THE BRATISLAVA PUBLIC TRANSPORT IN NETWORK ANALYSIS

Slavomír, ONDOŠ[1], Ivica PAULOVIČOVÁ[2], Lukáš, BELUŠÁK[3], Dagmar, KUSENDOVÁ[4]

Department of Human Geography and Demogeography, Faculty of Natural Science, Comenius University in Bratislava, Mlynská dolina, 842 15, Bratislava, Slovakia

[1]*ondos@fns.uniba.sk;* [2]*paulovicova@fns.uniba.sk;* [3]*belusak@fns.uniba.sk;* [4]*kusendova@fns.uniba.sk*

Abstract

This paper explores network approach possibilities to urban environment. Since the subject is too complex and difficult to cover in global, we keep focused on a selected sample of real-world spatially embedded network. Public transport is a likely candidate, due to increasing significance indicated in literature. First, we use the most relevant statistical properties of public transport network for the purpose of description. We are interested in whether edge weights have any significant effect on topology, or topology contains complete information on network; how the existence of transport hubs affects scaling behaviour; and if a specific topology creates artificial barriers within the city. We review and illustrate the analytical methods potentially valuable further. Second, we try to design a simple simulation, purpose of which is to estimate the strength of influence of basic distance based variables on the network topology, generated throughout decades on an evolutionary basis. Empirically, we use the multimodal Bratislava public transport system.

Keywords: public transport, network analysis, simulation, Bratislava

INTRODUCTION

Many real world complex systems can be described as networks with individuals in the real system represented by vertices and interrelationships among individuals represented by edges. Modelling of transport systems becomes important not only in large cities around the world. Public transport, being one of the spatial networks, consists of two basic elements, routes and stations. A route is formed along a number of stations. Under normal circumstances, passengers can travel from station A to station B along a certain route, and also travel from station B to station A along the same route in opposite direction. Therefore, public transport networks are usually treated as undirected graphs (Lu et al., 2007).

Public transport network is a result of a long-term evolution in a typical city. It reflects, similar to urban morphology, the history of built environment and current functional structure both in the same time. The information enclosed in the network topology itself seems to be richer than only documenting transport possibilities for everyday life of metropolitan inhabitants and visitors, despite services operated crucially depend on their transportation demands. Otherwise, transportation network would be inefficiently wasting public resources, which to some extent still can be the case. Our intention is to explore selected analytical possibilities concerning the network architecture built in public transport network and to find research paths leading potentially towards future topology improvements helping to make the network serving better than the network in its current evolution stage.

Traditional literature of transportation network is abundant. During the past few years, several public transport systems have been investigated using various concepts of complex networks. Most of previous studies have analyzed specific sub-networks of public transport networks in various large cities and in different parts of world. For instance, subway network analyzes of Boston by Latora and Marchiori (2000, 2002) who defined measures of local and global network efficiencies. They notably found that the small-world behaviors existed in that system. Seaton and Hackett (2004) calculated the clustering coefficient, path length and average degree vertex of the rail systems in Boston, US and Vienna, Austria. Also, Musso and Vuchic (1988), Vuchic and Musso (1991) focused on evolution and characteristics of subway networks. Sienkiewicz and Holyst (2005) have analyzed the bus and tram networks of Polish cities finding that some systems appeared to show a scale-free behavior, with scaling factors. A very similar analysis was offered by Xu et al. (2007) focusing the complexity of several bus networks in China. However, as far as the bus, subway, or tram sub-networks are not closed systems, the inclusion of additional sub-networks has significant impact on the overall network properties as has been shown for the subway and bus networks of Boston (Latora et al., 2001, 2002). Von Ferber et al. (2007) used complex network concepts to analyze the statistical properties of public transport networks of several large cities, looking at all technologies and accounting for the overlapping property of transit systems, notably finding a harness effect. They also attempted to model system based on number of stations and lines.

