Mobile phone data as source to discover spatial activity and motion patterns

(As the review process is double blind, please make sure that your submission does not include the author’s name) special focus: • Spatial information & society

Abstract

This paper discusses the application of mobile phone data for exploring the time-dependent population distribution and motion patterns, by detecting the dynamics within these changing distributions. Such pattern explorations make use of the mobile phone location information as proxy for personal location information. Activity and motion exploration of mobile phone data, achieving a rather high spatial resolution, requires a dense mobile phone network infrastructure. As urban environments are supplied with a more dense network of mobile communication transmitter stations then rural ones, urban areas are better suited for such an analysis.
The contribution describes the difficulties of mobile device data extraction, the uncertainties occurring within the data sets, gives an overview of the current possibilities to make use of these data for urban planning and mobility strategies as well as for social space studies, and finally provides an short outlook about analysis and visualisation possibilities which might occur in the future.

1 Introduction

Population distribution is since long detected through census activities. These data are important but provides today insufficient information on population distribution – as this is changing frequently during the day, because of the various activities making necessary to access various places. One simple example is the day and night population distribution as effect of different housing- and workplace locations.
But further population movements - education- , recreation-, tourist- and shopping activities, leading to steadily changing population distribution patterns during the day - could not be observed in detail, at least not for the entire population of an area or a city.
As census data give only a coarse image of the population distribution in a static way, mobile phone location data turn out as new and appreciated data source which promises a wide range of possibilities exploring these data.

Today, wireless mobile communication devices are creating new dimensions of interconnectedness between people, places, and urban infrastructure. The wish for ubiquitous accessibility let the population subscribe mobile device services which allow exploring and visualizing human activity patterns and collective motion traces (Ratti et al., 2006b). Those visualizations expose a wide range of urban dynamics: information and communication networks, movement patterns of people and layout of transportation systems, spatial and social usage of public space and neighborhoods. Not only communication device location is valuable information source but also the examined interrelations. Visualizing those dynamics provides information which may supports urban planning, traffic network improvement, public infrastructure adoption, emergency and disaster management and even studies dealing with social disparities and cohesion.

The contribution gives an overview about available data, data mining methods and approaches for data analysis, to encourage potential users and data providers to make use of these hidden information sources.

2 Mobile phone data exploration – a short application history

As the mobile communication technology is rather new and the mobile phone market penetration only accelerated during the last decade, research dealing with mobile phone location pattern is rather novel and thus only little literature is available. The first well known exploitations have been carried out by Massachusetts Institute of Technology (MIT) delivering quite fancy maps and animations depicting urban motion dynamics, merging science and art, to be presented on exhibitions and fairs like the Venice Biennale 2006 (*senseable.mit.edu/realtimerome/)* or the Graz Cultural City of Europe Exhibition 2005-2006 (*senseable.mit.edu/graz/)* (Ratti et al., 2005, 2006b). The figures are based on real world observations, but the results cannot be explored in detail as the maps frequently don't show legends. The emphasis was put on the way of presenting the information with different visual software tools to raise awareness from the audience. Fig. 1 presents one of those images as an example.

Fig. 1: Mobile device distribution in Graz, Austria – average afternoon distribution pattern (Source: MIT, SENSEable City Lab, 2006, http://*senseable.mit.edu/graz/)*

Around 2009 and 2010 a few articles were published in scientific journals or conference proceedings, presenting further tests, and initial exploitations (e.g.: Jinxing et al, 2009, Michalopoulou et al, 2010, Ratsameethammawong et al, 2010).

A first "serious" application has been carried out in 2010 supporting the disaster management efforts after the earth quake in Haiti: Scientists explored Haiti's cellular network to assess earth quake impact and mitigation action in Haiti. The rough exercise delivers quite detailed numbers and proofs that about 600.000 Haitians have left the nation's capital, Port au Prince (PaP), in the first three weeks after the earth quake (Bengtsson et al., 2011).

