

## CHARACTERIZATION OF URBAN TRANSPORTATION NETWORKS USING NETWORK MOTIFS

Lilla PELLEGRINI\*, Alexandru IOVANOVICI\*\*, Monica LEBA\*

\*Department of Computer and Electrical Engineering, University of Petrosani, strada Universitii, nr. 20, Romania,  
E-mails: pellegrinililla@gmail.com, MonicaLeba@upet.ro

\*\*Department of Computing and Information Technology, Politehnica University Timisoara, Vasile Parvan Boulevard no. 2, Romania,  
E-mail: iovanalex@cs.upt.ro

### ABSTRACT

*We use tools and techniques specific to the field of complex networks analysis for the identification and extraction of key parameters which define "good" patterns and practices for designing public transportation networks. Using network motifs we analyze a set of 18 cities using public data sets regarding the topology of network and discuss each of the identified motifs using the concepts and tools of urban planning.*

**Keywords:** *Intelligent Transportation Systems, network motifs, complex networks, public transportation*

### 1. INTRODUCTION

The public transportation system realizes the connection of the locality to the internal and international network and at the same time supports and influences the socio-economic evolution of one city. Accessibility, defined as the possibility of access to a desired destination, depends mainly on the extension and quality of the transport infrastructure and on the availability of services, being closely linked to that of connectivity, which is a usual term in network analysis.

Network analysis has become very popular in the last decades as it has proved to be immediately applicable in a large area of science. The Network approach has two benefits. One of them is that it simplifies and visualises the huge amount of data and the second benefit is that it has become very effective in picking out the most important elements and finding their most important interactions. Additionally, numerous techniques have been developed to discover the deeper topological structures of a network, such as community structure, core-periphery structure or small-world and scale-free properties [2]. These properties are usually the most common characteristic features of real-world complex networks.

Urban road network represents a spatial network as a result of the geographical features. The nodes and edges are fixed in space. The analysis and study of the topology structure could be the origin of the traffic state assessment and the optimisation of the traffic organization.

Numerous investigations of the transportation systems have been made in the last 20 years. The development of small-world networks and the appearance of modern graph theory lead to numerous studies about the topic of public transportation systems as complex networks. Many statistical characteristics have been published, for example the small-world property and scale-free distribution of various graph measures [5]. The public transportation is a network where the nodes are represented by the bus stations and the edges connect successive stations. Beside the above-mentioned characteristics, in a network there can be found some small recurrent substructures, so called motifs. The study of the motifs has become a regular tool of complex

network science, in order to accentuate the design basis of the structure of empirical networks.

### 2. NETWORK MOTIFS BACKGROUND

A network motif can be defined as a subgraph, usually with a small number of nodes, that appears significantly more frequently in the network than it does in a group of appropriately-chosen random graphs. The first mention of the motifs was done by Milo et al. who said that they are recurring, significant pattern of interconnection [1]. Milo et al. found them, among many others, in biochemical gene regulation networks and in the network of hyperlinks from the World Wide Web. They demonstrated that different sets of motifs are linked to diverse types of networks. Each motif could influence explicit functions, such as outlining universal network classes. Still, the existence and explanation of motifs in transportation networks has not been in the focus of researchers. The theoretic research on transportation system as complex networks characteristically concentrates on macroscopic structures like network diameter, or microscopic measures such as node centrality [1].

The importance of network motifs could be determined by the consideration that small subgraphs display special forms of links among network nodes and consequently they could have a regulatory or dynamic role. The frequency distribution of motifs in the network could be defined as a motif spectrum. This could be viewed as a pattern of the network structure and it permits to relate diverse networks and to create groups of networks with comparable significance profiles [2]. The analysis of motif spectra is an useful concept to unveil universal design principles underlying the structure of complex networks [6]. It is a tool to explain some properties of complex systems with relatively simple structures that led to an interest in network motifs in an expanding number of studies and across disciplines.

