Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 27 (2017) 1034–1041

Transportation Research Procedia

www.elsevier.com/locate/procedia

20th Euro Working Group on Transportation, EWGT 2017, 4-6 September 2017, Budapest, Hungary

Identification of communities in urban mobility networks using multi-layer graphs of network traffic

Mehmet Yildirimoglu [a,*], Jiwon Kim [a]

[a] School of Civil Engineering, The University of Queensland, St Lucia, QLD 4072, Australia

Abstract

This paper proposes a novel approach to identify the pockets of activity or the community structure in a city network using multi-layer graphs that represent the movement of disparate entities (i.e. private vehicles, buses and passengers) in the network. First, we process the trip data corresponding to each entity through a *Voronoi* segmentation procedure which provides a natural null model to compare multiple layers in a real world network. Second, given nodes that represent Voronoi cells and link weights that define the strength of connection between them, we apply a community detection algorithm and partition the network into smaller areas independently at each layer. The partitioning algorithm returns geographically well connected regions in all layers and reveal significant characteristics underlying the spatial structure of our city.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: community detection, city structure, public transport, bluetooth

1. Introduction

Understanding the properties and the evolution of complex networks is crucial for many different fields, e.g. urbanism, epidemiology, power systems, etc. Complex networks consist of the nodes that represent system elements and the links that indicate the interactions between the elements. Examples of such networks include the internet, social networks, genetic interaction networks. Recently, transportation networks, too, have drawn significant attention among statistical physicists and applied mathematicians. Understanding human mobility patterns from a complex network theory perspective has been an active research area and provided worthwhile insights about the way cities and people are organized (Barthélemy, 2011).

Spatial structure of transportation networks is essential to understanding the interactions between neighbourhoods, identifying the infrastructure needs and improving the mobility operations. The availability of big mobility data sets (e.g. GPS trajectories, public transport smart card data, Bluetooth observations) allows us to build spatial networks that accurately represent travel patterns and identify natural communities and borders in the system. *Communities* in the study of complex networks refer to groups of nodes, within each of which nodes are connected more densely and

* Corresponding author.
E-mail address: m.yildirimoglu@uq.edu.au

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.
10.1016/j.trpro.2017.12.070

between which nodes are connected more sparsely. *Community detection* is a process of identifying the underlying community structure in a given graph by dividing the network into groups of nodes with dense connections internally and sparser connections between groups. Community detection is an important area of study in the complex network analysis as it allows us to identify sub-structures within a network that represent natural partitions or functional units of the system.

The aim of this study is to investigate the community structure in an urban transport network considering different movement entities such as cars, buses, and public transport passengers. Using trajectory data obtained from three different sources, namely, bus GPS records, public transport smart card transactions, and roadside Bluetooth detector records, this study builds a graph of population movement for each of the three associated movement entities: buses, public transport passengers, and cars, respectively. The community structure is then identified for each of these graphs and compared across the three graph layers to understand similarities and dissimilarities in mobility patterns for different movement entities. The main contribution of this paper is the development of a unified multi-layer network that shares the same node set across different layers, thereby allowing the construction of mobility graphs that represent trajectories from disparate sources in a common form and thus enabling the comparison of community structures across layers.

The rest of the paper is organized as follows. In Section 2, we briefly describe the data and explain the processing step which is necessary to map different sources of data on the same network. In Section 3, we provide a comparative statistical analysis of different layers and present the community detection results. In Section 4, we provide a discussion and conclude the paper.

2. Methodology

A comprehensive analysis of human mobility patterns requires multiple layers, e.g. the flows of people and vehicles representing distinct transportation modes. A multiple layer analysis would allow us to detect the patterns for people with different modes (e.g. car drivers, transit users, etc.) and the interaction between supply and demand.

