Periodica Polytechnica
Transportation Engineering

Mobility Mapping Based on a Survey from the City of Berlin

44(1), pp. 35-41, 2016
DOI: 10.3311/PPtr.7587

Domokos Esztergár-Kiss[1*], Zoltán Koppányi[2], Tamás Lovas[2]

RESEARCH ARTICLE

Received 29 June 2014; accepted after revision 10 February 2015

Abstract

Analyzing mobility patterns helps to understand travel behavior of travelers, thus the transportation demands. A comprehensive mobility analysis was performed on a huge dataset from Berlin. This dataset contains answers from 41,000 volunteers, who reported their trips with departure and destination information. A modified graph representation is used to display locations and travel routes between them, thus traveling connections between the districts can be easily interpreted. We found that the destinations of the commuters are concentrated around some center locations, while entertainment based activity is more evenly distributed. Based on our results a realistic forecast of travel demand can be provided.

Keywords

mobility analysis, travel patterns, location based information, trip purposes

[1] Department of Transport Technology and Economics,
Faculty of Transportation Engineering and Vehicle Engineering,
Budapest University of Technology and Economics,
Stoczek u. 2., H-1111 Budapest, Hungary

[2] Department of Photogrammetry and Geoinformatics,
Faculty of Civil Engineering,
Budapest University of Technology and Economics
Műegyetem rkp. 3., H-1111 Budapest, Hungary

Domokos Esztergár-Kiss, Researcher ID: A-7930-2013

Zoltán Koppányi, Researcher ID: F-4485-2015

Tamás Lovas, Researcher ID: F-4491-2015

[*] Corresponding author, email: esztergar@kku.bme.hu

1 Introduction

The novel positioning solutions due to the penetration of social networking and mobile devices produce a huge amount of location information with different quality. The analysis of these patterns is a research topic since a long time (Gonzales et al., 2012; Timmermans, 2005), but significantly enhanced interest emerged in the last decade. Due to the increasing number of available data sets the researchers are able to track individuals and are able to understand the mobility of the city population. Improving public transportation quality requires the introduction of novel approaches and using cutting edge technologies for processing different data sources, and the application of the latest methods in Intelligent Transportation Systems (ITS) (Zhang, 2010). Using the results of mobility analysis from individuals, daily activity chains of travelers can be obtained and the patterns of their movements can be derived. The travelers' predicted routes can be beneficial for transport operators, as they can create such timetable schedules of buses and trams that represent the real travel demand of the passengers. On the other hand, understanding and describing mobility enables the simulation of the city life. These outputs are essential for urban planners and decision makers.

The contribution of this study is to use questionnaires type data acquisition reported by travelers to derive general mobility information. As an initial point in this paper, we focus on the mobility mapping of the data. Our dataset includes approximately 110,000 records from 41,000 volunteers and was collected in the city of Berlin. They reported on their daily mobility pattern with departure and destination district, the purpose of the trip, the vehicle they used and other attributes of the trip. This information allows geographical representation of the trips between the districts of Berlin and therefore, enables the analysis of the recent conditions of the traffic and transportation. In this paper we intend to answer the following questions:

- How can we derive location-based results from survey data?
- How can we display this location-based information?
- How can we interpret and what are the limitations of the proposed method and the questionnaire-based data acquisition?

The next section gives an overview on the related works dealing with the potential of mobility modeling in transportation. Then some general statements were derived regarding the households, persons and journeys. The Method section discusses the geocoding process on alphanumerical data and the mobility representation. Finally, the mobility graphs are presented regarding the purpose of the trip and the transportation mode in different time intervals.

2 Background, related work

Concerning mobility patterns, previous studies claim that particular person's daily routes are really similar to each other, and according to González et al. (2012) the chosen routes show a high degree of spatial and temporal regularity. That means the passengers are very likely to use the same routes and visit the same places frequently, which allows providing reliable forecast based on mobility data from the past.

The paper of Phitakkitnukoon et al. (2010) presents a mobility analysis method. The main goal of the study was to describe daily activity patterns of travelers, who usually visit the same working area. Using correlation analysis, the results were shown on an activity map. Individuals share similar activity patterns if they are close to each other.

Timmermans et al. (2003) reported on an international comparison of travel patterns. The aim of the research was to understand the connection of urban structure and travels of individuals. They used data from different data collections with different methods and compared them applying a unified methodology. The results indicated that travel pattern is independent of the urban structure.

