Mobile Phone Data as Source to Discover Spatial Activity and Motion Patterns

Wolfgang LOIBL and Jan PETERS-ANDERS

Abstract

This article discusses the application of mobile phone data for exploring population distribution and motion patterns and detecting the dynamics of the distributions. The location information comes from the communication network stations – the cell towers – to which the mobile devices are connected. The kind and accuracy of the location information may vary, depending on the data storage principles of the mobile communication companies.

A short history of mobile device data application for geo-location introduces the research field. Issues dealing with mobile phone data – data representativeness, location accuracy, and data processing as well as privacy concerns – are discussed. Further some recent applications making use of these data for urban planning as well as transportation infrastructure improvement are presented. The conclusion finally provides an outlook about the expected future progress in data processing and analysis.

1 Introduction

Population distribution is since long detected through census activities. These data are important but provide today insufficient information on population distribution – as the distribution is changing frequently during the day, because of the various activities making it necessary to access various places. One simple example is the day and night population distribution as effect of different housing- and workplace locations. So, mobile phone location data turn out as a new data source which promises a wide range of possibilities.

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 help 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 a valuable information source but also the examined relations, interrelations and impacts. Visualizing those dynamics provides information which may support urban planning, infrastructure layout, traffic network improvement, public infrastructure adoption, emergency and disaster management. The article may encourage other researchers making use of this formerly hidden information source.

Jekel, T., Car, A., Strobl, J. & Griesebner, G. (Eds.) (2012): GI_Forum 2012: Geovizualisation, Society and Learning. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-521-8.

2 Mobile Phone Data Exploration – A Short Application History

As the mobile communication technology is a common feature since a decade and the mobile phone market penetration accelerated only since a few years, research dealing with mobile phone location pattern is rather novel, but the number of publications dealing with this issue is steadily increasing. Early applications in the 1990ies and early 2000s explore location tracking indoor, using WIFI-, infrared – or other radio signals as described by RATTI et al. (2006a). One of the first city wide outdoor applications, also mentioned by RATTI, was the "Amsterdam Realtime Project" (www.waag.org/realtime) in 2002, where personal digital assistants (PDAs) with GPS receivers were used to track motion patterns of a sample of test persons in Amsterdam, who were equipped with such devices.

The first well known exploitations of standard mobile phone data, making use of cellular networks for geo-positioning, have been carried out 2004 by the Massachusetts Institute of Technology (MIT) in cooperation with an Italian telecom provider, using data covering the city of Milan to explore the capability of mobile phone data as "contribution to urban understanding" resulting in the mobile landscapes project (RATTI, 2006a). Then several studies where carried out by the MIT-research group, delivering quite fancy maps and animations, depicting urban motion dynamics, 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).

Later the MIT research group continues conducted applications for different cities with different project partners, extracting urban population distribution numbers or visitor flows using mobile phone data as "digital footprints" (e.g. GIRARDIN et al. 2008). Around 2008 to 2010 a growing number of articles were published by a growing number of authors dealing with population distribution tracing, mobility profiles, traffic planning and even social behaviour (GONZALES et al. 2008, MOHAN 2008, JINXING et al. 2009, MICHALOPOULOU et al. 2010, STOICA et al. 2010). An important application on disaster management was conducted in 2010: Haiti's cellular network traffic was observed, assessing the 2010 earth quake impact and rescue efforts. The fast and rough exercise delivers quite detailed numbers and proofs that about 600.000 Haitians were able to leave Haiti's destroyed capital, Port au Prince, in the first three weeks after the earth quake (BENGSSON et al. 2011).

Although every day a large amount of mobile device data is generated in nearly all countries, which could be used for mobility and population distribution investigations, relatively few scientists make use of these possibilities. Major reasons might be the mistrust applying these data violating privacy – from the public side, the scientists and the science funding agencies, and little efforts of the service providers to organize their data in an appropriate way which allow easy handling and analysis.

