

Discovering events in the city via mobile network analysis

Laura Ferrari · Marco Mamei · Massimo Colonna

Received: 29 May 2012 / Accepted: 14 November 2012 / Published online: 24 November 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract Can we automatically identify relevant places and events happening in the city from the analysis of mobile network use? In this paper we present a methodology to discover events from human mobility patterns as recorded by mobile network usage. Experiments conducted over an extensive dataset from the main Italian telecom operator show that the proposed approach is effective and can be applied to a number of different scenarios. These results can have a strong impact on a wide range of pervasive applications ranging from location-based services to urban planning.

Keywords Mobility models and patterns · Human mobility · Urban and environmental planning · Social dynamics

1 Introduction

The widespread diffusion of mobile phones provides a practical way to collect geo-located information from a large user population in real time. The analysis of such collected data is a fundamental asset in the development of pervasive and mobile computing applications, including

online map services, automatic gazetteers, urban planning, and disaster response.

In particular, data on human mobility allows us to identify hotspots, patterns and routine behaviors happening in the city. For example, if a lot of people frequently get together at a given place, we can infer that the area is a popular hot-spot of the city. If an unusually large number of people clusters in a given area for a certain time, we can infer that some kind of event is happening in there (e.g., a protest is underway). Similarly, if the distribution of people across all the city is anomalous, we can infer that there is something special associated with that day (e.g., it is a holiday).

Identifying such kind of peculiar events would have a strong impact in a number of application scenarios. Local governments and city planners would gain notable advantages from such kind of knowledge, for example to prioritize resources to areas and events attracting more people, or to shape the city to better reflect citizens behavior.

Given these opportunities, the central question we are trying to answer is:

Can we identify events happening in the city from the analysis of mobile network use?

To answer this question, we developed a data mining methodology to extract patterns from data summarizing mobile network usage. The proposed approach consists in creating a model of the average network usage pattern in a given location, and in identifying temporal statistical deviations from such model. In this approach, temporal deviations will represent *events* happening in that location.

To evaluate the proposed methodology, we analyzed the Telecom Italia (main Italian telecom operator) mobile network usage over 10 months in two administrative regions of Italy (Piemonte and Emilia Romagna inhabited by 9 millions people).

L. Ferrari (✉) · M. Mamei
Dipartimento di Scienze e Metodi dell'Ingegneria,
University of Modena and Reggio Emilia,
Reggio Emilia, Italy
e-mail: laura.ferrari@unimore.it

M. Mamei
e-mail: marco.mamei@unimore.it

M. Colonna
Telecom Italia Lab, Torino, Italy
e-mail: massimo.colonna@telecomitalia.it

In particular, we make the following contributions:

- We try to systematically identify all the events happening in the city. Previous works in the area (Calabrese et al. 2011; Vieira et al. 2010) instead visually analyze mobile network data to describe only episodic events (e.g., the World Cup final match on July 9, 2006, or Madonna’s concert in Rome on August 6, 2006 (Calabrese et al. 2011)) or focus on dense areas with a minimum critical mass of individuals (Vieira et al. 2010). By focusing on all the events and not only on the episodic ones, we are able to propose and test a general methodology that can be potentially applied to any kind of events and any city.
- We present some insights on the historic data necessary to identify events in the city. We first analyze the amount of data required to compute reliable statistics with the aim of assessing how much training data has to be collected before running event detection. Second, we try to understand if the proposed methodology is able to work in real-time and detect events that are underway.

The remainder of this paper describes the methods and the results obtained with respect to the above contributions. In particular, in Sect. 2 we illustrate related works in the area and discuss how this work compares with them. Section 3 presents the methodology we developed to extract events from the analysis of mobile network traffic. Section 4 illustrates experiments and results to validate our proposal. Finally, Sect. 5 concludes and presents future works in the area.

2 Related work

This section presents an overview of some works related to the analysis of data about human movement. In particular, we focus on: (i) the analysis of mobile network usage and (ii) the detection of patterns and events using social networks. These kind of analysis are those that best reflect our approach.

2.1 Mobile network usage analysis

Some pioneering works (e.g., Ratti et al. 2006; Calabrese et al. 2011; Vieira et al. 2010) analyze cell phone activity to get information about people whereabouts. In particular, in Ratti et al. (2006) authors introduce the potential of mobile network usage for the urban planning community. With a case study on the city of Milan, Italy they show the intensity and the evaluation of urban activities in space and time. In Vieira et al. (2010) authors apply scan statistics to mobile network data to monitor the movement of vehicles

and pedestrians as well as to automatically detect dense areas in the city. Another group of works deals with the problem of characterizing patterns of mobile phones traffic. By applying cluster analysis and eigendecomposition (Calabrese et al. 2010) authors aim to match usage characteristics to urban space utilization.