Transportation can be divided in short-term, medium-term and long-term processes, where trips belong to the first group, and transportation mode choice to the medium-term processes, which is influenced by mobility demand. Thus, if the mode choice is known, the mobility demand can be predicted, and thus, if the demand is known, the traffic can be controlled. Ficzere et al. (2014) developed a system to help policy makers to create long-term transport strategies based on travel time.

In the paper of Péter and Fazekas (2014), the processes in large-scale road traffic networks were described, where a new dynamical model was developed considering real circumstances.

Concerning the demand some measures can be realized and recommendations can be addressed to the passengers according to Juhász (2013) (e.g. infrastructural investments, tax regulation, time-based fares, enhanced information).

Markovits-Somogyi and Aczél (2013) analyzed the human behavioral patterns and they found that during a trip people behave following a specific logic. Their focus of the research was rather put on economic aspects, i.e. on the effects of travel costs regarding transport mode choice of the individuals.

Song et al. (2010) investigated the limits of predictability in human behavior. They also used routes based on mobile phone information derived from the mobile networks, and using this information they were able to predict journeys. They claim that there are many different travel patterns for the passengers, but these patterns are well predictable, furthermore the results are independent of the distance.

The paper of Buliung et al. (2008) describes the spatial properties of trips focusing on the differences between weekdays and weekends. The results indicated a high degree of regularity, which was measured by a spatial repetition index. For all activities, the repeated trips reached 72% of the total. Additionally similarities were reported concerning many surveys across Europe.

The daily variability was investigated in the study of Kang and Scott (2010). They claim that the travelers may choose different routes on different days, which means it is not sufficient to analyze only a single day survey. Especially changes of weekdays and weekends are more different. The differences in rural areas were investigated by Hine et al. (2012) concerning weekly activity behavior. Travel patterns were analyzed for different area accessibilities using questionnaires and travel diaries. The influence of this factor was calculated through regression analysis. They found that travelers of higher accessibility to transportation are more integrated into local communities.

In the study of Bradley and Vovsha (2005) the interactions between members of the households were analyzed. A group decision-making process was taken into account, and patterns were derived for the individuals. Furthermore, the paper discusses the statistical analysis of interactions and a choice model based on the utility (with person-specific components) of travelers.

Analyzing data and presenting its results is crucial according to Kamruzzaman et al. (2011). They used GIS environment to provide travel behavior of students. Their aim was to detect transport demands based on visited locations, daily distances and activity types.

Young travelers' trips were also analyzed by Kerr et al. (2007), and they found that after categorizing groups by demographic aspects, the walking distances show high correlation with income and number of cars of the household, although in this survey only walking was analyzed, no other transportation modes.

Mukherjee, Pate, Krishna (2014) have developed a heterogeneity index for assessment of the relationship between land use pattern and traffic congestion.

It can be clearly seen that numerous surveys and analyses were conducted in the past years in the field of travel activity, however, many of them concentrate on some specific aspects of behavior analysis.

3 Dataset

Our dataset in this study is the "Mobilität in Städten – SrV 2008" survey conducted by a researcher group of TU Dresden in Germany. The aim of this survey is to examine the users' demands, to analyze daily travels and to monitor changes of

behavioral patterns. Generally, the survey consists of household interviews covering several German cities, which is repeated every 5 years. Our dataset contains data from 41,000 persons from Berlin in 2008. The range of interviewees covered the demographic diversity of Berlin based on statistics reports, thus the survey contains 3-5% of the entire population. Through 6 months, the interviewer chose households randomly and each person from the chosen household had to report all of his or her trips on that specific day. These answers with household's information and metadata were stored and organized in a database.

Table 1 General statistics from the database

Name	Total	Average	Std deviation
Households	19.354	-	-
Persons in the household	41.050	2,12	1,12
Cars in the household	14.283	0,72	0,77
Bicycles in the household	29.358	1,51	1,63
Trips per person	111.228	2,69	1,79
Distance of trips	1.024.984 km	9,21 km	42,84 km
Duration of journeys	2.765.401 min	24,86 min	31,9 min
Speed of journeys	-	15,43 km/h	16,48 km/h

The database consists of 3 tables: households, persons and travels. The household table contains the location (city, district, street), basic information about households (number of persons, number of cars), and reachability of public transportation modes (walking time to the next bus/tram/metro stop). The person table stores personal information (age, gender, type of occupation, qualification, income) and other personal attributes (visually handicapped, physically disabled). Some general statistics can be seen in Table 1. The data analysis proved that the examined data follow logarithmically, or generally, power-law distributions, which can be the major of the explanation of the high STD.