2.1. Study Network and Data

The study network investigated in this paper is the Brisbane metropolitan area, Australia. Three trajectory datasets are obtained from this region as follows:

- bus trajectory data containing GPS tracks of 13640 buses between 6am-10am on Tuesday, 22/09/2015
- bus passenger trajectory data obtained from smart card transaction records between 6am-10am on Tuesday, 19/03/2013
- car trajectory data obtained from roadside Bluetooth detection records between 6am-10am on Tuesday, 22/03/2016

The bus trajectory data were obtained from archived data of real-time bus tracking information in General Transit Feed Specification - Realtime (GTFS-RT) format, provided by TransLink (Brisbane's sole transit agency). For the bus passenger trajectory data, *go* card (Brisbane's smart card system) transaction records were used, which contain information on boarding and alighting locations and times of individual passengers' trips. The car trajectories were extracted from Bluetooth detection records collected along state-controlled roads in Brisbane, provided by Queensland Department of Transport and Main Roads (TMR).

2.2. Multi-layer Graphs of Movement Trajectories

Graphs derived from the large mobility data sets have been analysed in the context of complex networks (Levinson and Krizek, 2007; Zhong et al., 2014; Saberi et al., 2016). Representing movement data as a graph has been also actively studied in visual analytics Adrienko and Adrienko (2011); Andrienko et al. (2010); Andrienko and Andrienko (2008); Guo (2007); Guo et al. (2010); von Landesberger et al. (2016). Inspired by these studies, recent work by Kim et al. (2016, 2017) has presented a procedure of dividing a large geographic area into small cells and constructing a graph of trajectory data using these cells as nodes. Following this approach, this paper first attempts to segment the study area into cells. First, data points in three trajectory sets are combined and grouped in space to form spatial point groups with an approximately 1km radius. The centroid of each point group is then estimated by finding the mean of the data points within the group. Once the centroids of all point groups are obtained, a *Voronoi* diagram is constructed using these centroid points as *seeds* for the Voronoi diagram. The Voronoi polygons are then used as cells that segment the study area. For more details on the network segmentation using a Voronoi diagram, readers are referred to Adrienko and Adrienko (2011); Kim et al. (2016).

After constructing the Voronoi digram based on the aforementioned three trajectory datasets, we obtained 319 cells that divide the Brisbane metropolitan area. Once the network is divided into cells, these cells are used to define the node set of a graph of aggregated flows, where nodes represent different regions in the city and edges represent traffic flows between two regions. Given one set of nodes, three different graphs (or edge structures) can be constructed: one for each trajectory dataset as shown in Figure 1. Having such a *multi-layer network* structure (i.e., defining separate graphs for different trajectory layers that are based on the same node set) allows trajectories with different resolutions to be mapped onto the same graph structure, providing a common ground for comparing flow patterns in different mode layers.

The first layer in Figure 1 represents the public transportation service; bus trajectories collected from GPS devices enable us to create a spatial network where the nodes are the neighbourhoods of the city and the links indicate the level of accessibility between them through public transport. This layer portrays the supply in the system, and the detected communities are due to the operation policy of the public transport authorities. The second layer in Figure 1 exhibits the movements of public transport users in the network or the demand in the system. The automatic fare collection system records the time and the location where a traveller taps on and off the readers. Such data provide a very high resolution of mobility patterns and covers almost all neighbourhoods in the city. The third layer in Figure 1 mainly describes the movements of private cars in the network; Bluetooth detectors installed in the roadway network record the identifiers of Bluetooth devices and enable us to construct an approximate trajectories for a subset of users in the network.

2.2.1. Bus Trajectory Graph

In order to produce the bus layer graph, we do not solely employ origin and destination nodes of a certain bus trajectory. We simply create a link between all the pairs of stations that are along the bus trajectory, which gives a better representation of service that the bus provides. In other words, we assume there is one passenger traveling between each pair of stations across the path. Note that the total link weight that is computed through this approach can be much higher than the actual number of passengers who would use the bus. The resulting graph gives an indication of how tightly different parts of the city are connected through the public transportation system, and it accounts for the frequency of bus service across the cells through its link weights.

