

Complex network analysis of public transportation networks: a comprehensive study

Andor Háznagy*, István Fi*, András London[†] and Tamás Németh[‡]

*Department of Highway and Railway Engineering, Budapest University of Technology and Economics

Email: fi@uvt.bme.hu, haznagy.andor@epito.bme.hu

[†] Department of Computational Optimization, University of Szeged

Email: london@inf.u-szeged.hu

[‡] Department of Computer Algorithms and Artificial Intelligence, University of Szeged

Email: tnemeth@inf.u-szeged.hu

Abstract—In this study, using the network approach, we analyzed the urban public transportation systems of 5 Hungarian cities. We performed a comprehensive network analysis of the systems with the main goal of identifying significant similarities and differences of the transportation networks of these cities. Although previous studies often investigated unweighted networks, one novelty of our study is to consider directed and weighted links, where the weights represent the capacities of the vehicles (bus, tram, trolleybus) in the morning peak hours. In particular, we calculated descriptors of global network characteristic and various centrality measures of the network nodes in both the weighted case and unweighted case. By comparing the results obtained for the different cities, we get a highly detailed picture of the differences in the organization of the public transport, which may due to historical and geographical factors. Also, by comparing the results obtained from the weighted and unweighted approaches, we can identify which are the most sensitive routes and stations of the network pointing out some organizational inconsistencies of the transportation system.

Keywords—Public transportation networks, Network analysis, Empirical analysis, Small-world networks

I. INTRODUCTION

Network analysis and graph-based data mining have become very popular in the last decade as these tools have proved to be readily applicable in a wide range of fields including biology, economy and the social sciences [1]. The network approach is not only useful for simplifying and visualizing the huge amount of data, but it also effective in picking out the most important elements and finding their most important interactions. Furthermore, several techniques have been developed to explore the deeper topological features of a network, such as community structure [2], core-periphery structure [3] or small-world [4] and scale-free properties [5]. These properties are usually the most common characteristic features of real-world complex networks.

A great deal of attention has been paid to investigating transportation systems for many decades because of its practical importance. In the past decade, partly due to the development of small-world networks and the appearance of modern graph theory, several studies have treated public transportation

systems as complex networks, and several statistical properties have been discovered, such as the small-world property and scale-free distribution of various graph measures [6], [7], [8], [9]. In most of these studies, the public transportation network (PTN) model represents nodes as stations and stops of a public transportation system, and edges that connect consecutive stations along a route.

Here, by adopting the network approach, we will study the urban public transportation systems of 5 Hungarian cities. The choice of the cities was based on the following criteria: (i) we are especially interested in cities with a population between 100,000 and 250,000; (ii) the characteristics (like land-use and economic role) and the organization of the public transportation of these cities are similar; but (iii) the geographical conditions (relief, hydrography, size of the area) are different. Although previous studies often used unweighted networks, one novelty here is to consider directed and weighted edges, where the weight of a link refers to the *morning peak hour capacity* of this link obtained by using the capacities of the vehicles (bus, tram, trolleybus) and schedules of the lines that go though that link. Our analysis was based on the capacities of the PT in order to get a detailed picture about the existing PT networks. We note, that the modal split and the real number of passengers in the PT vehicles are the main descriptors of public transport systems from the optimization point of view. However, in the study, we present an alternative approach which requires a smaller number of data, but gives a global picture about the PTNs. We performed a comprehensive network analysis of the systems with the main goal of identifying similarities of and differences between any two transportation networks of these cities. In particular, we calculated various centrality measures of the network nodes in both the weighted case and unweighted case (such as weighted in- and out-degree, betweenness, closeness, local average connectivity, PageRank centrality) and global characteristics (like diameter, average path length, degree distribution and community structure). By comparing the results obtained for the different cities, we get a detailed picture of the differences in the organization of the public transport, which may have arisen for historical, geographical and economic reasons. However, by comparing the results obtained from

TABLE I. CITIES ANALYZED IN THIS STUDY. LINKS-SIMPLE REFERS TO THE NUMBER OF LINKS IN THE SIMPLIFIED GRAPHS, WHERE A LINK IS DEFINED BETWEEN TWO NODES THESE NODES ARE THE CONSECUTIVE STATIONS OF AT LEAST ONE LINE. LINKS-MULTIPLE REFERS TO THE NUMBER OF LINKS IN THE MODELS, EACH LINE BETWEEN TWO STATIONS IS REPRESENTED BY A LINK. THE CODES OF THE VEHICLE TYPES ARE AS FOLLOWS: B: BUS, E: ELECTRIC TROLLEYBUS, T: TRAM

