

# Urban Mobility Analysis With Mobile Network Data: A Visual Analytics Approach

Hansi Senaratne, Manuel Mueller, Michael Behrisch, Felipe Lalanne, Javier Bustos-Jiménez, Jörn Schneidewind, Daniel Keim, and Tobias Schreck

**Abstract**—Urban planning and intelligent transportation management are facing key challenges in today’s ever more urbanized world. Providing the right tools to city planners is crucial to cope with these challenges. Data collected from citizens’ mobile communication can be used as the foundation for such tools. These kinds of data can facilitate various analysis tasks, such as the extraction of human movement patterns or determining the urban dynamics of a city. City planners can closely monitor such patterns based on which strategic decisions can be taken to improve a city’s infrastructure. In this paper, we introduce a novel visual analytics approach for pattern exploration and search in global system for mobile communications mobile networks. We define geospatial and matrix representations of data, which can be interactively navigated. The approach integrates data visualization with suitable data analysis algorithms, allowing to spatially and temporally compare mobile usage, identify regularities, as well as anomalies in daily mobility patterns across regions and user groups. As an extension to our visual analytics approach, we further introduce space-time prisms with uncertain markers to visually analyze the uncertainty of urban mobility patterns.

**Index Terms**—Visual analytics, mobile network data, urban dynamics, spatial and temporal patterns, mobility, intelligent transportation system.

## I. INTRODUCTION

WITH more than 50% of world population living in urban areas today,<sup>1</sup> it is imperative to foster careful planning of city infrastructure, especially urban transportation. Ubiquitous data such as location-enabled *GSM (Global System for Mobile Communications) mobile communication data* can be utilized as alternatives to officially surveyed data, and therefore used as digital footprints to understand the human movement patterns in an urban environment and the urban dynamics [1]. Deriving actionable, useful knowledge from such large amounts of complex GSM data require appropriate

data analysis methods and models. Visual-interactive data analysis methods can help to support data analysis by effective integration of automatic data analysis methods, interactive steering of the analysis process, and result visualization [2].

In this paper we explore a GSM internet usage dataset collected from 358 users over seven months from Santiago de Chile, and introduce a suite of visual analytics methods for the following tasks: (1) extract movement patterns of users out of GSM internet usage data, and (2) explore spatio-temporal patterns through the derived movements. In case of such mobile internet usage data collected from users, the location can be approximated by the location of the closest receiver antenna. Unlike geographical coordinates that are extracted from accurate GPS devices, this results in higher positional uncertainty. Therefore, we extend the developed visual analytics approach to (3) further assess the *uncertainty* of extracted user movements.

A suite of analytical views are defined to foster data exploration within the specified tasks. These are: (i) *map view*, to represent the spatial references of the selected data points, such as the movement trajectories (Sections IV-A and IV-E). (ii) *matrix view* helps in the analysis of user movements. This view is utilized for example to compare the similarity of users’ movement profiles (Sections IV-A, IV-D, and IV-E). As we consider each user and each location as a dimension of data, compact and ordered matrix visualizations were helpful to efficiently and visually explore the sample-sample relationships, as opposed to the use of scatterplots or parallel coordinate plots that are inefficient particularly with high dimensions or in terms of the efficient use of the display [3]. (iii) *tabular view and tree node view* shows the unique antenna IDs, user IDs, area codes and other attributes such as the signal strength, to investigate each antenna or user individually. The tree node view with its inherent hierarchical structure was a coherent choice as opposed to a geographic map, for analysing the administrative hierarchies in Santiago (Section IV-C). (iv) *raster map view* helps to differentiate urban areas in terms of an aggregated value, such as the internet usage patterns over the Western, Eastern, Northern, and Southern communes of Santiago (Section IV-D). (v) *small multiples view* displays difference slices of a selected attribute in the dataset. To avoid over-plotting and confusion, this view aligns the data in an orderly manner to enable comparisons across different variables (Section IV-E).

The sum of analysis methods developed in this paper demonstrates how an encompassing visual interactive analysis system can facilitate the exploration of GSM mobile internet

Manuscript received August 9, 2016; revised January 22, 2017 and May 26, 2017; accepted July 2, 2017. Date of publication August 25, 2017; date of current version May 2, 2018. This work has been supported by the German Research Foundation DFG within Strategic Research Initiative SPP 1335, and the advanced human capital integration program of the Chilean science and technology commission (CONICYT), as well as the STIC AmSud Research Program with the MOSAIC project. The Associate Editor for this paper was W. Chen. (Corresponding author: Hansi Senaratne.)

H. Senaratne, M. Mueller, M. Behrisch, and D. Keim are with the University of Konstanz, 78457 Konstanz, Germany (e-mail: hansi.senaratne@uni-konstanz.de).

F. Lalanne is with Inria, 2827 Santiago, Chile.

J. Bustos-Jiménez is with the Universidad de Chile, 1058 Santiago, Chile.

J. Schneidewind is with Telefonica Germany, 80992 Munich, Germany.

T. Schreck is with the University of Graz, 8010 Graz, Austria.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2017.2727281

<sup>1</sup><http://www.demographia.com/db-worldua.pdf>

usage data. We demonstrate the usefulness of our system by validating a prototype through anecdotal use-case findings [4]. We have further incorporated the feedback of an expert from the telecommunications domain, that discusses the usefulness of the proposed visual analytics system. In his expert view, the spatial and temporal pattern extraction from GSM data is an interesting and an important task for exploring the urban dynamics in a city. The expert feedback has been appropriately reflected in the development of our approach.

The contribution of this paper is therefore the development of a suite of visual analytics methods to support urban movement analysis in five aspects: (1) similarity of user movements (Section IV-A), (2) home and work area classification (Section IV-C), (3) origin-destination based region partitioning (Section IV-D), and (4) temporal change detection (Section IV-E). Furthermore, at the end of this paper we introduce (5) space-time prisms with *uncertain markers* to model the positional uncertainty of movement (Section V).

The insights obtained enable novel applications ranging from transportation management to general urban planning and smart city management. The remainder of the paper unfolds as follows: In the following section we briefly review related work, and in Section III we describe the dataset in detail. In Section IV we present the main analysis. In Section V we present an extension to the suit of visual analytics methods for visually analyzing the uncertainty of movement data. Section VI presents a discussion followed by a conclusion of the presented methods in Section VII.