The travel table contains the starting location and time, the purpose of the trip, arrival location and time, transportation mode and weather. Here we define the single trip as a travel with a certain purpose. During a trip, more transportation modes can vary. Thus, a trip from the workplace to home interrupted with shopping means two trips, one with the purpose of shopping, one with the purpose of getting home. These are stored separately as two unique trips in the database.

The location information is represented by location codes. These codes are the 195 statistical area codes of Berlin, which correspond to the borders of the districts and it reflects social and economic aspects. These district IDs are the location descriptors that gives the departure and the destination of the trips. The daily travels of a single individual can be obtained as a vector concatenating the consecutive rows.

4 Method

4.1 Acquiring district location

The participants reported the departure and destination locations of their trips. In order to describe their mobility, it is essential to acquire the coordinates of these locations. The reported locations are represented with the ID of the districts and sub-districts. Berlin is divided to 12 districts (Table 2), and the districts are further divided into 187 smaller areas (here we refer them as sub-districts).

Table 2 District IDs and their names

ID	Name	ID	Name	ID	Name
01	Mitte	05	Spandau	09	Treptow-Köpenick
02	Friedrichshain-Kreuzberg	06	Steglitz-Zehlendorf	10	Marzahn-Hellersdorf
03	Pankow	07	Tempelhof-Schöneberg	11	Lichtenberg
04	Charlottenburg-Wilmersdorf	08	Neukölln	12	Reinicken-dorf

The relationship between the name of the districts and their numbers is unambiguous. Knowing the names of the districts, their centroids (i.e. center point of the area) can be geocoded into WGS84 geographical coordinates using Google Geocoding API. Figure 1 shows the sub-districts' coordinates inside the border of Berlin. Note that these coordinates are only the estimation of the real district centers due to the geocoding error.