In a similar work (Candia et al. 2008), authors study both the spatio-temporal human dynamics and the detection of events. This work applies an event-detection methodology similar to the one we propose: it builds a model of the “normal” network activity, then detects outliers with respect to that model. Apart from some technical differences, the main improvement of our approach is that we evaluate our methodology by trying to measure detection performance with respect to actual ground truth data.

In Andrienko et al. (2012), authors have developed a suite of visual analytics methods for reconstructing past events from activity traces. In particular, they have developed a peak detection algorithm to extract events happening in the city.

One limit of all these approaches, in the case of event detection, is that they focus on episodic and visual-only identification of some specific events (e.g., the World Cup final match on July 9, 2006, or Madonna’s concert in Rome on August 6, 2006 (Calabrese et al. 2011)). Our work instead proposes a methodology to identify all the events in an automatic way: starting from the data, we detect *all* the events that happened in the environment and then we try to measure the resulting detection performances.

2.2 Patterns/events detection from social networks data

A variety of research studies have been recently conducted utilizing geo-referenced social networking platforms. These works apply data mining algorithms to such data to extract information about city dynamics. A first group of works is based on photo-sharing sites. Researchers have been able to analyze a global collection of geo-referenced photographs (e.g., taken from Flickr) with the goal of identifying hot spots, events and tourist routine behaviors (Rattenbury et al. 2007; Mamei et al. 2010). Another group of works is based on location-based social networks (e.g., Twitter and Foursquare). A spatial analysis of the aggregate activity generated by such networks (see for example (Sakaki et al. 2010; Ferrari et al. 2011)) shows how social activity in a city is distributed, revealing spatial patterns. Moreover, in Lee et al. (2011) authors propose methods for detecting geo-social events and patterns based on crowd moving behaviors.

The key point of our work is that the presented methodology is tested upon a huge dataset of mobile network usage. Even though location-based social networks are

continuously growing, they haven't yet reached a mass of data enabling detection of fine grained events. More in general, the presented approach can be extended also to these kind of social network-based sources of data.

3 Methodology

In this section we describe our approach. We first present the dataset at hand, then we show our methodology to extract temporal events from data describing mobile network use.

3.1 Dataset acquisition

Data about mobile network usage has been obtained collecting and processing the cellular network information via *CityLive*, an ad-hoc software platform developed by Telecom Italia (TI). In TI's GSM cellular network, the number of protocol messages being exchanged during the different procedures (such as calls, short message service (SMS) transmission, handover etc.) is logged by the network equipment—which is mainly composed by the base station controllers (BSCs)—and is used for both network monitoring and performance evaluation. Since the network comprises BSCs from multiple manufacturers each with their own counter logs and data formats, in order to have an efficient implementation of network's operation and maintenance (O&M), the cellular network is connected to an ad-hoc system, called performance export system (PES). It periodically reads all these counters and organizes them into homogeneous and human-readable graphs and tables. The PES is organized on a regional basis and consists of 15 servers, distributed throughout the territory, each covering one or two administrative regions and managing the counters stored in all the BSCs of its territory. A PES Gateway installed in a central site periodically connects to these servers and downloads the counters of interest for the different applications. It is important to remark that such counters record an aggregate activity of the network usage without any reference to particular users. Using the above data, *CityLive* produces every 15 or 60 min (depending on the network equipment configuration) aggregated traffic maps in a raster form. *CityLive* splits the area under analysis into contiguous square pixels (with a size that ranges from 150×150 to 250×250 m in urban areas depending on the analysis requirements) and allocates the cell counters among all the pixels considering the actual radioelectric cell coverage maps provided by the radio network planning tool. The latter are obtained through sophisticated propagation models that take into account all the propagation phenomena involved (path loss, shadowing, etc.), as long as the area's topography and the building

characteristics. For the proposed data mining procedure, we used the *CityLive* matrices related to the pixel's total traffic expressed in Erlang collected from November 2010 to August 2011. The Erlang is the measure of telecommunications traffic density and it is a dimensionless "unit" representing a traffic density of one call-second per second (or one call-hour per hour, etc.).

This measure roughly correlates to the concentration of people in the area. This uncertainty derives from the fact that: (i) In Italy, cell phone coverage is about 156 % of the total population (97 millions)¹ and Telecom Italia has a market share of 33.3 % (32.277 millions).² We do not "see" all the other people. (ii) most importantly, our measures cover only cell phones that are generating network traffic at a given time, (iii) Erlang measures masks the actual number of people. Despite these clear limitations, in this study we *assume* that our measures are a proxy for the concentration of people in the area.