## II. RELATED WORK

Reference [5] developed a visual analytics approach for analyzing collective human mobility based on spatio-temporal patterns. For the visual analysis of their data (e.g., using time series visualizations) they used Tableau<sup>2</sup> together with SPSS Statistics<sup>3</sup> software, and for the spatial visual analysis of the data (e.g., using heat-map visualizations) they used ESRI ArcGIS<sup>4</sup> commercial software. Similar work of [6] uses visual analytics to explore collective behavior of humans. In [7] MobVis is presented, a tool that incorporates spatial and social data heterogeneously into one network to analyze individual and group behavioral patterns. They use interactive time charts and ontology graphs to support the temporal and semantic filtering of their data. They further introduce a *behavior rings* visualization method to represent and compare individual and group behaviors of people.

Furthermore, [8] utilizes one week's mobile phone data to analyze the structures of spatial interaction communities of a city in China. Their approach mainly utilizes agglomerative clustering techniques together with heat map and graph visualizations to discover the structures of spatial interaction in communities. Further the work of [9] used taxi data to explore urban activities in New York. They developed a visual analysis system that allows the users to visually query various taxi trips around the city. Tasks such as origin-destination

of human mobility were performed by enabling adequate spatio-temporal querying, using mainly heatmap and time-series visualizations. Similar work has also been carried out in [10]. Reference [11] presented the MobilityViewer, a suite of visual analytics methods to analyze the crowd flows in a city within a geographical and a social network, using mobile phone data. Unlike our approach where all views and methods are integrated into one system, [11] have their methods as isolated views.

Further, [12] presented a visual analytics system to explore co-occurrences in human mobility. To make these observations they develop six different views- each view visualizing the human mobility in different perspectives and the co-occurrences. Our approach however, utilizes each view to perform different spatio-temporal analytics and thereby uncover urban patterns. Reference [13] utilized public transportation data to make home and work area classifications. Their developed visual analytics system analyzes complex travel patterns and activity locations of each individual commuter, learns these classifications and then trains a classifier to derive commuter dynamics. To achieve the same goal we utilize the movement trajectories traversed during established core working hours and stay-at-home hours to distinguish between work and home areas respectively. A comprehensive survey on visual analytics approaches for urban analyses is in [14].

In this paper in addition to movement patterns, we further use GSM mobile internet usage attributes to determine urban patterns and show correlations. Furthermore, we introduce a method to address the positional uncertainties in the movement trajectories that are derived from approximated antenna locations.

## III. THE DATASET

The dataset comes from the Adkintun Mobile project [15]. Adkintun Mobile performs *passive* monitoring as a means for data gathering, implying that it only listens to changes in the network and device status over time. This means that it can provide a sustained measurement at regular intervals for a long period of time, improving the chances to perform temporal analysis, such as those described in the current work. The data was collected on a voluntary basis by 358 users, who were required to install an application on their Android phones. The data was recorded in a period of seven months from October 2012 to May 2013, where the mobile application logged device and network usage data every ten seconds. Each record contains information about the network traffic (received/sent bytes), the antenna operator, the Location Area Code of the currently logged-on antenna (LAC), the Cell Identifier (CID), the signal strength (measured in dBm) and a measure which indicates whether the phone was in active use at the time of the measurement. A second table contains coordinates of over 60,000 antennas across Chile along with their respective LAC and CID. However, within the scope of this study we use data for only the city limits of Santiago. The data is stored in a PostgreSQL 9.3 database with the spatial extension PostGIS.

<sup>2</sup><http://www.tableau.com/>

<sup>3</sup><http://www-01.ibm.com/software/analytics/spss/>

<sup>4</sup><http://www.esri.com/software/arcgis>

#### IV. LEARNING URBAN DYNAMICS THROUGH MOVEMENT PATTERNS

Exploration of mobile phone datasets can provide us insights into the urban dynamics of a city that help various stakeholders to cater better decisions and services to their targeted audiences. We carry out our analysis tasks in two tiers. Firstly, we focus on the *spatial changes* of mobile internet usage patterns. For this we extract user movement trajectories through the digital footprints left behind by mobile users, and use these changing movements of users to uncover spatial patterns in the metropolitan area of Santiago, Chile. Secondly, we focus on the *temporal changes* of mobile internet usage and use these to uncover spatial patterns and events.

##### A. Movement Trajectory Extraction from Mobile Internet Usage Data

A trajectory is a trace generated by user movement in geographical space, usually represented by a series of chronologically ordered points  $(p_1, p_2, \dots, p_n)$ , where each point consists of a geospatial coordinate set and a time stamp such as  $p = (x, y, t)$ .

The monitor data allows us to approximate trajectories of users with antenna positions as locations. This means that a user can be anywhere within the operational range of an antenna. This low spatial granularity of locations aids us in analysis through general regions as opposed to exact locations. This aspect is important also for the anonymization of the data. Our trajectory extraction algorithm consists of two steps. First, we define and detect a *movement session*. A session is defined as a time interval of continuous Internet usage of one user, that contains a set of  $n$  tuples  $(t|s)$  where  $t$  is a point in time and  $s$  a point in space. A session is considered to be over when the time difference of two consecutive monitor records exceeds a user defined threshold. Therefore, a session  $S$  can be defined as follows:

$$S = (t_1|s_1), (t_2|s_2), \dots, (t_n|s_n)$$

In the second step we extract the trajectory out of these sessions where each session gets divided into one or more trajectories. To decide where one trajectory ends and another starts, the spatial element of the  $(t|s)$  tuples is examined. Whenever a sequence of tuples without spatial change exceeds a certain duration threshold, the current trajectory is considered to be over and a new one is created. This is expressed in the following trajectory extraction algorithm:

Consecutive tuples of a session which have the same spatial component are grouped together. These trajectories can be extracted at different granularity. Highest granularity is when antenna locations are used as spatial elements of the tuples (Fig. 1), and lowest when the cluster centers are used as spatial elements (bigger clusters therefore result in coarser trajectories).

We have used these extracted movement trajectories to identify patterns in the surrounding urban environment. These are presented in the following sections.