To understand the network topology we employ on one hand state-of-the-art description techniques, presented in the first part of the text, as well as an attempt to lay foundations of simulation capable of informing us on principles hidden beyond pure topological properties of the network, shown in the second part of the paper. Motivation for our research is in limited understanding of how a public transport network in our case in a moderate size city can be built and operating most efficiently with a specific heritage of historical infrastructure unlikely to be transformed too radically from technical or financial reasons. Tightening public budgets over latest period only support relevance for this research motivation.

DESCRIPTION OF DATA

In the first part of this paper we will focus on the real data. We will perform our study according to the following scheme: first choosing public transportation system for analysis, second collecting the data, then representing them and finally calculating chosen basic statistics.

Choosing and collection of data

The first step was choosing a public transport network. We will be interested in various characteristics of chosen network. Therefore to allow general properties of a network structure to manifest themselves, the network analyzed should be large enough in terms of numbers of nodes and edges. The results presented below are based on an analysis of public transport network of Bratislava with 584 stations and 1502 lines of 94 different public transportation routes. Bratislava as the biggest and the fastest growing city in the Slovakia has sufficiently developed public transport network. In our assumptions the public transport in Bratislava suffers from inappropriate structure based on historical background and is highly influenced by natural conditions. Thus we have chosen this, not completely efficient public transportation network.

We want to use minimum input information possible, in form of the list of currently existing service lines provided. These are easily transformed into a directed graph alternatively weighted by time needed to cover specific distance between two adjacent stations. Bratislava is covered by a web of 1502 directed station dyads, which means that only 0.44% of all 340,472 potential links are operated by the service provider, formally leaving us with a sparse network problem. Even further generalization is possible in form of simplification into a non-directed graph where to a large extent double entered information is discarded since majority of services run symmetrically in both directions between adjacent stations.

The second step towards analysing public transport in Bratislava is therefore to collect the data by the given way. This is typically done with the aim to represent the network as a graph with nodes/vertices and links/edges. By having stations/stops all linked by lines, public transport system are in fact physical networks. Nevertheless, there exists several ways to define them as graphs. For this analysis is important each station and line, not only in the city area, bud also station which are located outside of the city (in the case of Bratislava they are located in Hainburg (Austria) and Rajka (Hungary)). The schedules of public transport for Bratislava were downloaded from the provider. We used data publicly available from webpage of "Dopravný podnik Bratislava" (Transport Company Bratislava, 2012). Consequently, obtained data was brought into appropriate format to construct the ordered lists of stations serviced by each line from schedules (hereafter we do not make any differences between bus, electric trolleybus or tram public transport lines). From consecutive stations of routes where built directed edges according to real accessibility. These serve as a background for the network structure analysis. At last, we added the shortest time needed to travel from one station to next to every edge.

Representation

As presented in Barthélemy (2011), there are many possible ways of representing public transportation system as a network. For the aim of this paper we used the space-of-stops or the L-space representation as defined in von Ferber et al. (2008). Each station is represented in this graph by a node and an edge between two nodes indicates that these are consecutive stations of at least one route. As follows from previous, neighbours of a node are only those stations, which can be reached from it within single station trip. For purpose of this analysis we use directed graph so it better reflects real conditions of selected PTN. As weights we use inverse travel time between stations.

Public transportation network statistics

From all existing network analysis we have chosen identifying communities and modularity as characteristics shown in this paper. In our expectation, identified communities in the real network will be visible also in the simulation, as they there will be more links inside them as between them.

There are many possible algorithms for identifying communities in networks. The fastest way uses greedy algorithm for optimizing modularity (Clauset et al. 2004) which can lead to producing super-communities. Another way is by using simulated annealing (Guimera et al. 2004). We used algorithm presented in Blondel et al. (2008) consisting of two passes repeating iteratively until modularity cannot be further increased. Each pass is represented by one phase, where modularity is optimized by local changes, and second phase, where new composition of communities is made by aggregation. For better outcome, more measures were introduced to algorithm using structural as well as dynamical properties (Lambiotte et al. 2009). Modularity of a partition in a weighted network is calculated as (Blondel et al., 2008):

$$Q = \frac{1}{2*m}\sum_{i,j}\left[A_{ij} - \frac{k_i * k_j}{2*m}\right]\delta(c_i, c_j), \tag{1}$$

where m is half of the sum of all weights; A_{ij} is the weight of edge between i and j; k_i is sum of weights of the edges connected to i; $\delta(u,v)$ is 1 if $u = v$ and 0 otherwise; c_i is community to which node i belongs. Modularity compares the density of links inside communities to links between them and can acquire values between -1 and 1.

