S. K. Chang and P. M. Schonfeld (1991) *Transportation Res*. **25**, 453.

[27] H. Chang, B.-B. Su, Y.-P. Zhou, and D.-R. He (2007) *Physica A* **383**, 687.

[28] T. A. Chua (1984) *Transportation* **12**, 147.

[29] A. Clauset, C. R. Shalizi, and M. E. J. Newman (2009) *SIAM Review* **51**, 661.

[30] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin (2000) *Phys. Rev. Lett.* **85**, 4626.

[31] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin (2001) *Phys. Rev. Lett.* **86**, 3682.

[32] R.Cohen, D. ben-Avraham, and S. Havlin (2002) *Phys. Rev. E* **66**, 036113.

[33] R. Cohen, S. Havlin, and D. ben-Avraham (2003) *Phys. Rev. Lett.* **91**, 247901.

[34] P. Crucitti, V. Latora, M. Marchiori (2004) *Physica A* **338**, 92.

[35] M. R. Cutkosky, A. B. Conru, and S-H. Lee (1994) *AIEDAM* **8**, 1.

[36] J. Dall, M. Christensen (2002) *Phys. Rev. E* **66**, 016121.

[37] C. Domb (1996) *The Critical Point*, Taylor & Francis, London, Bristol.

[38] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin (2000) *Phys. Rev. Lett.* **85** 4633.

[39] S. N. Dorogovtsev and J. F. F. Mendes (2002) *Adv. Phys.* **51** 1079.

[40] S. N. Dorogovtsev and S. N. Mendes (2003) *Evolution of Networks*, Oxford University Press, Oxford.

[41] S. N. Dorogovtsev, J. F. F. Mendes, and J. G. Olveira (2006) *Phys. Rev. E* **73**, 056122.

[42] E. J. Doyle and R. J. Vaughan (1981) *Transportation Res*. **15**, 149.

[43] P. Erdös and A. Rényi (1959) *Publ. Math. (Debrecen)* **6**, 290; (1960) *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17; (1961) *Bull. Inst. Int. Stat.* **38**, 343.

[44] J. W. Essam (1980) *Rep. Prog. Phys.* **43**, 833.

[45] M. Faloutsos, P. Faloutsos, and C. Faloutsos (1999) *Comput. Commun. Rev.* **29**, 251.

[46] M. Y. Fawaz and G. F. Newell (1976) *Trasportation Res*. **10**, 111.

[47] C. von Ferber, Yu.Holovatch, and V.Palchykov (2005) *Condens. Matter Phys.* **8**, 225.

[48] C. von Ferber, T. Holovatch, Yu. Holovatch, and V. Palchykov (2007) *Physica A* **380**, 585.

[49] C. von Ferber, T. Holovatch, V. Palchykov Physical (2008) *Papers of the Taras Shevchenko Scientific Society. (in Ukrainian)* **7**, 100.

[50] C. von Ferber, T. Holovatch, Yu. Holovatch, and V. Palchykov (2009) *P Eur. Phys. J. B* **68**, 261.

[51] C. von Ferber, T. Holovatch, Yu. Holovatch, and V. Palchykov (2009) *Traffic and Granular Flow '07. Springer*, 709.

[52] C. von Ferber, T. Holovatch, and Yu. Holovatch (2009) *Traffic and Granular Flow '07. Springer*, 721.

[53] R. Ferrer i Cancho and R. V. Solé, arXiv:cond-mat/0111222; S. Valverde, R. Ferrer i Cancho, and R. V. Solé (2002) *Europhys. Lett.* **60**, 512; R. Ferrer i Cancho and R. V. Solé (2003) *Lecture Notes in Physics, Springer, Berlin* **625**, 114.

[54] R. Ferrer i Cancho and R. V. Solé (2003) *Proc. Natl. Acad. Sci. USA.* **100**, 788; R. Ferrer i Cancho (2005) *Physica A* **345**, 275.

[55] L. C. Freeman (1977) *Sociometry* **40**, 35.

[56] A. Fronczak, P. Fronczak, and J. A. Holyst (2003) *Phys. Rev. E* **68**, 046126.

[57] M. T. Gastner and M. E. J. Newman (2006) *Eur. Phys. J. B* **49**, 247.

[58] M. Girvan and M. E. J. Newman (2002) *Proc. Natl. Acad. Sci. USA* **99**, 7821.

[59] M. S. Granovetter (1973) *Am. J. Sociol.* **78**, 1360.

[60] M. Guida, F. Maria (2007) *Chaos Solitons & Fractals*, **31**, 527.

[61] J.-L. Guillaume and M. Latapy (2006) *Physica A* **371**, 795.

[62] R. Guimera, L. A. N. Amaral (2004) *Eur. Phys. J. B* **38**, 381.

[63] R. Guimera, S. Mossa, A. Turtschi, L. A. N. Amaral (2005) *Proc. Nat. Acad. Sci. USA* **102**, 7794.