Moreover, the matrices used for the proposed data mining procedure presents different characteristics. In particular, the matrices for the Emilia-Romagna region consist of square pixels of 150×150 m size and they have been sampled every 60 min whereas the matrices for the Piemonte region consist of square pixels of 250×250 m size and they have been sampled every 15 min.

3.2 Overview

The approach we developed to analyze telecom input matrices takes inspiration from multidimensional database technology (Pedersen and Jensen 2001) and views data as a multidimensional data-structure—*cube* spanning spatial and temporal dimensions. In particular, we stacked telecom input matrices at different times to obtain the cube data structure. It is important to remark that these operations are only important for the practical aspects of the computation. However, we think that it is important to underline that we ground our results on a multidimensional database structure. With this regard, we conducted two main kind of operations on this cube:

- *Slice-and-dice* operations make selections to reduce the cube by considering only some time intervals and/or some pixels. For example, a result of these operations could be the subset of the whole data that refers only to the city center during Sunday afternoon.
- *Roll-up* operations perform aggregations over some dimensions of the cube. For example, for each given pixel these operations might return the mean value of weekdays mornings.

¹ Report Agcom 2011, <http://www.agcom.it>.

² Report Telecom Italia 2011 <http://www.telecomitalia.com>.

From a general perspective, the basic idea to detect events consists of the following three steps. As an exemplary scenario, let us assume that we want to identify unusually crowded situations at shopping centers during Sunday afternoon (e.g., possibly representing special openings).

1. We used slice-and-dice and roll-up operations to determine the mobile network usage for the time intervals and pixels under investigation. In the example, we selected data associated to each shopping mall's pixel on each given Sundays afternoon (e.g., Sunday, December 5th 2010, 3–9 pm). Then, for each selected pixel, we retrieved the corresponding mobile network usage at that time interval. From now on, we call the mobile network usage for the given time interval *observed* behavior.
2. Using slice-and-dice and roll-up operations on data spanning an extended period of time, we determined the *base* behavior on which to perform comparisons. In the example, we computed the distribution of mobile network usage for each shopping mall's pixel on an average Sunday afternoon (e.g., any Sunday at 3–9 pm).
3. Once such behaviors are established, comparisons can be performed between the *base* and the *observed* behaviors. If a *observed* behavior is very different from its *base* behavior (i.e., it is an outlier), we marked the element as one in which the target event is present. In the example, we will identify those days that are unusually crowded on Sunday afternoon.

3.3 Statistical measures

The *base* and the *observed* behaviors have the goal of creating useful measures to be compared in order to identify events. Accordingly, each behavior is described by the statistical distribution of its values on the basis of percentiles. We represented our measures via box-and-whisker plots (or simply boxplots) (see Fig. 6): the “box” represents the 25th ($Q1$), 50th, and 75th ($Q3$) percentiles. The inter quartile range (IQR) is the distance between the lower ($Q1$) and upper ($Q3$) quartiles. Finally, the “whiskers” are located at a distance $k_{bottom} \times IQR$ below $Q1$ and $k_{top} \times IQR$ above $Q3$.

Given these measures we can define:

- **Overcrowded events:** those in which the median value (50th percentile) of the *observed* behavior falls above the top whisker ($Q3 + k_{top} \times IQR$) of the *base* cube (e.g., a special opening is underway).
- **Underpopulated events:** those in which the median value (50th percentile) of the *observed* behavior falls below the bottom whisker ($Q1 - k_{bottom} \times IQR$) of the *base* cube (e.g., it is a holiday and people are out of the city).

Following the literature (Wilcox 2012), we chose the median value as a measure to determine if an event is “overcrowded”, “underpopulated” or none of the above. Moreover, k_{top} and k_{bottom} are the key parameters in our methodology. In general, the greater the value of k_{top} , the more crowded a place has to be to result hosting an event. The greater the value of k_{bottom} , the more underpopulated a place has to be to result hosting a (“repulsing”) event. Among all the possible values that k_{top} and k_{bottom} can assume, in the next section we show how to select the optimal ones that minimize the number of uncorrect events detected. In particular, in our experiments we found that the optimal value of k_{top} is usually higher than the optimal value of k_{bottom} . In contrast with other methodologies (see Section 4.3), our approach is based on two parameters and thus generates whiskers of different length since the top whisker depends on k_{top} while the bottom one depends on k_{bottom} (see above formulas). The choice of using two different parameters strongly depends on the fact that, as above mentioned, k_{top} and k_{bottom} usually present different values. Figure 1 compares the results obtained using respectively two different k parameters or a unique k parameter (in this case, the optimal k value is the one that minimizes the number of uncorrect events detected considering together overcrowded and underpopulated events). Results have been averaged over all the venues under investigation. In particular, in this experiment we have found that underpopulated events are the most affected by the unique k . This happens because in this case the lower threshold is higher, thus producing a lower number of correct events and a higher number of false negative events.