Using last described algorithm on directed weighted representation of real network resulted in 57 communities shown in figure 1 with only two hierarchical levels with modularity 0,885. 22 larger communities, bordered by a line, correspond to continuously urbanized fragments of Bratislava.

All the calculations and visualizations provided in this chapter were made in an open source interactive platform called Gephi (Gephi.org, 2012), version 0.8.1-beta. For geographical layout was used plugin Geo Layout. The information about positions of stations in geographical space was downloaded from Open Street Map (OpenStreetMap.org, 2012).

SIMULATION

The simulation exercise is based on a direct analogy with frequently used generalized linear modeling of raster represented urban land (Hu and Lo, 2007). One way of extracting hidden regularities, offering itself in our situation, but not commonly used is binomial logistic regression model predicting occurrence of transport service independently for each oriented dyad, irregularly spread across the urbanized territory of the city as a single input. The model actually predicts the probability of connection between theoretical boundaries 0.00 and 1.00, instead of occurrence itself, only based on a few theoretically relevant predictors. To some extent our exercise finds inspiration in Koenig and Bauriedel (2009).

Fig. 1. Communities in PTN presented as nodes in shades of grey with boundaries of bigger communuties.

We start from a set of localized stations defined by the geographic coordinates extracted from Open Street Map database, reflecting the location of transport service users and their transport destinations. This allows us to assume that location of stations plus their mutual location include also a relevant information on the purpose of their existence, which is the movement along unknown service lines running across unknown network drawn among already given locations. A preceding step of the more mature future analysis could be estimating the location for each station endogenously. Actual spatial distribution of locations is far from random. They appear in spatial clusters, linked together by strips of stations between clusters. They also reflect topography of the area since we deal with a city spread on both sides of Danube, one of the largest European rivers, meeting Carpathian mountain slopes, ending in a systematic spatial structure.

The purpose of any transport service is to remove the effect of geographical separation of distance by means of movement across space. Therefore, first driving force considered seems to be logically distance between two stations itself. Increasing length is expected to decrease the probability of connection. We employ log value of distance. Since point locations are scattered across a plane we also consider neighborhood relationships among them. The space between stations can be easily divided in zones of land, from which only one location is preferred on a distance basis. The points which do not prefer any location form borders between these zones. Geography then enters our consideration in an additional way. The Thiessen polygons constructed upon Delaunay principle allow us to construct the spatial contiguity matrix identifying neighbors from the universe of all possible connections. The resulting dummy predictor is expected to influence probability in positive direction. Being neighbors in space should increase probability of connection between these regions. This factor is related, but not necessarily the same as distance itself. The linear model will attempt to estimate how the two differ in effect on actual links distribution. The third factor considered is inspired by market potential constructed as an average distance to all other stations of the system increasing probability of connections distributed under expectations with a higher density in central city decreasing towards peripheries. Each dyad then gets average log value of two connected potentials, both directions same.

The estimation of a three-variable model allows to document direction, magnitude and significance of assumed linear effects on links distribution in empirical network. All of them have expected signs, more distant stations are less likely to be connected, neighbors are more likely to be connected, and central connections tend to be more likely than peripheral. So far logistic model based simulation works fine. But resulting network distribution uncovers limited applicability at this stage with only 738 dyads simulated correctly pointing at 49.13% success. The comparison between a real and a simulated network topology reveals that the model gives many dense disconnected clusters of very short links. Naturally, such a network is not a functional one, but it suggests how the regularities considered shape a real world network. In an extreme way it articulates the principle given already in our introductory description of point pattern in the data. To get closer to real world network one must find opposite directed influences capable of dispersing links out of dense clusters and of connecting them in one single network instead. Model might clearly benefit from inclusion of additional factors aiming at neighborhoods of individual locations.