---

#### Algorithm 1: Algorithm for Extracting Trajectories From Mobile Internet Sessions

---

**Data:**  $S$ , threshold  $k$   
**Result:** A set of trajectories  
 Start a new trajectory from  $tuple_i$  ( $i = 0$ );  
**foreach** following tuple **do**  
   Find closest cluster  $C_i$ ;  
   **if**  $|t_i - t_{i+1}| < k$  **then**  
     add  $tuple_{i+1}$  to the current trajectory;  
   **else**  
     end current trajectory and start a new trajectory  
     from  $tuple_{i+1}$ ;  
**end**  
**end**

---

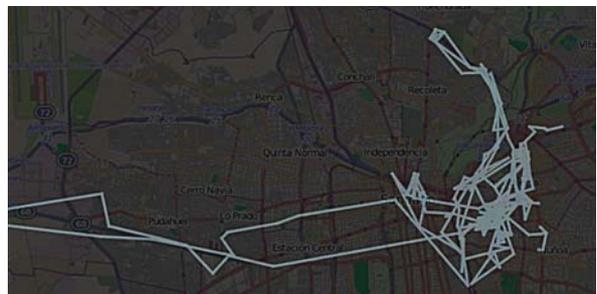


Fig. 1. Trajectory of a selected user with antenna locations taken as the consecutive user locations.

##### B. Spatial and Temporal Movement Similarity of Users

Similarity between two objects is said to be estimated with a *similarity measure* which may contain a cost of changing one feature into another, or the *distance between the two objects* [16]. Several works have introduced and discussed approaches for the analysis of similarity in movement trajectories. The work of [17] presents a two-tiered approach for assessing the similarity between movement trajectories. As a first step they decompose trajectories into segments based on parameters such as speed, acceleration, or directions of movement. In a second step they modify the Levenshtein edit distance technique [18] to assess the similarity of decomposed trajectory segments.

Reference [19] on the other hand present a taxonomy of approaches for similarity analysis based on the different parameters of a moving object, such as the speed, spatial path, and time duration. Similarity analysis within two or more movement trajectories help in various demographic studies for focus group marketing, urban planning, or city traffic flow management – for example to detect which areas in a city are most connected. Reference [20] suggested several ways in which two or more movement trajectories may be similar. These are: (1) *Similarity of overall characteristics* (2) *Co-location in space* (3) *Synchronization in time* (4) *Co-occurrence in space and time*. The authors further gave suggestions of computational and visualization techniques for detecting similarities between movement trajectories.

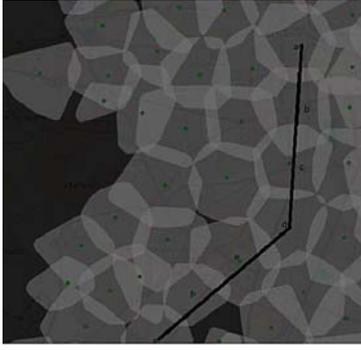


Fig. 2. Example of two edges (a-d and d-g) of a trajectory segment passing through several locations (b,c,e,f).

To assess and compare the spatial and spatio-temporal similarity between users' movement trajectories, we introduce a similarity-based matrix visualization. This extends the classification of techniques suggested by [20] for similarity analysis in movement trajectories. We consider *co-location in space* and *co-incidence in space and time* for similarity assessment in our derived movement trajectories of mobile users. To calculate the spatial similarity of users, we compare the movement trajectory of each user for possible co-locations. To determine the locations that the edges of the trajectory passes through (as demonstrated in Fig. 2), we calculated for each edge the intersection with other locations of the same level of granularity. The results are stored in a separate table. Using this table, we built a spatial profile for each user by extracting their distinct locations.

Accordingly, the spatial profile  $P$  for a user  $x$  at granularity  $g$  can be denoted as  $P(x, g) = a, b, c, d, e, f, g$ . Then, the similarity of two users  $x$  and  $y$  at granularity  $g$  can be calculated as:

$$\text{sim}(x, y, g) = \frac{|P(x, g) \cap P(y, g)|}{\min(|P(x, g)|, |P(y, g)|)}$$

To visualize similarity values of each user pair, we map them to the size of a circular gradient (Fig. 3 a, b). The size of the gradients is scaled between 0 (smallest) and 1 (largest) to visualize the normalized low-to-high similarity values respectively. These circular gradients are then drawn in the corresponding matrix cell. The adjacent map view visualizes the movement profiles of the corresponding user pair (Fig. 3 c).

To analyze if the movement of two users coincide in space and time, we check if any two spatially similar users (denoted by  $L(x, y, g)$ ) have as well an overlap in time (using a time interval of 5 minutes) at the given location. If there is an overlap, we add the location to the set  $O(x, y, g)$ . We calculate temporal similarity as follows:

$$\text{sim}(x, y, g) = \frac{|O(x, y, g)|}{|L(x, y, g)|}$$

We add this temporal component to the similarity matrix. The temporal similarity values are mapped to the opacity of the corresponding matrix cell. The results are shown in Fig. 4. The visualization of temporal similarity with the spatial similarity helps city planners for user demographic analysis.

### C. Place Classification Based on Home and Work Area Detection

Place classifications help to identify geographic space in terms of its unique characteristics. Thus far semantic technologies have been very popular to identify *place* using topological and semantic descriptions [21], [22]. Furthermore, the movement of people can also give us hints to the characteristics of the underlying geographic space. For example, three main forces have been identified by [23] for what drives humans to move within space and time. These are (1) *attendance*: decision to attend an event at a certain time and place, (2) *existence*: for day-to-day survival, and (3) *imitation*: following what someone else does. If we can identify one, or all of these cues of human movement we can generate and identify characteristics of space that can lead to a place classification.

In our work we use movement trajectories derived from the mobile Internet usage dataset to identify characteristics of geographic space. For this we use [23]'s *existence* as a movement cue. Since humans are creatures of habit, we can exploit their daily movement patterns to observe when they are travelling to work, and when they are travelling home. With these identified patterns, we want to classify geographic space according to residential and industrial areas. One possible application of this place classification is for real estate developers to identify their target customer groups within residential (home) and work (industry) areas. In the following we demonstrate how we achieve this for the city of Santiago, using visual analytics methods.

Forming user groups is helpful to use as an abstraction to conduct analysis on, and to help speeding up analysis in contrast to the long computation times required for ungrouped data. For this analysis we used an interactive tree node view.

Following the works of [24], we first extract the home and work locations of users based on their spatio-temporal mobile internet usage patterns. We performed the detection using a 500m clustering granularity. For each user, the following steps were performed.