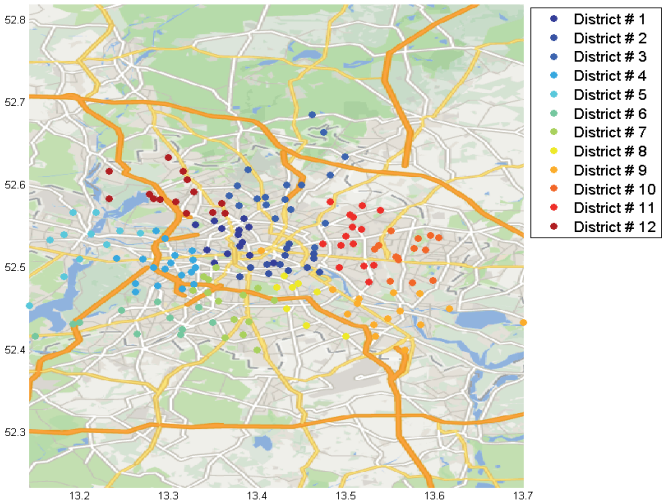

Fig. 1 Centroids of sub-districts derived from Google Geocoding API

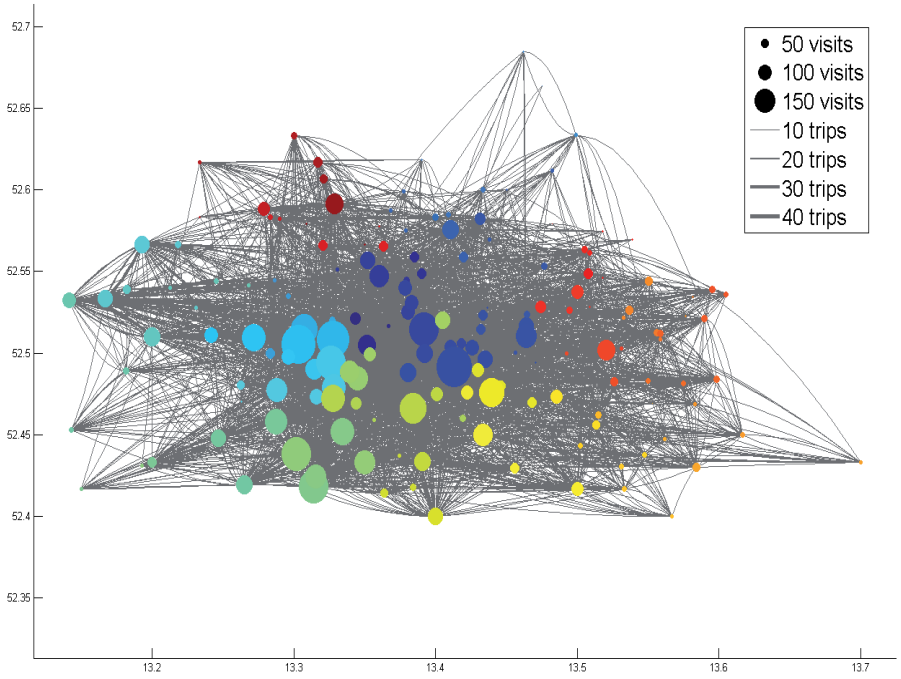

Fig. 2 Mobility of Berlin from a partial dataset

4.2 Representation of the trips

The departures and destinations are also available using geocoding, thus the directed adjacency matrix can be composed of the reported trips. Using adjacency matrix the traffic connection between districts can be visualized as a graph. The graph node locations are derived from the WGS84 coordinates of the districts.

Figure 2 represents a graph from a smaller dataset, which shows the connection between the sub-districts using this visualization method. Although numerous trip points of the city can be detected, the single trips are not detectable. It can be claimed that most trips are performed in the Districts 2, 6, 7 and 8 (Friedrichshain-Kreuzberg, Steglitz-Zehlendorf, Tempelhof-Schöneberg, Neukölln). In order to enhance the output quality, the dataset was filtered and the spatial resolution was decreased.

Spatial resolution is decreased by considering only trips between districts instead of sub-districts, thus transitions become more distinguishable for the visual comparison. Note that the dataset became also smaller, because only those trips have been investigated, where destination and departure districts are different. Additionally, to filter unreliable data, those persons were omitted from the analysis, who reported less than 3 trips or filled the forms partially. Furthermore, trips with departure or destination location outside Berlin are also removed. The final dataset after filtering contains 4,462 persons and 27,888 reported trips that are approximately 1/5 of the total dataset. In order to show as much information as possible in a single figure, some other visualization tools are used: the size of the node represents the number of visits, and the width of the edges reflects the number of trips between two districts. The edges are represented as a curve and their direction can be identified by their curvature that follows clockwise direction (Fig. 3).

$$A_{i,j} = \begin{cases} n_{i,j}, & \textit{if there is connection between the ith and jth districts} \\ 0, & \textit{otherwise} \end{cases}$$

where $n_{i,j}$ is the number of trips between the i th and j th districts.

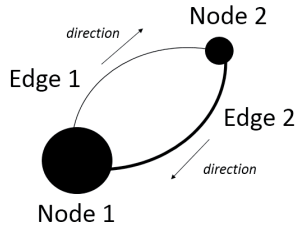

Fig. 3 Graphical representation of the graph

In the example, the sizes of the nodes show that the number of visits is greater in Node 1 than in Node 2. The two curves (Edge 1 and Edge 2) between the nodes represent trips in both directions. The direction can be determined by following the clockwise directed curvature. The width of the edges represents the number of trips between the nodes. In this case, Edge 2 contains more trips than Edge 1.

5 Discussion

5.1 Analyzing purposes

Figure 4 shows the visit frequencies sorted by trip purposes. The X axis shows the hours of the day and the Y axis shows the corresponding number of trips that is in the indicated time interval, respectively. The time resolution is 1/2 hour (i.e. the bins cover half hour interval). The different purposes are depicted by colors. The amounts of the trips build on each other, which means that the Y value of one selected purpose graph is the