3 Mobile Device Data – Content and Data Processing

3.1 Data representativeness

As MICHALOPOULOU et al. (2010) have proven a spatial relationship between mobile device activities and population distribution, the mobile device volume can be taken as proxy 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 estimated in the EC member states today up to 150 %. (www.mobileisgood.com gives an overview of the subscriber numbers for 2006 and 2007). Exemplary numbers for 2011 show still a further increase: in Austria 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%. M-tel, Bulgaria's market leader, covers even 80% of the market (personal information from A1 and M-tel).

These numbers proof a sufficient representativeness of the data as proxy to describe the entire population distribution, letting assume that the mobile device distribution pattern of the 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, the companies have attracted certain population groups as customers, depending on the tariffs, the image, the signal coverage or special services. But during the last years these biases have been declining as today usually all companies provide various tariffs and products, matching all user needs and budgets, show similar signal quality and coverage. Nevertheless there might be still a (small) social bias within the data of certain providers, which has to be explored and considered when applying the results for planning purposes (see chapter 3).

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 customers in the particular cell. The transmitter stations link the mobile devices with the provider's communication network. (The terms mobile, mobile phone, cell phone and mobile device denominate telecommunication equipment with similar functionality – mobile device will be used as general term.) As the telecom infrastructure must be prepared to handle coming activities from each customer (talking, texting, e-mailing, web-browsing), the technical 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: position information comes (today mostly) from the cell towers to which the mobile devices are connected. 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 deliver positions for cell sections which are based on antenna direction information or on position triangulation making use of signal quality differences of neighbouring cells towers.) The mobile devices frequently send/receive signals to stay connected to cell towers. Usually mobile devices are connected to the nearest cell tower, with best signal quality. If the signal quality declines due to movement of a device, the device will be redirected to a neighbouring cell providing better signal quality, which triggers a location update. The volume of mobile devices connected to a cell tower is restricted to a certain number, because of technical limitations on data transmission. Thus, in areas where more users are expected, cell towers are built more densely making smaller cells. Cell sizes may vary from a diameter of a few 100 meters in cities to several kilometres in rural areas, as shown in various cell tower maps (e.g. www.senderkataster.at, www.funksender.ch). As in urban areas the cells are smaller, the location accuracy is better than in rural landscapes, so urban areas are better suited for detailed spatial activity pattern analysis using mobile phone data.

In some cases the mobile devices stay not connected to the nearest cell tower: if too many mobile devices try to connect to a certain cell, the devices which cannot be served, are redirected to a neighbouring cell and a location update is reported which let assume the customer has moved to a different place. Cell connections may also oscillate between cell towers if customers stay at places with similar signal quality from more than one cell. Such "pseudo movements" of customers must be explored and considered.

3.3 Location accuracy, location information storage and delivery

As all mobile telecom service providers require customer activity monitoring for technical, for accounting and recently for security reasons, they log time –, location – and further activity-information for each device. A new log file entry is done, when a device state changes, because of personal activities (phoning, texting, moving, etc.) and technical reasons (being connected to a new cell, update cell information).

It is important to be aware that location and time stamp information is stored only for active mobile devices (of users talking, texting, e-mailing, and web-browsing). Locations of mobiles in "standby"-mode are just observed by the technical infrastructure, but not regularly logged. The location of devices in "standby"-mode are logged only when a device moves over greater distances, leaving a large "telecom service district" (controlling a set of smaller cells) and entering a different one. In this case the logged coordinates show the location of the cell tower, when approaching the new telecom service district or the coordinates of the service district centre. But finally all mobile devices, in working or standby mode, are logged at least once a day.

In the near future mobile devices will be expected to be used more often for non-oral communication – consuming location based services, e-mailing, web-browsing and for relations via social media – which will lead to an increasing number of mobile device actions and thus more log-file entries. Further, a growing volume of smart phones will improve location accuracy by delivering exact GPS based position. Today several non-telecom companies have started collecting "open signals" from active smart phone users, who (unwillingly?) publish their individual location by using location based services, where the coordinates result in (vague) "signal heat maps" (e.g. www.opensignalmaps.com).