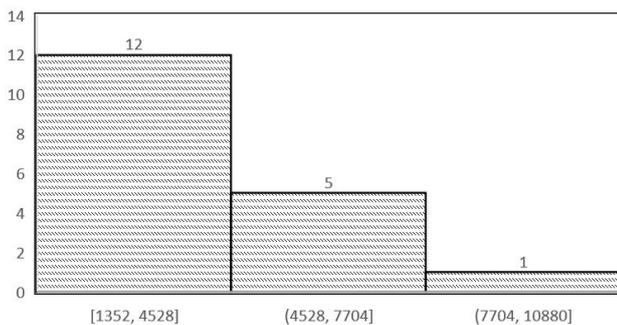
There are many ways to define a motif. Network motifs are the building blocks that profile the dynamic performance of a network or patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks [4]. This definition could have been a little confusing mostly the random

ensemble terms generated the confusion. Additionally, the terms subgraph and motif have been changed many times between them [6]. To be more specific, we choose to describe a motif as a group of topologically equivalent subgraphs of a network

### 3. DATA AND METHODOLOGY

Our investigation is targeting the public transportation network and consequently we have used one of the most extensive data-sets consisting of topology layouts and route traces available to the scientific community. The data-sets published by Kujala et al. in [8] consists of 25 cities form around the world for each of them having provided data regarding the layering of the network consisting of various "travel modes" as defined by the GTFS feed and for each of the layers we have access to a detailed description of it's constituents ("bus" stops and routes represented as poly-lines). This level of details itself allows investigation in the realm of geo-information systems and urban geography but we were more geared towards the network properties which could be extracted form there. The authors used publicly available data fed trough General Transit Feed Specification (GTFS) but done the significant work of curating the data and preparing them for further study

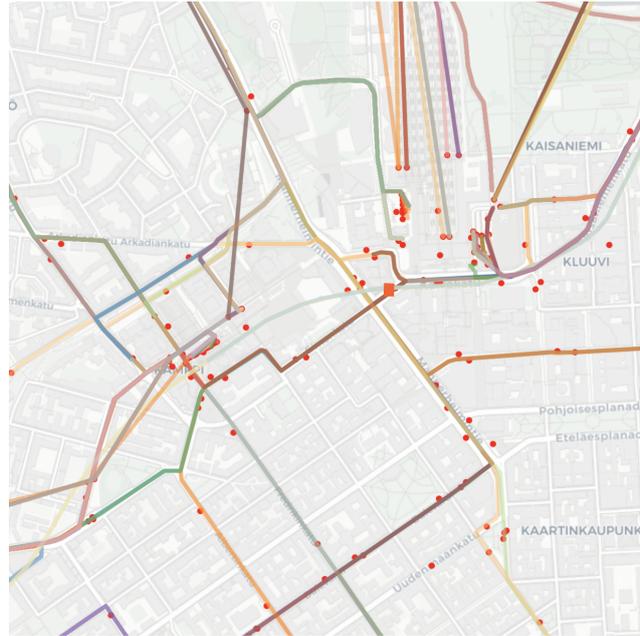
We have took into account in our investigation cities of diverse size (Fig. 1, on different continents and with a variety of local characteristics (separated by a river, on the seaside, having an old city center). Out of the bulk of data provided by the repository we have used the *networkNodes.csv* file which consists of public transportation stops which are going the function as network nodes and also the *network[Mode].csv* file which was selected only for bus and tram as discussed above.



**Fig. 1** Distribution of the city sizes seen as number of nodes in the corresponding network of public transport infrastructure show a balanced composition of the cities took under scrutiny in our study.

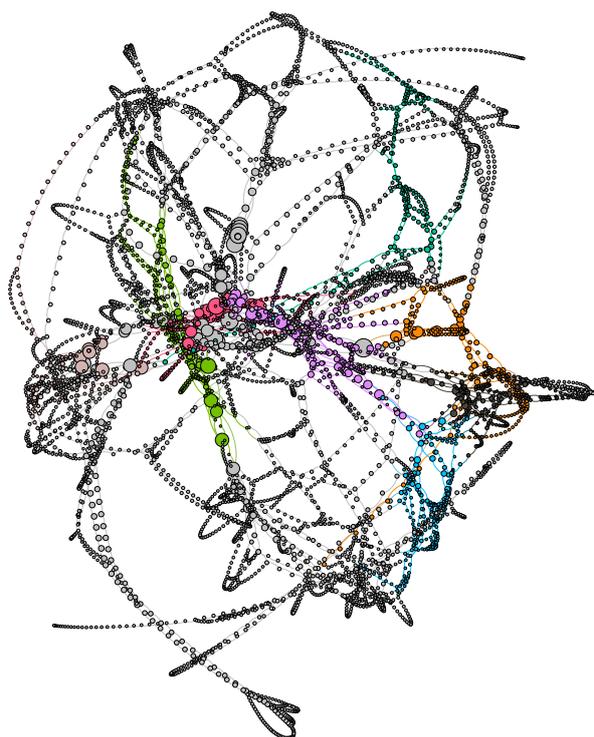
#### 3.1. Public transportation and complex networks

