

Activity-specific time profiles from *Foursquare* check-in data: An improved basis for mapping population dynamics?

Christoph Aubrecht ^{1*}, Joachim Ungar ², Sérgio Freire ³

¹ AIT Austrian Institute of Technology, Energy Department; Vienna, Austria (christoph.aubrecht@ait.ac.at)

² EOX IT Services GmbH; Vienna, Austria

³ European Commission - Joint Research Centre, Institute for the Protection and Security of the Citizen; Ispra, Italy

In this short discussion paper we assess the potential of volunteered geographic information (VGI), specifically *Foursquare* check-in data (www.foursquare.com), to provide an improved basis for mapping spatio-temporal population dynamics. The presented study refers to work-in-progress that is carried out in the framework of the EU CRISMA project where mapping dynamic population distribution patterns serves as essential input for human exposure assessments and subsequent modeling efforts, including evacuation planning and casualty estimations (Aubrecht et al., 2014).

Aiming at refining established population models that represent 'static' residential distribution patterns, current existing population dynamics models can be broadly categorized in (1) approaches applying *simplified binary distinction* between *daytime and nighttime* (e.g. Freire & Aubrecht, 2012) and (2) *multiple time-slice approaches* that try to account for the *continuous variation in human activities* occurring in particular during the daytime period (e.g. Leung et al., 2010). While the first category usually mainly refers to commuting and work statistics to identify basic daytime patterns, the second category additionally considers statistics derived from time use surveys (TUS) showing more refined activity patterns and their evolution during the day (that is also the current status of development for the CRISMA *DynaPop* approach). In fact, there is a third category of models that, however, follow an entirely different approach in a sense that *near-real-time distribution patterns* are analyzed using cell phone data logs. Due to inherent privacy constraints and restricted data availability, the latter approach is still not considered feasible for wide-scale implementation.

While activity-specific time profiles as derived from classical time use surveys have been applied in the mentioned multiple time-slice approaches to map population dynamics, it is a more recent development to analyze additional sources representing such human activity patterns. Due to the rapid increase in volume and spatial density of VGI data in recent years, there are now first attempts to use *empirically-derived 'facility occupancy curves' from location based social network (LBSN) data* such as Twitter for characterizing local-scale population dynamics (Bhaduri et al., 2014). The idea of characterizing spatio-temporal patterns of collective user activity is not entirely new, but earlier studies often struggled with limited data densities particularly outside the 'densely monitored regions' that are foremost in the US (Rösling & Liebig, 2013) and consequently issues of representativeness. It therefore proved beneficial to aggregate data tied to venue and thus activity categories (e.g. Noulas et al., 2011) which provides an excellent basis for linking to target zones in population disaggregation models.

In our study we extract venue and corresponding user check-in data for a sample region in the Lisbon metro area (Portugal) via the *Foursquare* API. This is a follow-up to previous work carried out in the same study area (Aubrecht et al., 2011) so that eventually also the expected significant increase in *Foursquare* user activity can be illustrated and analyzed in detail (being launched in 2009, *Foursquare* now - as of May 2014 - reports over 50 million users globally compared to 8 million in 2011). In our current efforts several improvements have been applied in the data extraction approach using *hierarchical hexagonal request grids*. The *Foursquare* API limits requests to 5,000 points per developer account (each request point identifying the respective 30 nearest venues). In performance testing using optimized hexagonal grids, 5,000 requests took less than 30 min to be completed during peak periods. During extended crawling, however, only around 2,000 requests could be successfully executed per hour. Parallel multiple user requesting did not increase the performance. Checking further, it is likely that the *Foursquare* API limits response feedback per IP address, not per developer account. Computing hardware and internet bandwidth do also not seem to have significant influence on the response times. Request frequency is eventually set to one hour, also in order to create a database applicable and comparable to the time use survey based multiple time-slice approaches described earlier on.

