HAL
open science

Big Data Analytics – Geolocation from the Perspective of Mobile Network Operator

Antonin Pavlicek, Petr Doucek, Richard Novák, Vlasta Strizova

To cite this version:

Antonin Pavlicek, Petr Doucek, Richard Novák, Vlasta Strizova. Big Data Analytics – Geolocation from the Perspective of Mobile Network Operator. 11th International Conference on Research and Practical Issues of Enterprise Information Systems (CONFENIS), Oct 2017, Shanghai, China. pp.119-131, 10.1007/978-3-319-94845-4_11 . hal-01888632

HAL Id: hal-01888632

https://inria.hal.science/hal-01888632v1

Submitted on 5 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Big data analytics – geolocation from the perspective of mobile network operator

Antonin Pavlicek[1[0000-0002-1230-5982]], Petr Doucek[2], Richard Novák[3], Vlasta Strizova[4]

[1, 2, 3, 4] University of Economics, Prague, Czech Republic
{antonin.pavlicek, doucek, strizova}@vse.cz
[3] T-Mobile Czech Republic, Prague, Czech Republic
richard.novak@t-mobile.cz

Abstract. The article demonstrates the possibilities of big data analysis of geolocation data of mobile network operator in the Czech Republic. Covers theoretical background of geolocation and then presents case studies conducted in the last four years: National Park visitors' distribution analysis, mountain ski resort usage, use of mobility data for the preparation of city territorial and development plan and use of mobility data for efficient tourism management at Vaclav Havel Airport.

Keywords: big data, geolocation, mobile network operator, human mobility, tourism

1 Introduction

Modern smart phones have become ubiquitous communications tools—now used not only for phone calls and text messages but also for accessing the Internet, taking pictures and recording videos with integrated camera, navigating with GPS or watching videos and playing games. The proliferation of mobile phones amongst general population is immense. The percentage of active mobile subscriptions within the population reached 96% in 2014 [3]. In developed countries, the number of mobile subscribers has surpassed the total population, with a penetration rate now reaching 121%, whereas, in developing countries, it surpassed 85% and keeps growing.

Analyzing the spatiotemporal distribution of phones geolocated to the base transmitting towers (BTS) may serve as a great tool for population monitoring. With data being collected by mobile network providers, the prospect of being able to map changing human population movement and distributions over relatively short intervals (while preserving the anonymity of individual mobile users) paves the way for new applications and a near real-time understanding of patterns and processes in human geography. [2]

1.1 Mobile phone location technology

Mobile network operator (MNO) must be aware of the geographic location of each mobile phone in the network in order to be able to route calls to and from them and to seamlessly transfer a phone conversation from one base station to a closer one as the user is moving. This originally technical necessity was transformed into a commercial opportunity to increase the Average Revenue Per User (ARPU), through what is now known as 'Location Based Services' (LBS). LBS are all services that use the location information of a mobile device to provide a user with location-aware applications and services. Such location information can be provided by the mobile network operator, the mobile phone device, or a combination of both, but this article focuses on the former.

LBS applications initially proposed were very broad, creative and raised quite a lot of expectation. For example, users were offered the possibility to make requests of the type of 'where is the nearest…?' (hospital, gas station, bank, restaurant, etc.), identify friends that walk nearby (Foursquare), ask for navigation instructions when lost (Google maps), locate lost phone (device locator), or receive a promotion from a familiar store when walking past it (location based 'spam'). [4] Nevertheless LBS failed to deliver its promises at the turn of the century, and its huge forecasted market potential did not come to reality [9]. This is partly because early services have been very restricted due to the poor location accuracy available, and the limited capabilities of both the handheld hardware (screen size and quality, processing power and storage capacity) and the network data transfer speeds and bandwidth[1] [4]. However, the second wave of geolocation services comes right now, and this time it seems that it is here to stay.

Mobile networks compose of cells around a base transmitting tower (BTS). Each active mobile phone, therefore, can be located by triangulating the geographic coordinates of its BTS. This network-based positioning method is simple to implement, phone and user independent and its accuracy depends directly upon the network structure; the higher the density of towers, the higher the precision of the mobile communication geolocalization [4].