Data associated with each of the cities usually can be plotted as seen in Fig. 2 but beyond mere visual observation which can yield some insight into clustering of the station and routes there is no much data which can easily be obtained so we relied on transforming these data into a graph-like structure.



**Fig. 2** Public transportation routes in center of Helsinki, red dots representing stops. Network is constructed by considering directed edges between each pair of consecutive stops along a route.

Before applying the motif finding techniques and extracting relevant patterns we have first inspected the PT networks we have selected using the tools of complex networks. Having a large enough graph we hypothesise that there should be emergent properties which stem form the structure of the network. In Fig. 3 we show a network-centric representation of the transportation network for city of Athens in Greece. Each stop is associated to a node in the network while each directed edges are used to depict the travel route of the vehicle. The nodes are not placed based on their geographical positioning but the graph is rendered in Gephi using the Force Atlas 2 rendering algorithm which uses an approach stemmed for attraction an repulsion forces in physics to place the nodes at an equilibrium. Colors and sizing is going to be explained later on.



**Fig. 3** Network representation of the public transportation in Athens, Greece. Nodes are representing stops and edges are directed towards the travel routes of the vehicles. Nodes are sized by the betweenness metric and colored by the community they are part of. Rendering layout is Force Atlas 2 as implemented in Gephi 0.8

On the complex network side we measured classical metrics and their distributions. In Table 1 we show most important metrics computed for the cities we took under scrutiny. Number of nodes and edges provide a sense of the scale of the city and in Fig. 1 one can perceive the almost equal melange between small, medium and large cities.

The other important metrics which describe a complex network from its topological point of view are depicted in Table 1. The *average degree* is smaller then when we analyze the road network as a whole because in the case of transportation networks most nodes are represented by stops along a route and each node shall have exactly one directed connection (towards the next stop). The values are greater than one because in the case of some stops there are transfer stations or more than one line share the same stop and consequently there is going to be seen as "fork" in the network. Next important metric, which also can be considered a metric of size when we are dealing with physical cities is represented by *network diameter* which is simply the length (number of edges) of the longest of the shortest paths computer over each pair of nodes. One can perceive in this data-set that this is linearly dependant with the number of nodes, which is to be expected, but there are slight abnormalities such as the case of Paris where for 10880

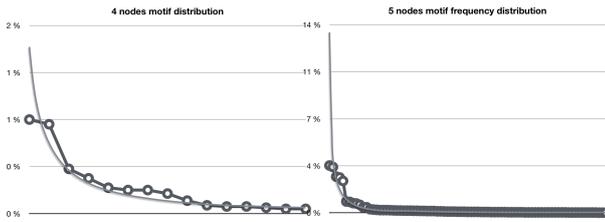
transport stops the the network diameter is merely 193, way lower than Detroit for example, which has 348 edges on its longest path while having only 4361 stops. It's worth noting that from a user's (commuter's) perspective the seemingly large numbers for the diameter are obtained from taking into account also the transit stations and not that a single line has more than 200 stops for example. The main reason for such disparities is stemming from the city topology and the way public transport is organized. Paris public transport relies heavily on "on-demand stops" where the commutes can press a button inside the vehicle the signal the driver for requesting to stop while other scenarios use fixed stops where the car actually stops even when there is nobody to get in or out.