The kind of data storage and of data delivery depends on the mobile service provider. Some providers deliver all logged technical information in an individual file format and the investigating teams have to extract the necessary geographic coordinates and time stamp information from the sequential log entries by their own. Other providers deliver only aggregated data – e.g. numbers of active users per network cell or per raster cell for pre-defined time steps. In this case no motion information can be extracted from the data and only user distribution or "user density" patterns can be depicted. Common motion patterns can be extracted if the location and time stamp information is provided for single users, which allows identifying location changes over time resulting in motion information.

As far as we know is the Austrian telecom provider A1 currently the only one – at least in Europe – who delivers such data as a commercial product to business customers and research organizations (www.a1net.business/a1traffic). Currently there are only few data customers which are able or willing to use those data, as data handling is complicated due to data format and data volume (approx.100 GB raw data daily for 1 mill. customers).

3.4 Data privacy issues

Data privacy refers to the relationship between technological possibilities and the legal right of privacy in collection and sharing data from individuals. There are more demands to be considered than not to extract data from single persons or merge those data with other content. There are complex questions of who controls the data about people (individually and collectively) and how the data is used. What rights do users have to preserve their privacy? What are the rights of data owners to exploit information about users without their permission? (http://digitalenterprise.org/privacy/privacy.html)

So – mobile data privacy is a severe issue. If providers deliver data where single mobile device positions could be extracted, the ownership of the devices is kept hidden and thus the individuals' privacy is respected. Even the companies' internal log files do not store information from individual subscribers, but so called IMSI attach codes (IMSI=International Mobile Subscriber Identity), encrypted codes, containing a random user number and further information for accounting and technical reasons. The encryption key is renewed daily. Further details can be found in IMSI (2010). Since no individual data are stored, the privacy of the subscribers is secured, but there is still a discussion that privacy can be overruled by tracing single routes. Nevertheless the researchers must be careful when working with those data and should accept some self-defined ethical guidelines – e. g.:

- data of single individuals must not be depicted,
- collective behavior pattern data must avoid identifying individual behaviour,
- location information must be fuzzyfied to hinder identifying accurate positions

All data mining and analysis efforts making use of data which rests to some extent on individual activities – what population distribution and motion pattern do, although they do not harm privacy issues – should produce public benefit.

4 Mobile Device Data Analysis – Spatial Application Examples

The authors are currently working on data mining and exploration methods making use of mobile device data covering some European cities: Vienna, Sofia, Bologna and Vitoria-Gasteiz. Data providers from Italy and Bulgaria deliver only aggregated data related to network or grid cells. The Austrian provider A1 delivers raw data which allows more detailed analysis. The analysis examples presented below refer to initial work which has been conducted for the Vienna application making use of sample data of some days.

The public (population, stakeholders and planners) may benefit in several ways – e.g.:

- improving traffic infrastructure after discovering spatio-temporal commuting patterns,
- upgrading accessibility or attractiveness after detecting and investigating locations, occupied by less 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 data (with its time stamps and geographic coordinates) can be a valuable source to explore common public motion and infrastructure adoption patterns.

4.1 Mapping mobile phone user distribution and distribution dynamics

The easiest way to map mobile phone user distribution is, to aggregate all customers with identical coordinates in a certain time range. These locations are either the cell towers or somehow triangulated positions based on cell tower locations and signal quality. Therefore the data must be sorted by time stamp, coordinate pairs and IMSI-codes and all double counts of IMSI-codes for the observed time range must be erased. Finally the IMSI-codes are erased.