| City | Area (km ²) | Pop. ($\times 1000$) | Density (inhab./km ²) | Nodes | Links-simple | Links-multiple | Lines | Diameter | Avg. path length | Vehicle types |
|----------|-------------------------|------------------------|-----------------------------------|-------|--------------|----------------|-------|----------|------------------|---------------|
| Debrecen | 461 | 204 | 442.5 | 306 | 711 | 1772 | 53 | 41 | 11.7 | BET |
| Győr | 174 | 129 | 741.4 | 230 | 529 | 1391 | 43 | 30 | 10.8 | B |
| Miskolc | 236 | 161 | 682.2 | 257 | 535 | 977 | 35 | 45 | 14.5 | BT |
| Pécs | 163 | 147 | 901.8 | 256 | 569 | 1960 | 55 | 36 | 13 | B |
| Szeged | 281 | 162 | 576.5 | 242 | 558 | 1192 | 40 | 35 | 11.8 | BET |

the weighted and unweighted approaches, we can highlight inconsistencies, organizational problems and identify which are the most sensitive routes and stations of the network.

The paper is organized as follows. In Section II we briefly overview the related literature. In Section III we give a brief introduction to the basics of network (graph) theory including the main definitions and the notions that we applied in our study. Then, our results are presented for a comprehensive study of the transportation networks. In Section IV we provide a summary and draw some pertinent conclusions.

II. RELATED WORK

In one of the earliest network-based studies of urban transportation systems, Latora and Marichiori investigated the Boston subway network and highlighted the small-world properties and local and global efficiency of the network [6]. Sen et al. studied the Indian railway network and also discovered small-world properties, like small average distances, and exponential degree distribution [10]. Seaton et al. compared the urban train line networks of Vienna and Boston by extending the usage of random bipartite graph models from social networks to technological ones [11]. Angeloudis and Fisk analyzed the world's 20 largest subway systems and found that the high connectivity and low maximum vertex degree are characteristic features and highlighted the robustness of the networks against random failures [12]. Chen et al. investigated the urban transportation networks of four major cities in China and showed that the distribution of the number of routes that a stop joins follows a power-law with exponential decay, while the distribution of the number of stops in a bus route follows asymmetric, unimodal functions [13]. Xu et al. defined spaces P and L as the network of stations connected by links and the network of stops as nodes with links that only exist between two nodes if they are consecutive stops on a route [14]. They analyzed three bus transportation networks in China and also examined the small-world behavior of them in both spaces P and L . Sienkiewicz and Holyst collected and analyzed the data of the PTNs of 22 cities in Poland and found that the degree distributions in space P follow a power-law, while in space L they are exponential [15]. Ferber et al. studied the PTNs of 14 major cities of the world and found that the degree distributions follow the power-law with various exponents [16]. They also proposed an evolutionary model of growth of PTNs. Later, they performed a comprehensive complex network analysis of these networks [7]. Derrible and Kennedy investigated 33 metro systems in the world and found that most of them are indeed scale-free (the degree distributions follow a power-law), but the presence of transfer hubs (stations hosting more

than three lines) results in relatively large scaling factors [17]. By analyzing 19 subway systems worldwide they found a close relationship between the number of passengers and network design by using new graph theory concepts. Zhang et al. measured the topological characteristics and functional properties of Shanghai's subway network and pointed out that the network is robust against random attacks, but weak for targeted attacks and the disconnection of the nodes with the highest betweenness values can cause serious damage in a network [18]. Roth et al. studied the temporal evolution of the structure of the world's 14 largest subway networks and showed that these networks converge to a shape that share similar generic features despite their geographical and economic differences [8].

III. METHODS AND RESULTS

A. Collecting data

We selected 5 Hungarian cities (Debrecen, Győr, Miskolc, Pécs, Szeged) based on the following criteria. The size and population of the cities are similar (the populations are between 100,000 and 250,000, the sizes are between 162 and 462 km², so these are medium-size), but their urban morphology is different. In Miskolc and Pécs the land undulates, while in Győr and Szeged a river that crosses the city is the main factor that determines the shape of the city. In Debrecen, there are no restricting factors on the morphology. We should also add that the railway tracks can have a similar role to that of the rivers in the separation. This effect appeared in all the cities investigated. The above-mentioned characteristics have had a high impact on the development of the cities and also on the organization of the public transportation systems. In order to perform a comprehensive network analysis of the public transportation systems of these cities, the first step was to generate the transportation networks (i.e. the representing graphs). This was done by modelling stations/stops as nodes and lines that connect them as directed links. If a line runs between two stops in both directions, as is usually the case, we can decompose the link that represents this line into two directed links due to the orientation. Furthermore, we can also assign weights for each node and each edge by using the capacity of the vehicles. This can be performed as follows:

- 1) Assign the lines to the stations where they stop by using the transport schedules.
- 2) Classify the stations that belong together.
- 3) Determine the morning peak hour capacity of each vehicle using the types of the vehicles (the data