#### Work area detection

1. Determine the set of distinct antenna clusters visited by the user during weekdays between the time 13:00 and 17:00 (which is a core working time for most people according to [24]).
2. Rank clusters by the number of days when the cluster was visited. If the ranking is not clear, take the total number of records into account.
3. Pick the cluster with the highest rank as work cluster.

#### Home area detection

1. Determine the set of distinct antenna clusters visited by the user during weekdays between the time 00:00 and 06:00 and between 20:00 and 00:00, or during weekends.
2. Rank clusters by the number of days when the cluster was visited. If the ranking is not clear, take the total number of records into account.
3. Pick the cluster with the highest rank as home cluster.

In a subsequent step, we used a reverse geotagger which takes a geo-location as input and outputs information such

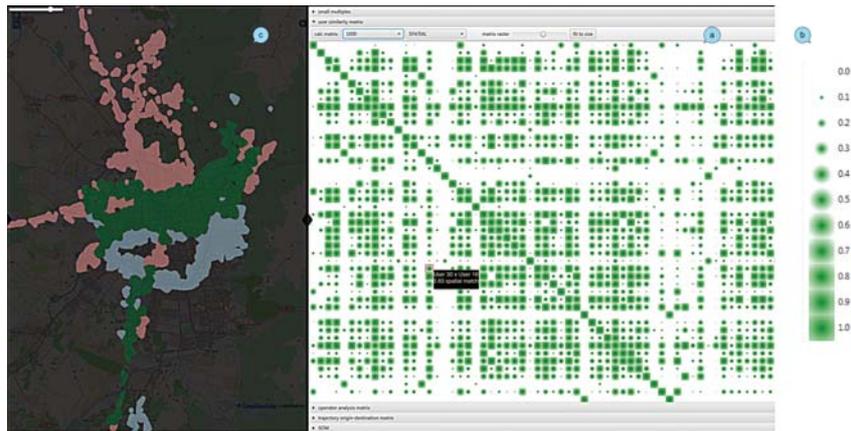


Fig. 3. Spatial similarity matrix of users. (a) The matrix overview facilitates the analysis of unique movements which are represented through blank lines. The larger the circular gradient, the higher is the similarity. (b) Legend shows the scaling of the circular gradient. Specific user similarities can be analyzed by zooming in. (c) Each matrix cell can be clicked to visualize the movement profile of the corresponding user pair on the map view. The movement profiles of the selected User 30 is shown in red and User 16 is shown in blue. The intersection of both user movements with a spatial similarity value of 0.63 is shown in green.

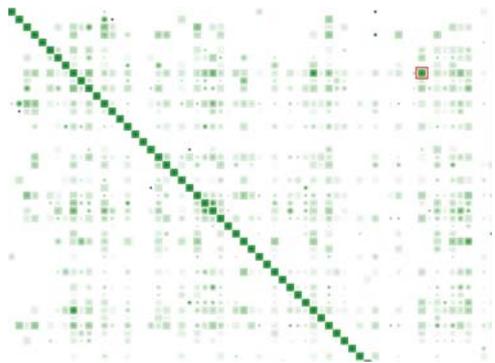


Fig. 4. An excerpt of the spatial similarity (circular gradient) and temporal similarity (opacity) matrix of users. The highlighted cell depicts a user pair with high spatial as well as temporal similarity. The diagonal cells only reflect the self-similarity, and therefore ignored in the analysis

as administrative hierarchies of a location, and the name of the closest city. The analyst can either use the home location or work location as input. These hierarchies of locations are then mapped to the tree nodes as seen in Fig. 5, administrative hierarchies are shown through ID values. To classify these areas according to the home/work locations we color the tree nodes according to the work place density for each area. For every area, we count the amounts of user homes ( $h$ ) and the amount of user work locations ( $w$ ) (as derived earlier in this section). We then divide ( $w$ ) by ( $h + w$ ) to get a ratio value which indicates whether the area is more a residential area (values  $< 0.5$ ) or more an industrial area (values  $> 0.5$ ). We map the value for each node on a diverging color scale from green (0) over orange (0.5) to red (1). The result are shown in Fig. 5. Accordingly, the area Barrio Brasil (value of 0.92, shown in red in Fig. 5) which is quite central in Santiago is a work/industrial area, popular for its student and night life, and the area Santa Carolina (value of 0.29, shown in green in Fig. 5) which is 20 km outside of the city center is a more residential suburb of Santiago. Both verify our findings.

#### D. Regional Partitioning With Origin-Destination Analysis

In Section IV-C, we have seen how geographic areas can be identified in terms of their unique characteristics by analyzing the movement behavior of people. On the other hand works such as [25] has shown that the geography around us also has an effect on the human interactions.

In our work we want to explore the *effects of geography on the movement of humans*. To achieve this we successively partition the geographic areas based on the *origin and destination* of movement trajectories. For example, the recent work of [26] used mobile data collected for weekdays over a period of one month with static map visualizations to observe the origin-destination of movements within Leicestershire, UK. The users' origin-destination movement analysis can help to uncover the attractiveness of an area, number of different places that people come from, the number of traversed trips for any time of day, or to capture the patterns of urban mobility in different areas of a city. Such analyses are useful for example in use cases within the city planning, transportation management, or emergency response domains [27].

Within our approach we have developed an interactive *origin-destination matrix* to *effectively* explore and analyze where user movement trajectories (described in Section IV-A) start and end. To achieve this task we utilize the table view - to choose individual users (if needed), the matrix view - to view the origin and destination of users, additionally a raster map view - to partition an area into multiple spatial segments and indicate the corresponding origin / destination locations.