Table 1. Coefficient estimates for three logistic model versions.

Model I	B	S. E.	t	Sig.	Exp(B)
Intercept	-7.949	0.593	-13.407	0.000	0.000
Log distance	-2.481	0.058	-42.488	0.000	0.084
Neighbors	1.559	0.095	16.441	0.000	4.756
Log potential	-0.086	0.191	-0.448	0.654	0.918
Log neighbors count	-0.399	0.251	-1.590	0.112	0.671
Log local potential	1.355	0.090	15.004	0.000	3.876

Model II	B	S. E.	t	Sig.	Exp(B)
Intercept	-7.927	0.591	-13.420	0.000	0.000
Log distance	-2.479	0.058	-42.628	0.000	0.084
Neighbors	1.557	0.095	16.451	0.000	4.744
Log neighbors count	-0.360	0.236	-1.530	0.126	0.697
Log local potential	1.328	0.068	19.512	0.000	3.774

Model III	B	S. E.	t	Sig.	Exp(B)
Intercept	-8.718	0.289	-30.173	0.000	0.000
Log distance	-2.493	0.058	-43.331	0.000	0.083
Neighbors	1.530	0.093	16.493	0.000	4.620
Log local potential	1.310	0.067	19.524	0.000	3.707

Offering itself as a generalization of previous forces automatically is an average distance towards neighbors instead of all remaining locations in the network. This having positive additional effect could help in dispersion of linkages out of dense clusters, giving better chance of missing connections between clusters to appear among 1,502 most probable links. The number of neighbors is another potentially helpful predictor, which can prefer connecting points with smaller number of neighbouring locations. These seem to be again those out of dense clusters of points. Five explanatory variables version (Model I) basically repeats regularities known from the previous, but there are some differences among the effects. Positive significant effect of log potential is lost in expense of positive significant local potential version. Also, size of the location neighborhood proves to be an insignificant predictor. This model version improves prediction to 778 correctly simulated links, giving a 51.80% success. The Figure 2 enclosed suggest that more sophisticated driving forces must be employed in further versions to redirect remaining links from dense clusters in space linking them into one connected network.

Fig. 2. Probability surface illustrating 5,000 most probable links in the network (a) and real network (c) to simulated (b) comparison. Distribution of average probabilities of connections among network communities including prominent internal connections on matrix diagonal.

Full model containing five variables on the right hand side of the equation is in the first section of the table, followed by two generalized versions (Model II) removing the least significant log potential, and then (Model III) removing log count of neighbors in third section. Second one has improved performance correctly simulating 780 links as our most successful, third version performs slightly worse giving 774 correct predictions. The Figure 2 illustrates how probability surface translates into a disconnected web of short linkages missing most of longer network elements. The contrast between Figure 2b and Figure 2d reflects clearly the space left for improvements in a more mature future modelling attempt. The Figure 2d documents a conceptual connection of the results from our simulation exercise with the descriptive network community detection approached in final part of this analytical section.

The remaining question is, whether distribution of probability values connects somehow to modularity-based network communities used so far for descriptive purposes. The answer can be found in average probability values constructed for all possible inter-community connections appearing in the matrix. Given all theoretical connections, the diagonal elements (7.57%) appear 53-times bigger than the non-diagonal elements (0.14%). Focusing now Delaunay neighbours only, the diagonal elements (40.39%) still appear 8-times bigger than the non-diagonal elements (5.29%). This means that our logistic model correctly assigns much more probability to the linkages falling inside different network communities found by Louvain method than those between them. Whether this contrast is a significant one requires a specific regression test, which would have to take into account spatial network autocorrelation present in our problem.

