Exploring Population Distribution and Motion Dynamics through Mobile Phone Device Data in Selected Cities – Lessons Learned from the UrbanAPI Project

Jan Peters-Anders, Wolfgang Loibl, Johann Züger, Zaheer Khan, David Ludlow

(Mag. Jan Peters-Anders, AIT-Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna AT, jan.peters-anders@ait.ac.at)
(Dr. Wolfgang Loibl, AIT-Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna AT wolfgang.loibl@ait.ac.at)
(Johann Züger, AIT-Austrian Institute of Technology, Giefinggasse 2, 1210 Vienna AT, johann.zueger@ait.ac.at)
(Dr. Zaheer Khan, UWE – University of the West of England, Fenchay Campus, BS161QY Bristol UK, zaheer2.khan@uwe.ac.uk)
(David Ludlow, MUP, UWE – University of the West of England, Fenchay Campus, BS161QY Bristol UK, david.ludlow@uwe.ac.uk)

1 ABSTRACT

The paper discusses experiences of development and implementation of public motion explorer (PME) tool as part of the EU FP7 project urbanAPI. This tool has been applied to three EU cities with the objective to investigate population distribution dynamics and anonymous population movement patterns within urban environments as an instrument to map shapes of urban attractiveness and accessibility and as a support for transportation and infrastructure planning. The paper describes technical details of the Motion Explorer application by demonstrating the different applications for the City of Vienna, Bologna and Vitoria-Gasteiz and it discusses the results of the first round of the user evaluation using the Criteria Indicators and Metrics methodology. The initial results indicate that the application is intuitive and highly useful for city planning and provides the evidence-based information, which is either expensive or difficult to collect using other approaches.

2 INTRODUCTION

Population distribution is since long detected through census activities. These data are important but provide today insufficient time specific information as population distribution is changing steadily during the day because of the various activities making it necessary to access various places. Those population movements – e.g. for education- , working-, recreation-, visiting- and shopping activities - lead to steadily changing population distribution patterns during the day. To observe these motion patterns, mobile phone location data turn out as a new and appreciated data source which promises a wide range of possibilities exploring these data. (Loibl & Peters-Anders, 2012)

Not only communication device location is a valuable information source but also the examined relations, interrelations and impacts. All information is helpful for urban planning, urban design and infrastructure layout. Visualizing those dynamics provides information which allows supporting urban planning, transportation infrastructure improvement etc.. This article provides an overview of approaches for data analysis to explore motion patterns. The pattern detection refers to geographic coordinates and time stamps extracted from mobile communication device location data.

3 PRINCIPLES OF MOBILE DEVICE DATA FOR POPULATION DISTIRBUTOIN AND MOTION ANALYSYS

3.1 Data representativeness

As Michalopoulou et al. (2010) have proven the spatial relationship between mobile device activities and population distribution, the mobile device volume can be taken as proxy data, in order to spatially describe population distribution-, activity - or motion - patterns. The mobile phone market penetration (the share of mobile devices related to adult and teenage population) can be observed in the EC member states between 80 and 130%. (www.mobilethinking.com gives an overview of the subscriber numbers: 2012 6,8 billion subscriptions have been contracted which is a market penetration of 96%. In Austria (which data are used here) 8 million citizens hold around 12 million mobile device subscriptions, resulting in a subscriber/population ratio of 150%. A1, Austria's largest mobile communication service provider, supplied in 2012 around 5,3 million subscribers, resulting in a market share of around 45%. (http://en.wikipedia.org/wiki/List_of_mobile_network_operators_of_Europe#Austria, observed -3/2014)

These numbers proof a satisfactory representativeness of the data as proxy to describe the entire population distribution, letting assume that the mobile device distribution pattern of large companies matches quite well with the population distribution pattern. Nevertheless a certain social bias within the data might still be possible: In the early days of the mobile communication market, people/population groups were attracted by certain companies depending on their tariffs, their image, and special services they provided. Austria's largest telecom company was e.g. known as "business customer provider". But during the last years these biases have been declining because of the various changes of tariffs, services and image of all providers. Today usually all companies provide various tariffs, matching all user needs and budgets. Thus the users observe all competitors and change their subscriptions more or less frequently (which is even easier, since phone numbers can be kept when changing the provider), or use phones from several providers (thus the national mobile subscription volume today often exceeds the national population numbers).