In the 2011 study, we extracted hourly data for one week in May for more than 10,000 venues in the study area. In order to guarantee best possible comparability, in 2014 we again extract one week in May. Work days as well as the weekend are thereby covered, which allows deriving differentiated time profiles accounting for variations going beyond the standard working day patterns.

There are several aspects in the *Foursquare* data that are useful for mapping population dynamics. First, relative *venue category-specific time use profiles* (or occupancy curves) could theoretically be applied directly in the course of the disaggregation process to account for spatio-temporal human activity variations. Due to the clearly still existing issue of VGI representativeness (despite the mentioned overall usage increase), another interesting option is to apply the *Foursquare* activity profiles to *calibrate existing survey-based time use statistics*. E.g., one factor that quickly becomes obvious refers to the characterization and representation of ‘dinner time’. While classical survey-based TUS show the relative number of a population sample that has dinner at a certain time, there is no indication on the spatial location. While on working days lunch is commonly taken in restaurants (nearby the work place), many people prefer to have dinner at home. The *Foursquare* venue category-based TUS clearly show that distinction and therefore represent a significantly improved basis for the population disaggregation. As mentioned above, due to the inherent nature of VGI data, only relative numbers of *Foursquare* user activities are considered directly applicable for population dynamics mapping. Therefore it seems most feasible to focus on ‘mobile’ population, thus mainly people that are out for work and study, rather easily identifiable via commuting, work, and education statistics. In that context it is crucial to account for study area-specific characteristics in terms of regular working hours, e.g. in Portugal indicating a 2-h lunch break between around 12 am and 2 pm. That kind of information can also help to *understand and potentially calibrate Foursquare data-inherent temporal shifts and uncertainties* that are due to irregular check-in dynamics (e.g. a user may check in at the lunch place when he is actually already leaving and not check in back at work afterwards). Another promising parameter of the *Foursquare* venue data is the absolute number of check-ins. We are currently testing if those could provide an indication on the *relative ‘importance’ of a venue* within a certain activity category, thus potentially approximating the housing density parameter that commonly serves as proxy for (residential) population density in dasymetric mapping approaches.

Next steps after successful extraction include analysis of the changes in user activity since 2011 as well as eventually the integration of VGI-derived activity curves in the CRISMA *DynaPop* approach and consequently in the production of *prototype dynamic population distribution grids* for Lisbon Metro as well as CRISMA pilot studies in France and Italy. Considering the above aspects, VGI data clearly show a strong potential to increase in importance in the field of population dynamics mapping.

This study is carried out in the framework of the EU CRISMA project (www.crismaproject.eu). CRISMA is funded from the European Community's 7th Framework Programme FP7/2007-2013 under grant agreement no. 284552.


REFERENCES

- Aubrecht, C., Ungar, J., Freire, S. (2011) Exploring the potential of volunteered geographic information for modeling spatio-temporal characteristics of urban population: A case study for Lisbon Metro using foursquare check-in data. In Pinto et al. (eds.), *Proceedings of the 7th Int. Conf. on Virtual Cities and Territories*. Lisbon, 57-60.
- Aubrecht, C., Steinnocher, K., Huber, H. (2014) DynaPop - Population distribution dynamics as basis for social impact evaluation in crisis management. In S. R. Hiltz et al. (eds.), *ISCRAM 2014 Proceedings, 11th Int. Conf. on Information Systems for Crisis Response and Management*. University Park, PA, 319-323.
- Bhaduri, B., Bright, E., Rose, A.N., Cheriyyadat, A. (2014) Occupancy Curves for Characterizing Population Dynamics. *AAG Annual Meeting*. Tampa, FL.
- Freire, S., Aubrecht, C. (2012) Integrating population dynamics into mapping human exposure to seismic hazard. *Natural Hazards and Earth System Sciences*, 12(11), 3533-3543.
- Leung, S., Martin, D., Cockings, S. (2010) Linking UK Public Geospatial Data to Build 24/7 Space-Time Specific Population Surface Models. *GIScience 2010, 6th Int. Conf. on Geographic Information Science*. Zurich.
- Noulas, A., Scellato, S., Mascolo, C., Pontil, M. (2011) An Empirical Study of Geographic User Activity Patterns in Foursquare. *ICWSM-11, 5th Int. AAAI Conference on Weblogs and Social Media*. Barcelona.
- Rösler, R., Liebig, T. (2013) Using Data from Location Based Social Networks for Urban Activity Clustering. In D. Vandenbroucke, B. Bucher, J. Crompvoets (eds.), *Geographic Information Science at the Heart of Europe. Lecture Notes in Geoinformation and Cartography*. Springer, 55-72.