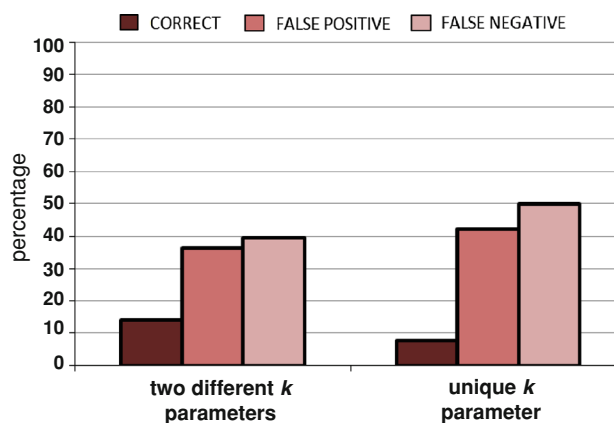


Fig. 1 Percentage of correct, false positive and false negative events detected respectively using two different k parameters (left) and a unique k parameter (right). Results have been averaged over all the venues under investigation

3.4 Choosing k values

To select the optimal values for k_{top} and k_{bottom} we identified two complementing approaches.

3.4.1 Intra-class variance optimization

We plot the number of events (overcrowded/underpopulated) being detected for various values of k . Given its simplicity, among the several ways that automatically set this parameter to an optimal value (e.g., finding local maxima and minimum, k-means variation clustering, mixture modeling, etc.), we adopted the Otsu algorithm (Otsu 1979). This algorithm has been originally proposed for image processing to automatically perform histogram shape-based image thresholding. The algorithm assumes that the image to be thresholded contains two classes of pixels (e.g., foreground and background), then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal. Compared to the other alternative methodologies, this algorithm depends only on the difference between the means of the two clusters, thus avoiding having to calculate differences between individual intensities and the cluster means.

To adapt this algorithm to our scenario, we considered the graph with the number of possible events as a function of k (see Fig. 3a). For each possible threshold k , we compute the intra-class variance between relevant and not-relevant events. The threshold minimizing intra-class variance is the optimal one. The algorithm consists thus in computing, for each threshold k :

$$\sigma^2(k) = \omega_1(k) \cdot \sigma_1^2(k) + \omega_2(k) \cdot \sigma_2^2(k)$$

where ω_1 and ω_2 are the probabilities of the two classes (events and not-events), and σ_1^2 and σ_2^2 are the variances of these classes. This approach can be applied separately to overcrowded and underpopulated events to identify k_{top} and k_{bottom} respectively.

3.4.2 False-positive, false-negative optimization

We use a list of past events (*ground truth*) that happened in that place as a training set to optimize the value of k .

While the previous approach does not require the availability of a training set, if such training information is available we can set k_{top} and k_{bottom} so as to minimize false positives (i.e., events found by our approach, but that are not in the ground truth: for example, the approach marks the 10th of January as an event, but such a day is not in the list) and false negatives (i.e., events that are in the ground truth, but that our approach did not find: for example, the 25th of December, the Christmas Day, is in the list of past events but the approach doesn't mark such a day as an

event). These type of errors can be reduced by a correct fine-tune of the k_{top} and k_{bottom} parameters. In particular, a too low value of k_{top} and k_{bottom} will cause the detection of higher number of events simply because the threshold is very low and consequently even though an event did not attract a wide audience, it is recognized as an event in any case. A too high value of k_{top} and k_{bottom} will produce the opposite case, where no events are recognized because the threshold is too high.

Accordingly, we computed the number of false positives and false negatives for different values of k . Then we selected the k value that minimizes the sum of false positives and false negatives.

It is finally worth noticing that a complementary option is to choose k on the basis of application-specific requirements. Thus, an application only interested in major overcrowded events, could decide to set k_{top} to a very high value to detect only extreme outliers. In these cases k_{top} and k_{bottom} could be set explicitly by a domain expert.

4 Experiments

In this section we present experiments applying the proposed methodology to our data. In particular, we focused on the cities where authors live (namely, Modena and Turin). It is worth emphasizing that these are rather different cities: Modena is a medium-size city of about 200,000 inhabitants, while Turin is a large city of about 1 million inhabitants. Evaluation in such a different settings is important to assess the generality of our approach.