2.2.2. Bus Passenger Trajectory Graph

Smart cards used in public transportation system provides a massive source of activity data in Brisbane. As it works in the principle of 'tap on - tap off', the fare collection system records passengers' boarding and alighting records. A passenger trajectory in this study is defined as a single origin-destination (OD) trip without any transfer or a sequence of multiple trips including transfer points between trips. As such, the first and last points of each passenger trajectory correspond to the origin and destination points of the whole journey, allowing us to distinguish between transfer points and actual origin and destination points. To produce the passenger layer graph, we connect an edge from the first (origin) to the last (destination) points of each passenger's trajectory and aggregate the edges for each origin-destination (OD) pair to represent the magnitude of OD flows as edge weight.

2.2.3. Road Traffic Trajectory Graph

Brisbane transportation network includes a significant number of bluetooth enabled intersections. These bluetooth devices record Media Access Control (MAC) IDs of the discoverable devices within their zone of coverage. MAC ID is a unique identifier for each device. Therefore, we can easily construct the trajectories of vehicles by matching their IDs across the detected intersections. This data set is limited by the available bluetooth detectors and the likelihood of missed detections. Hence, origin and destination cells are defined as the first and last bluetooth detections of vehicles in the network within a certain time frame. Mapping onto the same cell structure as before, we build the car layer graph where the link weights denote the number of cars traveling between the nodes.