Going further we measured the *modularity* metric for each network. As the name says it's a measure of how well the network can be separated into smaller "chunks" later called communities. Our previous work on this subject [9] provided significant results related to road networks as a whole and cities having a good modular structure. Higher values for modularity signify a better and easier division while the lower value impose that the network does not have a clustered structure (such as artificial grid networks). It's worth noting that in this case, even if we deal with urban networks we don't take into account the geo-location of the nodes while relying only on the topological properties (network connectivity). Here we can perceive very high values for the modularity which is a strong indication of the community structure. This is consistent with both the literature [10, 11] and the empirical observations. Cities are organized and evolved around the geographical features and landmarks while the neighborhood are the micro-societal way or organizing living. The public transport evolved around these factors and most of the time communities in public transportation networks are consistent with neighborhoods while the connections between communities are represented by metro lines or other forms of "long-links".

*Clustering coefficient* is another metric we have determined for our data set. This is similar to the previous notion of modularity, while both being empirical metrics of measuring *clustering* i.e. the way nodes are grouped together. Clustering coefficient uses the number of triangles which can be found (some simple form of *motif* in relationship with the 2-paths (used by the modularity algorithm). So, in our case low values for clustering coefficient are caused by the fact our networks have a lot of long 1-D paths along the travel route of our vehicle, consequently almost no triangles.

Putting all together we run the community detection algorithm, keeping each time the same constant value for the resolution parameter to 1.0. Due to algorithm's non-deterministic nature, for each set we have run the algorithm five times and taken the average value of communities rounded to the nearest integer. The lowest value of 22 communities is obtained for Venice with its very particular geographical features while cities with high values (Detroit-72, Paris-60) are either geographically big, or have a qualitatively good public transport infrastructure [12, 13], such as Dublin, Grenoble, Helsinki or Toulouse. For the example of Athens we have in Fig. 3 each node colored according

to the community it was identified to be part of. One final aspect we took under scrutiny is the *betweenness* metric. This is a classical one in the realm of complex networks, is associated with a specific node and loosely is defined as the number of shortest paths we traverse the node when computing all the shortest paths between all the pairs of nodes in the network. Usually is perceives a centrality metrics, measuring the importance of the node to the networks, because empirically a node which is part of numerous paths should be an "important one" [14]. In Fig. 3 the nodes are sized according to their betweenness from largest to smallest.



**Fig. 4** Distribution of the frequency of 4 and 5 nodes motifs: one can observe the heavy tailed, almost power-law distribution, suggesting there is a small number of frequently occurring motifs which could represent patterns of good practice in designing transportation networks.

### 3.2. Motif discovery and analysis

The main direction of our investigations is represented by *motif* analysis. Seen as subgraph or particular regular structure these represent structural patterns in the transportation network and can provide a static insight into the *good practices* of designing such systems. For all the datasets we have used FanMod [15] for motif discovery. For each of these we have searched for motifs of 3 to 5 nodes. In Fig. 5 we show the most prevalent motifs of 3 to 5 nodes, directed, among the 18 cities we took under scrutiny. There, above each snapshot representation of the associated subgraph, the number indicates the unique ID associated with the FanMod database, for further referencing. Since the three nodes motifs are trivial to observe in numerous maps of city transportation, in Fig. 4 we have shown the frequency distribution for the 4 and 5 nodes motifs, drawing attention to the power-law characteristic of it, which allows us to hypothesise there are a small number of "popular" motifs, which need to be investigated in order to extract patterns of good practices. We are also going to present in section 4 a few avoidable motifs either caused by the inability to serve a purpose or they fact they are not appropriate for public transportation.

#### 3.2.1. 3-node motifs

First we took under consideration all the motifs which can be made out of three nodes. In this case ranking the findings among the 18 data sets, the top ones are depicted in Fig. 5. Even if they might look simplistic and obvious some considerations can be made based on them. The most prevalent is motif 12 which is simply any sequence of 3