Using the map, an arbitrary area of interest can be selected for investigation, which is then rasterized by partitioning the map into multiple spatial segments. These segments are coordinated with the rows and columns of the matrix. Each matrix cell is associated with an origin (rows) and a destination area (column). The origin-destination analysis for aggregated trajectories of all users is shown in Fig. 6 component (a). The cells are colored according to the amount of trajectories traversed within a particular origin and destination. As we

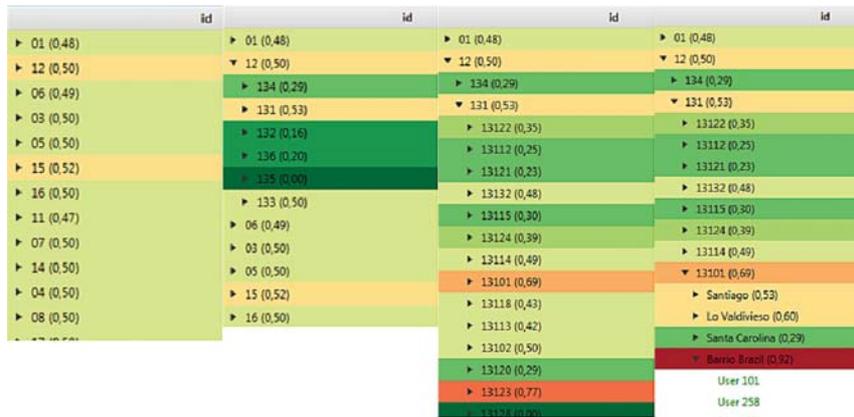


Fig. 5. User classification by home and work areas. The tree node areas are colored according to the home (green) and work (red) area ratio.

can observe from Fig. 6 component (a), most traversed routes appear to be between the center and the North-West of Santiago, and fewer routes within the South-West regions in Santiago. While this analysis looks at aggregated trajectories, the same analysis can be performed to observe the origin-destination per user.

Characteristics of the surrounding geography that may impact such movement patterns are for e.g., disparities that one can observe within various modalities such as economic, health, sanitation etc. One aspect that explains the observed movement patterns within Santiago in Fig. 6 is the economic division in the city [28], which according to the communal human development index for Santiago shows a higher concentration of high standard public and private facilities in the central and Eastern parts of Santiago, than in the Southern and Western parts of Santiago. Reference [28] shows in their study that the Southern counties in Santiago have the highest inequality in terms of income. Detecting such disparities most often involve surveys and census studies in the state of the art research. The presented visual analytics approach helps the user to interactively explore such disparities as geographic effects on human movement behavior.

#### E. Temporal Change Detection of Usage Patterns

Detecting changes in usage patterns over time can be a central and demanding task for city administrators to observe unusual events occurring within city limits, and thereby install necessary changes to accommodate such re-occurring events. For this analysis we utilize the map view, small multiples view, and the matrix view.

Temporal changes in usage patterns are often analyzed with the help of sorting and clustering algorithms in combination with matrix visualizations, which can provide for potentially useful views. However, these algorithms depend on parameters which are hard to set a priori. In our case large amounts of matrix views can be generated by selecting an area of interest and varying the time intervals w.r.t. one or multiple measurements of interest. In the following, these matrix views are referred to as (temporal) *situations*. Our goal is to automatically detect *change points* (or *interest points*) and show these to the users for inspection of possibly interesting events

Fig. 7 shows our semi-automatic temporal change detection approach. Each situation is represented as a matrix image and arranged horizontally in a small multiple fashion in the first line of the visualization. The matrix here represents a grid overlaid on the map, and each cell represents the aggregated number of records per antenna in each map cell, but the analyst can select different attributes. The goal is to automatically find points in time, when there is a high amount of change in the matrix view from one time stamp to the next, indicating potentially interesting situations, due to strong changes.

In our approach, we compute a measure of change for each raster location (cell) from one time step to the next. We normalize the changes between -100% which indicates the most amount of change (also color coded in red) and +100% which indicates the least amount of change. We also compute the sum of change values for all cells in the map. This aggregated measure of change is visually represented in a line chart under the small multiples. Detection of change points can then be done by filtering for time steps with large aggregate change values.

As one can see from Fig. 7, this visualization allows perceiving data differences in regions of interest for several periods of time. For example, the amount of connections increased significantly during Feb, 08-11th 2013, and then from 13-14th 2013. The strong usage increase from 08-11 February can be assumed by the return of Santiago citizens from their annual two week vacation starting from the beginning of February. 13-14th, February increase can be explained possibly by the St. Valentine's day activities.

## V. EXTENSION FOR UNCERTAINTY ANALYSIS IN MOVEMENT TRAJECTORIES

As work in progress we have developed a preliminary approach to visually analyze the positional uncertainty of movement trajectories that are extracted in Section IV-A.

When it comes to the antenna location-based movement data that is used for movement analysis in this paper, the unavailability of antenna signals, for example due to indoor movements (such as moving through a tunnel, or a building), or the large operational range of an antenna where the user could

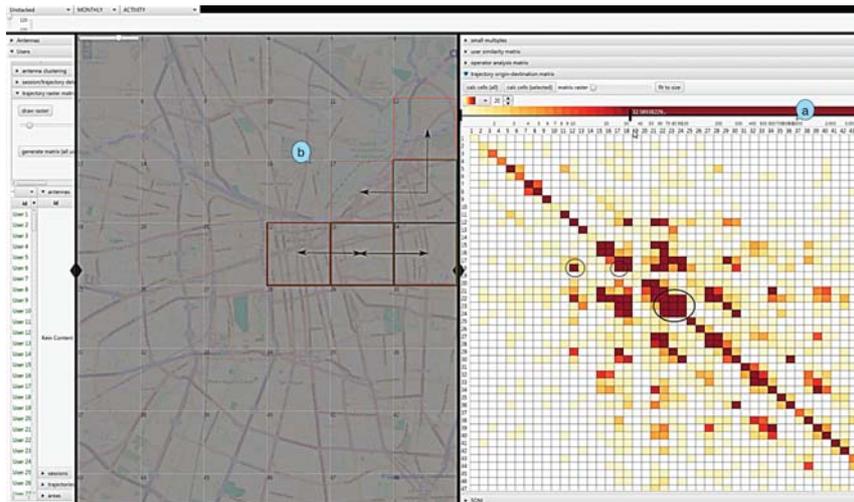


Fig. 6. Origin-destination analysis of aggregated user trajectories. (a) Each row and column represents an origin and destination respectively for the aggregated trajectories. The color scale represents low (lighter) to high (darker) no. of aggregated trajectories, (b) The selected cells in the matrix are coordinated with the rasterized map, as shown in the green (origin) and red (destination) spatial segments.

be anywhere within this area, cause positional uncertainty of movement approximations. Furthermore, as seen in the algorithm 1 the movement *approximations* consider interpolations of *crisp locations* in the form of chronologically ordered points,  $p = (x, y, t)$ . In our analysis we drew a straight line from one crisp point to another, assuming the user took a direct route from one position to another. However, in reality this may not be true, and further gives rise to positional uncertainty.