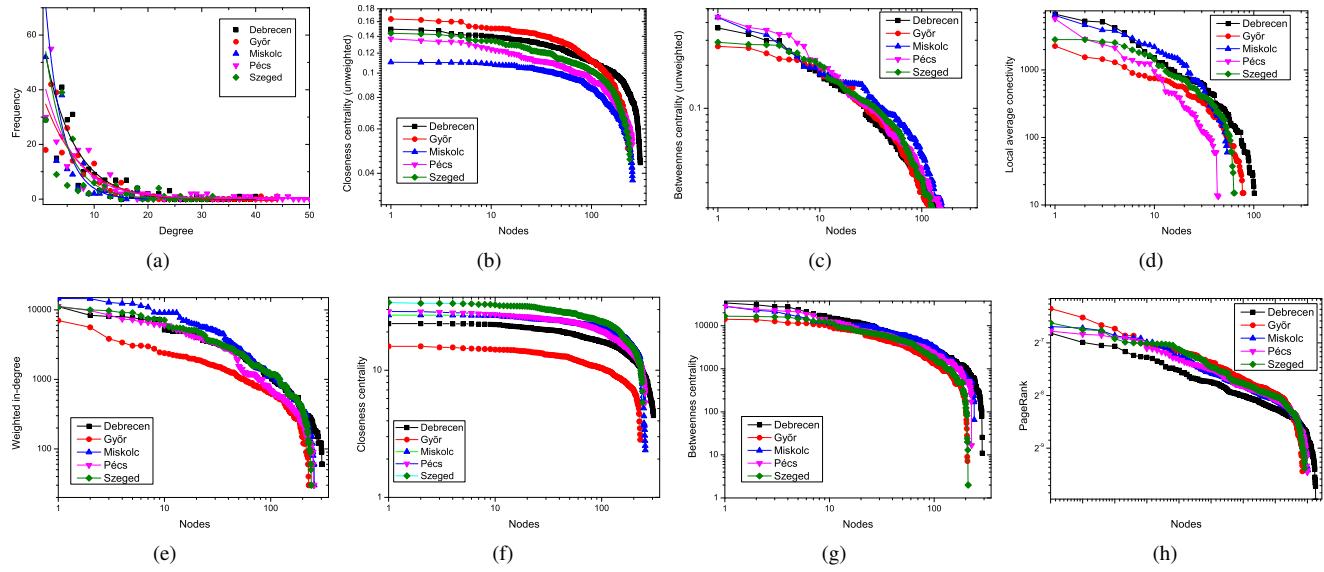


Fig. 1. **a-c:** Cumulative node degree distribution, Closeness centrality and Betweenness centrality of the unweighted networks, respectively. **d-h:** Local average connectivity, Weighted degree distribution, Closeness centrality, Betweenness centrality and PageRank centrality of the weighted networks, respectively.

provided by the public transport companies of the cities).

Merging the stations into a single one was necessary for the following reasons. It frequently occurs that stops belonging to the same node have different names. In a special case, it happens that there are four different names of the same stop in a 4-way crossroads. On the one hand these stops can be viewed as just one stop, while on the other hand this classification allows us to unambiguously cover the road network of the city with the PTN. In a big public transportation interchange or terminal where a high number of lines intersect, usually the lines have different stops. These stops were also merged. In the case where a line makes two stops in two stations that were merged, we consider only one stop of the line for this node. In the case where the route of the line is a one-way instead of a two-way between two consecutive stations, the stops were not merged.

A calculation of the maximal capacities of the different lines was performed based on the evaluation of the vehicle capacities¹ in the morning peak hours (6-8am). For every single line, we collected the follow-up interval of it and multiplied it by the capacities of the vehicles belonging to this line between 6-6:59am and 7-7:59am. By averaging the two

¹The following types of vehicles are considered: mini bus: 30 persons; normal bus/trolleybus: 60 persons; articulated bus/trolleybus: 100 persons. In the case of trams, the situation is more complicated. The types of trams are different for every city; moreover the passenger capacities are calculated in different ways by the different manufacturers. To calculate the tram capacities, we used the formula $3 \times (\ell w - sr^2)/10^6$, where ℓ is the length of the vehicle, w is the width of the vehicle (both in millimeters), s is the number of seats and r^2 is the area taken up by a single seat with $r = 500$ millimeter. Here, we got the following results: Debrecen: Ganz KCSV6: 142 (persons), CAF Urbos 3: 227; Miskolc: Skoda 26T: 214, Tatra KT8D: 187; Szeged: Tatra T6A2: 81, Tatra KT4: 101, Pesa 120Nb: 187.

values, we obtained the *average morning peak hour capacity* (AMPHC) of the line. For each node and link, we assigned the sum of AMPHCs of the lines that stop at that node or pass through that link between two consecutive stations at least once. By considering the morning peak hours, it can be seen that number of passengers that go from the outer districts to the inner city area is significantly higher than the number of passengers that go in the opposite direction. Based on this observation, we are able to identify traffic source and traffic sink districts. For those who are interested, all the data is available on <http://www.epito.bme.hu/uvt/dolgozok/feltoltesek/haznagy/ptncomplexanalysis.zip>