In the experiments we focused the analysis on: (i) shopping centers (3 in Modena and 10 in Turin); (ii) football stadiums (2 in Modena and 1 in Turin) and (iii) small residential suburbs (1 for each city as a control group for which we expected to find no event). This choice has been guided by the fact that we were able to collect some ground truth information about events in that places. In particular, we retrieved from the Web a list of 12 days where shopping centers made special openings (*overcrowded events*) as long as 8 festivities (*underpopulated events* i.e., days where shops are unexpectedly closed). For the football stadiums, we retrieved a list of football matches and concerts that have been played there: 11 events for Modena (8 for the first stadium and 3 for the second one) and 68 for Turin. It is important to note that the higher number of events for the stadium in Turin is due to the fact that there are two important teams in that city playing in several football leagues. The number of events for the shopping malls is the same for both cities since they correspond to Italian festivities.

We identified the proper resolution with which to operate on the time dimension by relying on a data

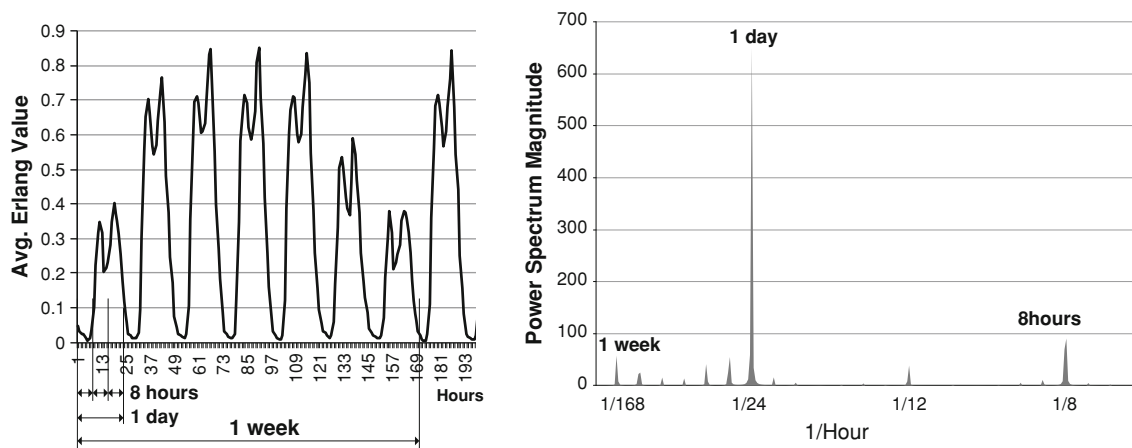


Fig. 2 A Fourier transform of the number of calls in the city averaged over all pixels. The transformation highlights the underlying datas most important cycles—1 week, 1 day and 8 h—by representing

processing approach. Since the urban behavior is characterized by periodic patterns at different time scales, we performed Fourier analysis on our dataset. In particular, Fig. 2 shows a Fourier transform of the number of calls averaged over all the city representing the signal as the sum of a set of sinusoidal frequencies multiplied by coefficients. Since the Fourier transform defines a relationship between a signal in the time domain and its representation in the frequency domain, the resulting power spectrum highlights some periodic patterns. In particular, as expected, our dataset presents strong periodicity on a weekly, daily and 8-hours (roughly corresponding to morning, afternoon, evening) scales.

On the basis of these results we decided to focus either on daily events, or on events involving a large portion of the day (8 h). On the one hand, ground truth for this kind of events is more easy to be collected. On the other hand—as a consequence of the periodicity highlighted by Fourier analysis—base (“normal”) behaviors tend to be more stable and reliable.

In the following of this Section, we report four kinds of experiments to evaluate our approach:

1. Evaluation of the approaches to select the best value for k ;
2. Evaluation of event detection performances;
3. Analysis of the performance and stability of our results with respect to different subsets of the data;
4. Comparison of our approach with other methodologies in the literature.