To model these positional uncertainties within the movement data in our visual analytics approach we utilise the *space-time cube* [29], where the horizontal plane is used for depicting the spatial extent  $(x, y)$  of the data and the vertical plane is used for depicting the temporal extent  $(t)$  of the data. These time-ordered sequences of samples called *markers* make up the *space-time path* for each of the users of the mobile Internet usage dataset.

An example of the space-time path of a selected user within a 24 hour time frame with the inter-sample time interval is shown in Fig. 9. As one can observe from this figure, the selected user is somewhat stationary during the morning and evening hours, but in between spends a longer duration of time also in the northern commune of Santiago. However, the data was collected at a 10-second sampling rate. Therefore ideally the inter-sample time duration should be 10 seconds at each path segment. This variation of the time interval duration can be due to a lack of antenna signal reception, or technical issues with the mobile phone devices. It therefore indicates already the uncertainties in the data.

These positional uncertainties can be represented as *volume* in the space-time cube. This volume represents all the space-time points within an area that any user *may* have traversed based on any given marker and a maximum travel velocity. Therefore the volume between two given markers is called the *space-time prism*. It's projection onto the horizontal geographic plane is called the *Potential Path Area* (PPA). Reference [30] demonstrated the use of these space-time prisms for

crisp markers. References [31] and [32] modeled the uncertainty of locations in between crisp markers on a road network.

Due to the inherent positional uncertainties in our derived movement data, we have *regions* instead of crisp markers. Those region-based markers denote several possible markers of user movement; and we call those here *uncertain markers*. Therefore in our work we define and adapt the space-time prisms for uncertain markers to visually analyze the positional uncertainties in the approximated movement data.

Based on the work of [30], a space-time prism  $Z$  between two crisp markers  $m_i = (p_i, t_i)$  and  $m_j = (p_j, t_j)$ , and a maximum velocity of  $v_{max} > \frac{\|p_i - p_j\|}{t_j - t_i}$  can be formulated as a parametric function of time. This is shown below.

$$Z_{ij}(t) = \begin{cases} SE_i(t) & \text{if } t \in (t_i, t') \\ SE_i(t) \cap SE_j(t) & \text{if } t \in [t', t''] \\ SE_j(t) & \text{if } t \in [t'', t_j] \end{cases}$$

$SE_i(t)$  is the spatial extent that can be reached from point  $p_i$  in the time interval  $t - t_i$  with a maximum velocity  $v_{max}$ . This spatial extent can be defined as a constrained set of points as follows:

$$SE_i(t) = \{p \mid \|p - p_i\| \leq (t - t_i)v_{max}\}$$

Similarly,  $SE_j(t)$  is the spatial extent that can be reached from point  $p_j$  in the time interval  $t_j - t$  with a maximum velocity  $v_{max}$ . This spatial extent can be defined as a constrained set of points as follows:

$$SE_j(t) = \{p \mid \|p_j - p\| \leq (t_j - t)v_{max}\}$$

$SE_i(t)$  and  $SE_j(t)$  have two possible topological relations where  $t'$  and  $t''$  are time points between  $t_i$  and  $t_j$ . These possibilities are, when at  $t'$   $SE_j(t)$  contains  $SE_i(t)$  and at  $t''$   $SE_i(t)$  contains  $SE_j(t)$ .

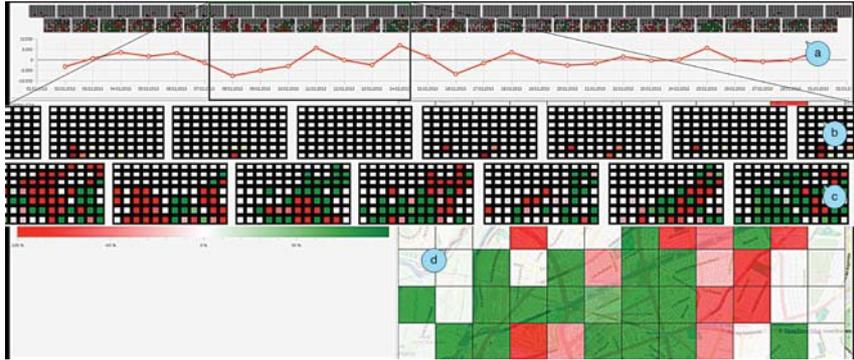


Fig. 7. Temporal change in usage patterns. (a) the overview of usage changes over a period of one month (b) the change of aggregated usage data between consecutive days (c) the aggregated mobile usage for each day (d) map to spatially reflect selected changes. Significant changes can be seen during 08-11, and 13-14 of February, 2013. The map corresponding to one of the small multiples, shows the temporal changes in usage between 10.02. and 11.02. in the Vitacura region of Santiago. Most changes are shown in red, and least changes are shown in green.

Further,  $t'$  and  $t''$  are calculated as follows:

$$t' = \frac{t_i + t_j - t_i^* j}{2}$$

$$t'' = \frac{t_i + t_j + t_i^* j}{2}$$

$$t_i^* j = \|p_i - p_j\| * v^{-1}_{max}$$

Given the velocity  $v_{max}$ ,  $t_i^* j$  is the minimum travel time from  $p_i$  to  $p_j$ . This parametric function is used for the space-time prism between uncertain markers as well, however the key difference is in the way we define the spatial extents  $SE_i(t)$  and  $SE_j(t)$ , and the maximum travel time  $t_i^* j$ . Therefore we determine the spatial extents for the uncertain markers by using the hull at the extreme points of the regions. For convex regions we use the convex hull.

Therefore, the spatial extents of two uncertain markers  $m_i = (region_i, t_i)$  and  $m_j = (region_j, t_j)$  can be represented as follows.

$$SE_i(t) = p | p \in CER(region_i, t - t_i)$$

$$SE_j(t) = p | p \in CER(region_j, t_j - t)$$

$CER(region, t)$  is the expanded region at time  $t$  and a given  $v_{max}$ . For defining the maximum travel time for uncertain markers, we divide the distance between the points by the assumed speed. Therefore this is expressed as follows:

$$t_i^* j = \frac{maxDist(region_i, region_j)}{v^{-1}_{max}}$$

For the space-time path example shown in Fig. 9, we have now estimated the potential path area and the corresponding space-time prism for two uncertain path segments. This is shown in Fig. 8.