[64] P. Hage and F. Harary (1995) *Social Networks* **17**, 57.

[65] F. A. Haight (1964) *Operations Res.* **12**, 964.

[66] P. Holme, B. J. Kim, C. N. Yoon, S. K. Han (2002) *Phys. Rev. E* **65**, 056109.

[67] Yu. Holovatch, C. von Ferber, A. Olemskoi, T. Holovatch, O. Mryglod, I. Olemskoi, V. Palchykov (2006) *J. Phys. Stud. (in Ukrainian)* **10**, 247.

[68] J. A. Holyst, J. Sienkiewicz, A. Fronczak, P. Fronczak, and K. Suchecki (2005) *Phys. Rev. E* **72**, 026108.

[69] N. Hwang, and R. Houghtalen (1996) *Fundamentals of hydraulic Engineering Systems*, Prentice Hall, Upper Saddle River, NJ.

[70] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási (2000) *Nature (London)* **407**, 651.

[71] H. Jeong, S. P. Mason, A.-L. Barabási, Z. N. Oltvai (2001) *Nature (London)* **411**, 41.

[72] T. Kalisky, R. Cohen (2006) *Phys. Rev. E* **73**, 035101(R).

[73] K. S. Kim, L. Benguigui, M. Marinov (2003) *Cities* **20**, 3139.

[74] P. L. Krapivsky, S. Redner, F. Leyvraz (2000) *Phys. Rev. Lett.* **85**, 4629.

[75] V. Latora, M. Marchiori (2001) *Phys. Rev. Lett.* **87**, 198701.

[76] V. Latora, M. Marchiori (2002) *Physica A* **314**, 109.

[77] W. Li, X. Cai (2004) *Phys. Rev. E* **69**, 046106.

[78] W. Li, Q. A. Wang, L. Nivanen, A. Le Méhauté (2006) preprint cond-mat/0601091.

[79] X. Li and G. Chen (2003) *Physica A* **328**, 274.

[80] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg (2001) *Nature* **411**, 907.

[81] Z. Liu, Y.-C. Lai, N. Ye, and P. Dasgupta (2002) *Phys. Lett. A* **303**, 337.

[82] M. Marchiori, V. Latora (2000) *Physica A* **285**, 539.

[83] N. Mathias and V. Gopal (2001) *Phys. Rev. E* **63**, 021117.

[84] M. Molloy, B. A. Reed (1995) *Random Struct. Algorithms* **6(2/3)**, 161; (1998) *Combinatorics, Probability and Computing* **7**, 295.

[85] A. E. Motter and Y. C. Lai (2002) *Phys. Rev. E* **66**, 065102.

[86] A. E. Motter (2004) *Phys. Rev. Lett.* **93**, 098701.

[87] G. F. Newell (1979) *Transportation Science*, **13**, 20.

[88] G. F. Newell and C. F. Daganzo (1986) *Transportation Res*. **20**, 345.

[89] M. E. J. Newman, S. H. Strogatz, and D. J. Watts (2001) *Phys. Rev. E* **64**, 026118.

[90] M. E. J. Newman (2001) *Phys. Rev. E* **64**, 016131.

[91] M. E. J. Newman (2002) *Phys. Rev. Lett.* **89**, 208701.

[92] M. E. J. Newman (2003) *SIAM Review* **45**, 167.

[93] M. E. J. Newman (2003) *Phys. Rev. E* **67**, 026126.

[94] B. Nienhuis (1982) *Phys. Rev. Lett.* **49**, 1062.

[95] R. Pastor-Satorras, A. Vespignani (2001) *Phys. Rev. Lett.* **86**, 3200.

[96] S. B. Pattnaik et al (1998) *JTE* **368**.

[97] D. J. de S. Price (1976) *J. Amer. Soc. Inform. Sci.* **27**, 292.

[98] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras (2004) *Phys. Rev. E* **70**, 036106.

[99] G. Sabidussi (1966) *Psychometrika* **31**, 581.

[100] K. A. Seaton, L. M. Hackett (2004) *Physica A* **339**, 635.

[101] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, S. S. Manna (2003) *Phys. Rev. E* **67**, 036106.

[102] A. Shimbel (1953) *Bull. Math. Biophys.* **15**, 501.

[103] J. Sienkiewicz and J. A. Holyst (2005) *Phys. Rev. E* **72**, 046127.

[104] J. Sienkiewicz and J. A. Holyst (2005) *Acta Phys. Pol. B* **36**, 1771.

[105] H. A. Simon (1955) *Biometrika* **42**, 425.

[106] R. J. Smeed (1963) *J. Institution Engrs*. **10**, 5.

[107] R. V. Solé, J. M. Montoya (2001) *Proc. R. Soc. Lond. B* **268**, 2039.