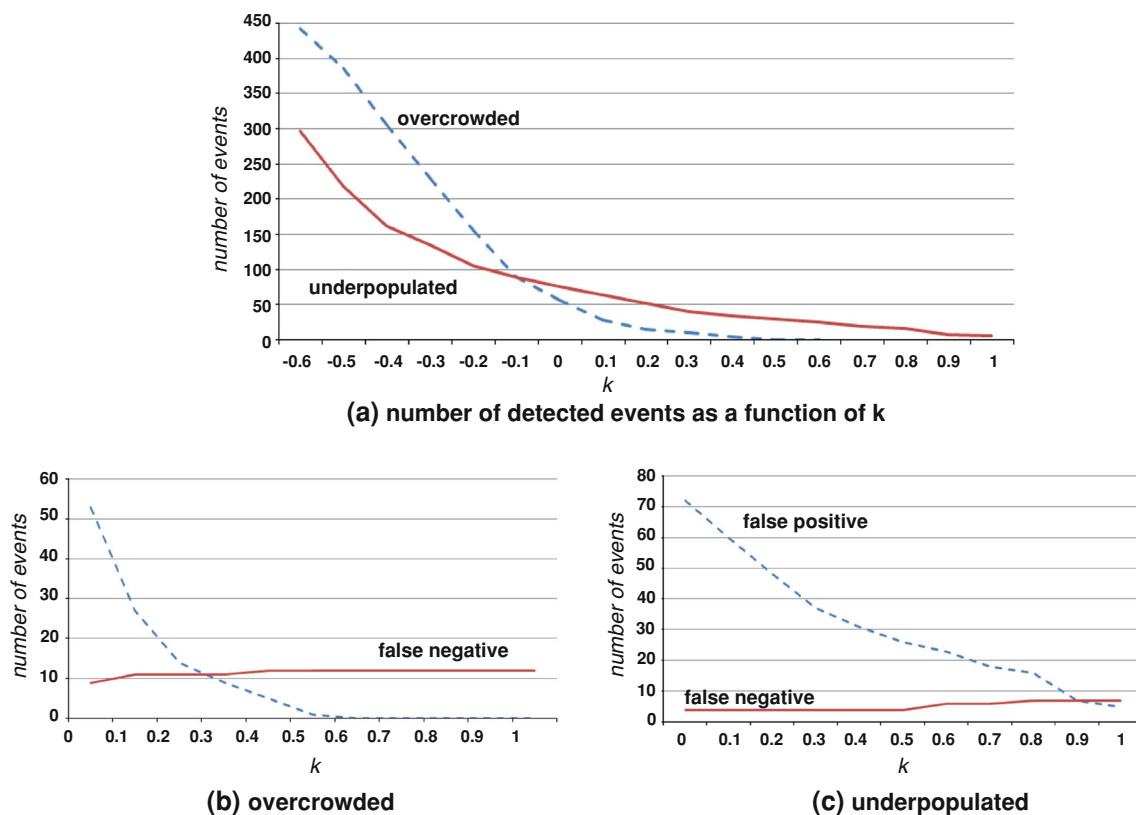
4.1 Selecting optimal k values

We applied the approaches described in Sect. 3.4 to the selected cases.

the signal as the sum of a set of sinusoidal frequencies multiplied by coefficients. *Left* time domain, *right* frequency domain

Figure 3a shows the number of detected events with different values of k for one of the shopping malls under investigation in Turin, Italy. Following the Otsu approach we can select both k_{top} and k_{bottom} equals to -0.1 indicating the point where the intra-class variance is minimized. Figure 3b, c show the number of false positive events and false negative both in the case of overcrowded and underpopulated events. Looking at these graphs we can refine our parameters. $k_{top} = 0.3$ and $k_{bottom} = 0.9$ to discover events in that shopping mall more accurately. The table at the bottom of Fig. 3 shows the confusion matrix obtained using $k_{top} = 0.3$ in order to discover overcrowded events and $k_{bottom} = 0.9$ to discover underpopulated events. More in detail, each row of the matrix represents ground truth data: *overcrowded* events, *underpopulated* events, and *normal* days in which we expect no peculiar patterns. Each column represents the kind of event discovered by our algorithms. For example, the first row of the matrix shows that our algorithms classify the 12 overcrowded events in the ground truth as 8 overcrowded, 4 normal and 0 underpopulated. It is possible to see that a large fraction of the result are on the main diagonal of the matrix representing correct classification. However, the algorithm has some troubles distinguishing between overcrowded and normal events.

Similarly, Fig. 4 shows the same analysis for the first stadium in Modena, Italy. From Fig. 4a it is clear that for football stadiums underpopulated events are of no significance (stadiums are *normally* empty when there are no matches). In addition, such a graph shows that only very few events register a high concentration of people. Figure 4b illustrates false positive and false negative overcrowded events at different values of k_{top} . Similarly, Figure 4c shows the number of false positive and false negative events in the case in which we consider in the



	OVERCROWDED	NORMAL	UNDERPOPULATED
OVERCROWDED	8	4	0
NORMAL	9	298	6
UNDERPOPULATED	0	1	7

Fig. 3 Top: number of events versus value of k for one of the shopping malls in Turin, Italy. Bottom: confusion Matrix for the same shopping mall obtained using $k_{top} = 0.3$ and $k_{bottom} = 0.9$. Each row represents ground truth labels, while each column represents computed labels

ground truth only the top most crowded three events. The table at the bottom of Fig. 4 shows the confusion matrix for the corresponding football stadium. Values are obtained using $k_{top} = 0.2$ (as suggested by Fig. 4b) to discover overcrowded events. Also in this case rows represent ground truth data: *overcrowded* events, and *normal* days in which we expect no peculiar patterns (as discussed above, we decided not to consider *underpopulated* events). Each column represents the kind of event discovered by our algorithms. This confusion matrix shows a rather good classification accuracy, however it shows some troubles distinguishing between overcrowded and normal events, as in the shopping mall example.