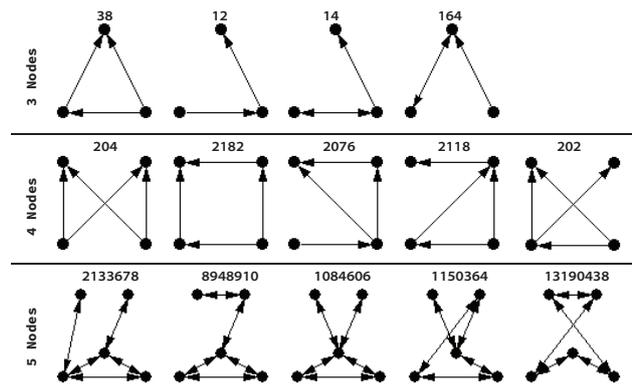
stops along a route, followed closely by motif 38 which represent a forking in the transportation network where from a stop one can chose two close stops of another line. Motifs with id's 14 and 164 are representations of a similar situation in which along a route there is a stop where there is a merging with another line shared with the first one, but which run in opposite direction.

#### 3.2.2. 4-node motifs

Increasing the number of nodes we gain access into a more complex view of the network, allowing us to see interactions between two or more transportation lines. The leading pattern, with id 204 is simply caused by to line which share the same two stops (an obvious frequent situation). Going further we have id 2182 which depicts the scenario where two lines share two stops, but in-between them there is another stop, individual to each line. Next are 2076 and 2118 which similarly describe a situation into which along the route of a specific line, there is a portion where stops are shared with another "smaller" route. Finally motif 202 describe the topology of a *hub* where three lines share a single unique stop, without sharing adjacent stops (they are taking different routes)

#### 3.2.3. 5-node motifs

Going further we examined the structures consisting of 5 nodes, where usually one can find so called *hubs*. Leading motifs are those with id's 2133678 and 8948910 there the visually central node is the one share among there major lines (the edge is bidirectional, so the car goes both ways along the same path). On third place (id=1084606) we have an even bigger hub where all the three lines are converging into the same stop, but this is less prevalent mostly because of the limited possibilities of finding topographical conditions into various cities. One can find such structures into two largely distinct cases: big cities such as Paris and Detroit where there are numerous inter-modal stations (switching among various means of public transportation) or in smaller cities (such as Nantes, Rennes, Luxembourg or Venice) where the same concept of inter-modality exists, but usually between inner city routes and outer city ones. The fist most popular structure is the one with id 13190438, where there is a complete *square* between four stations shared among two lines which run their routes back-and-forth.



**Table 1** Main complex networks metrics computed for the cities under study

City	Nodes	Edges	Avg. Degree	Net Dia.	Modularity	Avg. clustering coef	#Comm, r=1
Athens	6658	7870	1,182	206	0,922	0,007	46
Belfast	1917	2181	1,138	146	0,903	0,005	31
Bordeaux	3212	3805	1,185	125	0,9	0,006	31
Detroit	5683	5948	1,047	348	0,953	0,005	72
Dublin	4361	5276	1,21	127	0,913	0,019	38
Grenoble	1383	1514	1,095	123	0,898	0,002	29
Helsinki	6642	8611	1,296	116	0,921	0,01	43
Lisbon	6949	8730	1,256	153	0,933	0,007	41
Luxembourg	1352	3127	2,313	53	0,872	0,073	28
Nantes	2208	2587	1,172	125	0,901	0,009	32
Palermo	2176	2561	1,177	138	0,893	0,004	27
Paris	10880	12604	1,158	193	0,93	0,003	60
Prague	4476	5911	1,321	109	0,906	0,018	37
Rennes	1378	1643	1,192	85	0,881	0,009	22
Rome	7500	9669	1,289	143	0,924	0,013	41
Toulouse	3237	3668	1,133	171	0,925	0,003	40
Turku	1850	2341	1,265	92	0,887	0,008	26
Venice	1733	2398	1,384	82	0,885	0,026	22

**Fig. 5** Most prevalent directed network motifs of 3 to 5 nodes among the cities under study.

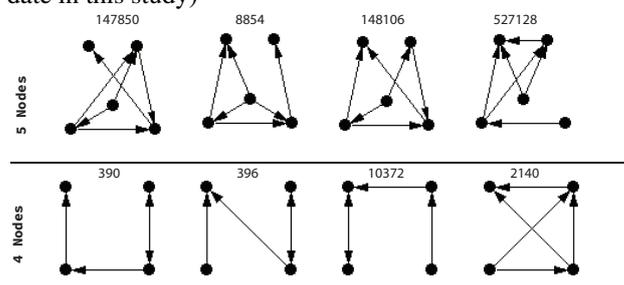
#### 4. ANTI-MOTIFS IN TRANSPORTATION NETWORKS

Inspired by the *anti-patterns* existent in the realm of software engineering we coined the term of *anti-motifs* to describe the less occurring structures we found-out in the transportation networks and which are a signature of avoidable situations.