B. Complex network analysis of the PTNs

Formally, an *undirected (directed) network* or *graph* $G = (N, L)$ consists of two sets N and L , where $N \neq \emptyset$, while L is a set of unordered (ordered) pairs of elements of N . The elements of $N = \{1, 2, \dots, n\}$ are called *nodes* and the elements of L are called *links*. A network is usually represented by its *adjacency matrix* $A = [a_{ij}]_{i,j}$, which is an $n \times n$ matrix with entries $a_{ij} = 1$ if there is an edge (directed edge) between i and j and $a_{ij} = 0$ otherwise. For undirected graphs if the (i, j) edge exists, then $a_{ij} = a_{ji} = 1$, i.e. A is symmetric. If a function $w : L \rightarrow \mathbb{R}$ is given that assigns a real number w_{ij} to each edge (i, j) , then the graph is *weighted*. The *degree* d_i of node i is the number of links that are connected to i . If the network is directed, we can define the *in-degree* d_i^+ and *out-degree* d_i^- of a node i , these are the number of incoming links to i and the number of outgoing links from i , respectively. The weighted degree of a node can be calculated similarly using $w(i) = \sum_i w_{ij}$ instead of d_i ($i = 1, \dots, n$). A *walk* between two nodes i and j is a sequence of edges $(i, k_1), (k_1, k_2), \dots, (k_m, j)$. The length of the walk is the number of edges on it. If all the nodes along the walk are distinct, then the walk is a *path*.

Global network characteristics

1) *Diameter*: Let ℓ_{ij} be the shortest path between nodes i and j . The diameter of the network is defined as the maximum of the shortest paths among all pairs of nodes, i.e.

$$D(G) = \max_i \max_{j \neq i} \ell_{ij}, \quad (1)$$

In practice, the diameter presents the longest route (number of stations along the longest route) in the network if a passenger uses the optimal routes, which means that she uses the shortest route between any two stations. In Table I we show the diameters for the PTNs. It is interesting to observe that the diameter does not correlate with the area of the city.

2) *Average path length*: The *average path length* is defined as

$$\langle \ell \rangle = \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} \ell_{ij}, \quad (2)$$

which exists only if there are no unconnected nodes in the network. We will restrict Eq. (2) to this case. The average path length corresponds to how many stations there are between two stations on the shortest route on average, if we choose these stations randomly. We can see in Table I that the the PTNs reveal a small-world feature from the average path lengths point of view, since $\langle \ell \rangle \sim \log N$, i.e. the average distance between the nodes is proportional to the logarithm of the number of nodes.

The number of shortest paths from a node i is defined as

$$\ell_i = \sum_{j \neq i} \ell_{ij}. \quad (3)$$

Fig. 2(a) shows that the distribution of the shortest paths is a normal distribution with mean that varies between 10.8 and 14.5 (Table I) and variance between 5.2 and 7.7.

3) *Eccentricity distribution*: The *eccentricity* e of a node i is the longest distance between i and any other node in the network; that is

$$e(i) = \max_{j \neq i} \ell_{ij}. \quad (4)$$

Here, it tells us how far a stop/station is from the most distant stop/station in the PTN. In Fig. 2(b) we plotted the eccentricity distribution of the PTNs. The shape of the function is quite different in the case of Debrecen, due to its extensive area and Miskolc, where many peripheral areas increase the distances between certain stops/stations.

4) *Degree distribution*: The list of the node degrees is the degree sequence of the network. The *degree distribution* $P(d)$ is defined as the fraction of nodes having degree d ; or, equivalently, it is the probability that a uniform randomly chosen node has degree k . In the case of directed networks, we can consider the in-degree and out-degree distributions.

Fig. 1(a) shows the degree distributions in the unweighted case, where multiple links are allowed, which has an exponential decay $P(d) \sim \exp(-d/\hat{d})$, where \hat{d} is of the order of the average node degree. In contrast, the weighted degree

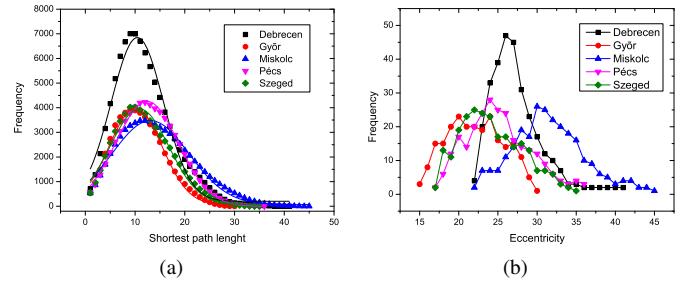


Fig. 2. **a:** Shortest path length distribution; solid lines shows a fit to the function. **b:** Eccentricity distribution.

distribution (Fig. 1(e)) of the (weighted) networks has a power-law decay $P(d) \sim d^{-\gamma}$, where γ varies between 1.05 to 1.2.