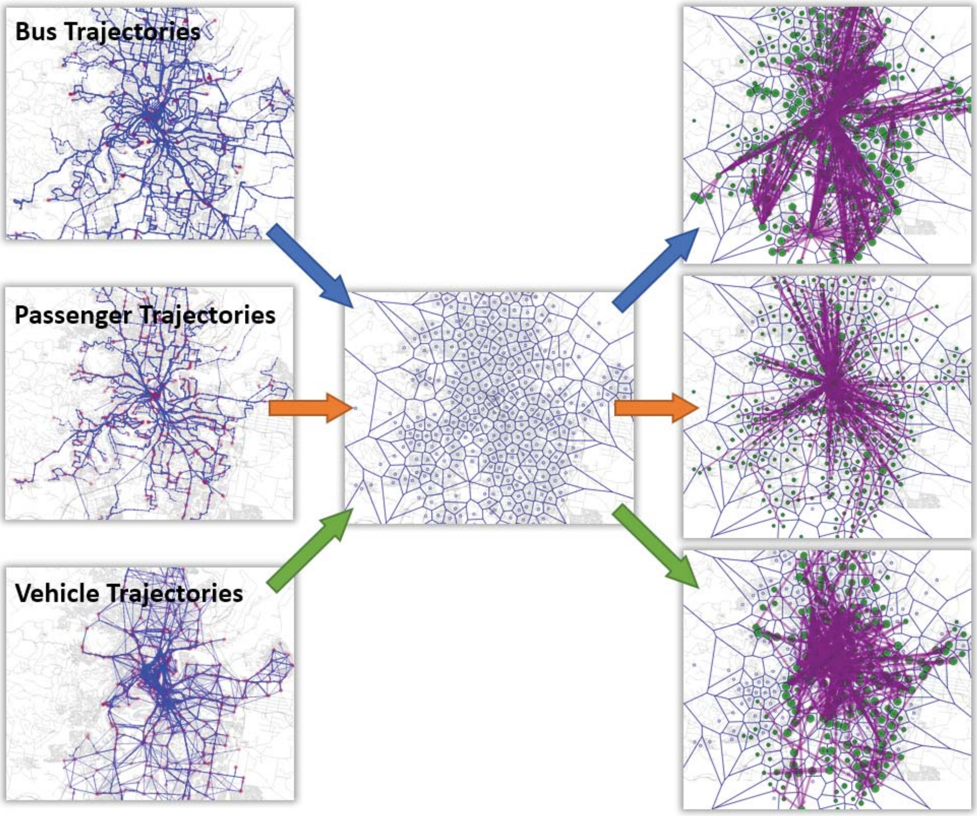

Fig. 1. Generating the graphs at different layers

2.3. Community Detection

Given a city network and the weight of link between nodes obtained from three trajectory layers, the next step is to identify the community structure in each graph layer. The goal of this step is to partition the city network into smaller zones in a way that maximizes intra-community interactions and minimizes inter-community connections. Although there are many different methods, there is no definitive approach to identify communities. One way to identify communities is through the maximization of a quantity called *modularity* (Newman and Girvan, 2004; Fortunato, 2010):

$$Q = \sum_{s=1}^{n_M} \frac{l_s}{E} - \left(\frac{d_s}{2E}\right)^2 \tag{1}$$

where n_M is the number of partitions, l_s is the number of links inside partition s, E is the total number of links in the network, and d_s is the total degree of the nodes in partition s. Maximization of Eq. 1 guarantees maximization of intra-connections inside the partitions, and the minimization of inter-connections between them. The first term in Eq. 1 is the fraction of links inside module s to total number of links, while the second term represents the expected fraction of links in that module if links were located at random in the network (and by keeping the same degree distribution). Note that the number of partitions n_M is also a variable that has to be calculated using an optimization procedure. High modularity values are obtained when the nodes in the same communities are strongly connected to each other and few connections exist across the communities. As precise formulations of this optimization problem are known to be computationally intractable, several algorithms have been proposed in the literature. Throughout this study, we apply the community detection algorithm developed by Blondel et al. (2008).

3. Results

Two commonly used measures in complex network studies are node degree k and betweenness centrality b. The node degree k is the number of links connected to a node in the network, while the betweenness centrality b represents the fraction of shortest paths in the entire network that goes through a particular node. Fig. 2 provides complementary cumulative distribution functions (CDF), i.e. $P(X \geq x)$, of node degree k and betweenness centrality b normalized by their means k_0 and b_0, respectively. The comparative illustration of the statistical properties in the complex networks reveal significant differences in the three layers. Fig. 2(a) shows that the distribution of node degree in passenger layer is completely different than two other layers. The complementary CDF values drop very rapidly in the passenger layer, which indicates that there are very few nodes with high degree, while this is not the case in the other two layers. In addition, despite the close alignment of the right tail values between bus and car layers, there is significant difference in the left end of the curves. Fig. 2(b) indicates a rather similar trend between the three layers in terms of betweenness centrality distributions, while one can still observe significant differences between left and right tails. This analysis suggests that the three layers do not exhibit similar behaviour and it is hard to argue the existence of a fundamental process in the way they are formed. This result is particularly interesting for the bus and passenger layers. Although they constitute supply and demand parts of the same system, the distribution of their statistical measures is significantly different. Note that the graphs that are used for this analysis are not weighted.