In conclusion, it is fundamental to understand that the procedure to identify the most effective value for k_{top} and k_{bottom} strongly depends on the identified ground truth information, and thus on what we consider an “event”. In

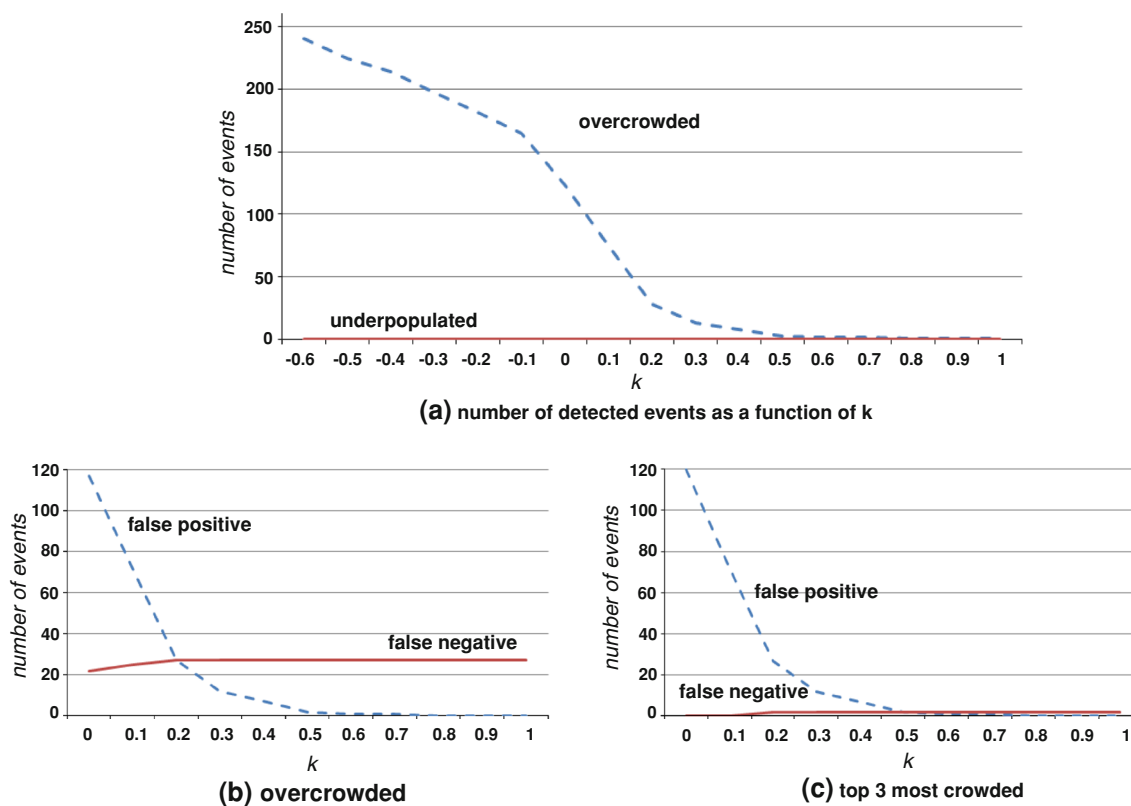
the stadium example of Fig. 4, if we consider events all the football matches (Fig. 4b), then the proper value for k_{top} is 0.2. If we consider that only major matches are events (Fig. 4c), then the proper value of k_{top} is 0.5.

It is also worth comparing these options with Fig. 4a. The fact that the knee of the graph (i.e., the point in which intra-class variance is minimized) is close to 0.2 supports the idea that the actual events expressed in the data are those corresponding to all the football matches as in Fig. 4b.

4.2 Detection of events

The goal of our experiments is to compare the events identified by our approach with the events in the ground truth.

The boxplot for a shopping mall in Modena (Italy) is displayed in Fig. 5. Statistics have been computed taking



	OVERCROWDED	NORMAL
OVERCROWDED	5	3
NORMAL	59	216

Fig. 4 Top: number of events versus value of k for the first stadium in Modena, Italy. Bottom: confusion Matrix for the same football stadium obtained using $k_{top} = 0.2$

into consideration a specific time-period (afternoon: 3–9 pm) for each day of the week. In particular, outliers that are under $Q1 - k_{bottom} \times IQR$ correspond to festivities (e.g., days where shops are unexpectedly closed, thus providing a low mobile network usage), while outliers that are above $Q3 + k_{top} \times IQR$ correspond to days where shops provided a special opening (thus providing an unexpected high mobile network usage).