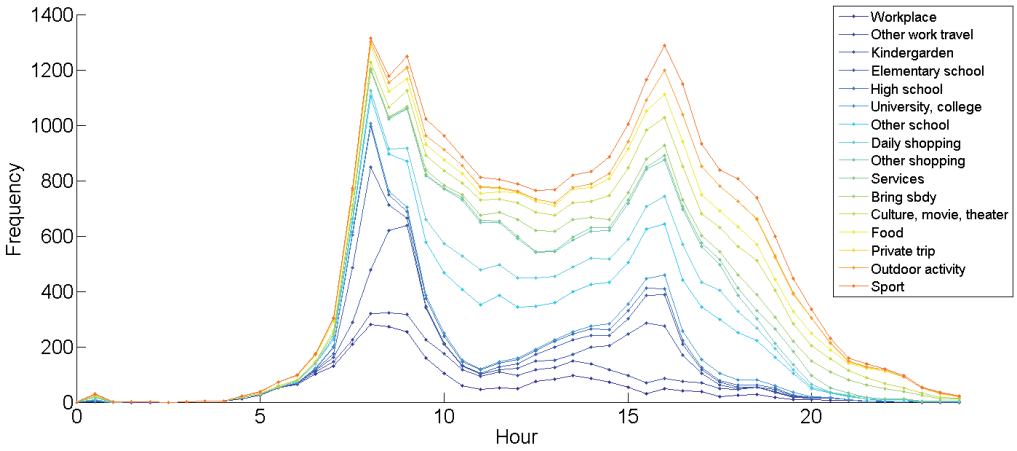

Fig. 4 Frequency of trips with particular purposes

summary of all graphs below. The number of the trips that corresponds to a single purpose is the difference of the Y value of the selected graph and the Y value of the previous graph. The aim of this representation is to present not only the purposes separately but the summary and ratio among the purposes.

In Figure 4 two peak hours can be detected, the first is between 7:00 and 10:00 and the second is between 15:00 and 18:00, which corresponds to the daily human behavior. Most of the trips regarding the early peak are the travels to the workplace, to the school, university or taking kids to the school/kindergarten. The frequencies of these trips decrease as time passes. Late afternoon and in the night the services and entertainment category trips have a major role. The least number of trips can be observed during late night hours.

5.2 Trips between districts

Figure 5 shows the reported trips between districts divided by time intervals of the day and the purposes of the trips. Two selected purposes are presented: work and entertainment (such as watching a movie, or going to the theater). The visit frequencies of the locations are normalized by 30 visits. The width of the edges also represents the visit frequencies, but only between two districts, and is also normalized. During the normalization process the highest frequency of the sub-figures was set to a predefined value (in our case 1), and all the others in the same sub-figure were calculated proportionally; this value is displayed in the legend of the sub-figures. Note that the size of the nodes can be comparable through all the sub-figures due to the same normalization constant (here 30), but the edges can be only comparable within the same sub-figure, not between two sub-figures due to different constants.

The sub-figures of the reported trips show that the central districts are playing a key role in the life of Berlin. District 1 (Mitte) is the central target area for working; in the morning passengers commute generally from the outside districts to District 1. In addition, other major working areas are District 2, 4 and 7 (Friedrichshain-Kreuzberg, Charlottenburg-Wilmersdorf, Tempelhof-Schöneberg), which are located around District 1.

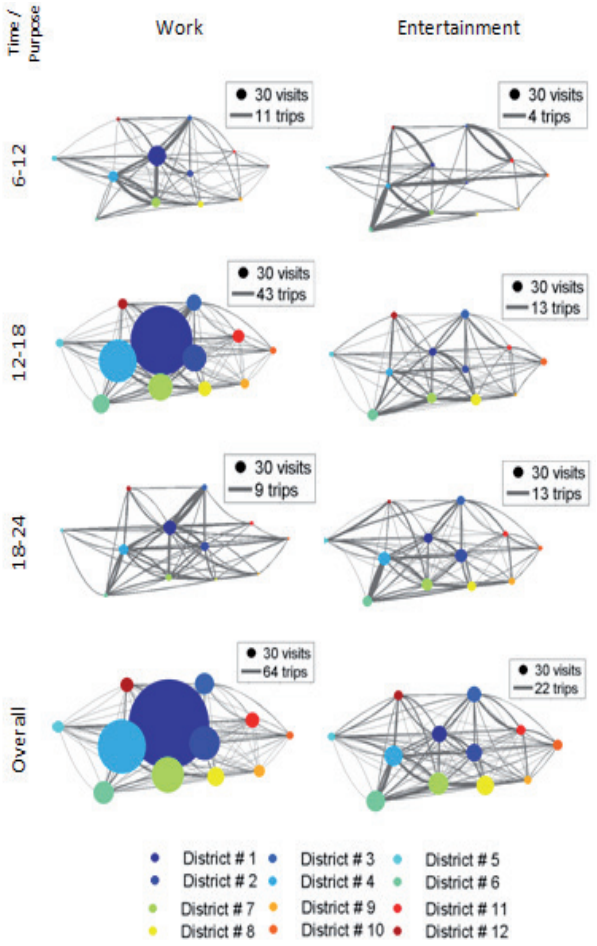

Fig. 5 Connection between the districts sorted by time and purpose

The afternoon period surprisingly shows more work trips; it is hard to decide whether this result comes from the unreliability or incompleteness of the dataset or this is the real situation.