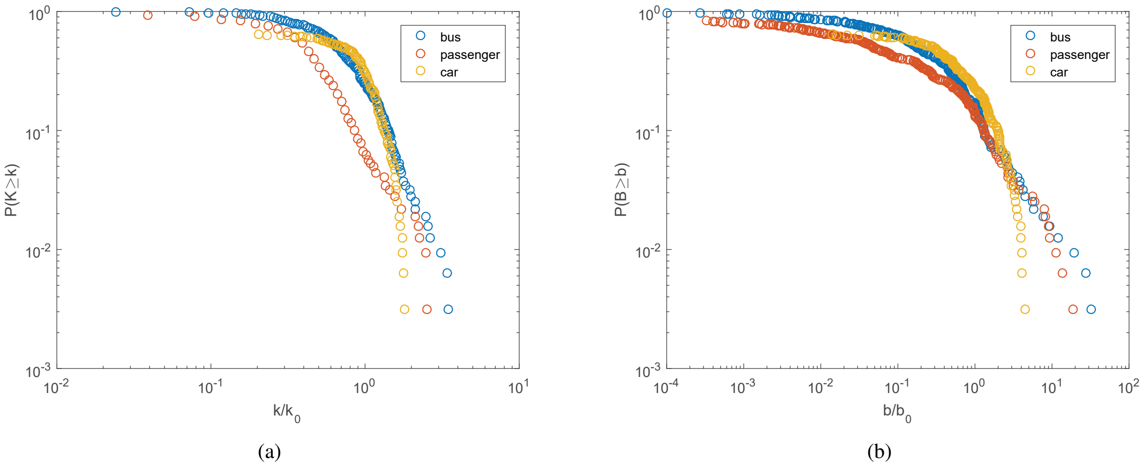

Fig. 2. Complementary CDF of (a) node degree k, (b) betwenness centrality b.

To compare the spatial structure in the three movement layers, we group the cells into larger zones and detect communities. This has been an important area in the complex network research where one looks for the best way to partition large complex systems, e.g. internet or social networks.

In the bus layer, community detection algorithm reaches the optimal modularity value after 3 iterations, and returns a surprisingly connected community structure (Figure 3(a)). Brisbane metropolitan area is divided into four distinct communities. Except few separated cells that belong to the third community (please see community IDs in Figure 3(a)), the algorithm returns perfectly connected components. Note that an adjacency constraint does not exist in the detection algorithm; resulting compact regions are due to the essential features of the public transportation system. Additionally, we observe significant implications of a monocentric city structure in the spatial division of communities. The area where the four communities intersect/meet is the city centre; there are nodes of all communities within this area. Note that a significant portion of bus routes in Brisbane is designed to serve to/from the city centre. Therefore, the sum of edge weights attached to these nodes is very high, and the algorithm seems to split them among communities in order to maximise their internal connections. Figure 3(b) presents the spatial structure that results from the passenger layer graph. The optimal modularity value is reached after 3 iterations. At this layer, Brisbane metropolitan area is divided into 5 communities, and two of them are fairly small. Although there are few scatters of communities in the network, the overall structure is again surprisingly compact and connected. Even though there are significant differences between Figure 3(a) and 3(b), the overall structure at the city scale looks quite alike. The city centre is split among 4 communities in the passenger graph as well due to the monocentric city structure. In case of car layer, the algorithm returns the optimal structure after 4 iterations (Figure 3(c)). Note that the number of total nodes in the network is lower than that of bus and passenger layers. The reason is simply the lack of bluetooth observations in the remaining cells. Similar to previously analyzed layers, the communities found in the car layer are compact and connected. Nevertheless, the structure topology is quite different from the bus and passenger layers. Most importantly, in the car graph, the city centre is not divided between communities; all the nodes in the city centre are assigned to community 1. Monocentric city structure seems to be less binding for the car drivers.

