Volunteered Geo-Dynamic Information for Health-Risk Exposure Assessment - A FRESHER Case Study

Christoph Aubrecht and Klaus Steinnocher

AIT Austrian Institute of Technology GmbH, Vienna, Austria

Abstract

In this paper we discuss the use of volunteered geo-dynamic information (VGDI) for assessing exposure to health risks and improving analysis of associated dynamic aspects in urban settings. VGDI is introduced as an alternative and complementary data source to traditional geodata whereby both spatial and temporal aspects are highlighted. Within the FRESHER project several health-related parameters are modelled, including air pollution and access to fast food restaurant locations. We discuss how Foursquare data (and VGDI in general) can benefit integrative smart urban analytics and provide sample illustrations for a test case area in Lisbon, Portugal.

Keywords:

foursquare, spatio-temporal analysis, population dynamics, health-risk exposure

1 Introduction

Volunteered geographic information (VGI) has attracted increased attention in the past couple of years in the context of integrative spatial analytics, especially in relation to urban areas. When the term VGI was coined by Goodchild (2007), widespread citizen participation and engagement in the voluntary creation of geographic information was still considered an "exotic" prospect. Things have clearly changed in that regard (Goodchild et al., 2016) and the notion of ubiquitous sensing (Gartner et al., 2007) is becoming more and more today's reality. Data from different sensors are analysed to extract additional multi-level information and improve our understanding of inter-relations in the environment. This approach, recently referred to as Urban Informatics – or Urban Geoinformatics when focusing mainly on spatial aspects – is driving the agenda forward regarding big urban data analysis and advancing the smart city vision. With regard to VGI, urban areas are of particular relevance as levels of use and contribution are highest there. However, the issue of representativeness is still considered one, if not the major, factor impacting applicability of VGI for operational use and input in decision support systems. Nevertheless, it has been widely agreed that the usefulness of VGI can in many ways outplay its inherent weaknesses. Addressing those weaknesses and capitalizing on the unique potential of VGI are crucial focuses of scientific research in the domain of VGI and Urban (Geo-)Informatics.

2 The concept of VGDI (Volunteered Geo-Dynamic Information)

In a recent paper (Aubrecht et al., 2016), we broke VGI down into its individual components and highlighted the conceptual lack of a formal acknowledgment of the temporal factor in VGI. We outlined that the temporal aspect should be considered one if not the main added value of VGI compared to standard "static" data sources, and we consequently introduced spatial dynamics to advance the concept to Volunteered Geo-Dynamic Information (VGDI). That dynamic nature is also the main concern of this study, where exposure to health risks is analysed in a spatio-temporal manner.

Following earlier proposed classification schemes of implicit vs. explicit volunteering (Craglia et al., 2012), we categorize VGDI into two classes, namely (1) unguided implicitly-collective information and (2) guided explicitly-collective information. The second category follows clear contribution guidelines and is often focused on one particular theme, and thus contextually pre-defined (e.g. infrastructure data in OpenStreetMap [OSM]). A lot of research focuses on such guided data creation and mining as a potential alternative to top-down governmental and authoritative data (Zielstra & Zipf, 2010). Whereas individual entries are highly relevant in this category, the main value of the first category lies in collective analysis of multiple entries (Noulas et al., 2011). Sources for category 1 include location-based social network (LBSN) services such as Foursquare (Swarm) and Facebook Places, as well as others like Twitter and Instagram, where spatially-explicit information is not the focus but can be extracted. Both classes of volunteered geoinformation have been integrated for a multitude of application domains with the most prominent examples in the literature probably being OSM and Twitter respectively. In the present study, we highlight the value of Foursquare venue and user check-in data as a potential input source for dynamic exposure modelling.

3 The FRESHER context

The FRESHER project (*FoResight and Modelling for European Health policy and Regulation*)[1] aims at analysing alternative futures, whereby the detection of emerging health scenarios will be used to test future policies to effectively tackle the burden of non-communicable diseases (NCDs). An empirically-based dynamic micro-simulation model is currently being developed, led by the Organisation for Economic Cooperation and Development (OECD) Health Division. The model captures various risk factors for NCDs and potential consequences and is applied to three European regions (Central-Eastern, Northern, Southern). To analyse potential future conditions and provide policy-relevant recommendations, first the present and the past need to be understood.

One aspect of the project in this context is to geospatially assess current levels and trends of exposure to various different health-related parameters at city-scale. Output of the assessment subsequently serves as one input component to the OECD micro-simulation model. Three case study areas were selected, representing one city in each of the three European regions. Those cities are Lisbon (Portugal), Vienna (Austria) and Tallinn (Estonia).