The number of entertainment purpose trips is slightly growing from morning to midnight. The frequently visited districts are distributed more evenly than in other cases. The overall graph of the entertainment trips shows that the outskirt districts tend to be as important as central districts.

Similarly, transportation modes were also examined in Fig. 6 regarding bicycle, car and public transportation. Cycling is popular in Berlin, the use of this transportation mode is located

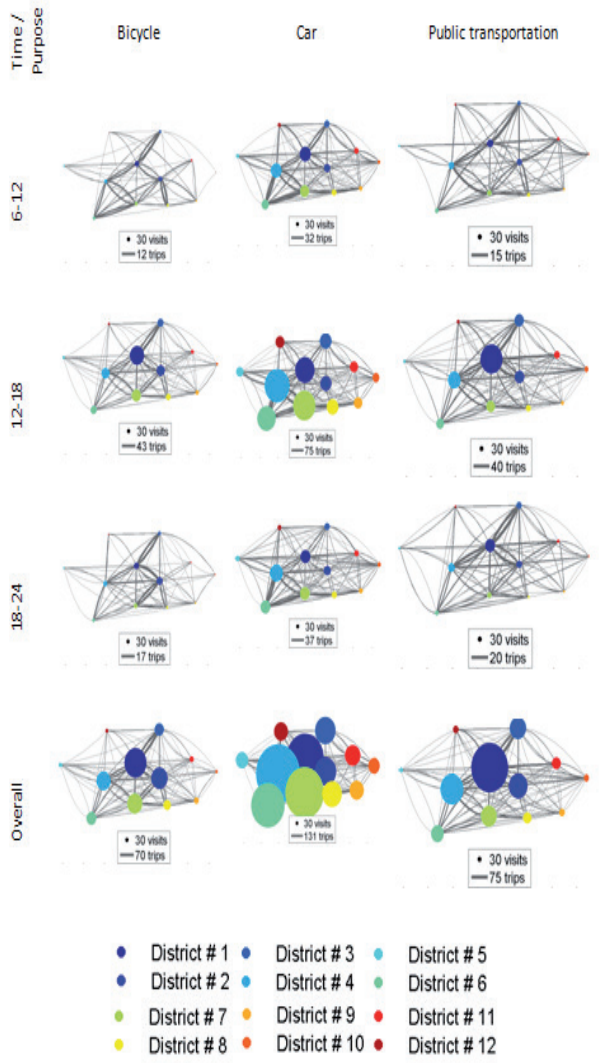

Fig. 6 Connection between the districts sorted by time and transportation mode

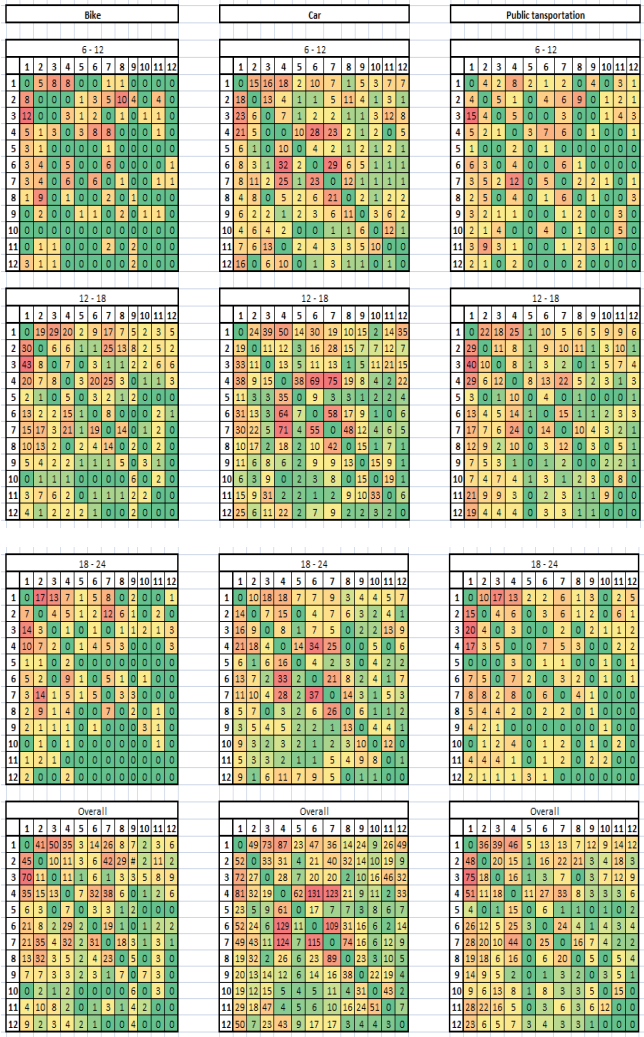

Fig.7 Trips between the districts sorted by time and transportation mode

mainly around the central districts. People living in the outskirts presumably prefer to use the car to reach their workplaces or the centrum due to longer distances. Also, note that cycling is more popular in the afternoon hours.

The dominance of the cars are not surprising, Berlin is a huge metropolis with well-developed road infrastructure. Comparing the sub-figures visits between districts show same frequencies through the day, which means that the mass points do not change in the 24 hours. The main destinations are the central districts, and the rate of car travels is high compared to other transportation modes. Sub-figures of the bicycle in the afternoon show similar frequencies as cars in the morning.

Concerning public transportation, travelers use this transportation mode very frequently in order to reach their destinations; in the afternoon hours nearly as many trips can be observed, as that of car travels. The importance of public transportation can be seen especially in the morning hours.

Concerning the trips, the values for each district are shown (Fig. 7) for the chosen transportation modes (bike, car, public transport). The diagonals are 0, as we eliminated trips within the zones, only between two different zones. In general, the same trends can be observed regarding the connection of districts during the day. Only in the case of biking the number of trips increase between district 1 and 7 in the afternoon, in case of cars it decreases in the evening between District 1 and 12.

Between Districts 4, 6 and 7 an extensive car usage is realized in the morning, while bikes are being used mainly between the Districts 1, 2 and 3, and public transport between 1, 3 and 4. The relatively high number of cyclists can be explained by short distances between the districts, congested road network and existing cycling infrastructure. The high frequency of public transportation usage can be due to the good connections of the lines in those specific districts. If the municipality and transport operators analyze the differences of trip modes during the day between the districts, the lack of infrastructural development can be identified to execute actions.