Records of the time and associated cell of anonymous mobile phone users are valuable indicators of human presence and offer a promising alternative data source for increasing the spatial and temporal detail of large-scale population datasets.[2] Mobile phone geolocation can be therefore used to:

- observe human mobility patterns at the individual level (police and security services only),
- monitor movements and activities of selected population using aggregated data,
- improve responses to disasters and conflicts,[7]
- plan epidemics elimination strategies,
- explore traffic flows and prevent traffic jams,
- study intensity of human activities at different times,
- identify seasonality in both domestic and foreign tourist numbers and destinations.

Legislation in USA and EU also requires mobile network operators to provide an accurate location for calls to emergency services.

2 Geolocation in T-Mobile Czech Republic

T-Mobile is the largest Czech mobile network operator, which is in regular contact with about six million terminals (40% market share) with an aggregate data rate of hundreds of millions of signal records generated daily. T-Mobile had decided to take full advantage of the big data and geolocation potential and over the last three years has developed a series of unique solutions that add value to the customer and provide a competitive edge for the company. In this paper (chapter 3) we present the sample of the most interesting solutions. But first, let's look into some definitions.

Data anonymization

Every geolocation project starts with anonymization. The legislation of the Czech Republic and EU stipulates that it is always necessary to make data anonymization before data processing, thus preventing the identification of individual end-users. T-Mobile uses sophisticated encryption algorithms to remove identification and uses aggregated data for processing, so only meta data arise in the calculations, which are the only ones used to interpret the results later.

Technological background

The source of T-Mobile's geomobile data is residual signaling data from mobile cell identification, which makes possible to know the approximate location of the mobile terminal and thus the distribution of the population in space and time. Further refinement of the position can be calculated if needed. Signaling data arises from typical mobile events such as call, data transmission, SMS message, terminal transfer between individual transmitters, or upon report call to the infrastructure in the so-called periodic specification when the terminal is periodically called for a signaling response. Data from signaling (after anonymization) can be stored in the data warehouse for further processing using classic business intelligence tools or special IT tools supporting large data. [5]

2.1 Continuous online monitoring system

The current distribution of mobile devices can be mapped through residual signaling data. Random but quite representative pattern of Czech population's mobility can be recalculated in real time into aggregated geodemographic matrix of mobility. Based on both global and local system calibration (according to control check-points) they are recalculated to represent the real number of persons in each area. Specialized software allows displaying the distribution of the population in nearly real time, as well as a historic time lapse sequences.

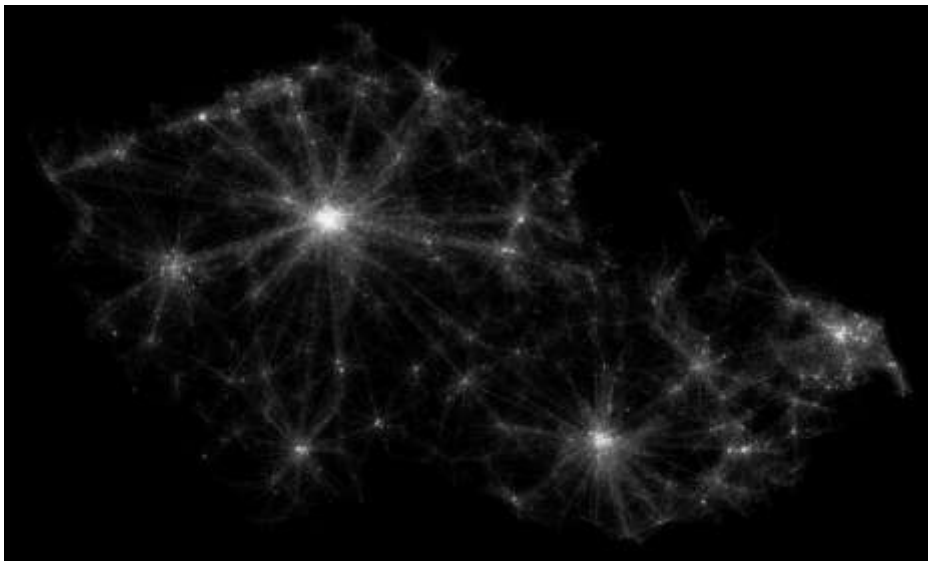

Fig. 1. Online monitoring visualization – movement of population in the Czech Republic

Fig. 2. Example of online monitoring visualization – detail

2.2 Business intelligence and big data tasks

Typical business intelligence and big data processing tasks that need to be handled when working with anonymized data exported from signaling to a data warehouse are as follows:

- Keep, search and archive records of terminals in a given area.
- Position these terminals in required geographic formats such as centroid, square, cadastre, or any given polygon.

