From Social Sensor Data to Collective Human Behaviour Patterns – Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments

Günther SAGL, Bernd RESCH, Bartosz HAWELKA and Euro BEINAT

Abstract

The digital traces that people continuously leave behind – voluntarily or not – while using communication devices such as mobile phones or interacting with social media platforms reflect their behaviour in great detail. In this paper we show examples of the spatio-temporal patterns of collective human dynamics, which we derived from 'social sensor' data. We used user-generated data in mobile networks and voluntarily published information at social media platforms, which both served as proxies for human activity and mobility. Results show collective human activity and mobility hotspots in selected European urban environments and thus provide additional insights into how collective social activity shape urban systems. We used different geo-visualisation techniques to effectively communicate the urban inherent spatial and temporal dynamics. Especially for education related purposes, this allows a better general understanding of collective human behaviour in urban environments.

1 Introduction

Today, research about large-scale human behaviour patterns is often based on user-generated data either within wireless communication networks such as mobile phone networks, or on social media networks like Twitter or Flickr. Within this research we consider user-generated data from such networks as 'social sensor data' and the actual network as 'large-scale sensor'. In combination with Geographic Information Systems (GIS), these inherently spatio-temporal data enable novel capacities to analyse and visualize large-scale human dynamics in a more integrated manner even close to real-time.

This can potentially have significant impacts on a variety of Geographic Information (GI) related research areas such as urban planning, traffic management, sustainable infrastructure establishment, or tourism. Also, aside from various geo-application domains, the more general understanding of social dynamics is necessary to get an overall impression of how urban environments behave in time and space. Such understanding can be enhanced by modern spatial data mining and geo-visualisation techniques.

Thus, coupling social sensor data sources with GIS and spatial data mining techniques is a necessary step towards a holistic assessment of urban dynamics. Discovering hidden implicit patterns in such data and mapping the behavioural relationships allows us to create a socio-technical view of the city. In this research we are not interested in extracting information from individual data sets, but from collective data. This allows for an

Jekel, T., Car, A., Strobl, J. & Griesebner, G. (Eds.) (2012): GI_Forum 2012: Geovizualisation, Society and Learning. ISBN 978-3-87907-521-8.

integrative assessment of urban processes from a high-level standpoint and prevents privacy infringements.

This paper illustrates selected examples of collective human behaviour in urban spaces from a 'social sensing' perspective. In the next chapter we briefly review relevant related work. Then we describe the data used and methods applied followed by the results and discussion. Finally, we give some conclusions and further research aspects.

2 Related Work

A substantial body of scientific literature is dedicated to the exploration of human behaviour patterns that are based on user-generated data in both mobile telecom networks and social media.

Examples based on mobile telecom traffic include the formal understanding of collective and individual human mobility (GONZÁLEZ et al. 2008), the segmentation of urban spaces (READES et al. 2009), or the exploration of social events (CALABRESE et al. 2010). PHITHAKKITNUKOON et al. (2010) created an 'activity-aware map' based on individual cellular data in order to understand the dynamic of inhabitants for urban planning and transportation purposes. RATTI et al. (2010) explored human interactions based on a large telecommunication database for Great Britain and show that these interactions significantly correlate with administrative regions. CALABRESE et al. (2011) invented the concept of co-location which was proved by behaviour of the telecom network users who called each other frequently and shared the same space at the same time in the city. GIRARDIN et al. (2008) explored the digital footprints left by people while moving within the urban spaces. By looking at cell phone network data and the georeferenced photos from Flickr, they tried to uncover the presence and movements of tourists in the city of Rome. SEVTSUK and RATTI (2010) confirmed that the regularity at different hours, days, and weeks in urban mobility could be significantly correlated with people's behaviour while using mobile phones in space and time.