Fig. 1 depicts such mobile phone user totals aggregated to the monitored locations in the Vienna city centre during an early morning hour. Bigger circles indicate more phone users linked to a cell tower. Small circles inside large ones are effect of local sub-cells (e.g. within buildings). The figure is based on 2009 data. Today mobile phone location information is provided with better detail, related to cell sections or even GPS coordinates.

Fig. 1: Mobile phone customers, aggregated to cell tower positions in the Vienna centre – early morning pattern, November 2009, Circle size indicates customer volume (Source: data: A1, processing: AIT – Austrian Institute of Technology GmbH)

The volumes of mobile phone users by cell or cell-section do not deliver a density pattern. Generating patterns requires relating user numbers to areas (e.g. districts, network cells, grid cells). Patterns can be depicted by taking any kind of spatial entity. The smaller the network cells and the smaller the analysis entities (census districts, traffic cells) the finer is the distribution pattern. To derive density patterns requires to aggregate mobile phone user numbers to certain areas. Figure 2 below depicts mobile phone user densities in Vienna related to 250 m grid cells for 2 morning hour time-slices as a part of a map series to examine the diurnal population distribution dynamics. The results allow examining not only effects of commuting to workplaces, but also effects of shopping, visiting etc.

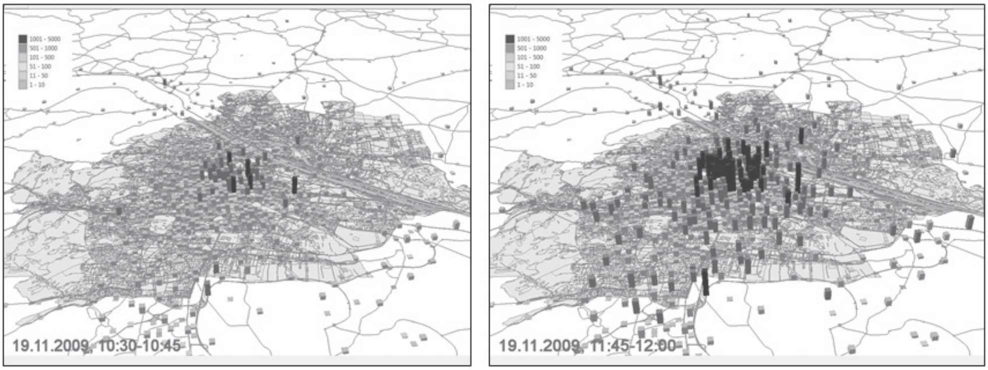

Fig. 2: Mobile phone user density dynamics in the Vienna Region during morning hours of a sample day 2009 (Source: data: A1, processing: AIT – Austrian Institute of Technology GmbH)

The maps show still mobile phone user distributions. To extrapolate the user volume to gain total population numbers requires examining the mobile user shares in the total population. A possible social bias must be explored with the help of a region-wise comparison of the provider's mobile device number with the region's total residential population number. If the regional shares are not evenly distributed, a regional bias is obvious, which has to be further quantified through statistics. Linear regression models can be applied to extrapolate the population totals.

4.2 Exploring motion or spatial relation patterns

To observe motion patterns in a certain area, the mobile device data set shall at least cover the commuter catchment as daily interaction extent). Spatial motion patterns can be detected by observing the changing locations of single mobile devices over time. Therefor the data must be sorted by IMSI-code, by time stamp, and by location. How double counts of identical locations and time stamps are treated, depends on the research question. To explore interaction extents allows erasing all log entries with identical position or identical time stamp of one IMSI code. To discover duration of stays at certain locations (e.g. to quantify occupation duration of infrastructure) requires to keep double counts. Identifying positioning errors requires a section-wise examination of the travel speeds. Unrealistic high travel speeds (> 100km/h for air-line distances) let expect location oscillation.