Nevertheless there might be still a (small) social bias within the data of certain providers, which has to be explored and calculated (Loibl & Peters-Anders, 2012).

3.2 Data location information

Mobile communication networks are organized in cells where each cell is equipped with an antenna, mounted on a cell tower, and with a transmitter station, supplying the mobile phone subscribers in the particular cell. (The terms mobile, mobile phone, cell phone and mobile device denominate telecommunication equipment with diverse functionality – mobile device will be used as general term.) The transmitter stations link the mobile devices to the mobile communication provider's network. As the telecom infrastructure must be prepared to handle incoming activities of each customer (talking, texting, e-mailing, web-browsing), the infrastructure has to observe continuously the connection between each mobile device and a respective network cell.

Each provider uses similar solutions for locating a mobile device: location information comes (today mostly) from the cell tower to which the mobile devices are connected to. The positions of the cell towers are taken as location proxies for the subscribers carrying mobile devices, a location accuracy which is -to some extent- sufficient. (Some telecom companies provide a position triangulation considering the neighbouring cell towers, which seems to be more accurate.) Movement of customers causes motion of their mobile devices. The mobile devices frequently send/receive signals to stay connected to the cell towers. If the signal quality declines (due to movement of a device), the mobile phone will be redirected to a neighbouring cell which provides the best signal quality.

The volume of mobile devices connected to a cell tower is restricted to a certain number, because of data volume limitations. Thus, in areas where a larger user number of devices is expected, cell towers are built more densely resulting in smaller cell extents. The cell sizes can vary from a diameter of a few 100 meters in city centres to several kilometers in rural areas, as shown in various cell tower maps (e.g. www.senderkataster.at, www.funksender.ch, www.cellreception.com,) As in urban areas the cells are smaller, the location accuracy is sufficient for detailed spatial activity pattern analysis using these mobile phone location data.

4 MOTION EXPLORATION TOOL DESIGN

4.1 Objectives

The motion exploration tool makes use of these data and it has been finally applied to the following three cities serving as case study areas and mobile device data have been acquired:

- for the Vienna Region where the log file data of all A1 mobile communication service subscribers have been collected during two weeks, containing geographic coordinates and time stamps .
- for Bologna where data for hourly time ranges aggregated for 100x100 m raster cells have been provided by TIM, a major Italian mobile communication service company, for a week and
- for Vitoria-Gasteiz where the geographic coordinates and the time stamps of mobile phone call initiation, have been provided by Telephonica, a major Spanish telecom company, for some days.

Thus for two cities – Vienna and Vitoria-Gasteiz - raw data have been explored, extracted and aggregated to raster cells to match the public motion exploration data interface to be used for tool application, for Bologna only pre-processed data were available.

4.2 Mobile Device Data pre-processing and analysis

Although every day a large amount of mobile device data is produced in nearly all countries and could be used for mobility and population distribution investigations, relatively few studies are carried out making use of these data. The most important reasons for this might be: (i) there is still a mistrust applying these data violating privacy - from the public side, the scientists and the science funding agencies, (ii) only a few mobile communication companies are willing to deliver these data, (iii) extraction of the appropriate content from the log files requires quite some efforts in terms of programming expertise and data storage capacity.

As described above, the Austrian provider A1 delivers raw data which allows a flexible exploration of population distribution pattern changes over time, aggregation of subscriber distribution to individual spatial entities and interaction and motion pattern analysis by aggregating single trips during the day. The applications presented below refer to current work which has been conducted for the Vienna region making use of sample data for a reference week.