In comparison to telecom data, the analysis of social media data is – due to its recent emerge –less scientifically explored. Since social media platforms can be seen as an additonal even real-time data sources at a global scale, they are of booming interest for various research domains. With respect to the social media users' behaviour, social media temporal pattern has been analysed already in global scale by DODDS et al. (2011). They uncovered and explained temporal variations in society 'happiness' by looking at the expressions posted on the online, global microblog and social networking service like Twitter. Moreover GOLDER and MACY (2011) proved that the mood of Twitter users has diurnal and seasonal patterns which could be correlated with work, sleep and day time. What is even more interesting, BOLLEN and MAO (2011) confirmed that measurements of collective mood posses also more practical implications, as it could be correlated to the value of the Dow Jones Industrial Average (DJIA) over time. On the other hand, spatial connotations of social media users seems to be still not enough explored, one of the few examples can be found in NOULAS et al. (2011).

In contrast to previous research we focus on the average patterns of collective human behaviour in selected environments at a macro scale, i.e. the more general activity and mobility patterns. The aim thus is to support an overall understanding of how such environments behave in time and space from a social activity point of view.

3 Data and Methods

The research presented in this paper is based on data from two sources, namely the mobile network, and the social web. Also, we address the different nature of the data: user-generated traffic in mobile networks is a kind of – from the mobile phone user's perspective – 'involuntarily' provided spatio-temporal data, regardless of its content. In contrast, the data and the content actively provided by individuals get published on social media platforms with the person's explicit agreement.

3.1 User-generated Data in Mobile Networks

A European telecom operator provided user-generated mobile network traffic data for the urban area of Udine in Northern Italy. The solely anonymized and aggregated data included directed handover data between the mobile networks' cells as vector data, and overall network traffic measured in Erlang[1] as raster data. We used the former for mobility, and the latter for activity analyses.

Similar telecom traffic data where provided by another European mobile phone operator for the city of Amsterdam. Here we focused on Erlang data in order to compare places based on their characteristic temporal signatures. The places were selected based on their functional context, for example recreational areas, shopping areas or business districts, and living areas.

In order to transform raw mobile network traffic data into multi-dimensional information layers reflecting human behaviour in time and space, we developed a fully automated workflow consisting of four consecutive steps: Data Pre-Processing: Conversion of ASCII files into vector and raster data; Data Management: data preparation and data structuring using GIS and geo-enabled DBMS; Data Integration: integration of data into geo-enabled software including GIS; Modelling & Analysis: Extraction of spatial and temporal information of humans' activity and mobility.

3.2 Volunteered Geographic Information from Social Media

In terms of Volunteered Geographic Information (VGI) we are currently facing a fast increase in available data sources. On the one hand, this results from established social media applications such as Twitter, Flickr or Facebook, which provide anonymised collective VGI through public Application Programming Interfaces (API). On the one hand, we are seeing an increasing willingness of people to actively contribute their own observations and comments in particular applications for a dedicated usage context, which can be subsumed under the broad term 'People as Sensors'.

---

[1] Dimensionless basic unit of telecom traffic intensity, named after A. K. Erlang: 1 erlang = 1 person calling for 1 hour, or 2 persons calling for 30min each and so on.

In the research presented we use collective data sources from Flickr, which provides user-geo-tagged images, to analyse social and touristic hot spots in different areas of a city or a region. While the general approach of using Flickr data to disclosing tourist dynamics has been proven (GIRARDIN et al. 2008), we present a new method of revealing hot spots in cities (Fig. 5 and Fig. 6) and seasonal variations in touristic activities in rural and urban areas (Fig. 7). One particular novelty in our approach is that all steps of the data analysis process (data re-projection, data pre-processing, joining spatial data sets, grid calculation, spatial averaging, indicator computation, snapping points to streets, statistical analysis etc.) are done in a fully automated workflow using highly specialised Geographic Information System (GIS) tools.

4 Results and Discussion

Fig. 1 shows the overall telecom network activity within the course of an average working day in the *city of Udine*, Italy. Although the screenshots shown herein clearly indicate the city centre, the entire and interactive Google Earth application enables an advanced understanding of when and where people are actively using the mobile network – and this activity is area-wide and at a high spatial (250m × 250m) and temporal (15min interval) resolution.