CONCLUSIONS

This paper describes and explores a public transport network in a typical, moderate size European city. The sparse network of service linkages among 584 stations scattered across the urbanized territory offers an exciting research problem. Public transport network is only a sample of spatial interaction networks that city offers as a documentation of its real-world functioning, potentially with a promise of efficiency focusing performance improvements in the future. Former part of the paper identifies hierarchy and barriers existing in the network topology. We have discovered a small number of prominent vertices, one of which serves as a hub. Their distribution corresponds perfectly with natural settings and barriers: mountain slopes, country borders, and river dividing the area in two. Unfortunately, those few nodes, including the most important "Račianske mýto", reflect more a path-dependent topology than correspondence with requirements of current Bratislava.

The modeling exercise in the later part tests the importance of several driving forces hidden behind specific topology of the network. Three predictors prove to be significantly changing probability of connection between two stations, namely geographic distance between them, neighbourhood status of a candidate connection, and an average distance to the station neighbours. Presented findings in their best alternative reach 51.93% success of simulation, pointing at about one half of network architecture inherited from basic distance based spatial relationships, but in the same time one half remaining unknown. Heterogeneous clustering of stations in space seems to be a concept worth further attention, especially knowing about a clear connection with the outcome of network communities detection.

ACKNOWLEDGEMENTS

This research was supported by VEGA grant 1/0562/12 with the name "New demographical analysis and prognosis of the population of Slovakia and its regions by using of progressive geographical applications".

REFERENCES

Barthélemy, M. (2011) Spatial networks. Physics Reports. 499, 1-101.

Blondel, D. V., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. (2008) Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment (10), P1000.

Brandes, U. (2001) A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology 25(2), 163-177.

Clauset, A., Newman, M., E., J., Moore, C. (2004) Finding community structure in very large networks. Physical Review 70(6), 1-6

Gephi.org (2012) The Open Graph Viz Platform, Available: https://gephi.org/, Accessed: 25.10.2012

Guimera, R., Sales, M., Amaral, L., A., N. (2004) Modularity from fluctuations in random graphs and complex networks. Physical Review 70(2), 1-4

Hu, Z., Lo, C. P. (2007) Modeling urban growth in Atlanta using logistic regression. Computers, Environment and Urban Systems. 31(6), 667-688.

Koenig, R., Bauriedel, C. (2009) Generating settlement structures: a method for urban planning and analysis supported by cellular automata. Environment and Planning B. 36(4), 602-624.

Lambiotte, R. Delvenne, J.-C., Barahona, M. (2009) Laplacian Dynamics and Multiscale Modular Structure in Networks.

Latora, V., Marchiori, M. (2000) Harmony in the Small World, Physica A 285(3-4), 539-546.

Latora, V., Marchiori, M. (2001) Efficient Behavior of Small-World Networks. Physical Review Letters 87(19), 198701-198704.

Latora, V., Marchiori, M. (2002) Is the Boston subway a small-world network? Physica A 314(1-4), 109–113.

Lu, H. et al. (2007) Complexity of Public Transport Networks. Tsinghua Science and Technology 12(2), 204-213.

Musso, A., Vuchic, V. R. (1988) Characteristics of metro networks and methodology for their evaluation. Transportation Research Record 1162, 22-33.

OpenStreetMap.org (2012) Open Street Map. Available: http://www.openstreetmap.org/, Accessed: 12.09.2012

Seaton, K. A., Hackett, L. M. (2004) Stations, trains and small-world networks. Physica A 339(3-4), 635-644.

Sienkiewicz, J., Holyst, J. A. (2005) Statistical analysis of 22 public transport networks in Poland. Physical Review E 72(4), 0461271-04612711.

Transport Company Bratislava (2012) Timeshedules of Transport Company Bratislava. Available: http://imhd.zoznam.sk/ba/cestovne-poriadky.html, Accessed: 10.09.2012.

Von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V. (2007) Network Harness: Metropolis Public Transport. Physica A 380, 585-891.

Von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V. (2008) Public transport networks: empirical analysis and modeling. The European Physical Journal B 68(2), 261-275.

Vuchic, V. R., Musso, A. (1991) Theory and practice of metro network design. Public Transport International 40(3), 298-325.

Xu, X.et al. (2007) Scaling and correlations in three bus-transport networks of China. Physica A 374(1), 441–448.