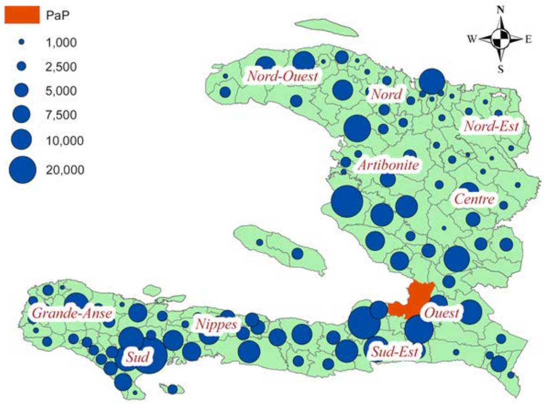

Fig. 2: Post earthquake population distribution after leaving Port au Prince (PaP) in Haiti, Jan 31, 2010. (Source: Bengtsson et al., 2011; doi:10.1371/journal.pmed.1001083.g001)

Today these data are still scarcely applied maybe because of several reasons: (i) there is still a mistrust applying these data violating privacy - from the public side, the scientists and the science funding agencies – although the data are fully anonymized , and therefore – (ii) only a few mobile phone companies are willing to deliver these "suspicious data", to avoid a privacy violation discussion, (iii) only few scientists are willing to take over exhausting, preparation steps, going through billions of logged records to make these datasets usable.

3 Mobile Device Data – Content and Data Processing

3.1 Data representativeness

As Michalopoulou et al. (2010) have verified the spatial relationship between mobile device activities and population distribution, the mobile device volume can be taken as proxy data which, in order to spatially describe population distribution or activity patterns.
The mobile phone market penetration (share of mobile device subscribers related to adult and teenage population) can be observed in the EC member states between 80 and 130% (http://www.mobileisgood.com gives an overview of the numbers from 2006 and 2007). Numbers for 2011 show still an increase, as the following examples show: In Austria the 8 million citizens hold around 12 million mobile device subscriptions, resulting in a subscriber/population ratio of 150% in Bulgaria, where the mobile device boom started little later, the share is nearly the same (147%) A1, Austria's largest mobile communication service provider, supplied in 2011 around 5 million subscribers, resulting in a market share of around 42%. MTEL, as Bulgaria's market leader, covers even 80% of the market (personal information by A1 and MTEL).

These numbers proof a sufficient representativeness of the available data which let assume, that the subscriber distribution provides a similar pattern as the entire population distribution, although a certain social bias has to be observed.

3.2 Data location information

Mobile communication networks are organized in cells where each cell contains a transmitter station with an antenna which supplies a certain area by providing mobile communication services to users in this particular area. The transmitter stations link the mobile devices of the users with the communication network of the respective provider. Mobile devices continuously send and receive signals - data parcels containing digitized speech or data. If the signal quality declines due to movement of a device, the mobile unit searches for neighbouring cells with a better signal in the area and connects to the cell location with the best signal quality. (Ratsameethammawong et al., 2010).

Mobile device location analysis makes use of the network's cell locations to estimate the mobile phone positions and movements. Areas with high population or visitor density require smaller cells, as the demand for services (phone calls, text messages, and internet connections) is expected to be higher and must not exceed the maximum data volume that can be transferred by one transmitter station due to limitations in radio transfer technology. Thus the cell sizes of mobile device networks vary from a diameter of a few 100 meters in urban areas up to several kilometres in rural areas (Ratti et al., 2005). As urban environments are supplied with a more dense network resulting in smaller cells to supply a larger number of people within the area, than in rural landscapes, urban areas are better suited for applying mobile phone data for higher resolution spatial activity analysis.

The technical solutions to locate persons with the help of mobile device connection data are quite homogeneous. The basic location information of the mobile device users comes from the communication network stations to which the mobile devices are connected. Some communication service providers also store sub-cell location information, estimating the relative position of the users, related to the antenna location, making use of radio beam direction or transmission time or similar. In the future it is expected (and e.g. announced by Austria's A1 mobile communication company) to store the smart phone locations, as far as they are active, making use of their GPS capabilities, which are applied for location based services like route navigation or accessibility of points of interest.