Relation patterns can be easily depicted by extracting the "home position" and the position of the farthest destination of each mobile user and drawing lines between the users' origin and destination. Aggregating trip origins and destinations to grid cells showing a position range instead of a single position will hide single trips and avoid too detailed motion information. Gridding of origin/destination (O-D) target areas requires generating O-D matrices, which finally allow quantifying common interaction patterns to calibrate traffic models or to identify temporal or spatial bottlenecks in transportation services for certain O-D relations.

The following map pair (Fig. 3) depicts interaction patterns of inhabitants (the A1 mobile users) in two municipalities: Lembach, a remote village in the northern border region of Upper Austria, and Purkersdorf, a little town bordering Vienna. The Lembach interaction spider map shows quite a number of local relations demanding local public transport service and a majority of trips in south-east direction – targeting the Upper Austrian central region with Linz as province capital. The Purkersdorf interaction pattern depicts some trip concentration targeting the western and central districts in Vienna but also various local relations targeting smaller centres in the rural neighbourhood, with a more remote focus on Tulln in the Northwest.

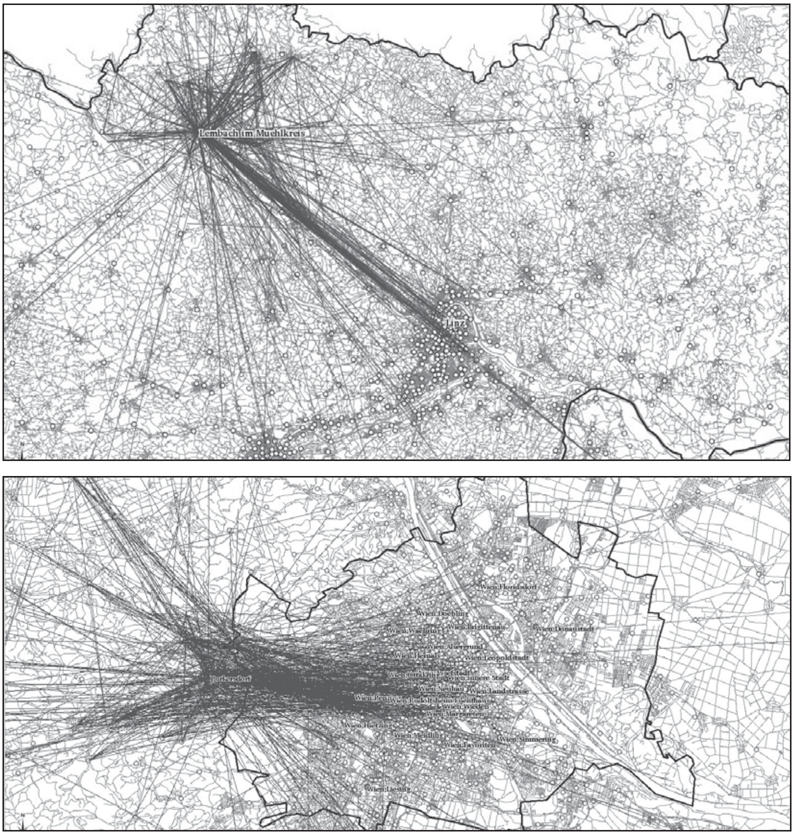

Fig. 3: Interaction patterns of mobile phone users originating in Lembach (top) and Purkersdorf (bottom) as observed on a sample day 2011 / same map scale (Source: data: A1, processing: AIT – Austrian Institute of Technology GmbH)

Of interest are also the distances of the daily trips in different regions to learn about common local mobility behaviour and local transportation demands. The two diagrams below (Fig. 4) show the differences in travel distances between the inhabitants (the A1 mobile phone users) from Purkersdorf (near the center) and (remote) Lembach to the various destinations (trips within an 80km circle). While 80% of the considered trips from Purkersdorf inhabitants are in a range below 20km distance from home, 50% of the Lembach trips are within 20 km, but the next 30% range till 40 km, stressing the higher travel demand in remote areas supplied with less public transportation services.