- Deal with signal skipping between neighboring BTS.
- Deal with the problems near international border areas (roaming).
- Store the number of people using the mobile phone in the given area in a specific time slot, together with time-lapse data.
- Manage algorithms to count unique terminal approaches versus cumulative access to all terminals.
- Identify the origin and destination matrix, which is important for determining the motion vector.
- Compute the whole population to allow other data layers calibration.
- Solve the non-homogeneity of data in some areas.
- Create enhanced models in locations where network topology does not meet the requirements in terms of precision.
- Modal split, that is to distinguish the movement of the population from the point of view of transport, such as public transport (train, boat or bus) or individual transport. [5]

2.3 Fields of application

Mobile geolocation is successfully used in a number of cases, the most common uses being:

- Crisis management (lost children, information on people in the area of fire, floods or chemical threats)[6].
- Detecting population mobility for state infrastructure and urbanization planning (new roads, P+R areas, public transport, land use plans).
- Commercial statistics (number of visitors to shopping centers, outdoor, festivals, tourism and city and area, replacement or supplementation of Czech statistical office research).
- Optimizing traffic flows.
- Location-based services such as a mobile ad for nearby services.

The list of examples would be unlimited with possibilities enriched by other external data (weather, social networks, CRM systems, etc.) taken into consideration.

3 Case studies: big data geolocation analysis

T-Mobile is very active in the big data field and, together with partners from the academic and commercial sectors, is involved in a wide range of research and commercial projects. Our paper will demonstrate above-mentioned possibilities on practical examples of implementation by the biggest mobile network operator in the Czech Republic.

3.1 Pilot case study of Šumava National Park

The objective of this pilot tourism-oriented project was to calculate the number of visitors in the Lipno, Kvilda, Modrava and Horská Kvilda regions at the turn of 2013 and

2014, to find out, where visitors came from, how long they stayed and which places they visited. Such data is useful for the national park administration, municipalities and local businesses - hoteliers, restaurateurs, sports facilities, and other entities. They get to know and possibly target the tourists of the Šumava Mountains. National Park, in cooperation with T-Mobile and KPMG, prepared the long-term geolocation analysis.

And what are the results of the case study? [8] Most foreign visitors arrived from the Netherlands (36%), Germany (35 %), Austria (6 %) and Russia (5 %). In total, 260,000 visitors arrived in Šumava during the monitored period, of which 24% were from abroad.

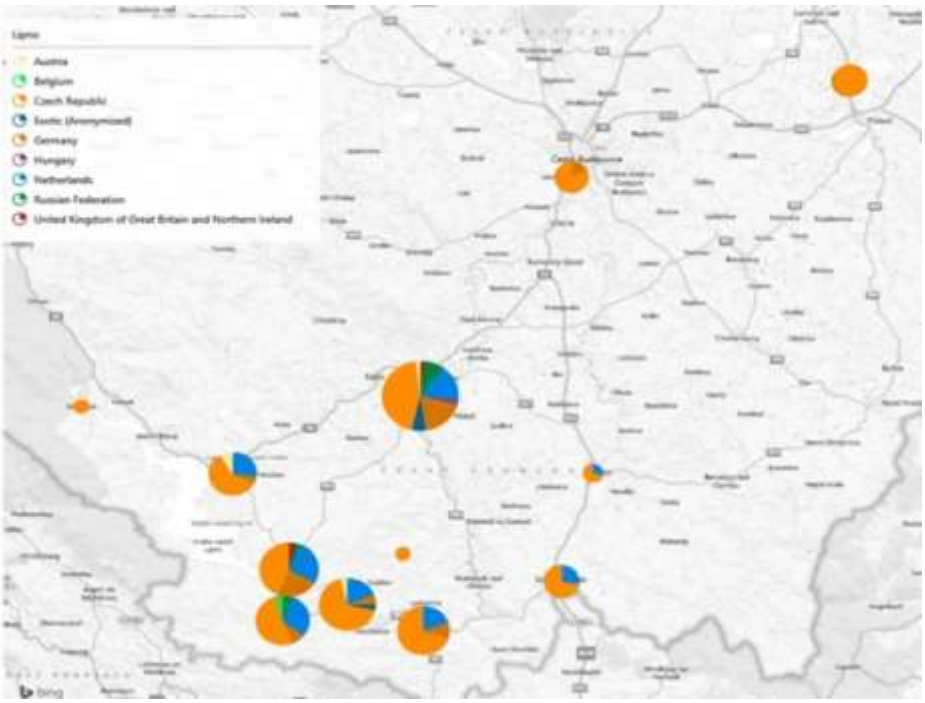

Fig. 3. Distribution of visitors in Sumava National Park (authors, based on data from T-Mobile)