3.3 Data storage and processing

Handling communication digitally requires the transfer of signals between the phones of the phone callers, the antennas to which they are connected, the transmitter stations and further via long range transmission through a company network, to transmitter stations, to antennas and finally of the receiving phones. As all mobile communication service providers must be prepared to handle phone activities they must somehow observe the mobile phone locations to identify which devices start/receive and stop a phone call, send or receive a text message or an e-mail, undertake a web-browsing session or whatever. The movements of the mobile devices from one cell to another - the so-called 'handover" - must also be recognized, even if the mobile devices are not active, to "inform" the transceiver stations about potential users (senders or receivers of messages) who have entered or left the area, covered by the respective antenna, to be prepared for connections between the mobile devices and the transmitter stations. Although not all communication service providers store this handover information. If such a log entry is missing, location change as motion information comes with some time delay, only when a mobile phone activity takes place. If a phone is just "travelling" with its owner during the day without being used, this phone does not serve as proxy information about people movement, if the service provider does not store this information.

But in the end all service providers must have some kind of data monitoring and some kind of data logging for technical, for accounting, as well as for security reasons. The kind of data storage depends on the mobile service provider and the information he wants to log. Time and location information of these log entries are the only data required from the communication service providers for this kind of spatial activity analysis. Usually the service providers deliver all logged information in a standard, compressed file format and the investigating teams, interested in population distribution and motion pattern analysis, have to uncompress the content and to extract the necessary longitude/latitude coordinates and time stamp information from the sequential records by their own.

3.4 Data privacy issues

“Raw data”, as logged by the mobile communication service providers, are strictly anonymous. Not the phone number is stored but a so called IMSI attach code (IMSI=International Mobile Subscriber Identity), an encrypted code, containing a random number instead of the mobile phone number and further codes providing information about the current service provider, the home country service provider (as roaming partner), the network area, the network cell, or similar technical details. The encryption key, including the random user number, is renewed on a daily basis which even hinders long time observations of anonymous single entities. Further details can be found in IMSI (2010). All companies store user related by making use of the IMSI code, to hinder personal identification of individual movement patters for unauthorized persons (c.f.: http://www.a1.net/privat/a1traffictechnologie). As no personal data are stored and the IMSI codes are encrypted, the privacy of the subscribers is fully secured.

4 Mobile phone data analysis results

Data of mobile phone locations allow to record time-specific population distribution information by mapping mobile phone subscriber volumes linked to certain antennas of the mobile communication network. Mobile phones are always linked to the nearest mobile phone antenna. User actions like phoning, texting, and internet access as well as user motion (provoking handovers between cell-phone antennas) trigger “events” that are recognized by the mobile communication system. Distinct time and location information of these events allow mapping the spatio-temporal distribution of the cell phone subscribers and applying them as proxy for time-specific population distribution.

There exist different ways of activity pattern and dynamics analysis, which will be shown below. The authors are currently working on applications within the urbanAPI project, funded within the EU FP7 program for several European cities: Vienna/AT, Sofia/BG, Bologna/IT and Vitoria-Gasteiz/ES. The applications presented below refer to current preparatory work which has been conducted for the Vienna application.

4.1 Mapping mobile phone user distribution and distribution dynamics

The easiest way to map mobile phone user distribution is to summarize all users, linked to the nearest mobile communication network antenna, for each requested time step, as carried out by MIT. Fig. 3 depicts such mobile phone user totals linked to the next antenna for the Vienna City centre during an early morning hour. Larger circles indicate more phone users linked to the certain antenna. Small circles inside large ones are effect of local sub-cells (e.g. within buildings). Ratti et al. (2005, 2006a,b) generated MIT mobile phone user distribution surfaces by interpolating the mobile phone user totals, as linked to the antennas.