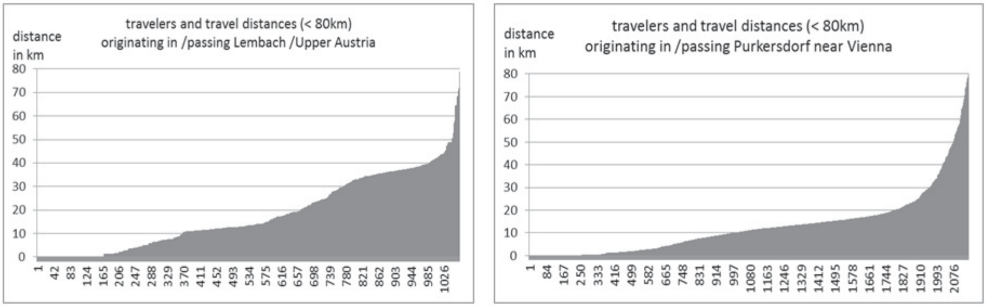

Fig. 4: Travel distance distribution of mobile phone users originating in Lembach (left) and Purkersdorf (right) as observed on a sample day in September 2011 (Travel distances limited to 80 km) (Source: data A1, processing: AIT – Austrian Institute of Technology GmbH)

Such investigations require some processing time, but can be easily carried out for all areas to allow detailed insights about local mobility habits and demands on transportation services. More detailed explorations may consider different time-slices identifying at which time the demands occur.

Again it has to be stressed, that the results are based on the common behaviour of the customers of one telecom provider. Extracting local interaction patterns which are representative for the entire population requires again extrapolating the phone subscriber numbers observed at certain areas. In this case not a general extrapolation ratio can be used but regionally different ratios which are able considering the social structure leading to different interregional interaction pattern.

5 Conclusion and Outlook

The presented examples give some insights into the exploration possibilities which can be carried out by using such mobile phone data sets. For the future faster processing and visualisation tools can be expected which allow the delivery of web-based dynamic maps on demand, to explore time slices on the fly. Day-time specific interaction matrices will therefore quantify diurnal origin-destination relation dynamics helping to better adapt transportation infrastructure and public service schedules to the demand.

Location accuracy might increase in the near future as the smartphone market volume accelerates, which allows frequent GPS-based positioning proving more individual user coordinates. More frequent usage of mobile devices for social network communication, for web-browsing and usage of location based services will generate additional location reporting which may reduce location errors and may finally allow monitoring traffic flow in real time in an entire transportation network.

Acknowledgement

Research presented her was partly carried out within the project "urbanAPI" (Interactive Analysis, Simulation and Visualisation Tools for Urban Agile Policy Implementation", starting in October 2011, funded by the 7th Framework Program of the European Union, call identifier: FP7-ICT-2011-7, under the grant agreement no: 288577.

References

BAYIR, M. A., DEMIRBASA, M. & EAGLE, N. (2010), Mobility profiler: A framework for discovering mobility profiles of cell phone users. Pervasive and Mobile Computing, 6 (2010), 435-454.

BENGTSON, 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.

GIRARDIN, F., CALABRESE, F., FIORE, F. D., RATTI, C. & BLAT, J. (2008), Digital footprinting: Uncovering tourists with user-generated content. Pervasive Comp., 7 (4), 36-43.

GONZALEZ, M., HIDALGO, C. & BARABASI, A. (2008), Understanding individual human mobility patterns. Nature, 453(7196), 779-782.

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

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.

MOHAN, P., PADMANABHAN, V. N. & RAMJEE, R. N. (2008), Rich monitoring of road and traffic conditions using mobile smartphones. SenSys (2008), 323-336.

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), 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. & 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.

STOICA, A., SMOREDA, Z., PRIEUR, C. & GUILLAUME, J.-L. (2010), Age, Gender and Communication Networks. Workshop on the Analysis of Mobile Phone Networks, Satellite workshop to NetSci 2010.