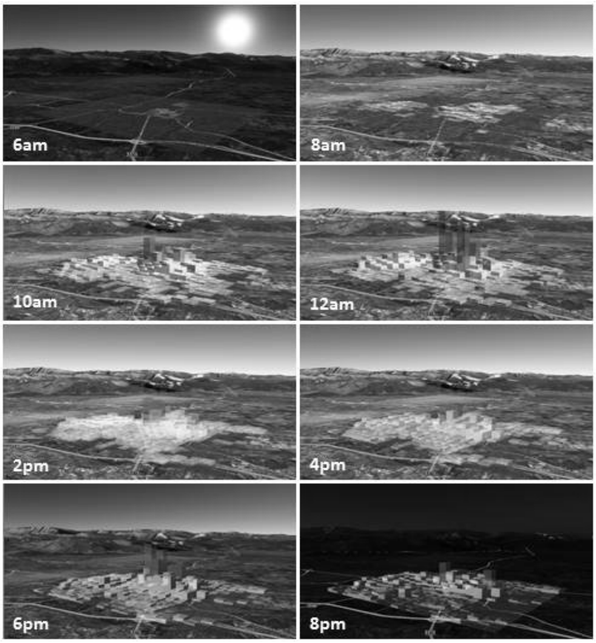

Fig. 1: Snapshots of the Collective Human Dynamics in Urban Udine – an Average Working Day as seen from a Mobile Networks Perspective

With respect to city-suburbs flows in Udine, Fig. 2 shows a higher density of handovers between western sectors (NW, SW) and the city centre as compared to the eastern sectors (NE, SE).

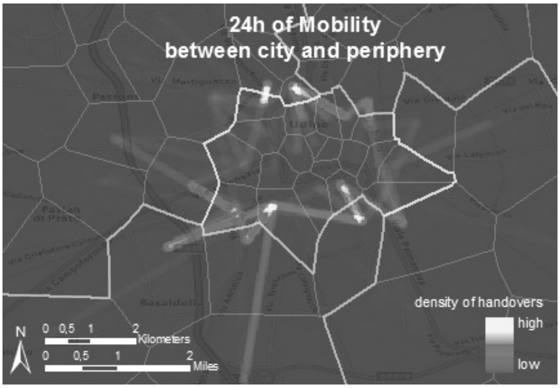

Fig. 2: Mobility in Space: Handover Density between the Udine City Centre and its Periphery

In addition to the spatial view, the temporal patterns shown in Fig. 3 confirm the overall higher intensity of human mobility between the western sectors and the city centre.

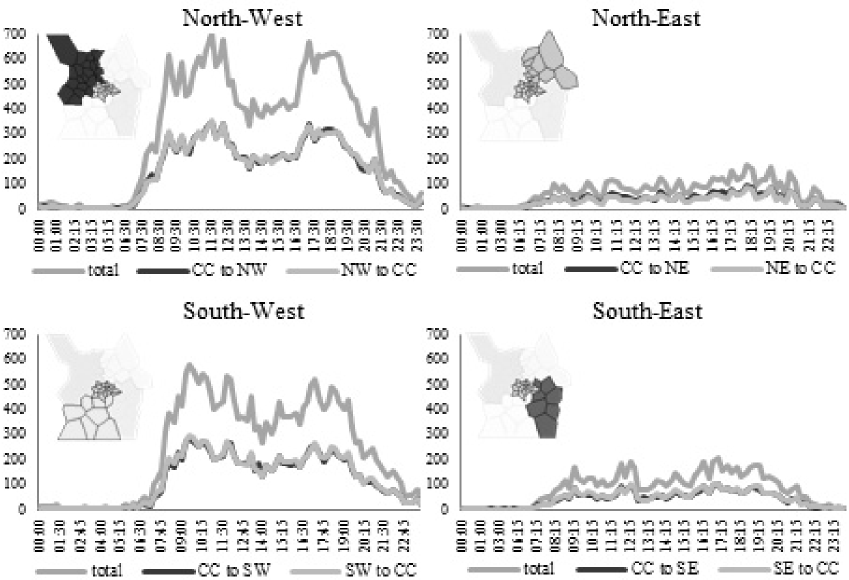

Fig. 3: Mobility over Time: Handovers between the Udine City Centre (CC) and its four peripheral sectors NW, NE, SW, and SE

Indeed, the western sectors – in particular the NW sector – show the distinct double-peak pattern, which is an indicator for working hours (SEVTSUK & RATTI 2010). In combination, the spatial and temporal patterns might be due to the highway crossing the western sectors and a major street entering the city centre from the North, which indicate mobility gateways to and from the Udine city centre.