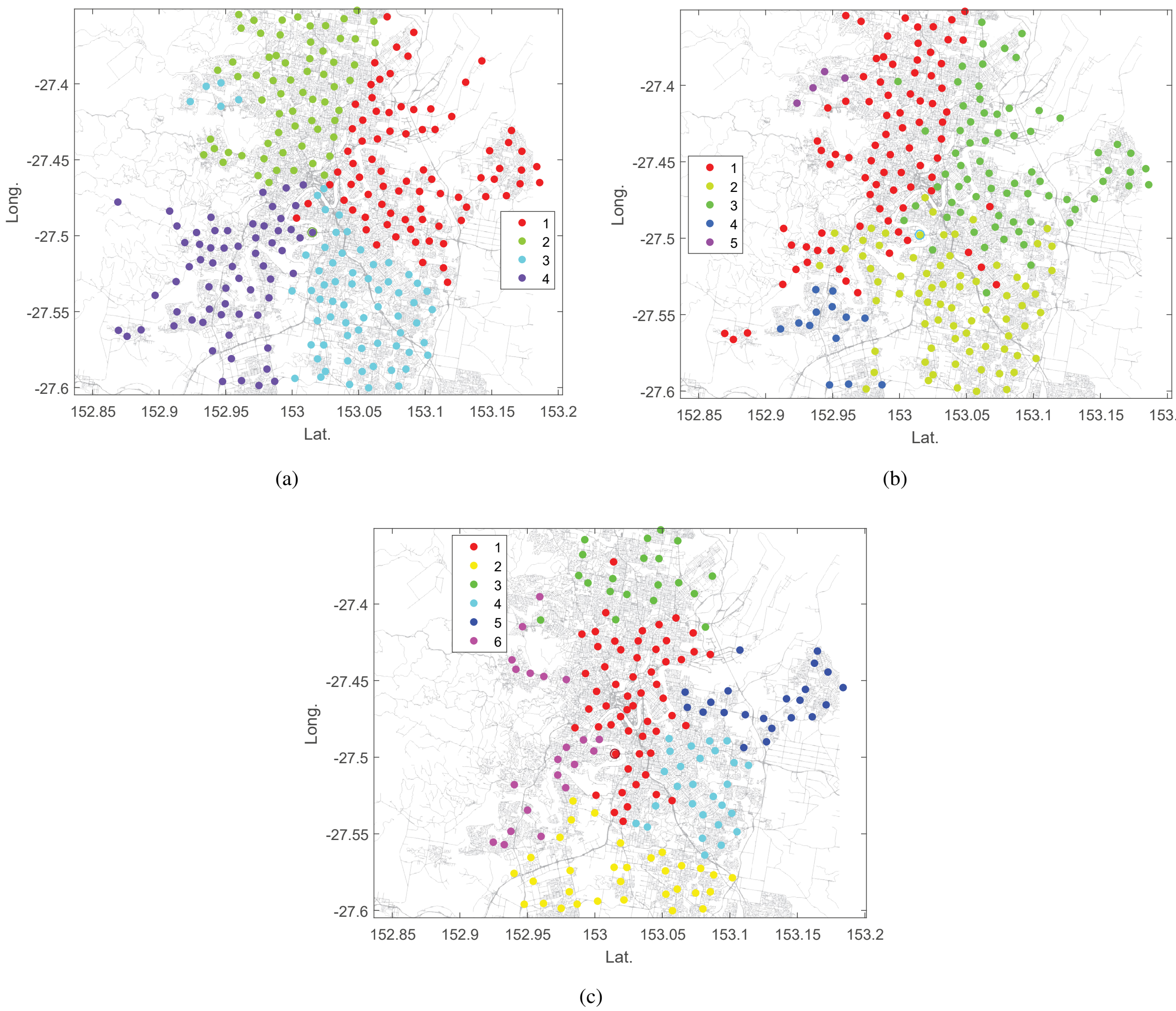

Fig. 3. Community structures for (a) bus flows, (b) passenger flows, (c) car flows.

In order to compare community structures at different layers, we define a similarity measure that accounts for the number of node (or cell) pairs that are correctly assigned to the same community in different layers. If p and s are different layers to be compared, the equation reads as follows:

$$\sigma_{ps} = \frac{\sum_i \sum_{i \neq j} \kappa_{ij}^p * \kappa_{ij}^s}{\min\left(\sum_i \sum_{i \neq j} \kappa_{ij}^p, \sum_i \sum_{i \neq j} \kappa_{ij}^s\right)} \tag{2}$$