## VI. DISCUSSION AND FUTURE PERSPECTIVES

In this paper several algorithms are presented together with various visual analysis approaches for (1) movement trajectory extraction out of GSM mobile internet usage data, (2) identification of patterns in urban environments based on the movement trajectories of users, (3) investigating the positional uncertainties in movement trajectories.

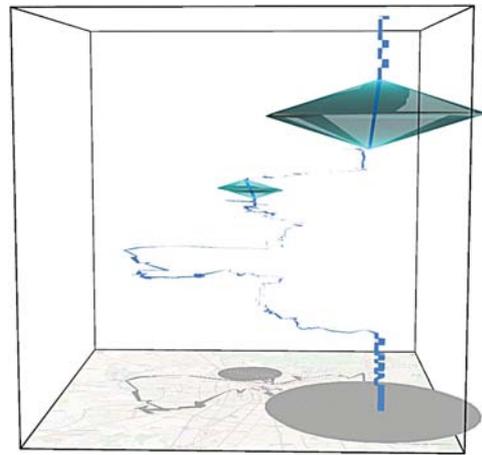


Fig. 8. Potential path area for two uncertain path segments and their corresponding space-time prism representations. The maximum velocity is assumed to be 50 km/h in this example.

Our findings obtained from our prototype on the data set from Santiago is indicative for a number of analysis cases of interest to mobile data analysis. While our anecdotal findings can be regarded plausible, we state that they are based on a sample of several hundred users from a much larger user basis, and may incur a bias. Therefore, in our future work we plan to evaluate our approach beyond what's based on anecdotal findings [4], to further test the generalisability of our approach to other similar, larger datasets.

Our analysis considers aggregates of mobile data consumed. More detailed analysis would be possible by enhancing the data collection mechanisms, e.g., distinguishing between data used for different applications such as email, web, voice, or video. The correlation of service usage with temporal and spatial aspects might lead to further insights of usage with great potential for applications such as transportation optimization or marketing.

An important aspect in any visual interactive analysis scenario is to guide the analyst to quickly find interesting and relevant views. Due to the large data volumes in mobile data analysis, a fully manual search through all possible

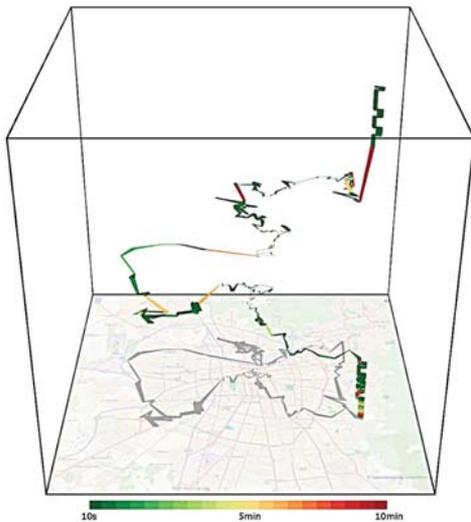


Fig. 9. The space-time path of a selected user with the path segments colored according to the inter-sample time interval duration.

data sections and views is practically not affordable. Hence, methods for automatic identification of interesting views are needed. An idea for future work is to interactively learn from the user which situations are interesting, and use such a learning scheme to suggest relevant situations in unseen data. To this end, approaches for visual relevance feedback [33] could be helpful.

Furthermore, our work in progress for uncertainty analysis would further extend in the future to investigate methods on how to *reduce* the uncertainties incurred by antenna jumps. By detecting the areas where such antenna jumps occur already paves path for uncertainty reduction in the data, and thereby reduce the uncertain space size for user locations. Further, the developed approach will be evaluated within a *user study* in future work, to mainly assess the usability of the chosen visualizations for the tasks.

## VII. CONCLUSION

This paper showcases how geospatial movement patterns help greatly to identify human behavioral patterns. Those derived human behavioral patterns can give us clues about the *arrangements* of the surrounding urban environments. To explore and identify these patterns this paper utilizes a GSM telecommunication dataset where the consecutive locations for movements are extracted from the antenna locations.

The encompassing visual analytics approach first introduces a matrix visualization to interactively explore the spatial and temporal similarity of movement between users. The movement patterns are further employed to classify the surrounding geographic space into industrial and residential areas based on the movement of users between home and work. Next, utilizing the same matrix view we have computed the origin and the destination of users' movement, thereby identifying the local spatial segregation patterns in the surrounding urban environment. Furthermore, this view and a small multiples view were utilized to automatically detect temporal changes in mobile internet usage, and further uncover seasonal patterns.

In the last section of this paper we define and adapt the space-time prism for *uncertain markers* as an extension to the space-time-cube method to assess and visualize the uncertainty of the approximated user movement patterns.