The main destinations for both domestic and foreign tourists are usually near (< 10 km) the place where they spent the night. One day trip was made by more than 1,000 (0,5 %) domestic tourists and 700 (1,2 %) foreign tourists. According to KPMG's analysis, tourists have spent more than 211 million CZK during the two-month survey period of the region, of which about 70% accounted for domestic and 30% for foreign tourists.

3.2 Case study – Czech mountain ski resorts

Based on a very positive response to the pilot study, a very thorough analysis of the behavior of visitors to Czech mountain resorts was conducted in 2015. Six top Czech

and Moravian ski resorts were analyzed: Harrachov (Giant mountains), Pec pod Sněžkou (Giant mountains), Rokytnice nad Jizerou (Giant mountains), Špindlerův Mlýn (Giant mountains), Kohútka (Javorníky), Lipno (Šumava)

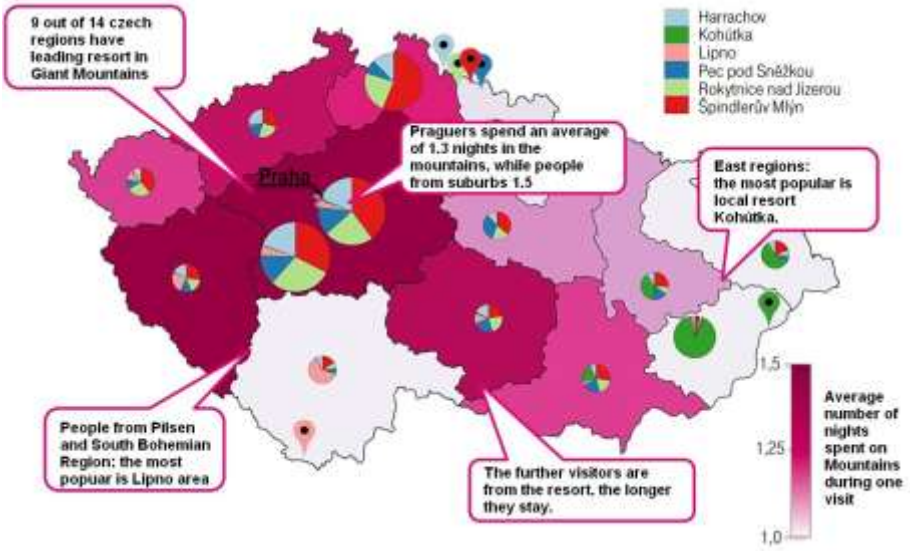

Fig. 4. Czech mountain ski resorts – origin of visitors (authors, based on data from T-Mobile)

The data from the mobile network was only one of the sources - it supplements the information about the profile of the visitors, obtained by the questionnaire survey, and the sample survey of the Czech population.

The results answer some important business related questions: What are the visitors of the mountains doing and what they expect? How many one-day visitors are there and how many tourists sleep in the mountains? Do they differ in behavior? When do the Slovaks begin to travel to the mountains? What services are missing the most numerous visitor groups? Those are valuable information, which would be otherwise hard to get. The most interesting findings are presented in Figures 2 and 3.