5) *Community structure*: Finding communities in a network means finding a way to partition the nodes into disjoint sets such that nodes in the same set are more densely connected to each other than to the rest of the network. Usually, a community in a network refers to the similarity and common features of the nodes that it contains.

In order to find communities, we use the *modularity optimization method* [19], which is based on the idea that a random graph is not expected to have a community structure. The key step is to compare the structure of the network with the structure of an appropriate null model graph using the modularity function

$$Q = \frac{1}{2|L|} \sum_{i,j} (a_{ij} - p_{ij}) \delta(C_i, C_j) \quad (5)$$

which have to be optimized. In Eq. 5 $\delta(C_i, C_j) = 1$ if and only if $C_i = C_j$, i.e the nodes i and j belong to the same community, and 0 otherwise. For an unweighted graph, $p_{ij} = \frac{d_i d_j}{2|L|}$ is the probability that nodes i and j are connected in a random graph with the same degree sequence (null model) as the original one. An extension of the model to weighted and/or directed graphs can be readily performed.

The communities of the PTNs are shown in figures 3(b), 3(d), 3(f), 3(h) and 3(j). The results show the following common features of the networks. On the one hand, for each city, the center of it contains one or two communities and most of the peripheral lines have different community classes. On the other, we observed that if the city lies in a valley (Miskolc) or is bounded by mountains on one side (Pécs) and hence the arrangement of the city is asymmetric, then it will have some special characteristics. The central core of the networks have been extended (figs. 3(e) and 3(g)) and this part of the transportation network can be partitioned into three or four communities.

Node centrality measures

In complex networks, centrality generally refers to the kind of measures that represent the most important and “central” nodes within the network from some given perspective.

6) *Degree centrality*: It is simply the degree d_i of node i (in the case of directed networks, the in- and out-degrees are used) and it tells us how big the neighborhood of i is. The weighted in-degree centralities of the 5 PTNs can be seen in Fig. 1(e). The distributions have a power-law decay, as we noticed earlier.

7) *Local average connectivity*: Let N_i be the set of neighbors of u and $G[N_i]$ be the subnetwork *induced* by the nodes in N_i . The degree of a node j in the subnetwork $G[N_i]$ is denoted by $d^{G[N_i]}(j)$. Next, the *local average connectivity* [20] of node i is defined as

$$LAC(i) = \frac{1}{d_i} \sum_{j \in N_i} d^{G[N_i]}(j) \quad (6)$$

and it describes how close its neighbors are. In a public transportation system it basically means that if a stop/station cannot be used for some reason, the neighboring stops become disconnected from each other. Nodes with high LAC values are the locally central nodes. Fig. 1(d) shows the distribution of LAC for the 5 PTNs. We observed that the distributions fit a power-law decay with degree exponent between 1.2 to 1.4.

8) *Closeness centrality*: The *closeness centrality* [21] of a node i is defined as

$$C(i) = \frac{1}{\sum_{i \neq j} \ell_{ij}}. \quad (7)$$

Here, the greater the value, the smaller the length of the shortest paths to all other nodes. The closeness centralities for the unweighted and weighted case can be seen in figures 1(b) and 1(f), respectively. The distributions display similar shapes for each city, but interesting observations can be obtained by comparing the the unweighted and weighted closeness values for one city. The centrality values in the unweighted networks tells us how central and important the nodes are according the to the structure of the network. Considering the schedules and capacities of lines in the PTN to assign weights to the links the nodes get closer or farther to each other from the transportation point of view, due to III-B11 (see later). The unweighted and weighted C values for each city can be seen (plotted in the same scale) in figures 4(a)–4(e). In the case where the centrality value in the unweighted network of a node is bigger than the value in the weighted case tells us, that, although the node has central position in the network, the stop that represented by this node may not be well exploited in the transportation sense. On the other hand, if the relation between the unweighted and weighted case is the opposite, the stop is overloaded according to the network traffic.

9) *Betweenness centrality*: Let σ_{ij} be the number of shortest paths between nodes k and j and $\sigma_{kj}(i)$ be the number of shortest path between them that pass through node i . The *betweenness centrality* [22] of node i is defined as

$$BC(i) = \sum_{k \neq i \neq j} \frac{\sigma_{kj}(i)}{\sigma_{kj}}. \quad (8)$$

In networks, the greater the number of paths that pass through a certain node (or edge), the greater the importance of this node

(or edge) and more central it is. The betweenness centralities for the unweighted and weighted case can be seen in figures 1(b) and 1(f) and display similar shapes as it was in the case of closeness. The unweighted and weighted BC values for each city can be seen (plotted in the same scale) in figures 4(f)–4(j). Similarly to closeness, if the BC value of a node in the weighted network is greater than its value in the unweighted case the represented stop may be overloaded in the PTN. The opposite relation refers to stop with spare capacity.