6 Conclusion

Analyzing traveler mobility patterns is of critical importance, especially in densely populated urban areas. Current mobile devices and cutting edge positioning methods enable the tracking of individuals during their trip, but as a result, these

techniques only provide location data. Questionnaire surveys, however, acquire rich attribute data besides the location. The current research proves that detailed analysis of questionnaire data provides useful information to public transport operators, urban planners and transportation authorities.

In this paper, we focused on mapping mobility from surveys. First, the text location information was transformed to coordinates using geocoding. Based on these positions the mobility and trips between the districts of Berlin were presented as a special graph representation. Using the acquired time-stamped and geocoded data

- the most frequented travel routes can be determined; this can be very useful for urban development, service marketing etc.
- the visit frequencies of each district can be observed and compared to each other; this can effectively support public transportation scheduling and infrastructure capacity optimization.

The main goal of the paper was to present a method for mobility data analysis and derive some spatial demands of passengers, which can be also useful for the operators. We defined the number of passengers traveling from one district to another district specified for the part of the day. This enables the definition of critical directions and times when the transportation network use is very high. The operators can use these outputs to discover frequently visited areas, and plan their network accordingly (e.g. more vehicles runs or building a new line for long term planning).

The presented work is the initial step of a comprehensive research. Based on the primary results provided by data mining, we intend to apply probabilistic automaton based methods (Linz, 2006:p.145; Stoelinga, 2002; Dupont, 2005; Virkar, 2014) to reveal further connections within the dataset. To get more detailed information on the travelers' behavior and to support their classification, probabilistic definition function based method is under development. The already accomplished task and results presented in this paper provide a solid basis for the further research.

Acknowledgement

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4. A/1-11-1-2012-0001 'National Excellence Program'.