$$\kappa_{ij}^{p} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same community in layer } p, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where σ_{ps} represents the similarity measure and takes a value between zero and one. The numerator in Eq. 2 counts the number of correctly located pairs in two layers, while the denominator represents the minimum of total number of 'same community' pairs in two layers. This value is close to one if all the node pairs are correctly assigned to the same community, and zero if they are located in different clusters. Table 3 depicts the similarity measures among the three layers we have analyzed in this paper. The values in the diagonal are always one, as they measure the similarity between the same layers. The similarity measures between bus-car and passenger-car layers are quite low, as it is supported by the visual evidence provided in Figure 3. On the other hand, the comparison between bus and passenger layers leads to a value of 0.68. That means, transit service and transit passengers form interchangeable community structures. Although such intercommunication is expected between the two layers, this study provides a rigorous approach to quantify the level of interaction. The high level of similarity between the two layers implies that transit passengers take into account the level of service provided by public transport authority while making accommodation decisions (where to live in the city). One may also claim transportation authorities design bus routes with respect to anticipated community structure in the city. However, the former seems more likely because of the dynamic nature of human beings.

Table 1. Similarity matrix

	bus	passenger	car
bus	1.00	0.68	0.45
passenger	-	1.00	0.46
car	-	-	1.00

4. Discussion

Our analysis has provided a novel approach to the study of city-scale mobility patterns through the newly available data sources that represent the movements of distinct entities in the network and the community detection techniques that identify the natural groups and borders in the system. City-scale movements of different entities, which are not easy to be collated in the original data mapping, are assembled in the *multi-layer network* structure that allows trajectories with different resolutions and characteristics to be mapped onto the same graph. The results show that, for each entity (i.e. bus, car and passenger), aggregated patterns can shape geographically well-connected communities. The resulting structure is quite alike for the bus and passenger layer, while the car movements produce a structure that is not line with the observations in other layers. In addition to the qualitative interpretation, we provide quantitative similarity measurement indices that compare the community structure across the layers, and these support the visual observations arising from detected communities. The future work should further investigate the similarities and dissimilarities between the movement layers. In particular, the mechanism that causes the swap of community assignments in the bus and passenger layers might provide further insights on the public transport structure and enable us to spot irregularities between the supply and the demand.

References

Adrienko, N., Adrienko, G., 2011. Spatial generalization and aggregation of massive movement data. IEEE Transactions on visualization and computer graphics 17, 205–219.

Andrienko, G., Andrienko, N., 2008. Spatio-temporal aggregation for visual analysis of movements, in: Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on, pp. 51–58.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S.I., Jern, M., Kraak, M.J., Schumann, H., Tominski, C., 2010. Space, time and visual analytics. International Journal of Geographical Information Science 24, 1577–1600.

Barthélemy, M., 2011. Spatial networks. Physics Reports 499, 1–101.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008, P10008.

Fortunato, S., 2010. Community detection in graphs. Physics reports 486, 75–174.

Guo, D., 2007. Visual analytics of spatial interaction patterns for pandemic decision support. International Journal of Geographical Information Science 21, 859–877.

Guo, D., Liu, S., Jin, H., 2010. A graph-based approach to vehicle trajectory analysis. Journal of Location Based Services 4, 183–199.

Kim, J., Zheng, K., Ahn, S., Papamanolis, M., Chao, P., 2016. Graph-based analysis of city-wide traffic dynamics using time-evolving graphs of trajectory data, in: Australasian Transport Research Forum 2016 Proceedings.

Kim, J., Zheng, K., Corcoran, J., Ahn, S., Papamanolis, M., 2017. Trajectory flow map: Graph-based approach to analysing temporal evolution of aggregated traffic flows in large-scale urban networks, in: 96th Annual Meeting of the Transportation Research Board. Washington, D.C.

von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., Kerren, A., 2016. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. IEEE transactions on visualization and computer graphics 22, 11–20.

Levinson, D.M., Krizek, K.J., 2007. Planning for place and plexus: Metropolitan land use and transport. Routledge.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical review E 69, 026113.

Saberi, M., Mahmassani, H.S., Brockmann, D., Hosseini, A., 2016. A complex network perspective for characterizing urban travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks. Transportation , 1–20.

Zhong, C., Arisona, S.M., Huang, X., Batty, M., Schmitt, G., 2014. Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science 28, 2178–2199.