Short CV – Dr. Christoph Aubrecht

April 2014

Personal Data	
Title, Name	Dr. Christoph Aubrecht
Office address	Giefinggasse 6, A-1210 Vienna, AUSTRIA
E-Mail	christoph.aubrecht@ait.ac.at
Office address	1850 I St. NW, Washington DC 20006, USA
E-Mail	caubrecht@worldbank.org
Residency	Sakarya, TURKEY
Date and place of birth	June 4, 1983 in Linz, AUSTRIA
Nationality	Austrian
Marital status	Married
Personal webpage	http://homepage.univie.ac.at/christoph.aubrecht



This short CV is a compilation of the most important milestones in terms of education, work and research experience. A detailed and up-to-date version including a full publication list as well as additional qualifications, a list of memberships to international scientific societies and a complete project management history is available on my personal webpage.



Short Bio

Christoph Aubrecht is a geospatial information scientist having a PhD in integrated GI Science and remote sensing from **Vienna University of Technology** and a prior Master's degree from the **University of Vienna**. He is affiliated as senior scientific consultant with the **AIT Austrian Institute of Technology** as well as with the **World Bank's** LAC-DRM team. He previously provided consultancy to the **Global Facility for Disaster Reduction and Recovery (GFDRR)** and held various visiting scientist positions at the U.S. **National Oceanic and Atmospheric Administration's (NOAA)** National Geophysical Data Center (NGDC), at **Columbia University's** CIESIN and the attached NASA Socioeconomic Data and Applications Center (SEDAC), as well as at the **University of Southern California**. For several years Chris has been adjunct lecturer in GI Science and remote sensing at University of Vienna. He served as scientific secretary of the **ISPRS** Technical Commission II and has been **editorial board member** of various international scientific journals including *Disaster Advances*, *Georisk*, and the *ISPRS International Journal of Geo-Information*. As former deputy editor-in-chief he currently serves as senior advisor for the IEEE and NASA funded online journal **Earthzine**. Christoph is Associate Member of the **International Congress on Disaster Management**.

Education (University Level)

10/2007-11/2013	Dr.rer.nat. (Ph.D. in Natural Sciences) 10/2007-11/2013 Doctoral Program in Natural Sciences; Surveying & Geoinformation at Vienna University of Technology (Integrative courses at the European Space Agency, Vienna University of Natural Resources & Applied Life Sciences, and University of Vienna) Dissertation (Evaluation 'excellent'): <i>A geospatial perspective on population exposure and social vulnerability in disaster risk research - Demonstrating the importance of spatial and temporal scale and thematic context</i>
10/2002-06/2007	Mag.rer.nat. (M.Sc. in Natural Sciences) 10/2004-06/2007 Specialization in Cartography & Geoinformation [on Master level] (Integrative courses at Vienna University of Technology, and Vienna University of Natural Resources & Applied Life Sciences) 10/2002-10/2004 Diploma Program in Geography at University of Vienna Master thesis (Evaluation 'excellent'): <i>Integration of remote sensing and socio-economic information for spatial modeling of landuse - Creation of a functional 3D building model of urban space</i>