---

[1] www.foresight-fresher.eu (accessed on 20 October 2016)

Various environmental and socio-contextual risk factors are considered for health-risk exposure analysis. These include directly measured parameters related to air pollution and temperature, as well as approximated adverse influencing parameters like excessive fast food consumption and restricted access to green urban spaces. In a geospatial sense, the aforementioned environmental variables refer to continuous fields (i.e. grids interpolated from point station measurement data). These continuous fields are then overlaid with disaggregated population distribution grids, in order to directly estimate exposure patterns. The approximated variables refer to exposure in the sense of modelling people's accessibility to discrete spatial features such as fast food restaurants and urban parks. Where in standard approaches census data and ground station network data as well as company registries and cadastral data are used for such assessments, in this study we discuss the potential integrative use of VGDI as an alternative or complementary input source.

4   Aspects of VGDI use for dynamic exposure analysis

Exposure terminology varies when applied to different scientific and topical domains. From a purely geospatial perspective, a certain number of elements (e.g. population or certain building structures) are exposed to a specific situation with potential adverse impact (e.g. a natural hazard, elevated air pollution level). Exposure in that context, then, basically describes the spatial intersection of distributional patterns. This kind of understanding is widely established in the domains of hazard and disaster risk assessment (UNISDR, 2009). In the health research domain as well as, for example, the climate change community, however, certain pollutants or adverse conditions (climate change effects) can themselves be considered and described as "exposures" (e.g. Thurston et al., 2015).

For either conceptual approach, both population distribution and potential adverse factors need to be modelled. In a dynamic assessment setting, temporal variations are integrated in addition to the spatial aspect, i.e. to address issues such as population moving out of a potential danger zone during the daytime, or a certain area featuring adverse conditions only during a specific time slot. The latter is commonly modelled in environmental impact analysis, thus illustrating varying conditions, e.g. regarding trends of air pollution or temperature patterns. For the population distribution element, however, mostly static census-based data are still consulted (Orru et al., 2009). In this article, we describe the potential use of VGDI for both parameters in a dynamic study setup and provide illustrative examples from the ongoing FRESHER Lisbon test case.
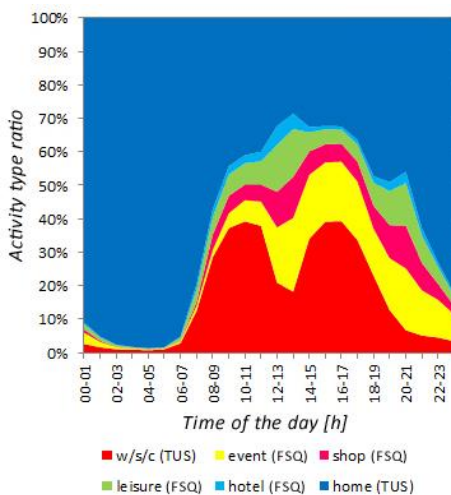
VGDI input for dynamic modelling of the "population" component

As on overarching objective, population dynamics models focus on the spatio-temporal refinement of traditional "static" representations of residential distribution patterns, thus moving from illustrating a night-time situation to a more complex picture of daytime human activity patterns. Considering the daily temporal cycle, current approaches can be broadly categorized as (1) simplified binary daytime/night-time models (e.g. McPherson et al., 2004; Freire & Aubrecht, 2012), or (2) multiple time-slice models that account for continuous

variation in human activities over the course of a day (e.g. Aubrecht et al., 2014; Martin et al., 2015).

While the first category typically refers to commuting and work statistics to identify average daytime patterns, the second category additionally considers time-use statistics and/or real-time information to identify refined activity patterns and their evolution during the day. The multiple time-slice approaches can thus include traditional time-use survey data, but also real-time mobile phone tracking data and data extracted from VGDI. In addition to mobile phone data, LBSN user activity has been used for the direct approximation of daytime population distribution patterns (Toepke & Starsman, 2015). While proven to be inadequate for estimating all-encompassing population distribution due to their lack of representativeness, the distributional patterns of certain activity categories, such as leisure and workplace clusters, are deemed to be retraceable using LBSN data. A recent study showed positive correlation patterns between georeferenced tweets and reference workplace census data for a test case in London, UK (Steiger et al., 2015).