10) *PageRank centrality*: The key idea behind the definition of *PageRank centrality* [23] is that nodes with the same in-degree (degree) may not have the same importance in the network. It treats an incoming link from a strongly linked node as more important than from a node with just a few connections. The PageRank scores (PR) of each node are calculated iteratively using the formula

$$PR(i) = (1 - \lambda) \frac{1}{n} + \lambda \sum_{j:j \rightarrow i} \frac{PR(j)}{d_j^-}, \quad (9)$$

where $\lambda \in [0, 1]$ is a free parameter with a value usually lying between 0.6 and 0.8. It was demonstrated that the PR scores will converge. Here, we used the value of 0.8 in our calculations. In [24] PageRank was used to identify the key nodes in a transportation system and also for traffic simulations [25] to find important nodes having a high impact on transportation efficiency. It is interesting to observe that the PageRank distributions are similar for all the 5 weighted PTNs (Fig. 1(h)) which is probably due to the organizational rules of the schedules being similar.

11) *Extending the centrality measures for weighted networks*: An extension of the definition of the centrality measures to weighted networks can be performed using the w_{ij} edge weights e.g. in the following way. The weighted degree of a node i is simply defined as $w_i = \sum_j w_{ij}$. In the case of PageRank, w_{ij}/w_i is used in Eq. 9 instead of d_j^- . The weighted closeness and betweenness can be defined by using $c_{ij} = 1/w_{ij}$ and $dist(i, j) = \sum_{(u,v) \in P} c_{uv}$, where P is a path between i and j and the weighted shortest path ℓ_{ij}^w is defined as the minimum of $dist(i, j)$.

IV. CONCLUSIONS

Today, complex network science is a rapidly growing part of physics, mathematics and computer science. The study of Public Transportation Systems treated as complex networks has received particular attention in the past decades. In our study, we considered directed and networks in both the weighted and unweighted case and we examined the PTN systems of 5 Hungarian cities. The cities have similar geographical properties (relief, hydrography, size of the area), density, population (between 100,000 and 250,000) and urban structure. In the weighted networks, the weight of a link refers to the morning peak hour capacity of this link obtained by using the capacities of the vehicles (bus, tram, trolleybus) and schedules of the lines that go though that link. Although the vehicle capacities are usually higher than the number of