Three levels of mobility can be identified: low (during late night and early morning hours 23pm – 6am), moderate (between 1pm – 3pm, and 7pm – 22pm) and high (between 7am – 12am, and 4pm – 6pm). In fact, the mobility 'hot-spots' indicated as yellow peaks (Fig. 2) represent in this case relatively high load of antennas with handovers.

The correlation of temporal signatures in Fig. 4 shows, on the example of the *city of Amsterdam*, The Netherlands, that areas with similar functional design also have similar mobile phone activity over time and vice-versa.

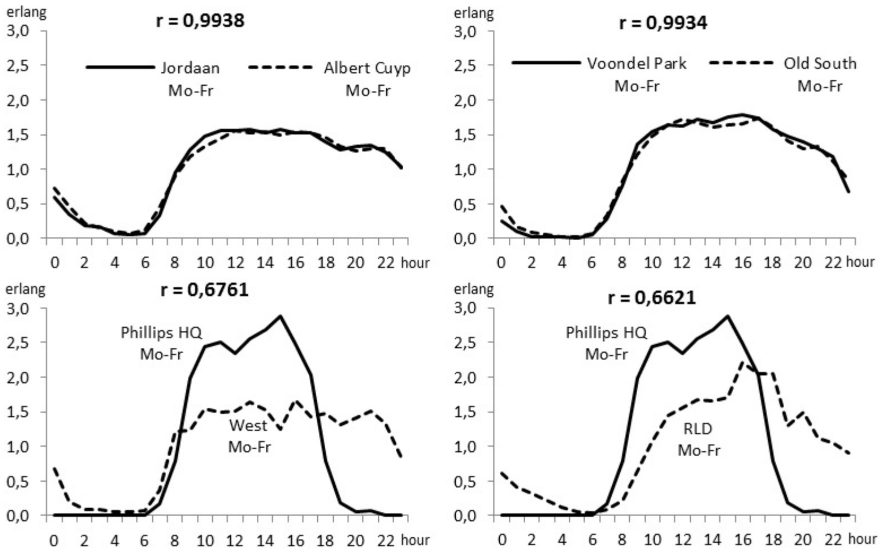

Fig. 4: Correlation of Characteristic Temporal Signatures of Places in the City of Amsterdam with Different Functional Design

For instance, 'Jordaan' as an exclusive residential district in the city centre and the 'Albert Cuyp' street market have – although ~3 km apart from each other – a very similar temporal signature (Fig. 4: upper left; r = 0.9938) .Since the functional design of an area strongly depends on its land use, this result indicate that some characteristics of the signature can be used to compare the de jure land use with the land use as formed by the people at a much higher spatial detail. The 'Phillips Headquarter (HQ)', as a distinct example for business areas, possess a rather steer slope in the morning and evening, and a double-peak at noon (Fig. 4: bottom). This temporal signature is therefore different from the 'West', which is a rather low income residential area (Fig. 4: lower left), and from the 'Red Light District (RLD)' shown in Fig. 4 (lower right).

Fig. 5 shows the density of Flickr points in the *city of New York*, USA, analysed in a variety of different methodologies. Fig. 5.a maps the original points in five density categories; Fig. 5.b shows the aggregated Flickr photo density in a 50×50 sqm vector grid; Fig. 5.c illustrates the photo density per street segment; and Fig. 5.d depicts the raw Flickr points snapped to the nearest street segment.

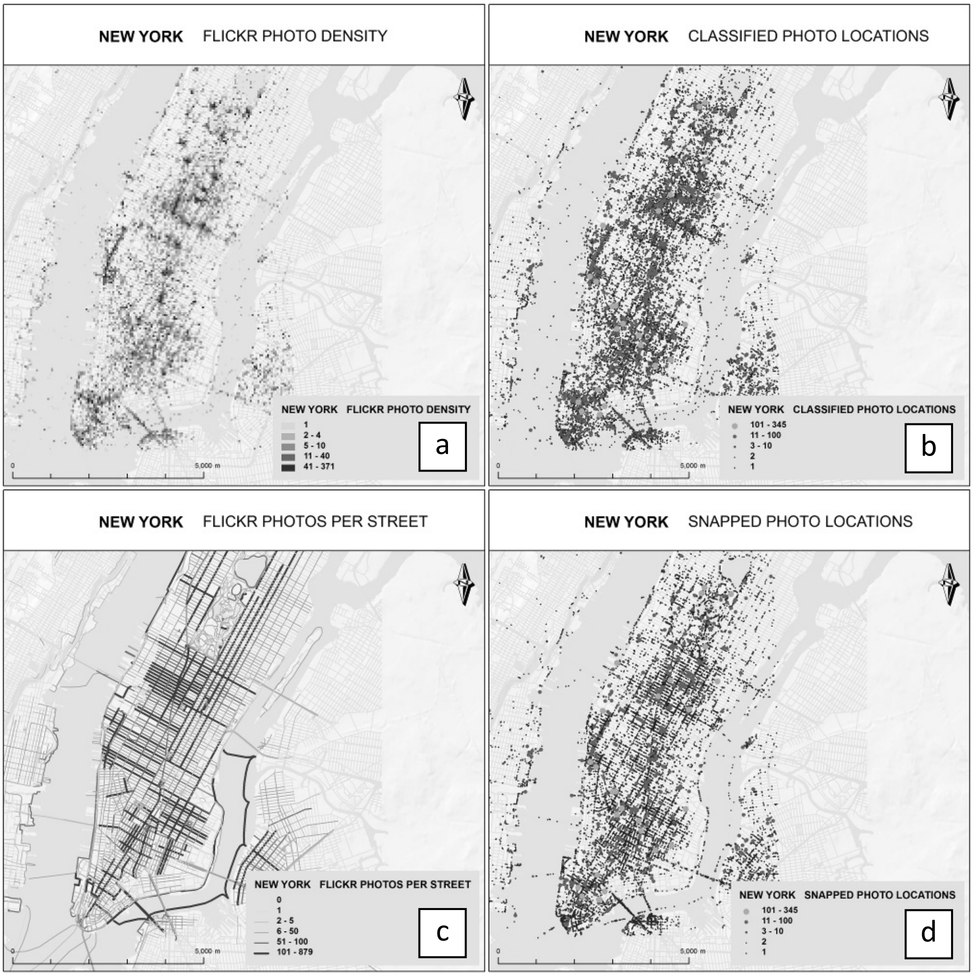

Fig. 5: Spatial Indicators for Tourism for New York Using Different Analysis Algorithms

Fig. 6 illustrates the density of social media usage in the *city of Madrid*, Spain. Preliminary results show that different quarters are characterized by different spatial indicators. For instance we can conclude from the spatial distribution Fig. 6 that the student quarter in the South-Western part of the city uses social media more intensely than other quarters with higher average age of the population. Similar conclusions can be drawn for other spatial indicators, which will be presented separately in a follow-up paper.

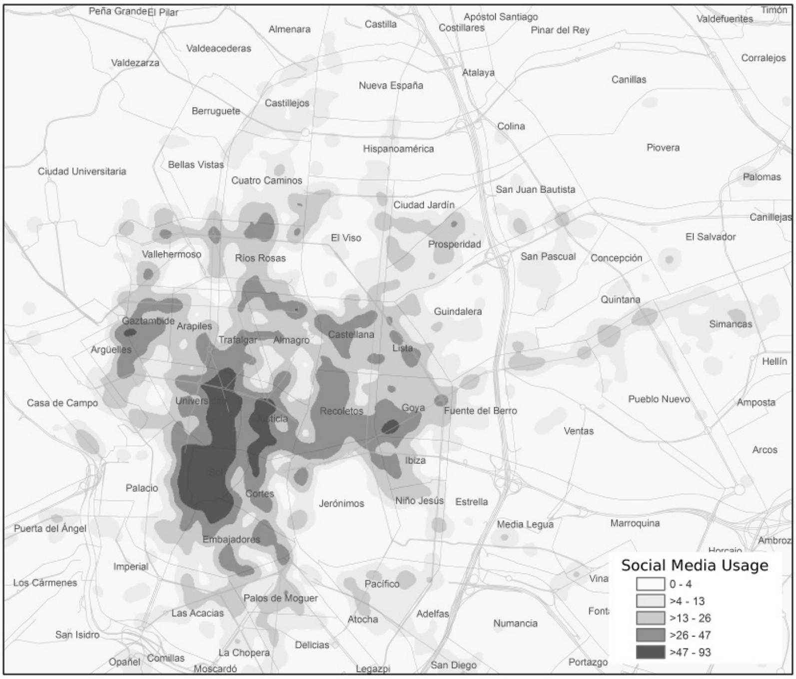

Fig. 6: Digital Usage Footprint in Madrid

In Fig. 7 we illustrate Flickr point densities for different seasons in the area of South Tyrol, Italy. The seasons are represented by the following time spans: winter – December to February; spring – March to May; summer – June to August; fall – September to November.

The interpretation of these maps yielded two main conclusions. First, as expected, the main tourist season is summer, i.e. the period from June to August. Second, touristic destinations vary heavily across the seasons. While the ski resorts in the dolomites are clearly dominating winter, the *city of Bolzano* and its immediate surroundings is the main attraction point during the other seasons. Furthermore, we can conclude that the Swiss Alps (particularly Davos and St. Moritz) induce more significant competition to the region of South Tyrol in winter and spring than in summer and fall, clearly indicating their importance as high-end skiing resorts.

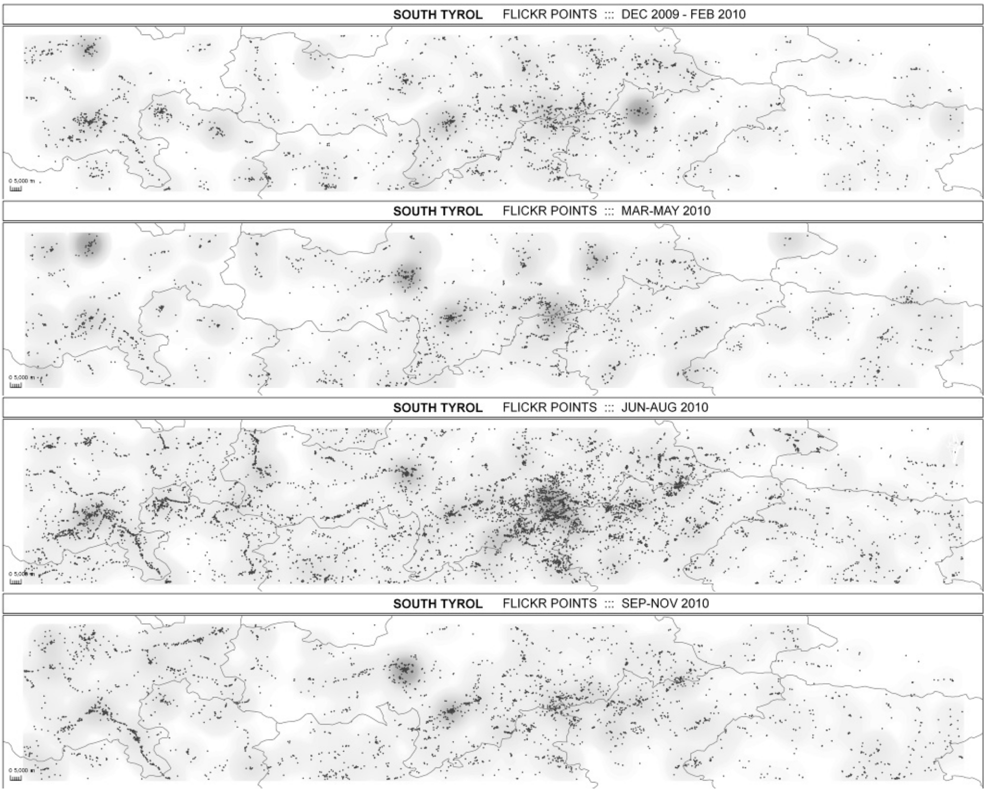

Fig. 7: Touristic Density in the Area of South Tyrol in Different Seasons

5 Conclusion and Outlook

In this paper we have shown several examples of spatio-temporal collective human behaviour patterns in selected European urban environments. We derived these patterns using geospatial methods from social sensor data including user-generated mobile network traffic and voluntarily published geoinformation on social media platforms.

We conclude from the results shown in this paper that spatio-temporal analysis of social sensor data in combination with geo-visualisation methods can contribute to a better understanding of urban systems in general and its inherent social dynamics regarding activity and mobility in particular. However, the behavioural patterns derived reflect the larger population only to some extent. Aspects to consider thus include the mobile operator's market share; the mobile penetration rate; data representativeness: which fraction and which social background do the Flickr data represent?; interpretation clarity: how can we ground-truth our conclusions?; the analysis algorithms applied: are our algorithms and spatial indicators suitable to draw comprehensive conclusions? These questions are currently addressed in follow-up research projects.

One major further research subject is, from our perspective, the impact of scale on the analysis results: Can the methods presented in this paper be used to extrapolate the collective human behaviour patterns to larger areas? Do the patterns vary at different temporal scales (e.g. weakly versus seasonal; or winter versus summer)?

The authors would like to thank all involved researchers and external contributors from the University of Osnabrück and the Massachusetts Institute of Technology who offered their valuable thoughts. Particularly, the authors would like to thank Fabien Girardin from the Near Future Laboratory.

References

BOLLEN, J. & MAO, H. N. (2011), Twitter Mood as a Stock Market Predictor. Computer, 44, 90-93.

CALABRESE, F., PEREIRA, F. C., LORENZO, G. D., LIU, L. & RATTI, C. (2010), The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events. Pervasive Computing (Lecture Notes in Computer Science). Berlin/Heidelberg, Springer.

CALABRESE, F., SMOREDA, Z., BLONDEL, V. D. & RATTI, C. (2011), Interplay between Telecommunications and Face-to-Face Interactions: A Study Using Mobile Phone Data. Plos One, 6.

DODDS, P. S., HARRIS, K. D., KLOUMANN, I. M., BLISS, C. A. & DANFORTH, C. M. (2011), Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. arXiv:1101.5120v4 [physics.soc-ph].

GIRARDIN, F., BLAT, J., CALABRESE, F., DAL FIORE, F. & RATTI, C. (2008), Digital Footprinting: Uncovering Tourists with User-Generated Content. IEEE Pervasive Computing, 7, 36-43.

GOLDER, S. A. & MACY, M. W. (2011), Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures. Science, 333, 1878-1881.

GONZÁLEZ, M. C., HIDALGO, C. A. & BARABÁSI, A.-L. (2008), Understanding Individual Human Mobility Patterns. Nature, 453, 779-782.

NOULAS, A., SCELLATO, S., LAMBIOTTE, R., PONTIL, M. & MASCOLO, C. (2011), A Tale of Many Cities: Universal Patterns in Human Urban Mobility. Proceedings of the National Academy of Sciences of the United States of America, 104, 20167-72.

PHITHAKKITNUKOON, S., HORANONT, T., DI LORENZO, G., SHIBASAKI, R. & RATTI, C. (2010), Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. In: SALAH, A., GEVERS, T., SEBE, N. & VINCIARELLI, A. (Eds.), Human Behavior Understanding. Springer Berlin / Heidelberg.

RATTI, C., SOBOLEVSKY, S., CALABRESE, F., ANDRIS, C., READES, J., MARTINO, M., CLAXTON, R. & STROGATZ, S. H. (2010), Redrawing the Map of Great Britain from a Network of Human Interactions. PLoS One, 5, e14248.

READES, J., CALABRESE, F. & RATTI, C. (2009), Eigenplaces: Analysing Cities Using the Space-Time Structure of the Mobile Phone Network. Environment and Planning B: Planning and Design, 36, 824-836.

SEVTSUK, A. & RATTI, C. (2010), Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks. Journal of Urban Technology, 17, 41-60.