[108] H. E. Stanley (1971) *Introduction to Phase Transitions and Critical Phenomena*, Clarendon Press, Oxford.

[109] D. Stauffer, A. Aharony (1991) *Introduction to Percolation Theory*, Taylor & Francis, London.

[110] Y. Tu (2000) *Nature (London)* **406**, 353.

[111] D. L. Van Oudheusden et al (1987) *Transportation* **14**.

[112] R. J. Vaughan and E. J. Doyle (1979) *Transportation Res*. **13**, 181.

[113] R. J. Vaughan (1986) *Transportation Res*. **20**, 215.

[114] D. Volchenkov and Ph. Blanchard (2008) *Phys. Rev. E* **75**, 026104; D. Volchenkov (2008) *Condens. Matter Phys.* **11**, 331.

[115] D. J. Watts, S. H. Strogatz (1998) *Nature (London)* **393**, 440.

[116] D. J. Watts (1999) *Small Worlds*, Princeton University Press, Princeton, NJ.

[117] D. J. Watts (2003) *Six Degrees: The Science of a Connected Age*, Norton, New York.

[118] J. G. White, E. Southgate, J. N. Thompson, and S. Brenner (1986) *Trans. Roy. Soc. London* **314**, 1.

[119] X. Xu, J. Hu, F. Liu, L. Liu (2007) *Physica A* **374**, 441448.

[120] R. Xulvi-Brunet, W. Pietsch, I. M. Sokolov (2003) *Phys. Rev. E* **68**, 036119.

[121] P.-P. Zhang, K. Chen, Y. He, T. Zhou, B.-B. Su, Y. Jin, H. Chang, Y.-P. Zhou, L.-C. Sun, B.-H. Wang, and D.-R. He (2006) *Physica A* **360**, 599.

[122] Portret of Leonard Euler from *http://www.nndb.com/people/954/000048810/*.

[123] Portret of Francis Guthrie from *http://www-history.mcs.st-and.ac.uk* .

[124] Portret of Gustav Robert Kirchhoff from *http://en.wikipedia.org/wiki/* .

[125] Foto of Paul Erdös from G. P. Csicsery documentary movie *"N is a number: a portret of Paul Erdös"* (1993).

[126] Foto of Alfréd Rényi from P. R. Halmos book *"I have a Photographic Memory"*, Americal Mathematical Society, Providence, RI, 1987.

[127] For links see *http://www.apta.com*.

[128] Due to an updated database numbers may slightly differ from those given in [48].

[129] Maps provided by *http://www.fahrinfo-berlin/Stadtplan* .

[130] Geocoding application on *http://developer.navteq.com/* .

[131] Accessible bus stop design guidance, Bus Priority Team technical advice note, BP1/06, Transport for London, 2006.

Abstract:

In this study, we have performed a comprehensive analysis of public transport networks(PTN) combining tools of complex network theory, computer modelling, and analytical calculations. We have started from an empirical analysis of the PTN of 14 major cities of the world and have determined their principal characteristics in terms of the complex network theory. Our empirical analysis gives a strong evidence, that the networks under consideration appear to be strongly correlated small-world structures with high values of clustering coefficients and comparatively low mean shortest path values. We further have introduced several PTN models. This was done both in 2d and 1d embedding spaces. In particular, our model of mutually interacting self-avoiding walks in 2d captures both special features of PTN as well as generating profiles of network characteristics in the various representations. We continued our analysis by studying the behavior of PTN under attacks. This enabled us to propose criteria that allow an a priori estimate of PTN robustness.

Keywords: complex networks, self-avoiding walk, random walk, vulnerability, resilience, targeted attacks, percolation, graph theory, harness effect, radial bus model, preferential attachment, public transport network.

Résumé :

Dans cette étude, nous produisons une analyse des réseaux de transport publics (acronyme PTN en anglais) en combinant des outils de la théorie des réseaux complexes, des simulations numériques et des approches analytiques. Nous avons commencé par une analyse empirique des PTN de 14 villes importantes dans le monde et en avons déterminé les principales caractéristiques en termes de réseaux complexes. Cette apporche empirique montre que les PTN apparaissent comme des réseaux ("small world") fortement corrélés avec des "coefficients d'agrégation" élevés et des "distances les plus courtes moyennes" comparativement faibles. Nous avons ensuite introduit divers modèles de PTN à 1 et 2 dimensions. Le modèle de marches aléatoires auto-évitantes (SAW) en interactions mutuelles capture certaines des propriété statistiques des PTN dans les divers modes de représentation. Nous avons poursuivi cette étude en examinant la résistance des PTN à divers scénarios d'attaques, ce qui permet de définir des critères de robustesse des réseaux considérés.

Mots clés : réseaux complexes, marche aléatoire auto-évitante, vulnérabilité, résilience, attaque ciblée, percolation, théorie des graphes, effet harnais, modèle du bus radial, attachement préférentiel, réseau de transport public.