References

Bradley, M., Vovsha, P. (2005) A model for joint choice of daily activity pattern types of household members. *Transportation,* 32(5), pp. 545-571. DOI: 10.1007/s11116-005-5761-0

Buliung, R. N., Roorda, M. J., Remmel, T. K. (2008) Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto travel-activity panel survey (TTAPS). *Transportation*, 35(6), pp. 697-722. DOI: 10.1007/s11116-008-9178-4

Dupont, P., Denis, F., Esposito, Y. (2005) Links Between Probabilistic Automata and hidden Markov Models: Probability Distribution, Learning Models and Induction Algorithms. *Pattern Recognition*, 38(9), pp. 1349-1371. DOI: 10.1016/j.patcog.2004.03.020

Ficzere, P., Ultmann, Z., Torok, A. (2014) Time–space analysis of transport system using different mapping methods. *Transport*, 29(3), pp. 278-284. DOI: 10.3846/16484142.2014.916747

Gonzalez, M. C., Hidalgo, C. A., Barabasi, A. L. (2012) Understanding individual human mobility patterns. *Nature,* 453, pp. 779-782. DOI: 10.1038/nature06958

Hine, J., Kamruzzaman, Md., Blair, N. (2012) Weekly activity-travel behaviour in rural Northern Ireland: differences by context and socio-demographic. *Transportation*, 39(1), pp. 175-195. DOI: 10.1007/s11116-011-9322-4

Juhász, M. (2013) Travel Demand Management – Possibilities of influencing travel behavior. *Periodica Polytechnica Transportation Engineering*, 41 (1), pp. 45-50. DOI: 10.3311/PPtr.7096

Kamruzzaman, M., Hine, J., Gunay, B., Blair, N. (2011) Using GIS to visualise and evaluate student travel behavior. *Journal of Transport Geography*, 19 (1), pp. 13-32. DOI: 10.1016/j.jtrangeo.2009.09.004

Kang, H., Scott, D. M. (2010) Exploring day-to-day variability in time use for household members. *Transportation Research Part A: Policy and Practice,* 44 (8), pp. 609-619. DOI: 10.1016/j.tra.2010.04.002

Kerr, J., Frank, L., Sallis, J. F., Chapman, J. (2007) Urban form correlates of pedestrian in youth: differences by gender, race-ethnicity and household attributes. *Transportation Research Part D: Transport and Environment*, 12 (3), pp. 177-182. DOI: 10.1016/j.trd.2007.01.006

Linz, P. (2006) *An Introduction to Formal Language and Automata*.4th ed. Jones & Bartlett Publications.

Markovits-Somogyi, R., Aczél, B. (2013) Implications of Behavioural Economics for the Transport Sector. *Periodica Polytechnica Transportation Engineering*. 41(1). pp. 65-69. DOI: 10.3311/PPtr.7101

Mukherjee, A. B., Pate, N., Krishna, A. P. (2014) Development of heterogeneity index for assessment of relationship between land use pattern and traffic congestion. *International Journal for Traffic & Transport Engineering*, 4 (4), pp. 397-414. DOI: 10.7708/ijtte.2014.4(4).04

Péter, T, Fazekas, S. (2014) Determination of vehicle density of inputs and outputs and model validation for the analysis of network traffic processes. *Periodica Polytechnica Transportation Engineering*, 42 (1), pp. 53-61. DOI: 10.3311/PPtr.7282

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., Ratti, C. (2010) Activity-Aware Map: Identifying human daily activity pattern using mobile phone data. In: *Human Behavior Understanding*, Lecture Notes in Computer Science*,* 6219, pp. 14-25. DOI: 10.1007/978-3-642-14715-9_3

Song, C., Qu, Z., Blumm, N., Barabási, A. L. (2010) Limits of Predictability in Human Mobility. *Science*, 327 (5968), pp. 1018-1021. DOI: 10.1126/science.1177170

Stoelinga, M. (2002) An Introduction to Probabilistic Automata. *Bulletin of the European Association for Theoretical Computer Science,* 78, pp. 176-198.

Timmermans, H., van der Waerden, P., Alves, M., Polak, J., Ellis, S., Harvey, A. S., Kurose, S., Zandee, R. (2003) Spatial context and the complexity of daily travel patterns: an international comparison. *Journal of Transport Geography*, 11 (1), pp. 37–46.

Timmermans, H. (2005) *Progress in activity-based analysis*. Elsevier Science Ltd.

Virkar, Y., Clauset, A. (2014) Power-law distributions in binned empirical data. *Annals of Applied Statistics,* 8(1), pp. 89-119. DOI: 10.1214/13-AOAS710