## REFERENCES

- [1] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.
- [2] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual Analytics: Definition, Process, and Challenges*. Heidelberg, Germany: Springer, 2008.
- [3] H.-M. Wu, S. Tzeng, and C.-H. Chen, "Matrix visualization," in *Handbook of Data Visualization*. Heidelberg, Germany: Springer, 2008, pp. 681–708.
- [4] T. Munzner, "A nested model for visualization design and validation," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 921–928, Jun. 2009.
- [5] G. Sagl, M. Loidl, and E. Beinath, "A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic," *ISPRS Int. J. Geo-Inf.*, vol. 1, no. 3, pp. 256–271, 2012.
- [6] Y. Zheng *et al.*, "TelcoFlow: Visual exploration of collective behaviors based on telco data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 843–852.
- [7] Z. Shen and K.-L. Ma, "Mobivis: A visualization system for exploring mobile data," in *Proc. IEEE Pacific Vis. Symp. PacificVIS*, Mar. 2008, pp. 175–182.
- [8] S. Gao, Y. Liu, Y. Wang, and X. Ma, "Discovering spatial interaction communities from mobile phone data," *Trans. GIS*, vol. 17, no. 3, pp. 463–481, 2013.
- [9] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.
- [10] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingiero, F. Pinelli, and R. Nair, "Allaboard: Visual exploration of cellphone mobility data to optimise public transport," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 2, pp. 1036–1050, Feb. 2016.
- [11] Y. Ma, T. Lin, Z. Cao, C. Li, F. Wang, and W. Chen, "Mobility viewer: An Eulerian approach for studying urban crowd flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2627–2636, Sep. 2016.
- [12] W. Wu *et al.*, "TelCoVis: Visual exploration of co-occurrence in urban human mobility based on Telco data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 935–944, Jan. 2016.
- [13] L. Yu *et al.*, "iVizTRANS: Interactive visual learning for home and work place detection from massive public transportation data," in *Proc. IEEE Conf. Vis. Anal. Sci. Technol. (VAST)*, Oct. 2015, pp. 49–56.
- [14] Y. Zheng, W. Wu, Y. Chen, H. Qu, and L. M. Ni, "Visual analytics in urban computing: An overview," *IEEE Trans. Big Data*, vol. 2, no. 3, pp. 276–296, Mar. 2016.
- [15] J. Bustos-Jiménez *et al.*, "How adkintunmobile measured the world," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct*, 2013, pp. 1457–1462.
- [16] C. Faloutsos, H. Jagadish, A. O. Mendelzon, and T. Milo, "A signature technique for similarity-based queries," in *Proc. Complex. Sequences*, 1997, pp. 2–20.
- [17] S. Dodge, P. Laube, and R. Weibel, "Movement similarity assessment using symbolic representation of trajectories," *Int. J. Geograph. Inf. Sci.*, vol. 26, no. 9, pp. 1563–1588, 2012.
- [18] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," in *Proc. Soviet Phys. Doklady*, vol. 10, 1966, p. 707.
- [19] P. Ranacher and K. Tzavella, "How to compare movement? A review of physical movement similarity measures in geographic information science and beyond," *Cartography geograph. Inf. Sci.*, vol. 41, no. 3, pp. 286–307, 2014.
- [20] N. Andrienko and G. Andrienko, "Designing visual analytics methods for massive collections of movement data," *Cartograph., Int. J. Geograph. Inf. Geovis.*, vol. 42, no. 2, pp. 117–138, 2007.
- [21] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *Int. J. Robot. Res.*, vol. 29, pp. 2–3, Dec. 2009.
- [22] A. Rottmann, Ó. M. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *Proc. AAAI*, vol. 5, 2005, pp. 1306–1311.

- [23] J. Hultman and O. Wärneryd, "What about nature?" in *Human Dimensions of Global Environmental Change*. Studentlitteratur AB, 2001, pp. 14–20.
- [24] S. Isaacman *et al.*, "Identifying important places in people's lives from cellular network data," in *Pervasive Computing*. Heidelberg, Germany: Springer, 2011, pp. 133–151.
- [25] C. Ratti *et al.*, "Redrawing the map of great britain from a network of human interactions," *PLoS ONE*, vol. 5, no. 12, p. e14248, 2010.
- [26] R. Tolouei, P. Álvarez, and N. Duduta, "Developing and verifying Origin-Destination matrices using mobile phone data: The LLITM case," in *Proc. Eur. Transp. Conf.*, Frankfurt, Germany, Sep. 2015.
- [27] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Comput.*, vol. 10, no. 4, pp. 36–44, Apr. 2011.
- [28] C. A. Agostini and P. Brown, "Geographic income inequality in chile," *Revista de Analisis Economico*, vol. 22, p. 1, Jan. 2007.
- [29] T. Hägerstrand, "What about people in regional science?" *Papers Regional Sci.*, vol. 24, no. 1, pp. 7–24, 1970.
- [30] H. J. Miller, "A measurement theory for time geography," *Geograph. Anal.*, vol. 37, no. 1, pp. 17–45, 2005.
- [31] B. Kuijpers, H. J. Miller, T. Neutens, and W. Othman, "Anchor uncertainty and space-time prisms on road networks," *Int. J. Geograph. Inf. Sci.*, vol. 24, no. 8, pp. 1223–1248, 2010.
- [32] B. Kuijpers and W. Othman, "Modeling uncertainty of moving objects on road networks via space-time prisms," *Int. J. Geograph. Inf. Sci.*, vol. 23, no. 9, pp. 1095–1117, 2009.
- [33] M. Behrisch, F. Korkmaz, L. Shao, and T. Schreck, "Feedback-driven interactive exploration of large multidimensional data supported by visual classifier," in *Proc. IEEE Conf. Vis. Anal. Sci. Technol.*, Oct. 2014, pp. 43–52.



**Hansi Senaratne** received the B.Sc. degree in computer science from the University of Bangalore, India, in 2008, and the M.Sc. in geoinformatics from the University of Münster, Germany, in 2011. She is currently a Doctoral Researcher with the University of Konstanz, Germany, since 2012. Her main research interests include visual analytics and uncertainty analysis in spatio-temporal data.



**Manuel Mueller** received the B.Sc. and M.Sc. degrees in information engineering/ computer science from the University of Konstanz. He is currently the Chief Information Officer with Refund Suisse AG.



**Michael Behrisch** received the M.Sc. degree in information engineering/ computer science from the University of Konstanz in 2011. He is a currently a Doctoral Researcher with the University of Konstanz, Germany, since 2011. His research interests include the visualization of relational data, pattern analysis in visualizations, and user-centric exploration approaches for large view spaces.



**Felipe Lalanne** received the Ph.D. degree from Telecom SudParis in 2012. He is currently a Research Engineer with Inria Chile. His main research interests and activities are Internet and network metrology, quality of service, and mobile sensing methodologies for smart city applications.



**Javier Bustos-Jiménez** received the Diplôme de Docteur d'Informatique from the University of Nice Sophia, France. He is currently the Head of the NIC Chile Research Labs, an institution affiliated to the University of Chile. His research interests include networking, Internet protocols, and mobile and distributing computing.



**Jörn Schneidewind** received the Ph.D. degree from the University of Konstanz. He is currently the Business Intelligence Portfolio Manager for O2 Telefonica Germany. For this paper he provided his expert feedback on the developed visual analytics approach.



**Daniel Keim** was the Associate Professor for databases and visualization with the University of Halle, and an Assistant Professor with the University of Munich. He is currently the Chairing Professor with the Data Analysis and Visualization Group, University of Konstanz. His research interests include databases, data mining, information visualization, and visual analytics.



**Tobias Schreck** received the Ph.D. degree in computer science from the University of Konstanz in 2006. He is currently a Professor with the Institute for Computer Graphics and Knowledge Visualization, Graz University of Technology. He was an Assistant Professor with the University of Konstanz and a Post-Doctoral Fellow with TU Darmstadt. His research interests include visual analytics and 3-D object retrieval.