The public (population, stakeholders, planners) may benefit from the applications in several ways (Loibl & Peters-Anders, 2012), e.g.:

- improving traffic infrastructure by discovering spatio-temporal commuting patterns,
- upgrading accessibility or attractiveness after detecting and investigating locations occupied by fewer people than other comparable areas
- better adaptation of public transportation time tables on demand by exploring the temporal variation of intensive infrastructure/open space utilization

To deal with such issues logged mobile device data (with its time stamps and geographic coordinates) turn out as a valuable source to explore public motion- and infrastructure adoption patterns. Distinct time and location information of the mobile device activities as well as the customer movements allow for a mapping of the spatio-temporal distribution of the cell phone subscribers and applying them as proxy for time-specific population distribution.

4.3 Mapping mobile phone user distribution and distribution dynamics

The easiest way to map mobile phone user distribution is to aggregate all customers observed at one location within a certain time range. These locations are either the cell towers or triangulated positions based on cell tower locations and signal quality. Therefore the data must be sorted by time stamp, coordinate pairs and so-called IMSI-codes (International Mobile Subscriber Identity) and all double counts of IMSI-codes for the observed time range must be erased. Finally the complex IMSI-codes have to be replaced by a serial number. Figure 1 depicts such mobile phone user totals aggregated to the monitored locations in the Vienna City centre during an early morning hour. Larger circles indicate more phone users linked to the particular cell tower. Small circles inside large ones indicate local sub-cells (e.g. within buildings).

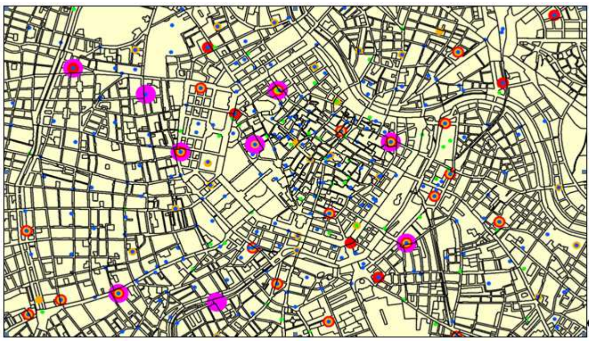

Figure 1: Mobile phone user totals, linked to the nearest A1 network antenna in the Vienna centre – single morning pattern 2009
Source: Data A1, processing: AIOT - Austrian Institute of Technology GmbH.

The numbers of mobile device users by cell do not deliver a density pattern. Generating density patterns requires aggregating user numbers within certain areas (e.g. districts, network cells, grid cells). The smaller the network cells and the smaller the analysis entities (census districts, traffic cells) the higher is the spatial resolution for visualizing the distribution pattern.

ISBN: 978-3-9503110-6-8 (CD-ROM); ISBN: 978-3-9503110-7-5 (Print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI

Such a spatial aggregation of mobile phone location data by time-slices has been conducted applying Vienna data for different time steps to depict the diurnal population distribution dynamics. Figure 2 presents so called "heat maps" showing the population distribution of mobile device subscribers (here the distribution at 10:45h and 13:00h, November 19th, 2011).