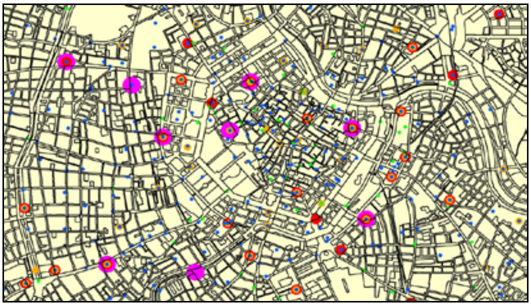

Fig. 3: Mobile phone user totals, linked to the nearest A1 network antenna in the Vienna centre – single morning pattern 2009 (Source: data A1, processing: AIT - Austrian Institute of Technology GmbH)

Figure 3 uses A1 data from 2009. Today the phones' location information is provided in a more detailed way (antenna cell sections or even individual GPS coordinates).

The following figure 4 shows again mobile phone user distributions in the Vienna Centre for a noon hour in 2009 and 2011 but not summarized for antennas. The red dots show the user positions in 2009 linked to antenna coordinates, the green dots show user positions in 2011 linked to network sub-cell sections. The green dots show a more even mobile phone user distribution, resulting in more position points, which allows a more accurate population density estimation.

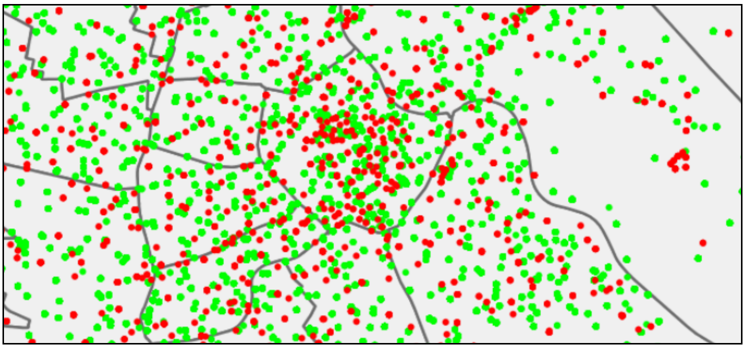

Fig. 4: Mobile phone user totals in the Vienna centre districts. Red dots: user positions 2009 linked to the A1 network antennas, Green dots: user positions linked to network sub-cell sections (Source: data A1, processing: AIT - Austrian Institute of Technology GmbH)

The numbers of mobile phone users by cell antenna or sub-cell centre- or individual phone positions do not deliver a density pattern. Generating pattern information requires summarizing user numbers within a certain area (e.g. districts, traffic cells, grid cells, etc.). The pattern can be depicted by taking any kind of spatial entity to which the mobile phone cells can be related. The smaller the mobile phone cells and the smaller the analysis entities (census districts, traffic cells, raster cells etc) the better is the location accuracy.

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 5 below shows the changes of the population distribution during the morning hours of an autumn day 2009 in the Vienna region as example for 2 time slices. The results allow examining not only effects of commuting to workplaces in the Vienna area, but also the effects of trading, shopping, visiting, etc., resulting in different population distribution patterns during the day, taking advantage of different transport infrastructure and thus causing different traffic loads along certain roads and public transport lines during certain hours.