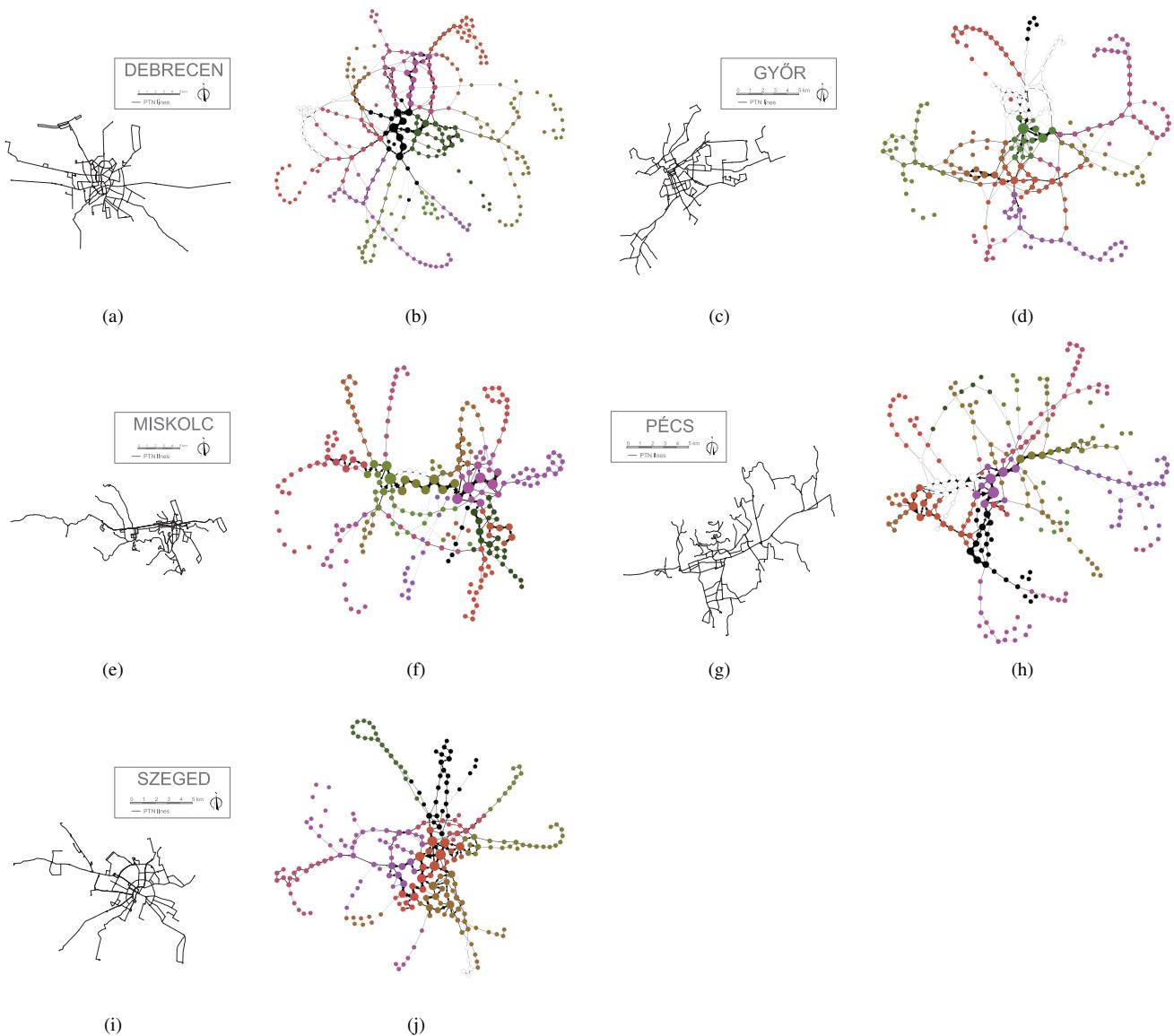


Fig. 3. **a,c,e,g,i:** Simple maps of the lines of the transportation system of the cities that were investigated. **b,d,f,h,j:** Partition of the PTNs into communities using the modularity optimization method. Nodes having the same color belong to the same cluster; the bigger a node, the higher its in-degree is; and the thicker an edge, the greater its capacity is

passengers, they can be used as an acceptable approximation to calculate the maximal loads of the links between consecutive stations. We compared weighted and unweighted graphs that model the PTNs and we discovered some interesting global characteristics, such as the degree distribution, distribution of various centrality measures and the shortest path lengths, local average connectivity and community structure with core-periphery features. Independent of the morphology of the cities the PTNs have a few high-degree nodes where many lines cross (these nodes are almost always in the network core and usually represent bus terminals, train stations and local transportation centers), but most of the nodes have a low degree (in both the

unweighted and weighted case) resulting in a fat-tailed degree distribution. Also, we identified some critical points in the transportation networks based on comparisons of the weighted and unweighted models. We considered the centrality values (Betweenness and Closeness) for the weighted and unweighted networks and checked the differences among them. The great value of a centrality measure of a node refers to the high importance, central position of this node and also refer to the transfer point role of it in the PTN. These points are usually referred as intermodal transport hubs, where passengers are exchanged between vehicles or between transport modes. Transport hubs include P+R points, important local and long-

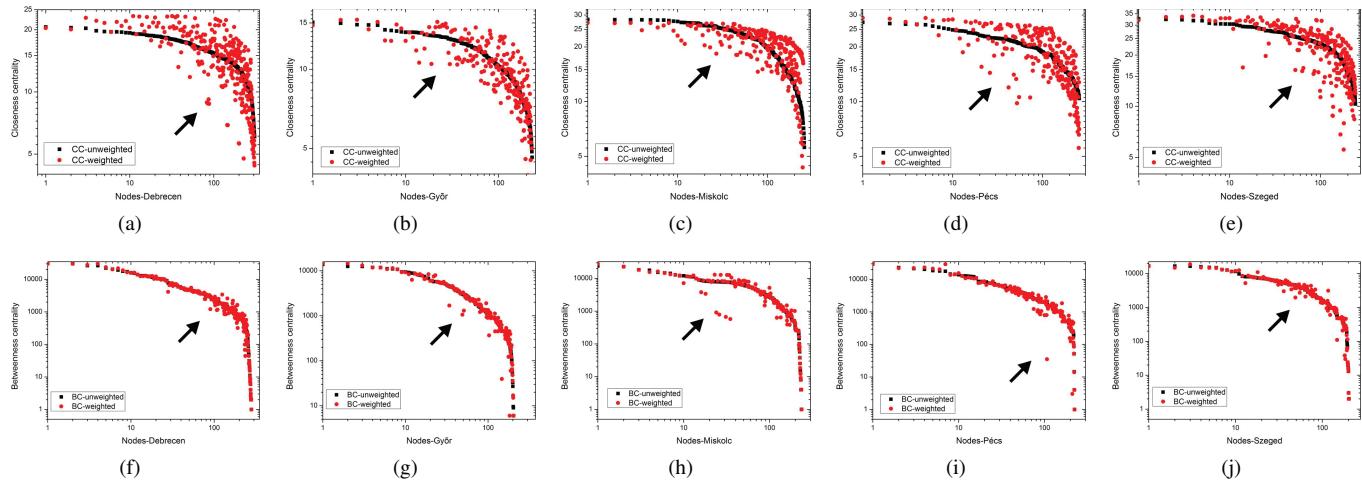


Fig. 4. The unweighted and weighted betweennes and closeness centrality measures for each city. The values are in the decreasing order of the centrality values for the unweighted networks.

distant transport terminals and train stations among others. The distribution of the centralities is different from the distribution of the capacities of the nodes. By comparing the centralities in the weighted and unweighted cases, we are able to identify overloaded stops and unexploited stops with spare capacity.