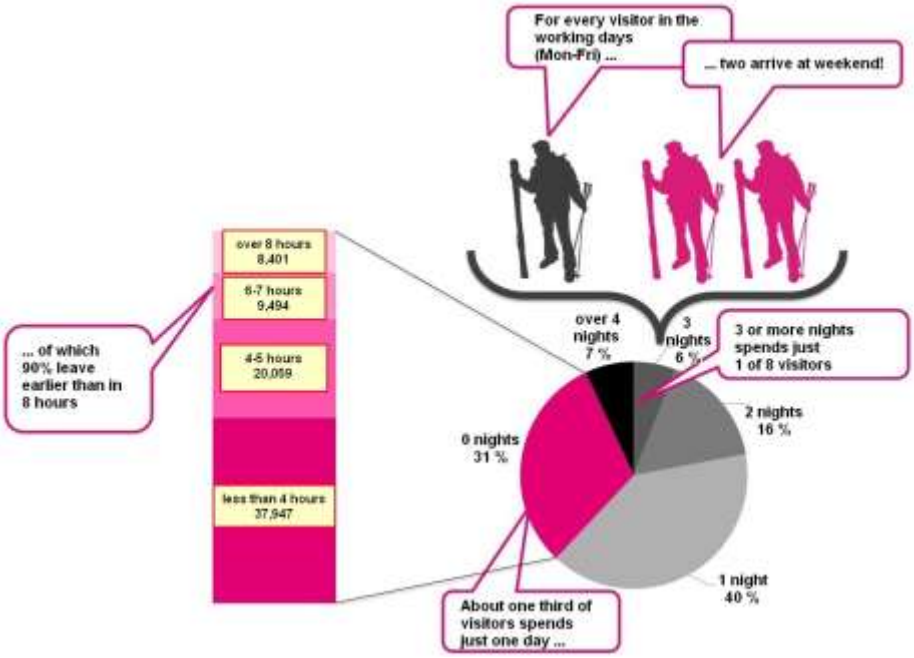

Fig. 5. Czech mountain ski resorts – length of stay (authors, based on data from T-Mobile)

3.3 Use of mobility data for the preparation of city territorial and development plan

The use of geolocation data is not limited to the commercial sector. Public administration also has a number of tasks, where geolocation can useful. The data make it possible to map the mobility of inhabitants in detail and better understand the mobility dynamics. Detailed monitoring of mobility enables to classify territory in terms of public needs and helps with the sustainable development of communities.

Fig. 6. Use of mobility data for city development coordination – density of population day, night, weekend (authors, based on data from T-Mobile)

Data can help quantify in detail mobility links between the city and its wider surroundings and help, for example, to plan accordingly public transport, housing or new schools and amenities (see Fig. 7).

Fig. 7. Average distribution of the inhabitants of Černý Most area (part of Prague) in the day-time (typical weekday). Purple - inhabitants in the place residence Orange - inhabitants traveling outside of his / her home

3.4 Václav Havel Airport

Václav Havel Airport Prague, formerly Prague Ruzyně International Airport (IATA: PRG, ICAO: LKPR) is the international airport in the capital of the Czech Republic. It is, with over 13 million passengers in 2016 (over 15 million expected in 2017), the busiest airport in the newer EU member states. It serves as a hub for Czech Airlines, Travel Service, SmartWings, Wizz Air, and Ryanair. Its 2 runways can handle 71,000 t of cargo and 137,000 aircraft movements per year.

The survey was conducted from July 2016 to May 2017 and recorded 12.2 million passengers - 6.218 million arrivals (of which 1.856 million Czech and 4.361 million foreigners) and 5.987 million departures (of which 2.071 million Czech and 3,915 million Foreigners). A significant seasonal component appeared in the airport's operations.

Days	Percentage
1	12%
2	20%
3	26%

4	15%
5	6%
6	4%
7	4%
8	2%
9	1%
10	1%
11	1%
12	1%
13	1%
14	1%
more	6%

Table 1. Length of stay in the Czech Republic

Although these basic statistical data can be gathered by counting arrivals and departures and aircraft occupancy capacities, the use of mobile geolocation brings additional, enhanced capabilities. It is possible to monitor the movement of passengers in the Czech Republic or the destination of Czechs abroad. We can, for instance, easily detect the length of stay of foreigners in the Czech Republic (see Table 1), or even check, which UNESCO monuments they decided to visit (see Table 2). The analysis can go even deeper – we can identify the day of UNESCO visit or even in what order they have been visited. For each monument, we can calculate its popularity by different nationalities (Russians seem to prefer Cesky Krumlov and Kutna Hora – they constitute 44.1% and 44.8% of foreign visitors there, Americans like to go to Lednice – 40.4%)

UNESCO site	%
Prague castle	85,87%
Cesky Krumlov	7,92%
Kutna Hora	3,98%
Telc	0,36%
Olomouc	0,35%
Tugendhat	0,35%
Holasovice	0,32%
Litomysl	0,30%
Lednice	0,26%
Kromeriz	0,10%

Table 2. UNESCO sites visited by airport passengers

Last but not least, it is possible to analyze also the movement of the visitors in the defined localities. Figure 8 displays the distribution of tourists from Russia, Germany, and Italy during 24 hours in Prague. Data allow measuring quantity and distribution of foreign visitors in the targeted area. Thanks to traffic data, it is possible to identify the potential for further development.