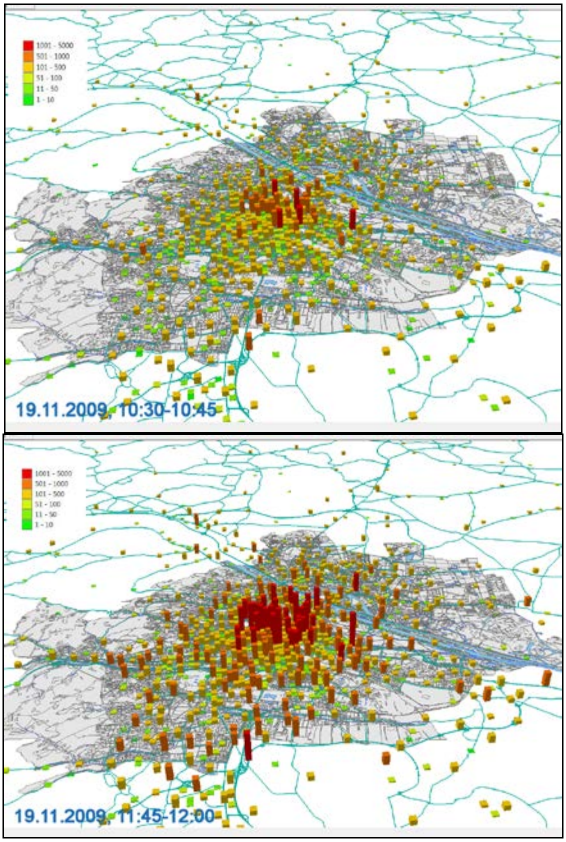

Fig. 5: Population distribution dynamics in the Vienna Region during morning hours (Nov 19, 2009) shown through A1 mobile phone user distribution patterns (Source: data A1, processing: AIT - Austrian Institute of Technology GmbH)

This information helps on the one hand to improve the transportation infrastructure in terms of e.g. public transport line intervals, or the efficiency of public transport interchange infrastructure. Other planning issues referring to time-specific and short term population distributions like taking advantage of open spaces for recreation purposes during weekends or for controlling large open air events can be also matter of consideration.

4.2 Exploring motion patterns and social spaces

Spatial motion patterns can be observed by observing the changing locations of single mobile phones over time. Movement tracks of anonymous users can be depicted by drawing poly-line sections with coordinate pair sequences by IMSI code.
To deliver such information not for all users, but for a selected area certain pre-processing steps are necessary. After sorting the log entries by user, timestamps and coordinates, the records can be treated as poly-line sections applied within a GIS to draw relation maps.
The following figure 6 depicts commuter interaction intensities within the Vienna Region by drawing the regional motion patterns of the A1 mobile phone users.

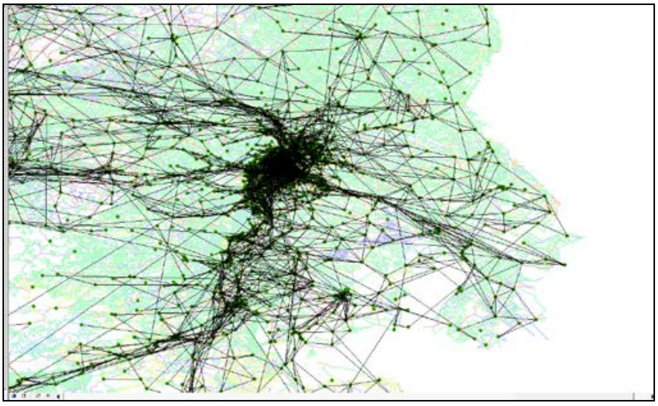

Fig. 6: A1 mobile phone user motion intensity in the Vienna Region as observed in November 2009. (Source: data A1, processing: AIT – Austrian Institute of Technology GmbH)

For detailed analysis the data may be still not accurate enough, Even in this small scale map, “leaping relations” – long lines passing large distances – can be observed, due to missing handover information when phones are not active. This issue is currently matter of investigation.

But such a rough relation pattern might be sufficient, to explore the spatial activity extent of certain social or ethnic groups in urban environments, as far as they live somehow concentrated in certain quarters: Selecting mobile phone users living in these quarters allows visualizing their motion patterns and identifying the extent of their social spaces where they are acting. The results can either proof manifest segregation or proof ongoing transitions towards mixed habitats. Opposite trends like beginning gentrification in certain quarters can be also identified by changing motion patters of mobile phone users living in traditional upper class quarters.

4.3 Exploring motion behaviour, travel speed and travel mode

Trip-chain analysis, travel speed- and travel mode analysis (car- versus public transport) requires to explore single user behaviour, which is a further issue that can be conducted by applying mobile phone user data.
An appropriate way for addressing single users is to transfer the raw logged mobile phone data into a database with the IMSI code, the coordinates and the related time stamp. After doing so, data base query statements allow extracting either single users which are active within a certain area and a certain time range. The queries still take hours as the number of user events per day (based on 5 million users) results in a billion records filling the data base.
The coordinate pairs between the trip origin and the (final or intermediate) trip destinations allow estimating trip lengths between the time steps. The calculations are the basis to carry out path time diagrams with the origin - destination distance along the vertical axe and the "arrival time" for the destinations as hours per day along the horizontal axe. The coordinates allow estimating the (Euclidean) distances between the home location as trip origin and the time specific trip destination location. Upward slopes show trips away from home (e.g. towards the workplace), the length of line sections indicate the duration of the travels, horizontal lines indicate duration of stays. Downward slopes show trips directing home. The steepness of the line sections' slopes indicate the travel speeds which allow assumptions about the travel mode (steeper slopes indicate higher speed through car usage).