While direct approximation of total-population distribution via VGDI remains unfeasible, there are several valuable aspects inherent in VGDI that are particularly useful when addressing population dynamics (Aubrecht et al., 2016). VGDI provides temporally and thematically fine-grained input information for target zone characterization in the course of the spatial disaggregation process. Human activity classes which are not available in traditional land use data (e.g. shopping, leisure) can thereby be mapped. Moreover, temporal variations can be differentiated, such as daytime leisure locations vs. nightlife spots. Furthermore, aggregating user check-in data based on activity categories enables time-use profiling. This can serve as an alternative to traditional survey-based time-use data and can also be useful for calibration and refinement purposes. Figure 1 illustrates a hybrid time-use profile where Foursquare data and time-use survey data were integrated. Foursquare is used in this context to map the mobile population, thus excluding people staying at home or at their workplace (contrary to findings using georeferenced tweets, Foursquare does not correlate with workplace patterns).

Figure 1:
Hybrid time-use profile using Foursquare (FSQ) data extracted for the Lisbon Metro Area, and Portuguese time-use survey (TUS) data (adapted from Aubrecht et al., 2016).

## VGDI input for dynamic modelling of the "adverse condition" component

As outlined above, in the FRESHER health-risk exposure context, relevant adverse conditions include air pollution and access to fast food restaurant locations.

Air pollution is commonly measured using ground station network data, which then serve as input for the creation of spatially interpolated grids. While the temporal resolution (recording frequency) is very high for such data compilations (e.g. hourly), the spatial resolution is limited to scattered point locations (e.g. 19 within the Lisbon Metro Area). Using VGDI can help overcome the spatial constraint, as done via mobile sensors in the Citi-Sense project.[2] Small portable devices are used by citizens in that project to record high spatio-temporal resolution air pollution data, which can supplement station network data (Jovašević-Stojanović et al., 2015). As Vienna is one of the Citi-Sense case studies, we aim at integrating that data also for the city's FRESHER health-risk exposure assessment.

Assessing access to fast food locations is not as straightforward. Traditional georeferenced address registries may still record usage of a certain location as a restaurant but they do not usually give any further detail. Company data sets such as yellow pages can complement such lists, but they are not usually publicly available. In any case, such data compilations merely record static locations, at the most including opening hours. Using VGDI enables disclosing the full spatio-temporal spectrum of a location via visitor numbers and frequency of visits, thus showing a specific location's relative importance as compared to others (see Figure 2), as well as its diurnal usage pattern. Within the FRESHER Lisbon test case, we link Foursquare fast food venue locations to a dynamic population model, thus estimating a time-dependent, shortest-distance parameter.

---

[2] www.citi-sense.eu (accessed 20 October 2016)
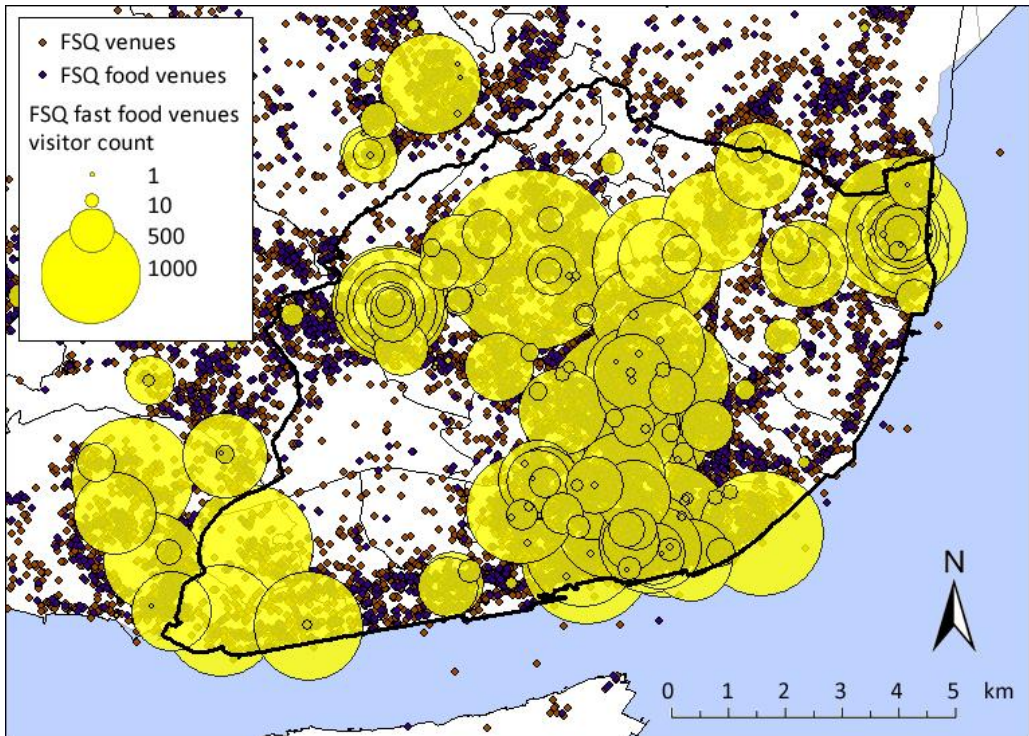
Figure 2: Foursquare (FSQ) fast food venues and visitor counts for the City of Lisbon.