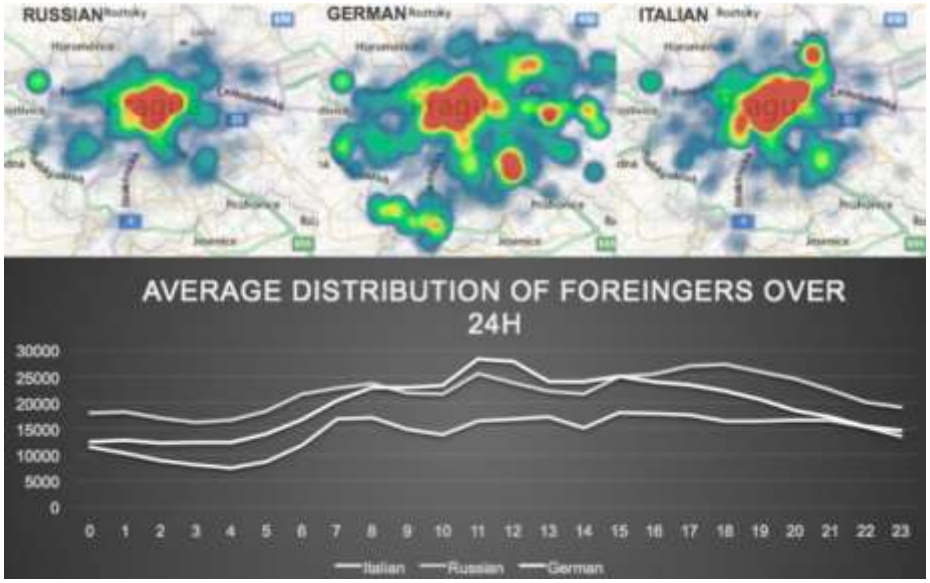

Fig. 8. Distribution of Russian, German and Italian tourists in Prague (authors, based on data from T-Mobile)

Mobile phone geomonitoring allows also determine catchment area of Prague airport – as shown in Fig. 9, it serves mainly central Bohemia region. Mobile monitoring can be (to some extent) used even abroad. Figure 10 pictures destinations of Czech travelers – flying from V. Havel airport – categorized upon the length of their trip.

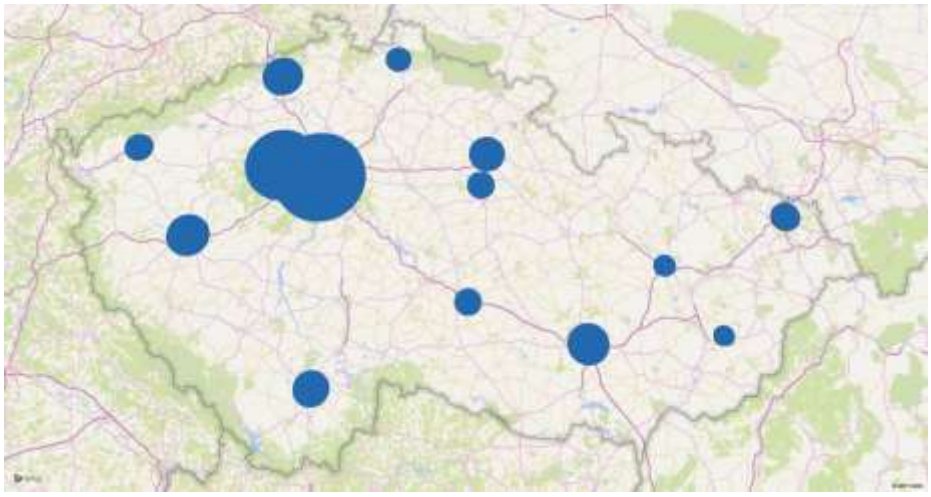

Fig. 9. Origin of Czech travelers – departures from V. Havel airport (authors, based on data from T-Mobile)