Finally, we conclude that the complex network methodology and graph theoretic measurements that we considered can reveal the main structural features and principles behind the organization of the public transportation systems. However, each city has a different modal split, i.e. the distribution of the usage of different types of vehicles (e.g. public transport, car, bicycle and P+R), and it could result different use patterns of the Public Transportation Network. Although the modal split effect did not appear in our investigation, we should mention that the cities in question have similar modal split. In the future, we would like to analyze bigger cities (with populations around 1-2 million) and also cities in different countries with similar layouts (population between 150.000 and 300.000, similar urban structure and land use) with network theoretic tools, and carry out dynamical studies based on more detailed data (where in addition to the schedules and capacities, the geographical distances are also given between the consecutive stations). We would also like to deal with the question of transfers between routes. The results of this study fit in well with the earlier studies in the field of classical PTN modeling. We think that these kind of methods applied here could assist in the planning of urban public transportation systems and could be integrated to the classical PT organization methodology.

REFERENCES

- [1] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
- [2] X. Li, P. Zhang, Z. Di, and Y. Fan, "Community structure of complex networks," *Complex systems and complexity science*, vol. 5, no. 3, pp. 19–42, 2008.
- [3] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi, "Structure and dynamics of core/periphery networks," *Journal of Complex Networks*, vol. 1, no. 2, pp. 93–123, 2013.
- [4] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [5] A.-L. Barabási *et al.*, "Scale-free networks: a decade and beyond," *science*, vol. 325, no. 5939, p. 412, 2009.
- [6] V. Latora and M. Marchiori, "Is the boston subway a small-world network?" *Physica A: Statistical Mechanics and its Applications*, vol. 314, no. 1, pp. 109–113, 2002.
- [7] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Public transport networks: empirical analysis and modeling," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 68, no. 2, pp. 261–275, 2009.
- [8] C. Roth, S. M. Kang, M. Batty, and M. Barthelemy, "A long-time limit for world subway networks," *Journal of The Royal Society Interface*, p. rsif20120259, 2012.
- [9] S. Derrible, "Network centrality of metro systems," *PloS one*, vol. 7, no. 7, p. e40575, 2012.
- [10] P. Sen, S. Dasgupta, A. Chatterjee, P. Sreeram, G. Mukherjee, and S. Manna, "Small-world properties of the indian railway network," *Physical Review E*, vol. 67, no. 3, p. 036106, 2003.
- [11] K. A. Seaton and L. M. Hackett, "Stations, trains and small-world networks," *Physica A: Statistical Mechanics and its Applications*, vol. 339, no. 3, pp. 635–644, 2004.
- [12] P. Angeloudis and D. Fisk, "Large subway systems as complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 553–558, 2006.
- [13] Y.-Z. Chen, N. Li, and D.-R. He, "A study on some urban bus transport networks," *Physica A: Statistical Mechanics and its Applications*, vol. 376, pp. 747–754, 2007.
- [14] X. Xu, J. Hu, F. Liu, and L. Liu, "Scaling and correlations in three bus-transport networks of china," *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 441–448, 2007.
- [15] J. Sienkiewicz and J. A. Holyst, "Statistical analysis of 22 public transport networks in poland," *Physical Review E*, vol. 72, no. 4, p. 046127, 2005.
- [16] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Network harness: Metropolis public transport," *Physica A: Statistical Mechanics and its Applications*, vol. 380, pp. 585–591, 2007.
- [17] S. Derrible and C. Kennedy, "The complexity and robustness of metro

- networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 17, pp. 3678–3691, 2010.
- [18] J. Zhang, X. Xu, L. Hong, S. Wang, and Q. Fei, "Networked analysis of the shanghai subway network, in china," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 23, pp. 4562–4570, 2011.
 - [19] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.
 - [20] M. Li, J. Wang, X. Chen, H. Wang, and Y. Pan, "A local average connectivity-based method for identifying essential proteins from the network level," *Computational biology and chemistry*, vol. 35, no. 3, pp. 143–150, 2011.
 - [21] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.
 - [22] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.
 - [23] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.
 - [24] G. Lianxiong, W. Jianping, and R. Liu, "Key nodes mining in transport networks based in pagerank algorithm," in *Control and Decision Conference, 2009. CCDC'09. Chinese*. IEEE, 2009, pp. 4413–4416.
 - [25] N. Mukai, "Pagerank-based traffic simulation using taxi probe data," *Procedia Computer Science*, vol. 22, pp. 1156–1163, 2013.