In Fig. 6 we present the least occurring motifs of 4 and 5 nodes as identified by FanMod using the data about the cities from our study. For the three nodes motifs we have no further insights to provide, being by themselves not many. Regarding the typical occurrences of interconnections between 4 nodes we have found out that motif with id **2140** is ranking on the lowest place, the reason from the urban planning point of view being represented by the fact this structure does not add any further value for the transportation network: the two lines already would share a large portion of the route, adding short-cuts only providing a higher reliability and fault tolerance to the system, but with not practical gains. Further motif with id **10372** has a really low probability of occurrence being actually a reiteration of the three nodes motif with id **12** which was discussed in section 3.2.1. On the same topic, motifs **396** and **390** are less encountered in this form from a topological point of view because they are simply a corner case of the network structure emphasized by the existence of bidirectional links as part of the motif. In practice this exists in numerous occasions (where the line uses the same boulevard for both back and forth movement), but a correct modelling of such a case would require the decomposition of the network into distinct paths for each of the directions. Consequently this

scenarios is not applicable.

Regarding the more convoluted structures consisting of 5 nodes we can observe in the upper section of Fig. 6 the web of interconnections which is less often found in practical urban planning scenarios. Considering the sheer size of the search-space when dealing with 5-nodes motifs, the lower echelon is represented by motifs which have a probability of occurrence between 0.015751 for motif with id **147850** and 0.018174 for the one with id **527128**. For all practical reasons these do not represent any interest. From an algorithmic point of view they rely onto the existence of bidirectional links which, for a correctly modelled network, can't represent valid scenarios (excepting underground metro lines or rail-shuttle services where the same car moves back and forth, but of which we don't have any date in this study)

**Fig. 6** Least prevalent directed network motifs of 4 and 5 nodes among the cities under study.

#### 5. DISCUSSION AND CONCLUSIONS

Public transportation networks provide an alternative to personal cars and other more polluting and expensive ways of commuting. Major cities involve a great deal of effort and budget for finding good solutions to reduce the congestion and transit times. Our investigation was geared

towards applying methods and techniques from the complex networks to identify key characteristics which link the topology of public transportation networks among major cities. We used the concept of motif to describe patterns of stops and routes at a microscopic level and their distribution among the cities we have investigated. Ranging the node-size of the subgraph from 3 to 5 we have examined the distribution of motifs of that size, putting into evidence the actual connection to urban planning situations and the solutions provided in each case by the specific pattern/motif. When staying at low sizes (3 nodes) we don't get much insight beyond the obvious patterns but we can use this case for validating the method because the simple structures are much easier to be cross-referenced with city transportation maps by domain experts.

When we switch to 4 and 5 nodes the structures which unfold can truly go beyond simple observations on the map and uncover patterns among routes which share same stops. We can observe hub-like structure mostly when we examine the 5-node structures, where the top 5 most frequent motifs are all variations of the hub topology. Going beyond 5 nodes was at this moment unfeasible because the limited geo-visualization workflow did not allow us to cross-reference with the actual terrain situation for further explanations.

Going further we wish to extend this investigation with a quantitative examination of the influence of the motifs to the quality of the public transportation. Until now we have extracted the relevant data, and made empirical correlations between the prevalence of various motifs and their corresponding situations in the terrain, but being to quantify the impact of each motif and link it to the overall quality of the service would allow designing better public transport infrastructure, avoiding "bad" patterns and favouring "good" ones.

Regarding the "bad" patterns we started the investigation by searching the less occurring patterns in the tail of the distribution, considering that "nature" through urban evolution and good design practices the field of architecture and urban planning has distilled the paramount structures which "do make sense" for practical reasons, but the scarcity of the less occurring patterns makes us consider relevant the top-occurring one, requiring other avenues of investigation for finding avoidable patterns.