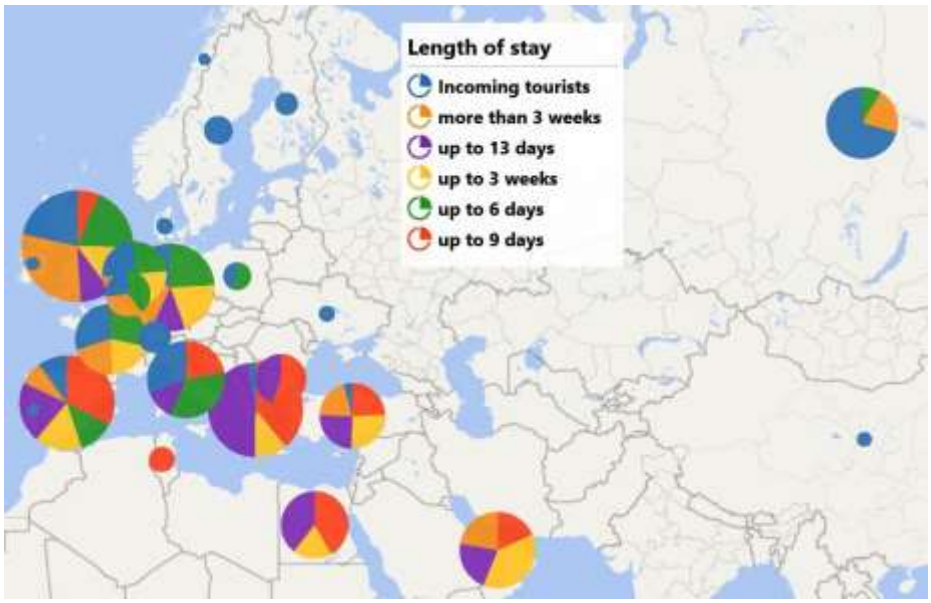

Fig. 10. Destinations of Czech travelers – departures from V. Havel airport and length of the trip (authors, based on data from T-Mobile)

4 Conclusion

This paper presented real-world studies conducted in the Czech Republic that combined mobile device geolocation and big data approach. Despite the relative youth of this field – first studies started less than 5 years ago – it is becoming a standard tool for a near real-time understanding of patterns and processes in human geography. It allows to observe mobility patterns and study intensity of human activities at different times and places. The data thus obtained are very accurate, up-to-date, and can bring previously unpredictable views and facts.

From the perspective of mobile network operators is this analysis also very interesting – the fact that they are the only ones with access to the necessary data puts them into very advantageous position.

In the future, further rapid development and massive expansion of big data geolocation analysis can be expected.

Acknowledgment

Paper was processed with contribution of long term support of scientific work on Faculty of Informatics and Statistics, University of Economics, Prague.

References

1. David Mountain, Jonathan Raper: Positioning techniques for location-based services (LBS): characteristics and limitations of proposed solutions. Aslib Proc. 53, 10, 404–412 (2001).
2. Deville, P. et al.: Dynamic population mapping using mobile phone data. Proc. Natl. Acad. Sci. U. S. A. 111, 45, 15888–15893 (2014).
3. International Telecommunication Union ITU: World Telecommunication Development Conference (WTDC-14): Final Report, http://handle.itu.int/11.1002/pub/809f5219-en.
4. Mateos, P., Fisher, P.F.: Spatiotemporal accuracy in mobile phone location: Assessing the new cellular geography. In: Drummond, J. et al. (eds.) Dynamic and Mobile GIS: Investigating Changes in Space and Time. pp. 188–211 Taylor & Francis (2006).
5. Novak, R., Kovarnik, L.: Big Data a T-Mobile. Computerworld. 2015, 6, (2015).
6. Skrbek, J.: New Possibilities of Information Services for Special Situations. IDIMT-2009 Syst. Hum. Complex Relatsh. 29, 123–130 (2009).
7. Skrbek, J., Kvíz, J.: Critical Areas of Early Warning System. In: Doucek, P. et al. (eds.) IDIMT-2010: Information Technology - Human Values, Innovation and Economy. pp. 193–202 Universitatsverlag Rudolf Trauner, Linz (2010).
8. TTG: Big data: Pomohou cestovnímu ruchu na Šumavě? | TTG - vše o cestovním ruchu, http://www.ttg.cz/big-data-pomohou-cestovnimu-ruchu-na-sumave/, (2014).
9. Zetie, C.: Location Services Find Their Way To The Enterprise, http://www.informationweek.com/location-services-find-their-way-to-the/26100784.