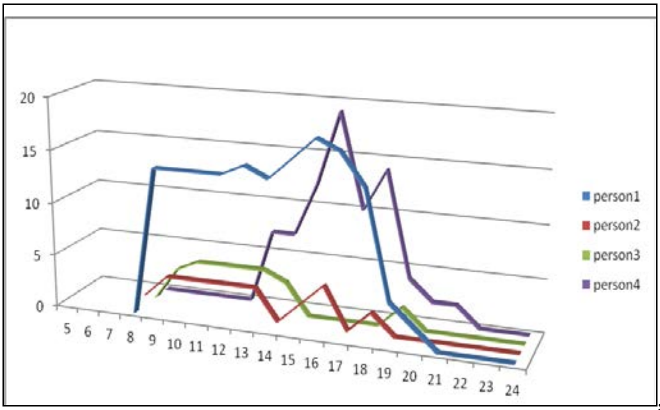

Fig. 7: Exemplary single user motion behaviour (Source: fiction data, processing: AIT - Austrian Institute of Technology GmbH)

Such a trip chain analysis allows rough assumptions about the trip purpose and is the basis to group trips of similar shape indicating similar motion behavior requiring similar traffic infrastructure regarding e.g. time schedule for the respective origin-destination relations.

5 Conclusion and Outlook

The current analysis examples give some insights into the exploration possibilities which may be carried out by using such data sets. For the future faster visualisation 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 to visualize real time data on the fly.

Location accuracy might increase in the near future as the Smartphone market accelerates, which allows frequent GPS-based positioning providing more and more frequent individual user coordinates. More frequent usage of mobile devices for social network communication, for web-browsing and making use of location based services would deliver additional events, which may reduce leap motion patterns and increase real route pattern

Thus applications for planning and transport control purposes are expected to increase and non-scientific and non-technical applications might occur in the future too.

References

AHAS, R. & ÜLAR, M. (2005), Location based services—new challenges for planning and public administration?, Futures, 37:6 547-561

BENGTSSON, L., Lu, X., THORSON, A., GARFIELD, R. & von SCHREEB., J., (2011), Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. PLoS Medicine – on- line. (http://irevolution.files.wordpress.com/2011/10/journal-pmed-mobile-phone-haiti.pdf)

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

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

JINXING, H., WENJING, C, LUO, J. and XIAOMIN, Y. (2009), Dynamic modeling of urban population travel behavior based on data fusion of mobile phone positioning data and FCD. In 17th International Conference on Geoinformatics. IEEE.

ATSAMEETHAMMAWONG, P & KULTHON KASEMSAN, M.L (2010), Mobile Phone Location Tracking by the Combination of GPS, Wi-Fi and Cell Location Technology. Communications of the IBIMA. Vol. 2010 (2010), Article ID 566928, 7 pages. DOI: 10.5171/2010.566928 http://www.ibimapublishing.com/journals/CIBIMA/cibima.html

RATTI, C., PULSELLI, R.M., WILLIAMS, S. & FRENCHMAN, D. (2006a), Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis. Environment and Planning B, 33, 5, pp. 727-748.

RATTI, C. et al. (2006b), Real Time Rome - Venice Biennale 2006 Contribution. http://senseable.mit.edu/realtimerome/

RATTI, C., SEVTSUK, A., HUANG, S. and PAILER, R. (2005), Mobile Landscapes: Graz in Real Time. Proceedings of the 3rd Symposium on LBS & TeleCartography, Vienna, Austria, 28-30 November 2005. http://senseable.mit.edu/papers/pdf/RattiSevtsukHuangPailer2005LBSVienna.pdf