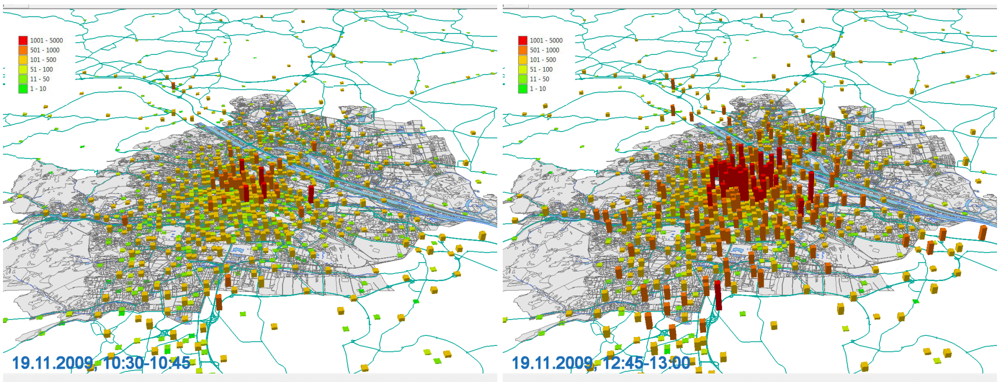

Fig. 2: mobile device subscriber distribution in Vienna related to 500m raster cells. Source: Data A1, processing and development: AIT - Austrian Institute of Technology GmbH (Loibl & Peters-Anders, 2012)

The data allow examining not only distribution of population within space but also the comparison of the variation of space occupation over time (Figure 3).

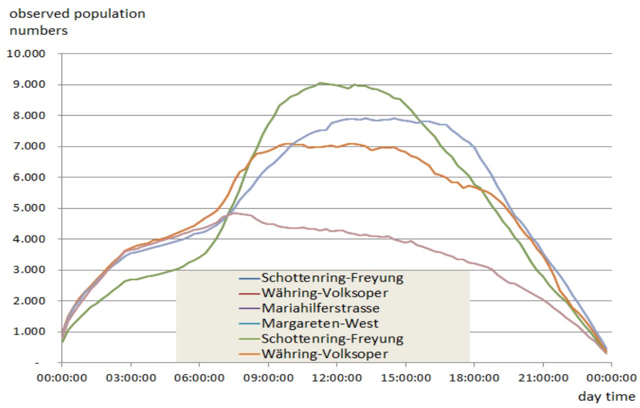

Figure 3: Visitor occupation characteristics during the day in selected places (1x1 km2 cells) within Vienna based on mobile phone log data. Source: Data A1, processing and development: AIT - Austrian Institute of Technology GmbH

	Destinations										
Origins	D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	totals O-D
O-1	x	10	4		14		10		12		50
O-2	4	x	33		44		4	15	39		139
O-3			x								
O-4				x							
O-5					x						
O-6						x					
O-7							x				
O-8								x			
O-9									x		
O-10										x	
Totals											x

Figure 4: Structure of a static O-D matrix

4.4 Motion exploration

Heat maps as shown above (Figure 2) present only cell occupation density pattern for time steps. Exploring motion requires data on interaction – on movement from one cell to another. The matrix below (Figure 4) shows the organisation of such interaction data – as an origin-destination (O-D) matrix – each matrix-cell contains the number of travelers moving from one origin to a destination cell.

But this matrix still provides static information: the general motion pattern of persons within a defined time range. To extract data on motion dynamics trip chains of person groups for origin cells during a day are collected by aggregating origin-destination interaction totals for regular time steps – e.g. 15 minute intervals, hours or a set of hours, depending on the required temporal resolution.

The required structure for examining motion dynamics is: for each origin cell rows per time step are provided, where the destination cells contain the traveler totals, targeting destinations. The tabular structure has to be repeated for each origin cell. Thus movements of the inhabitants of each origin cell can be visualized through animated maps. As a map can only show the movements starting from a single cell this origin cell has to be selected first. The map pairs below depict – as an example - the destination cells of moving inhabitant groups from two origin cells to show the local differences in movement behaviour (Figure 5).