5   Discussion

The objective of this short paper was to highlight the use and value of dynamic VGI or VGDI for health-risk exposure assessment, as applied in the FRESHER project framework. We showcased several aspects of Foursquare data applied for both population dynamics modelling and assessing conditions with potentially adverse health impacts. The additional temporal component of dynamic VGI is considered crucial to illustrate time-dependent aspects of exposure. Integrative use of VGDI and traditional data sources such as ground station networks and classical survey data proves an effective approach for smarter urban analysis. We see this as a step towards further discussions regarding big urban data analytics.

Acknowledgments

References

Aubrecht C., Steinnocher K., & Huber H. (2014). DynaPop - Population distribution dynamics as basis for social impact evaluation in crisis management. ISCRAM 2014. 11th International Conference on Information Systems for Crisis Response and Management. University Park, PA, USA (pp. 319-323)

Aubrecht C., Özceylan Aubrecht D., Ungar J., Freire S., & Steinnocher K. (2016). VGDI - Advancing the concept: Volunteered Geo-Dynamic Information and its benefits for population dynamics modeling. Transactions in GIS, early view (29 April 2016). doi:10.1111/tgis.12203

Craglia M., Ostermann F., & Spinsanti L. (2012). Digital Earth from vision to practice: making sense of citizen-generated content. International Journal of Digital Earth 5(5) (pp. 398-416)

Freire S., & Aubrecht C. (2012). Integrating population dynamics into mapping human exposure to seismic hazard. Natural Hazards and Earth System Sciences 12(11) (pp. 3533-3543)

Gartner G., Bennett D. A., & Morita T. (2007). Towards ubiquitous cartography. Cartography and Geographic Information Science 34(4) (pp. 247-257)

Goodchild M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal 69(4) (pp. 211-221)

Goodchild M. F., Aubrecht C., Bhaduri B. (2016). New questions and a changing focus in advanced VGI research. Transactions in GIS early view 1 Sep 2016 (doi:10.1111/tgis.12242)

Jovašević-Stojanović M., Bartonova A., Topalović D., Lazović I., Pokrić B., & Ristovski Z. (2015). On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. Environmental Pollution 206 (pp. 696-704)

Martin D., Cockings S., & Leung S. (2015). Developing a flexible framework for spatiotemporal population modeling. Annals of the Association of American Geographers 105(4) (pp. 754-772)

McPherson T. N., Ivey A., & Brown M. J. (2004). Determination of the spatial and temporal distribution of population for air toxics exposure assessments. 5th AMS Conference on Urban Environment. Vancouver, BC, Canada.

Noulas A., Scellato S., Mascolo C., & Pontil M. (2011). An empirical study of geographic user activity patterns in Foursquare. ICWSM-11 – Fifth International AAAI Conference on Weblogs and Social Media. Barcelona, Spain.

Orru H., Teinemaa E., Lai T., Tamm T., Kaasik M., Kimmel V., Kangur K., Merisalu E., & Forsberg B. (2009). Health impact assessment of particulate pollution in Tallinn using fine spatial resolution and modeling techniques. Environmental Health 8(7) (**DOI:** 10.1186/1476-069X-8-7)

Steiger E., Westerholt R., Resch B., Zipf A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. Computers, Environment and Urban Systems 54 (pp. 255-265)

Thurston G. D., Ahn J., Cromar K. R., Shao Y., Reynolds H., Jerrett M., Lim C. C., Shanley R., Park Y., & Hayes R. B. (2015). Ambient particulate matter air pollution exposure and mortality in the NIH-AARP Diet and Health Cohort. Environmental Health Perspectives, 24(4): 484–490.

Toepke S. L., & Starsman R. S. (2015). Population distribution estimation of an urban area using crowd sourced data for disaster response. ISCRAM 2015. 12th International Conference on Information Systems for Crisis Response and Management. Kristiansand, Norway

UNISDR (2009). Terminology on Disaster Risk Reduction. United Nations International Strategy for Disaster Reduction. Geneva, Switzerland

Zielstra D., & Zipf A. (2010). A comparative study of proprietary geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal