Mapping collective human activity in an urban environment based on mobile phone data

Sagl G., Delmelle E.M., Delmelle E.C. – January 2014.

Keywords: urban dynamics, mobile phone, geovisual analytics, geocomputation, human behavior.

Abstract

Identifying and characterizing variations of human activity – specifically changes in intensity and similarity – in urban environments provide insights into the social component of those eminently complex systems. Using large volumes of user-generated mobile phone data, we derive mobile communication profiles that we use as a proxy for the collective human activity. In this paper, geocomputational methods and geovisual analytics such as Self-Organizing Maps SOM are used to explore the variations of these profiles, and its implications for collective human activity. We evaluate the merits of SOM as a cross-dimensional clustering technique and derived temporal trajectories of variations within the mobile communication profiles. The trajectories' characteristics such as length are discussed, suggesting spatial variations in intensity and similarity in collective human activity. Trajectories are linked back to the geographic space to map the spatial and temporal variation of trajectory characteristics. Different trajectory lengths suggest that mobile phone activity is correlated with the spatial configuration of the city, and so at different times of the day. Our approach contributes to the understanding of the space-time social dynamics within urban environments.

1. Introduction

The increasing proliferation of a broad array of geographically referenced data derived from Global Positioning System (GPS) receivers, location-based services, or georeferenced user-generated data enables new opportunities in the analysis of human spatial behavior (Kwan, 2000). These emerging datasets offer the potential to extract collective human behavior patterns, enabling insights onto the social component of urban dynamics. User-generated mobile network traffic data is one such data source that may serve as a proxy to characterize society's behavior.(Ratti et al., 2006; Shoval, 2007; Sevtsuk and Ratti, 2010; Calabrese et al., 2011a; Sagl et al., 2012c; Yuan and Raubal, 2012).

The overall motivation of this research is to explore spatial and temporal variations in intensity and similarity of collective human activity at different times of the day and days of the week, thereby enabling an enhanced understanding of human behavior in the context of the city's spatial configuration. Such an enhanced understanding may be particularly useful to urban planners in facilitating sustainable decision making. First, information on daily human routines can inform public authorities for a more efficient allocation of rescue services in anticipation of increased interventions, distinguishing between critical and non-critical places. Second, urban planners may find supporting evidence that the planned (legal) zoning of an area coincide with its actual use, thereby reconsidering planning sustainability. We hypothesize that such variations in intensity and similarity of collective human activity can be revealed from mobile phone data.

Recent research employing mobile network data has sought to understand the temporal dynamics of these data across an urban landscape. In this regard, prior work has mapped the intensity of network data at various time increments producing map sequences to explore dynamics (Ratti et al., 2006; Pulselli et al., 2008). Sevtsuk and Ratti (2010) utilized network intensity values assigned to a geographic cell in a regression modeling framework. A series of dummy variables portraying hourly, daily, and weekly increments served as independent variables to test whether the time of day, day of the week, or week in the year could explain intensity and thus confirm the existence of a 'routine' in urban mobility. Andrienko et al. (2010a) and Sagl et al. (2012b) present visual analytic approaches to exploring temporal changes in urban mobile network data. Both studies demonstrate the effectiveness and the efficiency of such approaches in real-world analytic scenarios for providing complementary views on the temporal sequence of spatial conditions and, moreover, on the spatial distribution of local temporal variations. In this research, we build upon these initial visualization techniques and propose another method for depicting spatio-temporal trajectories of changes in mobile phone uses across an urban area. The visualization technique we employ further enables spatial statistical analyses to be performed on the output which aids in identifying 'outliers' or unanticipated patterns.

Specifically, this current paper advances existing research in investigating the rhythms of social urban systems by elaborating on spatio-temporal variations within collective human activity patterns based on mobile phone data. We propose innovative combinations of visualization and exploratory space-time analysis methods. Firstly, we use the Self-Organizing Map (SOM) as the underlying framework for the development of temporal trajectories of change in multi-dimensional mobile phone data across an urban area; a methodology first proposed by (Skupin and Hagelman, 2005) in the context of census data change. We expand upon this visualization technique by proposing subsequent analyses on the properties of the trajectories (e.g., length). Finally, trajectories are linked to the geographic space to identify clusters of similarity along with outlier trajectories through the use of local spatial autocorrelation statistics.

The paper is structured as follows. In Section 2, we provide a concise overview on the rationale of user-generated data and analysis methods in the context of urban social dynamics. In section 3, we introduce the data set used and the methodology developed for mapping collective human activity based on mobile phone data. Section 4 illustrates the results of the case study performed, which is followed by a discussion in section 5. We draw some conclusions in section 6 and provide insights for future research.

2. *User-Generated Data in Urban Social Dynamics Analysis*

Hägerstrand's (1970) seminal work on time geography stressed the importance of examining individual spatio-temporal movements and constraints in our understanding of urban and regional systems. The collective result of these space-time paths throughout a metropolitan region gives rise to the functional structure of the city. From a computational perspective, investigating these daily, spatio-temporal patterns of human mobility has been a challenge for decades. For instance, in 1984, Goodchild and Janelle collected 1500 individual travel diaries to characterize diurnal variations in the collective human behavior within urban social structures. In contrast to the 1980s, today's user-generated data include mobile phone data and data and information from social media such as Twitter, Instagram, or Facebook, which has provided a plethora of human generated data. That is, individuals leave behind – voluntarily or not – a number of traces when interacting with digital systems such as mobile communication networks.

Such digital traces can reflect individual as well as collective human behavior at various levels of spatial and temporal granularities. We distinguish between volunteered and in-volunteered user-generated data and information provided by individuals. In the former type the user explicitly agree and are fully aware of generating data and information, for instance when posting georeferenced pictures with some description on Instagram (a social media platform). This type of user-generated data and information is known as Volunteered Geographic Information (VGI) as coined by Goodchild (2007). However, VGI is generated by rather specific subgroups, increasingly via mobile devices (Perreault and Ruths, 2011). In contrast to VGI, the user-generated traffic in mobile phone networks is – from a user's perspective – typically provided involuntary and lacking in content: e.g. the number of text messages sent/received is known but not the text itself, or the number and duration of calls is known but not the topic of the talk itself.

Further reasons for our emphasis on user-generated mobile network traffic are, firstly, that this type of in-volunteered data provide a large and relatively unbiased sample across society (Shoval, 2007; Rubio et al., 2010). Secondly, when a mobile device is connected to the internet via the mobile phone network, the traffic generated due to posting VGI on social media represents an additional indicator since this user-generated traffic is an intrinsic part of the overall mobile phone network traffic. Thus, user-generated data from mobile phone networks has shaped the way we investigate spatio-temporal human behavior patterns – this is documented in a large body of scientific literature (Ratti et al., 2006; Krygsman et al., 2007; Onnela et al., 2007; Shoval, 2007; Candia et al., 2008; González et al., 2008; Reades et al., 2009; Noulas et al., 2012; Sagl et al., 2012c).

Urban environments are of particular interest due to the typically higher degree and complexity of social dynamics. Research on such dynamics on the basis of mobile phone data has been intensively conducted by the MIT's SENSEable City Lab[1] and partner institutions (Krings et al., 2009; Quercia, 2010; Calabrese et al., 2011a; Calabrese et al., 2011b; Di Lorenzo and Calabrese, 2011; Calabrese et al., 2013). Such data driven studies explore diverse characteristics of urban environments – from general urban activity patterns (Ratti et al., 2006; Sagl et al., 2012c) to individual mobility preferences (Calabrese et al., 2013) or characteristics of group behavior (Farrahi and Gatica-Perez, 2011). However, the variations in space and time within typical collective human activity patterns in urban spaces need more attention in order to provide insights into the dynamic of change.

Other individual data have been used to investigate the cycles and routine of everyday life. For instance, data from Bluetooth-enabled mobile phones in combination with cell tower locations (Eagle and Pentland, 2009), or Wi-Fi access points, Foursquare (a location-based social network) check-ins, Radio Frequency Identification (RFID) cards for public transport (Williams et al., 2012). Within the 'UrbanDiary' project, the GPS tracks of twenty individuals were compared with their statements made on how they experienced the city of London, UK (Neuhaus, 2010; Neuhaus, 2011). The results show that daily, weekly and monthly space-time patterns are influenced by the configuration of the urban environment, i.e. by the functional arrangement of the city.

[1] http://senseable.mit.edu/ (accessed on 2013/05/03)

3. Data and Methodology

3.1 Data

To explore variations in collective human activity we conduct a case study using user-generated mobile phone data for the city of Udine, a medium-size city located in Northern Italy (Figure 1). The overall mobile penetration rate, i.e., the ratio between subscribers of mobile networks and the population, in this region is 155%, or in other words, 155 subscribers per 100 people. The network operator that provided the data sample used in this paper has a market share of 34.2%. That is, every third subscriber of a mobile network in that region is registered by the mobile network operator that provided the data. Thus, the mobile network traffic data sample reflects a relatively high proportion of the mobile communication activity of the entire population.

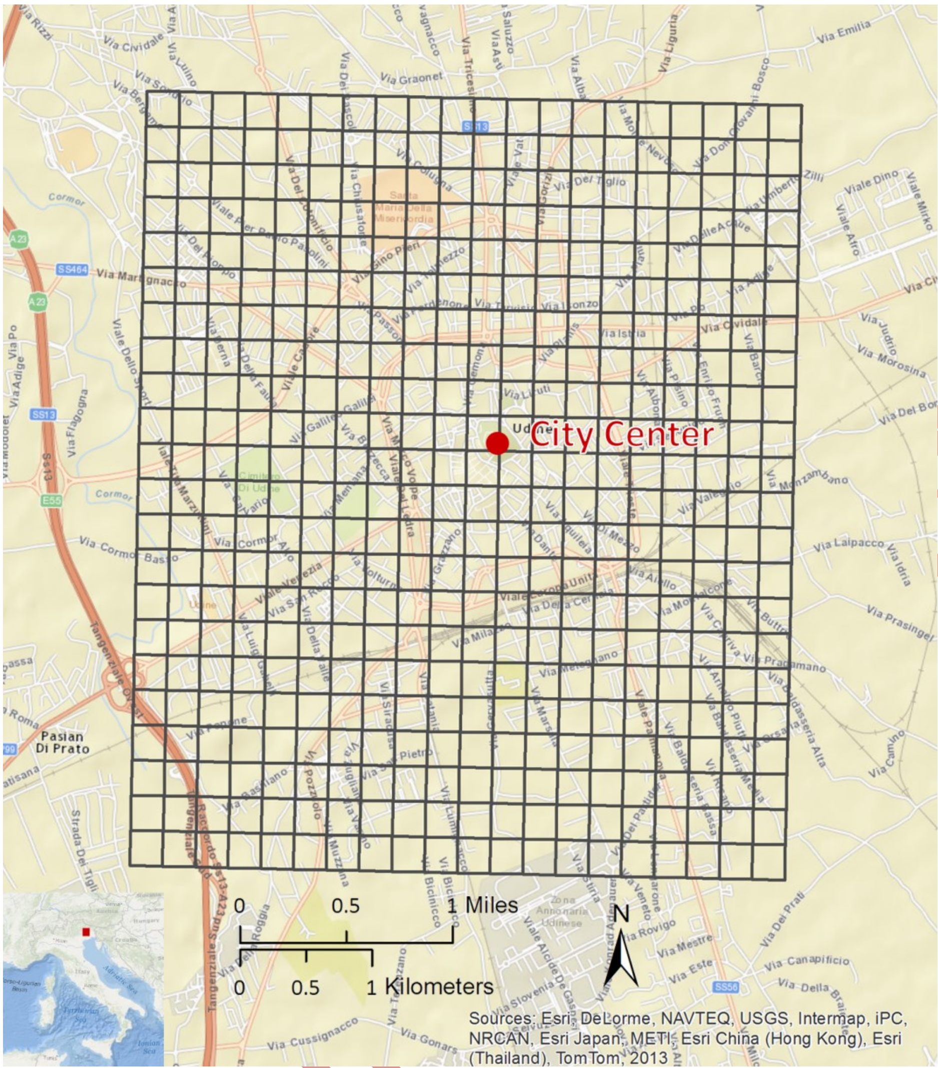


Figure 1: City of Udine (Northern Italy); the grid size represents the spatial resolution (250m x 250m) of the mobile network traffic dataset used (the cells in the regular grid do not relate to the network antennas' service areas.)

The user-generated mobile phone data set used fully covers the case study area; the time period of interest is from July 20th to September 30th 2009 due to the data have been validated to be consistent. From that data set, we selected five attributes as proxies for different human activity (Table 1). As provided by the network operator, each of the attributes is spatially aggregated to the corresponding 250m x 250m grid cell (refer to Figure 1) and temporally aggregated to a 15-minutes time interval. No identifier that might somehow relate to individual mobile phone activity was provided. The data set used is fully anonymized and does not violate any privacy aspects.

Table 1: Attribute Matrix of the User-Generated Mobile Phone Data Set used

	Incoming	*Outgoing*
Voice Calls	number of calls terminated in a given grid cell but originating from another, adjacent cell	number of calls originating from a given grid cell but terminated in another, adjacent cell
Text Messages SMS	number of text messages terminated in a given grid cell but originating from another, adjacent cell	number of calls originating from a given grid cell but terminated in another, adjacent cell
Total Traffic	The total network traffic is measured in Erlang[2] and implicitly contains certain types of VGI and is therefore used as a proxy for the overall collective mobile communication activity; the total traffic refers to a given single cell	

Figure 2 illustrates the space-time variation in the five variables of mobile phone usage – i.e., SMS in/out, voice calls in/out, and total internet traffic – in a space-time cube framework, following Andrienko et al. (2010a), and using the visualization package Voxler: the X and Y-axis denote the geographic space, while the Z-axis is the temporal axis. Volume rendering reflects the strength of mobile phone usage. Clusters of strong mobile phone usage are further reinforced using isovolumes. The five different variables tend to exhibit similar patterns: a larger cluster of cells in the center of town is particularly active around noon and after 6pm again. The second temporal cluster is more elongated for SMSs than phone calls, probably owing to the fact that several public companies or office close their business at night.

[2]'Erlang' is a dimensionless unit that refers to the ratio of number of persons to duration of calls: 1 Erlang = 1 person calling for 1 hour = 2 persons calling for 30 minutes each = three persons calling for 20 minutes each and so on.

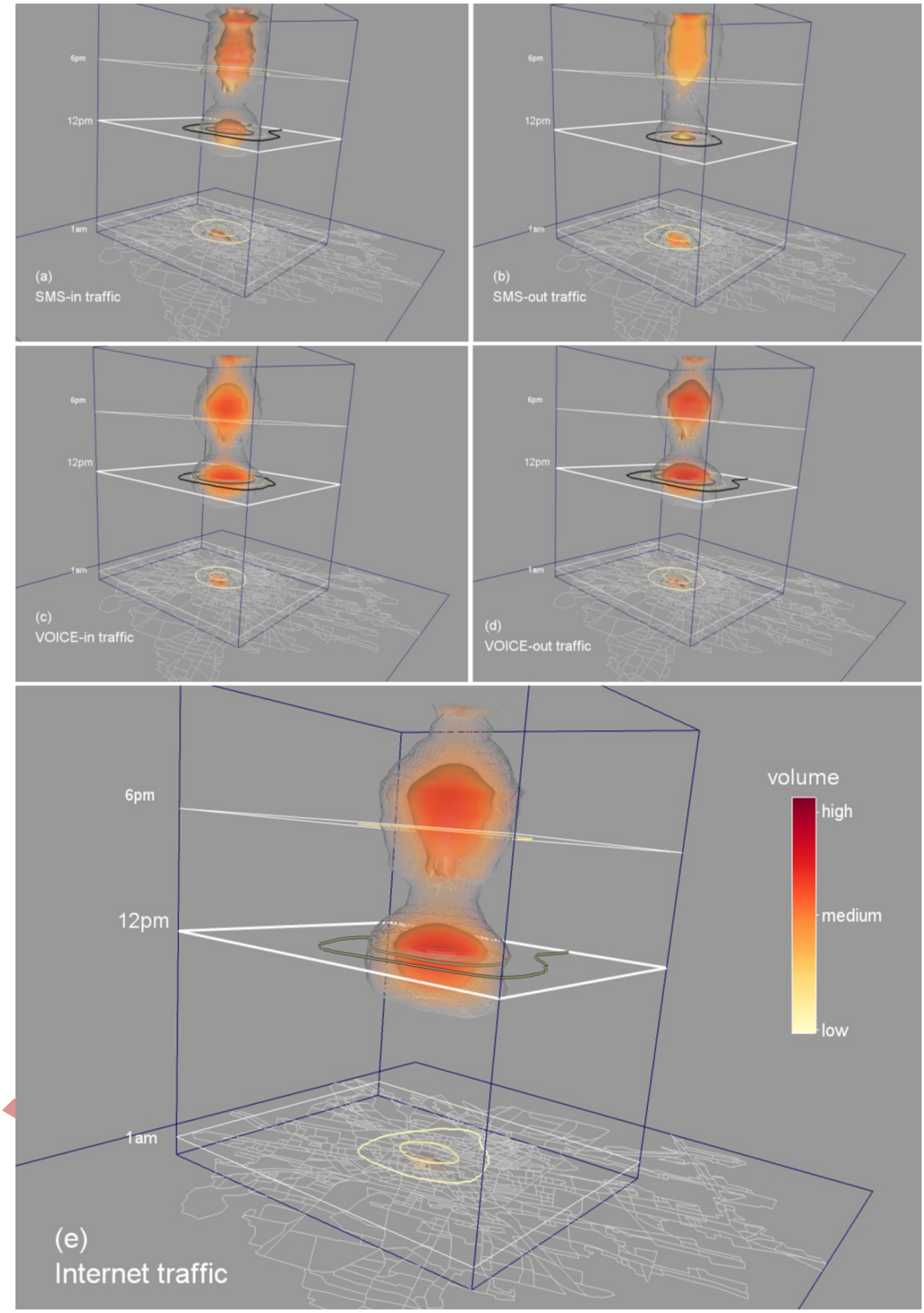


Figure 2: 3D visualization of mobile phone activity (categorized by received SMS, in a); outgoing SMS, in b); received phone call, in c); placed phone call, in d) and overall internet traffic in e). The shaded envelope denotes 99% of the activity, the triangulated envelope 95%. Voxels are red-colored to reinforce the overall intensity

Following Sevtsuk and Ratti (2010) and Sagl et al. (2012a), we further distinguished weekdays (Monday-Friday) from weekends (Saturday and Sunday) to take into account the two most evident differences in weekly mobile communication pattern. To strike a balance between temporal granularity and temporal representativeness, we aggregated the 15 minutes intervals to 1 hour intervals. As a result, for each grid cell (n=440), we obtain two groups of typical temporal signatures, one for typical weekdays, another one for typical weekend. This data preparation resulted in a data hyper-cube containing a total number of 440 (cells) x 2 (weekday/weekend) x 5 (number of attributes) x 24 (hours of the day) = 105,600 data values that represent the average mobile phone communication profiles and serve as input for further analysis.

3.2 Methodology

Geovisual analytics provides a suite of techniques that can be used in support explorative spatial data analysis (Andrienko et al., 2010a; Andrienko et al., 2010b). Geovisual analytics is thus particularly suitable for investigating user-generated mobile phone data in space and time (Andrienko et al., 2010c; Keim et al., 2010; Kohlhammer et al., 2011). As illustrated in Figure 3, Geovisual Analytics is used in all phases of the analysis approach – from raw mobile phone data exploration to the extraction of spatio-temporal information on variations of collective human activity. Figure 3 provides an overview of the methodology used for exploring mobile phone data, comprising three subsequent phases: first, computation of the mobile communication profiles derived from user-generated mobile phone data; second, generation of temporal trajectories reflecting changes in intensity and similarity within those profiles based on SOM; third, assessment of geospatial dimensions of these trajectories using LISA.

In order to construct temporal trajectories of variations in mobile phone usage according to the five attributes described in Table 1, we employ a visualization technique based upon the computational procedure of the Self-Organizing Map (SOM) (Kohonen, 1990). In essence, a SOM is a data clustering and projection procedure that takes input data of multiple dimensions and arranges it on an output space of a lower dimension (most often 2) so that observations that are most similar to one another across the initial variables are located in close proximity to one another on the output space (Skupin and Agarwal, 2008). The inherently visual output of this procedure makes it an ideal starting point for exploring temporal variations across multiple attribute dimensions (Delmelle et al., 2013).

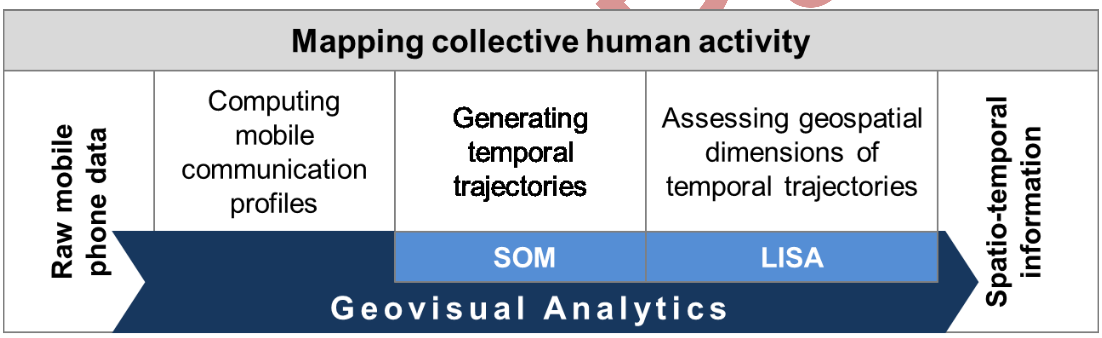
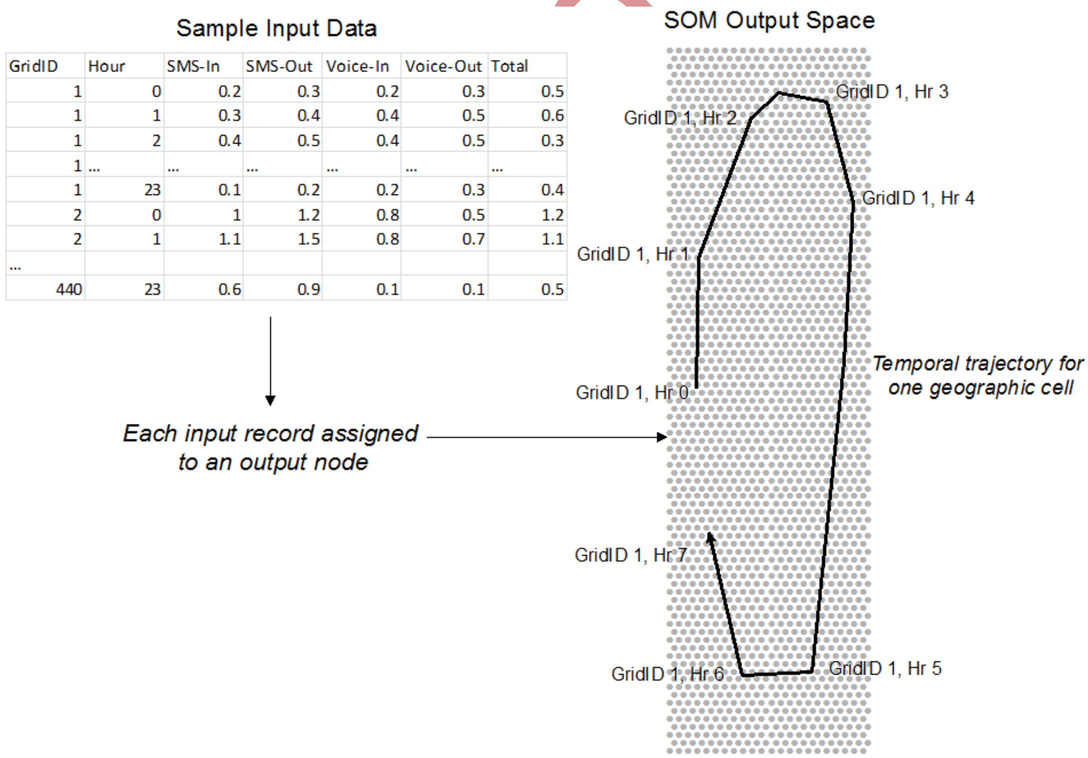


Figure 3: Overview of the Methodology: from raw mobile phone data (left) to spatio-temporal information on collective human activity (right)

A SOM consists of a set of input and output nodes arranged on an output space (see Figure 4 for an example of a 2-dimensional output space). The number of nodes is user defined and largely contingent upon the intended purpose of the SOM procedure: a small number of nodes make the algorithm more akin to a clustering procedure while a large number of nodes, more than the number of input observations, enable emergent structures in the data to be visualized. In this case, the objective is not to serve as a clustering technique per se, but to construct trajectories of attribute change atop this space. A variety of output space sizes were initially tested; our goal was to provide enough space to visualize changes across the spectrum of attribute values. A 2,992 node rectangular grid was finally decided upon (for an input dataset of 105,600 records); for our dataset, increasing the number of nodes continued to produce identical trajectory patterns. Each node contains an *n* dimensional vector of attributes where *n* is the dimensionality of the input space (in this case five). The SOM training procedure is an iterative process where at each step, a random input vector, *x*, is presented to the output space where nodes compete for *x* based on Euclidean distance similarity between *x* and all other vectors. When the best matching node is identified, the values of its attributes are updated to reflect the placement of *x* on the grid. The end result of this process is an ordering of the output space so that neighboring nodes have similar values across the *n* dimensions. For more details on the SOM algorithm, readers are pointed to Skupin and Agarwal (2008).



GridID	Hour	SMS-In	SMS-Out	Voice-In	Voice-Out	Total
1	0	0.2	0.3	0.2	0.3	0.5
1	1	0.3	0.4	0.4	0.5	0.6
1	2	0.4	0.5	0.4	0.5	0.3
1	...	...	...	...	...	...
1	23	0.1	0.2	0.2	0.3	0.4
2	0	1	1.2	0.8	0.5	1.2
2	1	1.1	1.5	0.8	0.7	1.1
...						
440	23	0.6	0.9	0.1	0.1	0.5

Figure 4: From multi-dimensional mobile communication profiles (sample input data) to temporal trajectories of change on the SOM output space

In order to construct trajectories of change across this output space, each observation, in this case, geographic cell and corresponding mobile phone attributes, is entered into the SOM multiple times; once for each time stamp. The observation's position on the output space is then traced for each temporal instance to establish a trajectory (Skupin and Hagelman, 2005; Delmelle et al., 2013). See Figure 4 for an illustration of the procedure. The movement of each trajectory therefore represents changes in the attribute space for a single geographic cell across the 24 hour time span; a greater distance between vertices implies a larger change in attribute values. Note that this approach differs from previous work employing SOM to classify and cluster movement trajectories (e.g.,Owens and Hunter, 2000) in that in our case, the geographic location does not change over time, but the attribute values of the mobile communication profile associated to a geographic location.

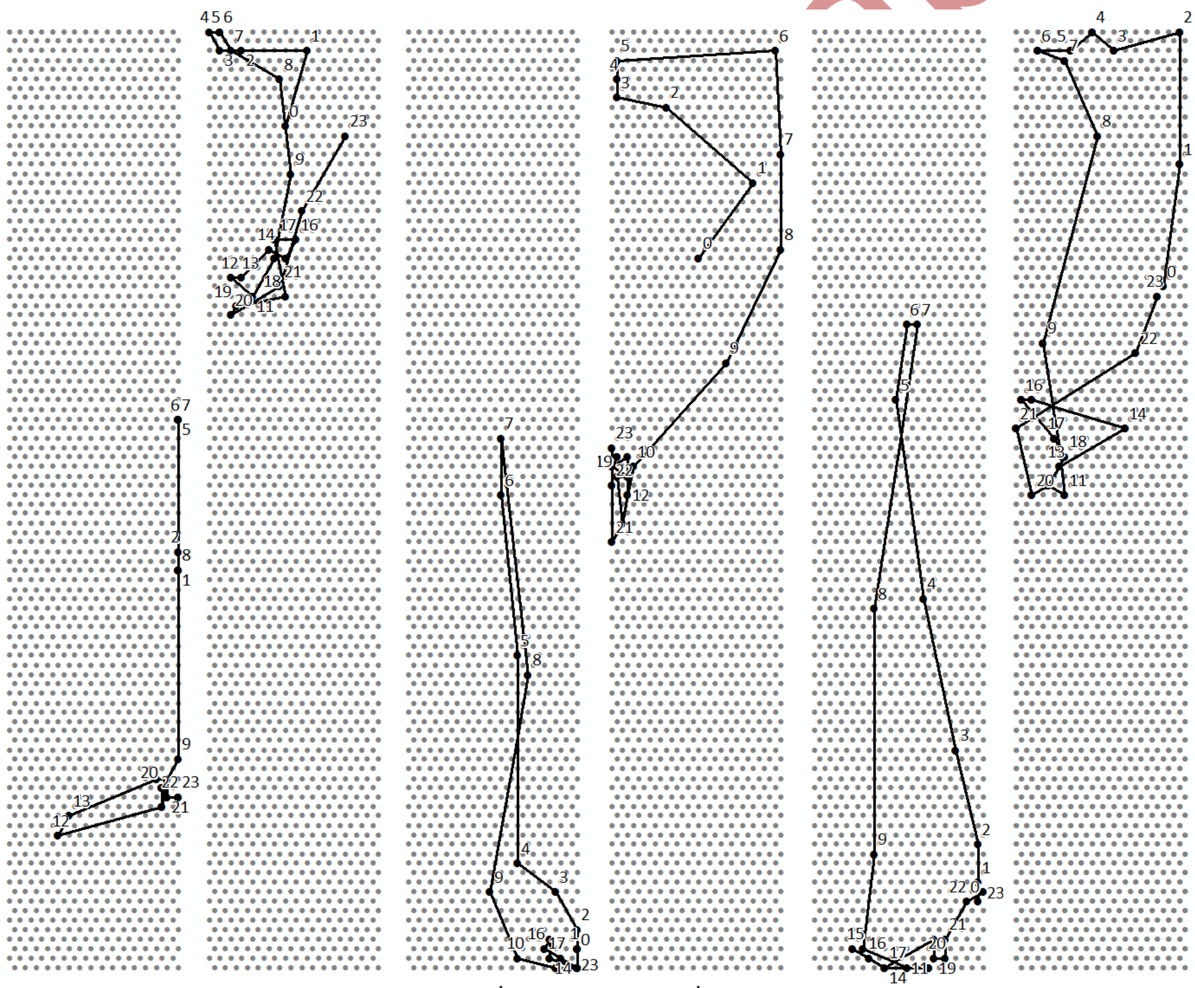
4. Results

4.1 Temporal Trajectories on Self-Organizing Maps

A temporal trajectory describes the typical daily spatio-temporal 'movement' of a given geographic grid cell within the SOM output space (not in the geographic space), thereby reflecting the degree of similarity among the five input variables of the mobile communication profiles (refer to Figure 4). We investigate the temporal dynamics of these mobile communication profiles using three distinct properties of the SOM trajectories as illustrated in Figure 5:

- *Length:* the length of the trajectory is used as an indicator of variation reflecting the changes within and among the five input attributes. We suggest that longer trajectories denote higher variability in the underlying communication profile, in turn reflecting a livelier and more dynamic real-world human activity. In contrast, very short trajectories indicate limited changes in mobile communication behaviour.

- *Location and Extent:* where is the trajectory is located within the SOM output space and how big its bounding box is indicate communication preferences, e.g. an overal predominance of voice over text communication.
- *Shape and Geometry:* the geometrical behaviour of a trajectory (simple versus complex) reflects different levels of fluctuation within the communication profile. A zig-zaged shape indicate, in contrast to a 'smoother' shape, higher variations among such levels.



a a b b c c

Figure 5: Temporal trajectories on the SOM output space (regular point pattern): the location of nodes, which reflect similarity among the five variables, serve as vertices for the corresponding hour of the day (0-23); a, b and c show examples of trajectories with a similar length but different location/extent and shape/geometry, respectively.

Trajectories with similar length but different location on the output space may indicate a similar underlying mobile communication patterns but at a different intensity level. In the remainder of the paper we focus on the ‘length’ property.

Three interesting common characteristics of the temporal trajectories can be identified from Figure 5. First, common overall patterns of movement in the two-dimensional SOM output space, i.e. vertices of the trajectories in chronological order: starting point of each trajectory is at the top, movement down, movement up, ending point at the top again. Second, a ‘big loop’, suggesting the daily circle of human activity in general. Third, a ‘small loop, which is interweaved in the big loop, reflecting the typical double-peak pattern around noon, indicating small variation at a high intensity level. This double-peak pattern is coherent with results from Sevtsuk and Ratti (2010) and Sagl et al. (2012c).

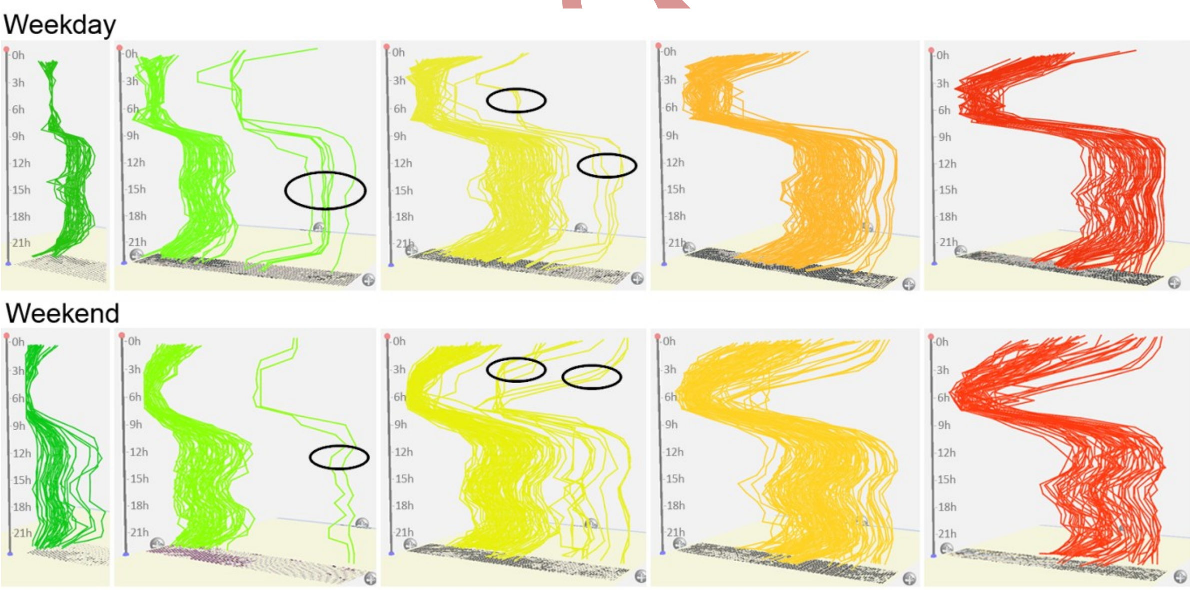


Figure 6: Temporal trajectories in the SOM output space classified by their length on weekdays (top) and weekend (bottom); shortest trajectories in green on the left, longest in red on the right; the black ellipses denote visually outstanding trajectories within the same class (created with GeoTime Software).

Figure 6 shows all 440 trajectories (one per geographic grid cell) on the SOM output space in a space-time cube over the course of a typical weekday (top) or a typical weekend (bottom). The time is shown on the vertical axis from top (early morning) to down (late evening). Emphasis is put on the relative similarity of a trajectory's shape and length among all other trajectories. Further, this visualization technique makes the aforementioned double-peak pattern around noon more recognizable.

We classify the length of SOM trajectories into five classes (following Jenks natural breaks classification method) in order to visually distinguish short and long from middle-length trajectories: from Figure 6 (left to right), the shortest trajectories are in green, while the longest trajectories are in red. All trajectories clearly comprise a similar basic pattern, corresponding to the typical 24 hour temporal signature including the double peak shape around noon (Sevtsuk and Ratti, 2010; Sagl et al., 2012c). In the second-shortest (light green) and the third-shortest (yellow) classes, the trajectories marked with black ellipses visually depart from the majority. The geographic locations of these visually outstanding trajectories are illustrated in Figure 7. The second-longest (in orange) and longest trajectories (in red) experience variability in the early morning (top of the figure): the pattern is more visually clustered on weekdays than on weekends.

4.2 Temporal Trajectories in Geographic Space

The trajectories' lengths are linked back to the geographic space of the test area to explore their geospatial aspects, which allows for identifying varying levels of human activities in diverse urban neighborhoods at different times. As shown in Figure 7, the geographic locations of these visually outstanding trajectories also tend to cluster – this indicates a tight relationship of trajectory characteristics between SOM output space and geographic space. Such a relationship suggests the influence of common underlying drivers, for example, the functional configuration of the urban test area.

Table 2 summarizes the Global Moran's I (Moran, 1948) calculation of the trajectories total length for a typical weekday and a typical weekend, suggesting significant spatial clustering of trajectory length, further underscoring that nearby neighborhoods tend to exhibit similar patterns of mobile phone usages.

Table 2: Global Moran's I summary on the trajectories' total length

	typical weekday	*typical weekend*
Moran's Index	0.713	0.688
Expected Index	-0.002	-0.002
Variance	0.002	0.002
z-score	14.661	14.152
p-value	<0.01	<0.01

To disclose the city structure with respect to collective human activity, we compared, for each grid cell, the trajectory length and the distance between the centroid of the trajectory's corresponding cell to the city center. Since a strong concentric city structure can typically not be assumed, we used Spearman's rho (ϱ) – a nonparametric correlation measure of statistical dependence between two variables based on their ranks of scores rather than their raw scores (Spearman, 2010) – to measure the monotonic rather than the linear relationship between trajectory length and distance to the city center. A relatively strong positive monotonic statistical dependence of ϱ = 0.61 on weekdays reveals that the functionality of neighborhoods located further away is different than neighborhoods located in the inner part of the city (Batty, 2008). The same relationship is observed for weekends, however, at a more modest value of ϱ = 0.56.

To identify regions where trajectory lengths are significantly clustered, we computed the Local Moran's I (Anselin et al., 1996) on the trajectories' total length and the Local Indicators of Spatial Association (LISA). LISA are herein named ‘HH’ = high values surrounded by high values (indicating a hot-spot), ‘LL’ = low values surrounded by low values (indicating a cold-spot), ‘HL’(‘LH’) = high (low) values surrounded by low (high) values (indicating a spatial outlier). Figure 7 shows two maps, one for a typical weekend (left) and another for a typical weekday (right), of Jenks-classified normalized total trajectories length. In addition, Figure 7 shows several hot-spots and cold-spots but, however, no spatial outliers. The spatial pattern of SOM trajectory length indicates that longer trajectories, which refer to higher variations of collective human activity, tend to form separate clusters on a typical weekend as compared to a typical weekday. As shown in Figure 7, the local variations of collective human activity are spatially more homogeneous on a typical weekday than on a typical weekend – this is reflected in the global Moran’s I measures shown in Table 2: 0.713 on a typical weekday as compared to 0.688 on a typical weekend.

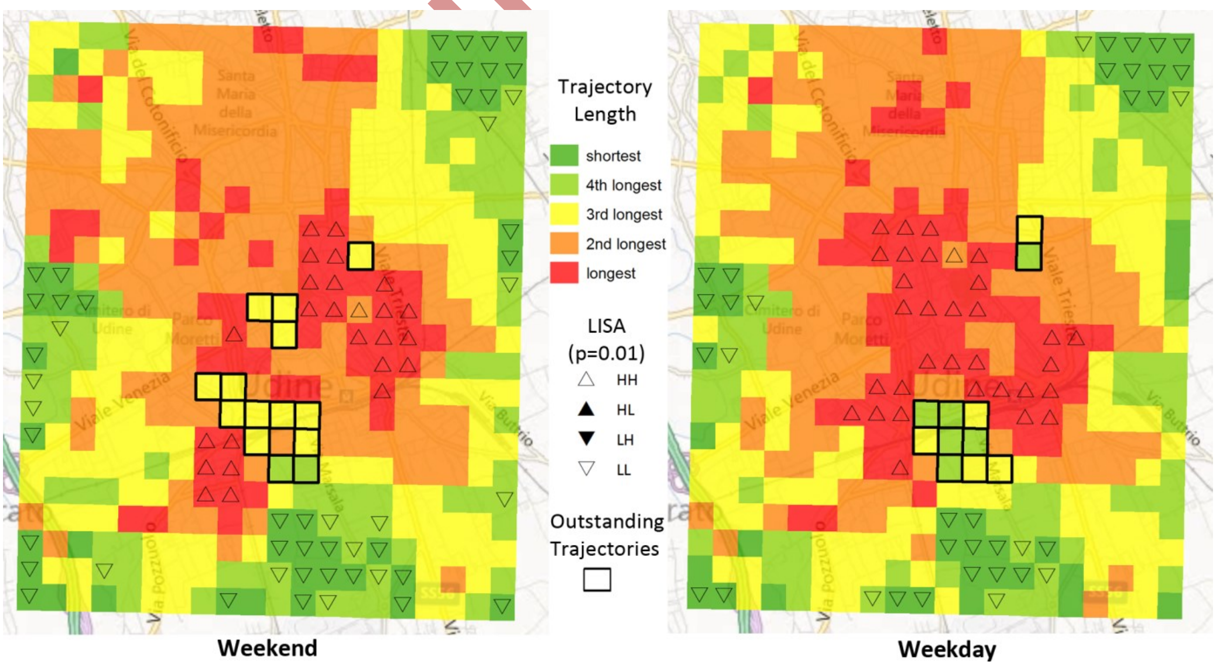


Figure 7: Spatial distribution of normalized trajectory lengths mapped in geographic space (left: typical weekend; right typical weekday); cells of visually outstanding trajectories from Figure 6 are highlighted

To elaborate on the hourly temporal progression of changes in intensity and similarity in mobile communication profiles, we use the length of the trajectory's segments as an indicator for the degree of that change. We split each trajectory into its segments (a trajectory consist of 23 segments, each segment corresponds to a one hour period) and calculated the segments' length. In analogy to the total trajectory length, we normalized all values been between 0 and 1.

Following Wood et al. (2011), Figure 8 summarizes the spatio-temporal pattern of variations of collective human activity on an hourly basis. The common pattern of the temporal progression at almost all cells it the first significant peak between 6 am and9 am on both weekend and weekday. However, a clear right-shift of about 2 hours for the weekend can be observed. In addition, Figure 8 shows places with similar or dissimilar weekend-weekday patterns of variations of collective human activity. Assuming a linear relationship between weekend and weekday variations, we correlate for each cell the temporal signature of a typical weekend with the temporal signature of a typical weekday. High (low) correlation coefficients suggest similar (dissimilar) mobile communication behavior on weekdays and weekend, which we use as an indicator for different functional areas in the city (e.g., residential versus business). Moreover, the resulting correlation coefficients of the temporal signatures, which are geographically referenced due to the cell grid, are then correlated in space in order to pinpoint local spatial associations (LISA) among neighborhoods (refer to Figure 7).

The map in Figure 8 shows, for each geographic cell, the time-graph of variations of collective human activity at weekend and weekday, their Jenks-classified correlation coefficient, as well as several spatial hot- and cold-spots and spatial outliers in the urban test area. Figure 8 further visualizes the locations of visually outstanding trajectories (refer to Figure 6 and Figure 7) that occur on weekend and weekday. These remaining locations show a relatively high correlation of 0.63 ≤ r ≤ 0.82, however, they do not coincide with hot- or cold-spots, or spatial outlier.

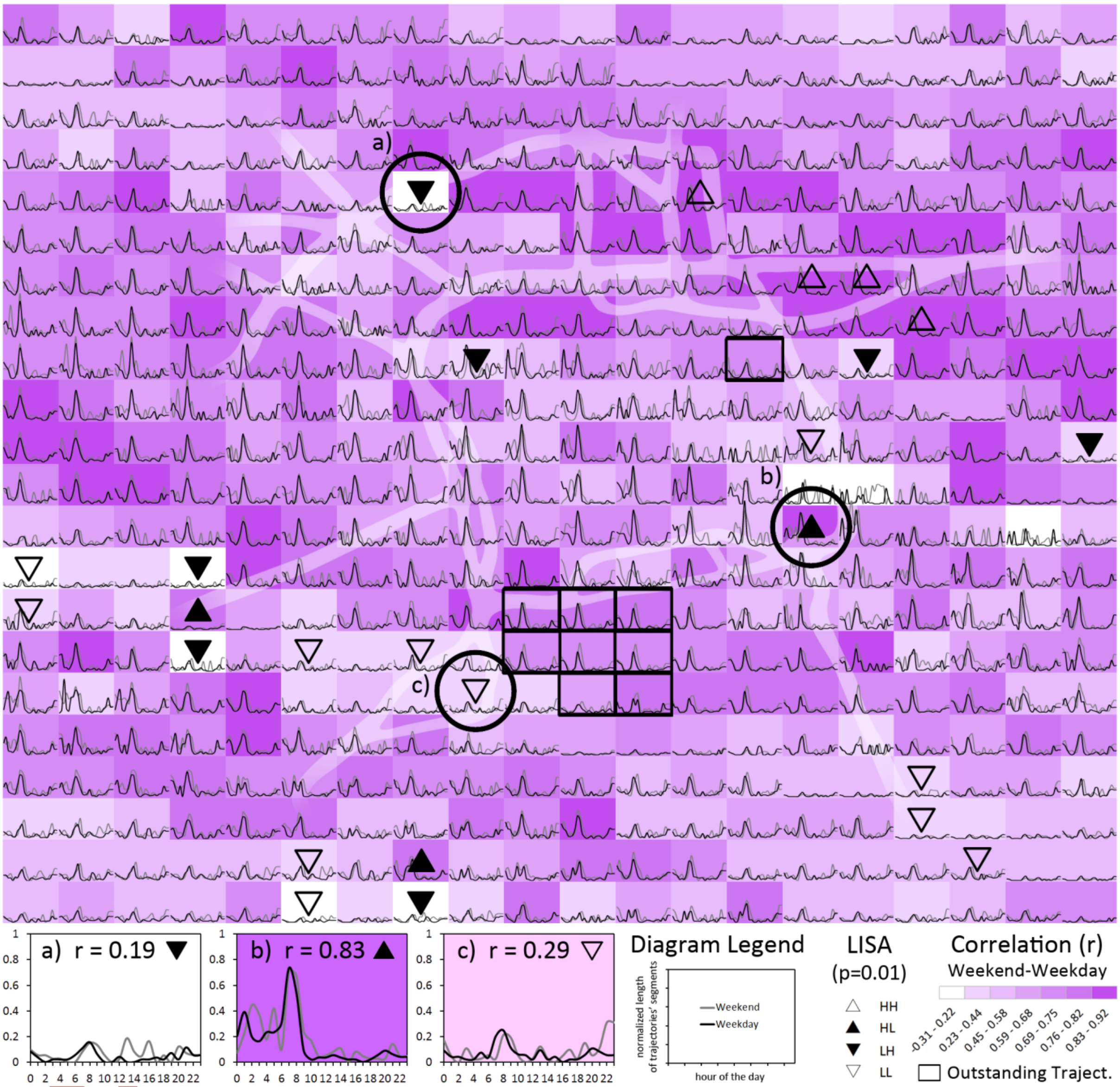


Figure 8: Spatio-temporal patterns of variations of collective human activity: weekend-weekday correlation and LISA of temporal changes in intensity and similarity mobile communication profiles per cell for 24 hours; the graph in each cell shows the normalized length of trajectories' segments for each hour of the day; details a, b and c show examples of two spatial outliers and a cold-spot, as well as their corresponding locations on the grid (major streets are indicated as transparent white lines for orientation purposes).

Three cells (a, b, and c) in Figure 8 are selected to exemplarily show some of their spatio-temporal characteristics in detail. Figure 8a: the cell's weekend and weekday variations are weakly positively related (r = 0.19) but the neighboring cells show a relatively high correlation of $0.58 \leq r \leq 0.89$, resulting in a spatial outlier; Figure 8b: are strongly positively related ($r = 0.83$) but its neighboring cells show correlations ranging from $-0,31 \leq r \leq 0.67$, which results in a spatial outlier; Figure 8c: the cell's weekend and weekday variations are moderately correlated ($r = 0.29$), as the neighboring cells show a moderate correlations of $0.31 \leq r \leq 0.64$.

5. Discussion

We used the length of the trajectory in the SOM output space as an indicator for changes in intensity and similarity, i.e., variations of collective human activity during the course of a typical weekday and a typical weekday, respectively. The classification of the trajectories' length for weekdays and weekends show that some trajectories within the same class comprise significantly different paths than others (Figure 5, and Figure 6 see black ellipses). These trajectories, which are visually clustered in the SOM output space, are also visually clustered in geographic space (Figure 7). When mapped back to geographic space, the characteristics of the trajectories' length reveal additional insights into the social dynamics of the city.

We addressed both the daily (Figure 7) and the hourly (Figure 8) spatio-temporal patterns of variations of collective human activity. Although the results from the global spatial autocorrelation computation (global Moran's I summary in Table 2) indicate a rather homogeneous mobile communication profile landscape, the spatial pattern on weekday shows a larger and more homogeneous hot-spot area as compared to the weekend nonetheless (Figure 7). At weekend the very central hot-spot area almost disappeared and lead to two quite disconnected hot-spot areas.

From a more overall point of view on the spatial distribution of the weekend-weekday correlations, Figure 8 reveals that the Northeastern part of the study area is dominated by relatively high correlations with four hot-spots (Locals Moran's I: HH). In contrast, the southern and south-western part is dominated by relatively low correlations and several cold-spots, as well as five spatial outliers. From the interpretation of these patterns shown in Figure 8, two hypotheses can be set forward:

1. Cells with a low correlation, i.e. different weekend-weekday variations of collective human activity, comprise some dedicated underlying functional drivers such as typical business where weekdays are working days and weekends are off.
2. Cells with a high correlation, i.e. a similar weekend-weekday variations of collective human activity, are assumed to comprise 'unspecified' or 'constant' functional drivers, e.g. recreational areas.

With respect to the detailed view on three selected cells (Figure 8 a, b and c), a 'virtual ground truthing' **Error! Reference source not found.**mainly using Google Maps Street View in combination with results from internet search engines revealed the following. The 'LH' spatial outlier shown in Figure 8a: the grid cell's underlying area is dominated by several parking lots that are next to a university medical center (Figure 9**Error! Reference source not found.**a); the use of such functional urban infrastructure typically follows business hours. In this context, weekends are different from weekdays, which could result in a low correlation value at the given cell while its neighboring cells comprise significantly higher correlation values. This is in line with the daily patterns shown in Figure 7: the greater area called 'Santa Maria della Misericordia' comprises higher variations of collective human activity on weekday than on weekend.

The ‘HL’ spatial outlier shown in Figure 8b: the grid cell’s underlying area is characterized by multistoried buildings with diverse businesses on the ground floor and apartments in the upper floors (Figure 9b), whereas the neighborhood is dominated by apartment buildings. The local mixture of business and living place could thus result in an outstanding high correlation of the hourly variation of collective human activity at the given cell, while the daily variation behaves similarly on weekend and weekday (Figure 7).

Figure 9: Virtual ground truthing: a) several parking lots next to a hospital and medical university in the street 'Via Forni di Sotto'; b) Multistoried buildings with both diverse businesses (ground floor) and apartments in the street 'Viale Ungheria'; c) Urban residential area in the street 'Via Palermo' (Source: Google Street View, last accessed 2013/08/16)

The cold-spot shown in Figure 8c: this cell and its neighborhood are characterized by typical urban residential area comprising multistoried apartment houses, as well as one- or two-family houses (**Error! Reference source not found.**Figure 9c). Residents probably work elsewhere on weekdays but may stay longer at home on weekends. Such a behavior could then result in low correlations of the hourly variation of collective human activity. This is confirmed by the more pronounced hot-spot of the daily variations on weekend as compared to weekday (Figure 7).

With respect to more general spatial patterns of variations of collective human activity, the slightly higher Global Moran's Index (Table 2) for a typical weekday (0.713) indicates a more spatially homogeneous pattern regarding spatial autocorrelation than compared to a typical weekend (0.688). In other words, the variations of collective human activity tend to be more clustered on a typical weekend than on a typical weekday, which implies that the city's social component responds to the city's changing functional configuration. The results are highly significant due to the given z-values and p-values resulting in a confidence level of > 99.9%. This pattern is also visually recognizable from the two maps shown in Figure 7: the spatial distribution of the normalized trajectory length appears more clustered on weekday than on weekends.

The visually outstanding trajectories identified within the SOM output space (Figure 6) are mapped in geographic space (Figure 7). These trajectories' geographic locations, however, do not coincide with hot-spots, cold-spots, or spatial outliers – neither on weekend nor on weekday (Figure 7). This is also true when considering only the locations that are common at weekend and weekday (Figure 8). So, referring to subsection 4.1, the ‘length’ property only of these visually outstanding trajectories does not reveal additional information for variations of collective human activity. These trajectories remain as an interesting starting point for investigating the ‘location/extent’ property.

6. *Conclusions and Further Research*

In this paper we demonstrated an approach for exploring variations in intensity and similarity of collective human activity by analyzing vast volumes of user-generated mobile network traffic data with Geovisual Analytics, Self-Organizing Maps and Local Indicators of Spatial Association. We started with the hypothesis that the mobile communication profiles derived from mobile network data reflect such variations in collective human activity.

Within the comprehensive data sample we distinguished voice from text and incoming from outgoing communication and took the overall network traffic, which includes data transfer, video communications, social media interaction etc., as a base reference. We then computed an average 24 hour mobile communication profile for every single 250m x 250m grid cell (the spatial unit to which the data were spatially aggregated and provided by the network operator) for weekdays (Monday to Friday) and weekends (Saturday and Sunday). These temporally and spatially fine-grained mobile communication profiles served as input for the analysis based on SOM. The resulting temporal trajectories were characterized within the SOM output space and linked back to geographic space, the urban environment of Udine, Northern Italy. The spatial distribution and spatial autocorrelation of the trajectory length represent variations of collective human activity.

Generally, the results tend to confirm the working hypothesis stated in the introduction. In fact, the virtual ground truthing shows that, e.g., the details of Figure 8a, as discussed above, tend to confirm the first generated hypothesis: parking lots as a dedicated functional infrastructure tend to imply different levels of variation within human activity on weekdays as compared to weekends. Further, the tendency that higher variations of collective human activity (i.e., longer trajectories) are located closer to the city center could – following Batty (2008) – indicate a mono-centric city structure in terms of social urban dynamics. It is therefore legitimate to assume that the city's functional configuration provides certain underlying drivers that influence the collective human behavior which is, to some degree, reflected in the way people use the mobile phone network. Consequently, certain spatial and temporal variations of collective human activity are, to some degree, reflected in changes in the mobile communication profiles – as we have demonstrated herein.

We conclude that different parameter of user-generated mobile phone data can serve as the basis for investigating variations of collective human activity. The methodology developed and used is rooted in GIScience and comprises interdisciplinary methods and techniques. Geovisual Analytics as an inductive reasoning framework can successfully be applied to explore large volumes of multivariate data and therefore supports all phases of the analysis. The mechanism of SOM as a foundation for visualizing temporal trajectories across multiple dimensions is shown to be a reasonable technique to exhibit the variations of collective human activity from mobile communication profiles. In geographic space, the use of LISA is demonstrated to be effective in assessing the geospatial aspects of these variations. To sum up, we demonstrated an innovative approach to explore variations in collective human activity based on communication profiles derived from mobile phone data. This approach advances existing research and provides additional insights into urban social dynamics.

The results confirm that variations of collective human activity in diverse neighborhoods at different times correlate with the functional configuration of the city. This has been demonstrated on the example of Udine, a medium-sized North-Italian urban environment. Our work adds a complementary perspective to look at spatial and temporal variations of typical collective human activity patterns. This research thus contributes to a better understanding of the dynamic social component in complex urban systems.

Our case study area of Udine presented some limitations in the SOM trajectory analysis: because of the sharp distinction between the urban and suburban area, the five mobile phone variables were highly correlated across the two-dimensional output space. Therefore, the identification of more nuanced clusters – i.e. locations with high incoming SMS and mobile phone calls, but low outgoing information could not be gleaned. A larger urban area and a smaller geographic division of the city would likely be necessary for such a distinction. However, the methodology illustrated here serves the potential to trace longer-term temporal transformations of the urban area from one functional description to another. Finally, further research should explore variations in virtual social interaction based on data from diverse social media sources, thereby explicitly taking into account VGI. The integration of VGI content extraction mechanisms such as natural language processing algorithms potentially enables a more context-sensitive perspective on urban social dynamics.

Acknowledgement

The authors thank Euro Beinat for his valuable thoughts with respect to social urban dynamics and collective sensing. The authors also thank Curtis Garton, Product Manager at Oculus Info Inc., for providing a test license of GeoTime V5.3 software. Finally, the authors thank the reviewers for their constructive and valuable feedback.

References

Andrienko, G., Andrienko, N., Bak, P., Bremm, S., Keim, D., Landesberger, T. v., Politz, C. & Schreck, T. 2010a. A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage. *J. Locat. Based Serv.,* 4**,** 200-221.

Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P. & Keim, D. 2010b. Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns. *Computer Graphics Forum,* 29**,** 913-922.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., Jern, M., Kraak, M.-J., Schumann, H. & Tominski, C. 2010c. Space, time and visual analytics. *International Journal of Geographical Information Science,* 24**,** 1577-1600.

Anselin, L., Bera, A. K., Florax, R. & Yoon, M. J. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics,* 26**,** 77-104.

Batty, M. 2008. The Size, Scale, and Shape of Cities. *Science,* 319**,** 769-771.

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D. & Ratti, C. 2011a. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems,* 12**,** 141-151.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J. & Ratti, C. 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies,* 26**,** 301-313.

Calabrese, F., Smoreda, Z., Blondel, V. & Ratti, C. 2011b. The interplay between telecommunications and face-to-face interactions - an initial study using mobile phone data. *Physics and Society [physics.soc-ph]*.

Candia, J., González, M. C., PuWang, Schoenharl, T., Madey, G. & Barabási, A.-L. 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical,* 41**,** 224015.

Delmelle, E., Thill, J.-C., Furuseth, O. & Ludden, T. 2013. Trajectories of Multidimensional Neighbourhood Quality of Life Change. *Urban Studies,* 50**,** 923-941.

Di Lorenzo, G. & Calabrese, F. 2011. Identifying human spatio-temporal activity patterns from mobile-phone traces. *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on* Washington, DC, USA.

Eagle, N. & Pentland, A. 2009. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology,* 63**,** 1057-1066.

Farrahi, K. & Gatica-Perez, D. 2011. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Trans. Intell. Syst. Technol.,* 2**,** 1-27.

González, M. C., Hidalgo, C. A. & Barabási, A.-L. 2008. Understanding individual human mobility patterns. *Nature,* 453**,** 779-782.

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal,* 69**,** 211-221.

Goodchild, M. F. & Janelle, D. G. 1984. The city around the clock: space - time patterns of urban ecological structure. *Environment and Planning A,* 16**,** 807-820.

Hägerstrand, T. 1970. What about people in Regional Science? *Papers in Regional Science,* 24**,** 6-21.

Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F. (eds.) 2010. *Mastering the Information Age - Solving Problems with Visual Analytics,* Goslar: Eurographics Association.

Kohlhammer, J., Keim, D., Pohl, M., Santucci, G. & Andrienko, G. 2011. Solving Problems with Visual Analytics. *Procedia Computer Science,* 7**,** 117-120.

Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE,* 78**,** 1464-1480.

Krings, G., Calabrese, F., Ratti, C. & Blondel, V. D. 2009. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment,* 2009.

Krygsman, S., de Jong, T. & Schmitz, P. 2007. Capturing daily urban rhythms: The use of location aware technologies. *Computers in Urban Planning and Urban Management, 10th International Conference.* Iguassu Falls, Brazil.

Kwan, M.-P. 2000. Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems,* 2**,** 85-90.

Moran, P. A. P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological),* 10**,** 243-251.

Neuhaus, F. 2010. UrbanDiary-A Tracking Project Capturing the beat and rhythm of the city: Using GPS devices to visualise individual and collective routines within Central London. *The Journal of Space Syntax,* 1**,** 315-336.

Neuhaus, F. 2011. UrbanDiary. *In:* NEUHAUS, F. (ed.) *Studies in Temporal Urbanism.* Springer Netherlands.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. 2012. A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLoS One,* 7**,** e37027.

Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. & Barabási, A. L. 2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences,* 104**,** 7332-7336.

Owens, J. & Hunter, A. Application of the self-organising map to trajectory classification. Third IEEE International Workshop on Visual Surveillance, July 1 2000 Dublin, Ireland. 77-83.

Perreault, M. & Ruths, D. 2011. The effect of mobile platforms on Twitter content generation. *Proceedings of the Fifth International Conference on Weblogs and Social Media.* Barcelona (Spain).

Pulselli, R., Ramono, P., Ratti, C. & Tiezzi, E. 2008. Computing urban mobile landscapes through monitoring population density based on cellphone chatting. *Int. J. of Design and Nature and Ecodynamics,* 3**,** 121-134.

Quercia, D. L., N.; Calabrese, F.; Di Lorenzo, G.; Crowcroft, J.;. Recommending Social Events from Mobile Phone Location Data. 10th IEEE International Conference on Data Mining (ICDM2010), December 13-17 2010 Sydney, Australia. 971-976.

Ratti, C., Pulselli, R. M., Williams, S. & Frenchman, D. 2006. Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design,* 33**,** 727-748.

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.

Rubio, A., Frias-Martinez, V., Frias-Martinez, E. & Oliver, N. 2010. Human Mobility in Advanced and Developing Economies: A Comparative Analysis. *AAAI 2010 Spring Symposia Artificial Intelligence for Development, AI-D 2010.* Stanford, USA.

Sagl, G., Blaschke, T., Beinat, E. & Resch, B. 2012a. Ubiquitous Geo-Sensing for Context-Aware Analysis: Exploring Relationships between Environmental and Human Dynamics. *Sensors,* 12**,** 9835-9857.

Sagl, G., Loidl, M. & Beinat, E. 2012b. A Visual Analytics Approach for Extracting Spatio-Temporal Urban Mobility Information from Mobile Network Traffic. *ISPRS International Journal of Geo-Information,* 1**,** 256-271.

Sagl, G., Resch, B., Hawelka, B. & Beinat, E. 2012c. From Social Sensor Data to Collective Human Behaviour Patterns: Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments. *In:* JEKEL, T., CAR, A., STROBL, J. & GRIESEBNER, G. (eds.) *GI-Forum 2012: Geovisualization, Society and Learning.* Wichmann Verlag, Berlin.

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.

Shoval, N. 2007. Sensing human society. *Environment and Planning B: Planning and Design,* 34**,** 191-195.

Skupin, A. & Agarwal, P. 2008. Introduction: What is a Self-Organizing Map? *Self-Organising Maps.* John Wiley & Sons, Ltd.

Skupin, A. & Hagelman, R. 2005. Visualizing Demographic Trajectories with Self-Organizing Maps. *Geoinformatica,* 9**,** 159-179.

Spearman, C. 2010. The proof and measurement of association between two things. *International Journal of Epidemiology,* 39**,** 1137-1150.

Williams, M. J., Whitaker, R. M. & Allen, S. M. Measuring Individual Regularity in Human Visiting Patterns. Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), 3-5 Sept. 2012 2012. 117-122.

Wood, J., Slingsby, A. & Dykes, J. 2011. Visualizing the Dynamics of London's Bicycle-Hire Scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization,* 46**,** 239-251.

Yuan, Y. & Raubal, M. 2012. Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data. *In:* XIAO, N., KWAN, M.-P., GOODCHILD, M. & SHEKHAR, S.

(eds.) *Geographic Information Science, Proceedings of the 7th International Conference.* Springer Berlin / Heidelberg.