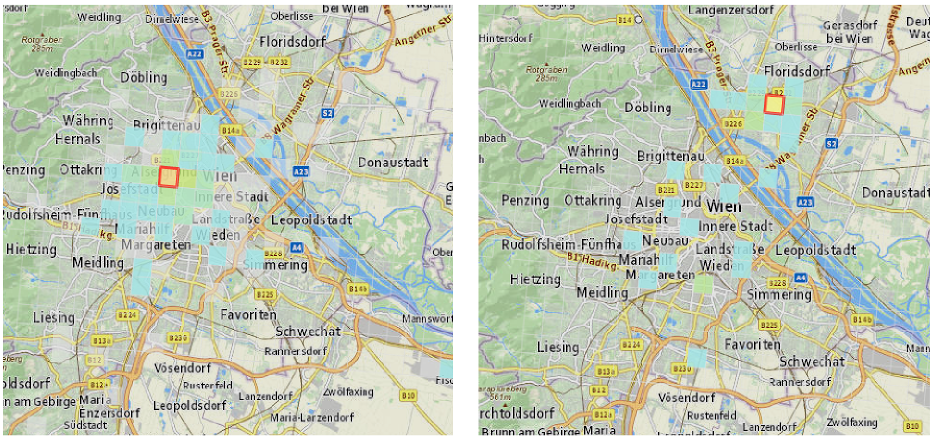

Figure 5: Comparison of target traffic from 2 different source cells (left versus right) – entries over time. Source: Data A1, processing and development: AIT - Austrian Institute of Technology GmbH

4.5 Tool development

A Web based tool has been developed providing 3 applications:

- Application 1 Mapping of diurnal population distribution patterns,
- Application 2 Mapping of motion dynamics for interactively selected cells (targets or origins), and
- Application 3 Depiction of the diurnal visitor occupation of an interactively selected cell.

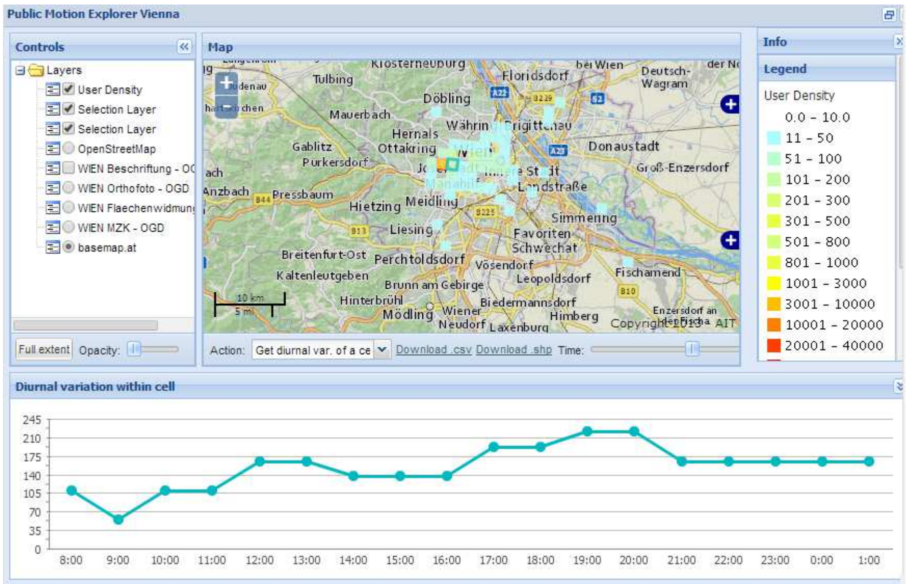

Figure 6: Mapping of motion dynamics for interactively selected cells (targets or origins), and presentation of the diurnal visitor occupation of an interactively selected cell. (Application 2 and 3). Source: Data A1, processing and development: AIT - Austrian Institute of Technology GmbH

5 PRELIMINARY EVALUATION RESULTS

Using CIM methodology (Khan Z, et al 2013a; Khan Z, et al 2013b) an evaluation design process was carried out that resulted in detailed evaluation criteria, sub-criteria (derived from ISO 25010 characteristics) and their respective assessment indicators and weighted questions. The objective of the interim evaluation was to assess the usability, functionality, benefits and relevance of urbanAPI applications against user requirements and pre-defined criteria. The overall evaluation participation rate was very promising as 16 expert users from three case study cities and stakeholder board, with different roles and expertise i.e. urban planners, policy makers, GIS experts, IT experts and others, participated for the interim evaluation of PME applications. Due to space limitations we briefly present usability and benefits results.