## REFERENCES

- [1] TIAN, Z. – JIA, L. – DONG, H. – SU, F. – ZANG, Z.: Analysis of urban road traffic network based on complex network. *Procedia Engineering*. **137** pp. 537–546 (2016).
- [2] HÁZNAGY, A. – FI, I. – LONDON, A. – NÉMETH, T.: Complex network analysis of public transportation networks: A comprehensive study.
- [3] WANG, D. – LI, B. – LI, J. – LI, C. – WANG, L.: Research on the public transport network based on complex network. *Indonesian Journal Of Electrical Engineering And Computer Science*. **12**, 4833–4839 (201).
- [4] SARAMÄKI, J. – ONNELA, J. – KERTÉSZ, J. – KASKI, K.: Characterizing motifs in weighted complex networks.
- [5] AGASSE-DUVAL, M. – LAWFFORD, S.: Subgraphs and motifs in a dynamic airline network. *Arxiv Preprint Arxiv:1807.02585*. (201)
- [6] PAULAU, P. – FEENDERS, C. – BLASIUS, B.: Motif analysis in directed ordered networks and applications to food webs. *Scientific Reports*. **5** pp. 11926 (2015).
- [7] MILO, R. – SHEN-ORR, S. – ITZKOVITZ, S. – KASHTAN, N. – CHKLOVSKII, D. – ALON, U.: Network motifs: simple building blocks of complex networks. *Science*. **298**, 824–827 (2002).
- [8] KUJALA, R. – WECKSTRÖM, C. – DARST, R. – MLADENVIĆ, M. – SARAMÄKI, J.: A collection of public transport network data sets for 25 cities. *Scientific Data*. **5** pp. 180089 (2018).
- [9] TOPIRCEANU, A. – IOVANOVIĆI, A. – UDRESCU, M. – VLADUTIU, M.: Social cities: Quality assessment of road infrastructures using a network motif approach.
- [10] BOTTA, F. – DELGENIO, C.: Analysis of the communities of an urban mobile phone network. *Plos One*. **12**, e0174198 (2017).
- [11] HELDEROP, E. – GUBESIC, T.: Streets, storm surge, and the frailty of urban transport systems: A grid-based approach for identifying informal street network connections to facilitate mobility. *Transportation Research Part D: Transport And Environment*. (2019).
- [12] MOUWEN, A.: Drivers of customer satisfaction with public transport services. *Transportation Research Part A: Policy And Practice*. **78** pp. 1–20 (2015).
- [13] CHICA-OLMO, J. – GACHS-SÁNCHEZ, H. – LIZARRAGA, C.: Route effect on the perception of public transport services quality. *Transport Policy*. **67** pp. 40–48 (2018).
- [14] TANG, J. – ZHANG, S. – ZHANG, W. – LIU, F. – ZHANG, W. – WANG, Y.: Statistical properties of urban mobility from location-based travel networks. *Physica A: Statistical Mechanics And Its Applications*. **461** pp. 694–707 (2016).
- [15] WERNICKE, S. – RASCHE, F.: FANMOD: a tool for fast network motif detection. *Bioinformatics*. **22**, 1152–1153 (2006).

Received February 2, 2020, accepted November 30, 2020

## BIOGRAPHIES

**Lilla Pellegrini** is a PhD. candidate in the field of Computer Science at Petrosani University in Romania with a solid background in mathematics and shifting her interests towards the application of complex networks theory into the field of intelligent public transportation systems. Email:

pellegrinililla@gmail.com

**Alexandru Iovanovici** holds a PhD in Computer Engineering at Politehnica University Timisoara, Romania and since 2017 is Lecturer at the Department of Computer Engineering and Information Technology at the same university. His research interests are in the fields of Intelligent

Transportation Systems and Embedded Systems. Email: iovanalex@cs.upt.ro

**Monica Leba** is Professor at Faculty of Mechanical and Electrical Engineering from University of Petrosani with strong background in the fields of Control Theory and Embedded Systems. Email: MonicaLeba@upet.ro