Employment Record and Research Experience (Active Positions in Red)	
Since 06/2012	The World Bank, Urban & Disaster Risk Management Unit for Latin America and the Caribbean Region; Washington, DC, USA Since 12/2013 Disaster Risk and GeoInformation Specialist 07/2013-12/2013 Disaster Risk Information Consultant 06/2012-07/2012 Disaster Risk Information Analyst
Since 10/2006	AIT Austrian Institute of Technology (formerly Austrian Research Centers - ARC); Vienna, Austria Since 08/2013 Senior Scientific Consultant 05/2012-07/2013 Scientific Consultant 02/2010-04/2012 Research Associate Level: Scientist 11/2007-01/2010 Freelance Research Associate Level: Junior scientist 05/2007-07/2007 Research Associate Contractor 10/2006-04/2007 Research Associate Master student
Since 10/2003	University of Vienna, Institute for Geography & Regional Research; Vienna, Austria Since 03/2008 Adjunct Lecturer 02/2004-03/2004 Research Assistant 10/2003-07/2007 Teaching Assistant
01/2007-10/2013	Vienna University of Technology, Department of Geodesy & Geoinformation; Vienna, Austria 11/2007-10/2013 Cooperative Doctoral Researcher (scientific collaboration with AIT) 01/2007 Visiting Scientist (within the scope of the Master thesis)
06/2012-09/2012	Global Facility for Disaster Reduction and Recovery, GFDRR Labs; Washington, DC, USA Geoinformation Technology Specialist
08/2010	University of Southern California, Spatial Sciences Institute, GIS Research Laboratory; Los Angeles, CA, USA Guest Researcher
07/2009-09/2009	Columbia University, The Earth Institute, CIESIN – NASA, Socioeconomic Data and Applications Center; Palisades, NY, USA Visiting Scientist
08/2006-10/2009	NOAA National Oceanic & Atmospheric Administration, National Geophysical Data Center, Earth Observation Group; Boulder, CO, USA 09-10/2009, 07/2008, 07-09/2007 Visiting Researcher, Foreign National Research Affiliate 08/2006-09/2006 Intern
03/2005-06/2008	ASFINAG Austrian Freeway Company; Vienna, Austria 03-06/2008, 10-11/2007, 01-06/2007, 03-07/2005 Project Associate
Scientific Record	
Since 2008	Earthzine (IEEE and NASA funded scientific online magazine) Since 2012 Senior Advisor to the Board of Directors 2009-2012 Deputy Editor-in-Chief 2008-2009 Associate Editor for Disasters 2008-2013 Special Issue Guest Editor (Disaster Mitigation & Response, Disaster Management, Wildfires)
2007-2008	ISPRS International Society for Photogrammetry & Remote Sensing Scientific Secretary Technical Commission II
Editorial Board Member for Disaster Advances, Journal of Geography & Regional Planning, Georisk, ISPRS International Journal of Geo-Information Special Issue Guest Editor for Physics and Chemistry of the Earth, Natural Hazards, CMES: Computer Modeling in Engineering & Sciences Book Volume Development Editor for Land Use: Planning, Regulations, & Environment Session and Track Chair and Convener at various international conferences incl. ComplIMAGE, CUÉE, EGU, GEOG-AN-MOD, GIScience, IAEG, IDRC, IDRIM, IGARSS, IMACS, ISCRAM, ISDA, ISG, VCT, VipIMAGE Reviewer for various international journals incl. Advances in Space Research, Computers, Environment and Urban Systems, Disaster Advances, Earthzine, Ecological Modelling, GIScience and Remote Sensing, International Journal of Health Geographics, Journal of Environmental Planning and Management, Journal of Geography and Regional Planning, Natural Hazards, Natural Hazards and Earth System Sciences, Physics and Chemistry of the Earth, PLoS ONE, Population, Space and Place, Population and Environment, Remote Sensing, Remote Sensing of Environment, Transactions on Geoscience and Remote Sensing – TGARS as well as international project proposals incl. to RIVM Dutch National Institute for Public Health and the Environment, BELSPO Belgian Federal Public Planning Service Science Policy	