The PME application is accessible via a web-based graphical user interface that shows visual maps and provides various interactive features. Most of the evaluation results indicate that the PME application is an effective source of information to identify population distribution and mobility patterns across the city that can potentially be used for urban and transport planning and policy making. PME application allows end users to interact with the city map and to explore population distribution and mobility patterns. It also allows showing dynamically changing space occupation for various zones of the selected city.

One of the major limitations identified is the lack of appropriate quality and granularity of GSM data (with necessary details) acquired from different mobile service providers. This data does not provide sufficient details to fulfil all necessary requirements of end users e.g. examining social biases, transportation mode, position accuracy, live/active and dead mobile device connections etc. Nevertheless, the cross-comparison of features supported by available data from different mobile communication providers indicate strengths of the application as well as identify additional data elements needed to support similar features for other cities. This indicates that application itself has the potential to contribute significantly in gaining insights of space usage and mobility analysis. However, the overall population distribution and mobility patterns represent an approximation but still are useful to gain this kind of information which otherwise is not available or too expensive to acquire at city scale. Further, it provides sufficient information to initiate new planning projects, decision making and policy making. Most of the evaluators found application intuitive and easy to understand and use but also indicated that resolution of maps can be enhanced and intuitiveness can be further improved e.g. by providing context sensitive help. In addition, it is recommended that suitable means should be adopted to build capacity of end users via sufficient training/guidance depending on the skill sets and IT background of end users.

6 CONCLUSION AND OUTLOOK

The presented examples give some insights into the exploration possibilities to be carried out by using mobile phone data sets. The evaluation results indicate benefits and potential of the application but also identify need for quality and granularity of data elements needed to support a variety of planning needs and user requirements. For the future faster visualisation and better elaborated tools can be expected which allow the delivery of web-based dynamic maps on demand, to explore single time slices and in a more far future real time data on the fly - given that there will be a closer collaboration with the mobile network providers to do this. Concerning 3D visualisation one of the biggest challenges will be to find appropriate representations of the data in the 3D scenario as well as appropriate interaction mechanisms. Furthermore, suitable pre-processing steps of the data towards these representations must be explored and implemented. Beyond that it would be interesting to link the (anonymised) motion data with additional data like age, sex or occupation to gain deeper insights into the behaviour and habits of the moving population of a city.

7 REFERENCES

IMSI, International Mobile Subscriber Identity - Assignment and Management Guidelines and Procedures. V.12.0. Telcordia Technologies, Inc. Piscataway, NJ, 2010

MICHALOPOULOU, M., Riihijärvi, J. & Mähönen, P. Studying the Relationships between Spatial Structures of Wireless Networks and Population Densities. RWTH Aachen University, Institute for Networked Systems, 2010

KHAN, Z., Ludlow, D., Rix, J. (2013a) Applying the Criteria Indicators and Metrics Evaluation Method on ICT Research: The HUMBOLDT Project, Research Evaluation Journal, Vol. 22, No. 1, 2013, p. 30-40.

KHAN, Z., Ludlow, D., Caceres, C. (2013b) Evaluating a collaborative IT based research and development project", Evaluation and Program Planning (2013). DOI: 10.1016/j.evalprogplan.2013.04.004, Vol. 40, p. 27-41

LOIBL, W., Peters-Anders, J. (2012) Mobile phone data as source to discover spatial activity and motion patterns, in: Jekel/Car/Strobl/Griesebner (Eds.), GI_Forum 2012: Geovisualization, Society and Learning. p. 524-533. Wichmann Verlag, Berlin & Offenbach.