Figure 6 shows the boxplots for a shopping mall, a football stadium and a small residential suburb in Modena. As above mentioned, the stadium provides few events compared to the shopping mall. For the small residential suburb we unexpectedly found two events, however rather close to the top whisker.

Similarly, Fig. 7 shows the boxplots for a shopping mall, the football stadium and a small residential suburb in Turin. It is important to emphasize the different scale of the graphs in Figs. 6 and 7. Such a difference is due to the fact that the two cities have very different sizes, thus events in

Turin attract a higher number of people. Moreover, it is possible to see that in this case football matches at stadiums can be identified much more reliably than in the previous case. This is because the football teams in Turin (i.e., Torino and Juventus) are much important and with a wider audience than the team in Modena. As an example, the event marked as an outlier in Fig. 7a with date 22 May 2011, is the match Juventus-Naples that attracted an audience of about 20,000 people. Unfortunately, the number of false negative events still remains high. From one side, the most important football matches attract more people than the ones in Modena. From the other side, even though the football teams are more important, not all the matches attract a wide audience (and in any case only a fraction of people use the phone during the game); this can explain the high number of false negative events.

Figure 8 shows the number of false positive and false negative events for all the places under investigation for both cities. It is worth noticing that—as expected—our

Fig. 5 *Boxplots* for a shopping mall in Modena, Italy. *Black dots* represent outliers

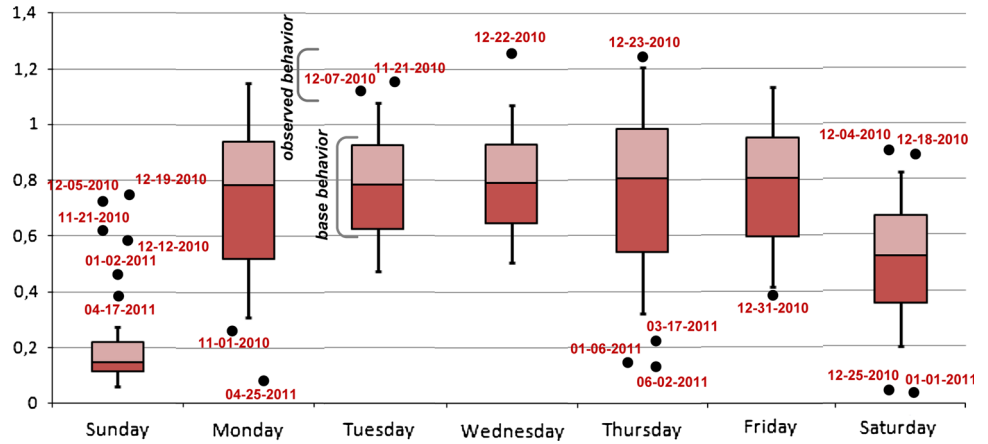


Fig. 6 *Boxplots* of Erlang data for three different areas of interest in Modena, Italy: **a** a city stadium, **b** a shopping mall and **c** a small residential suburb

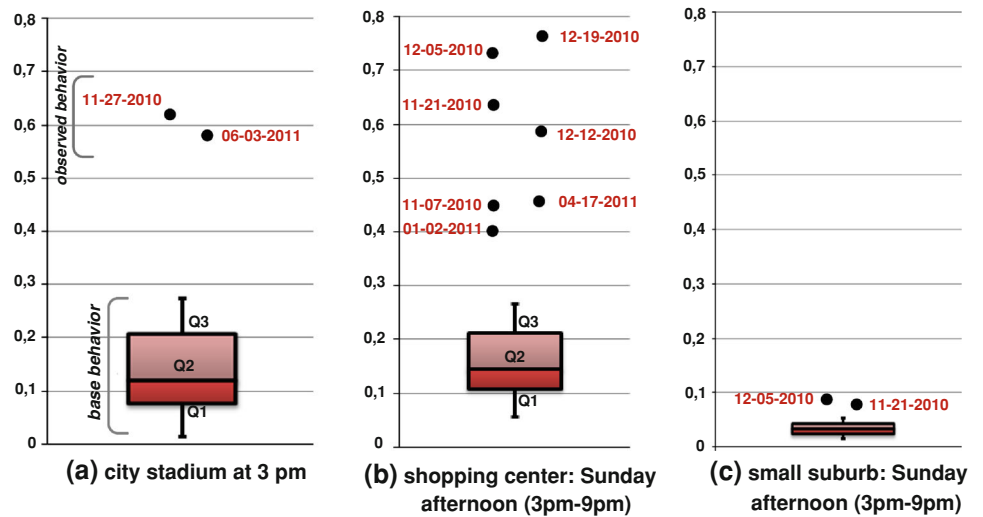
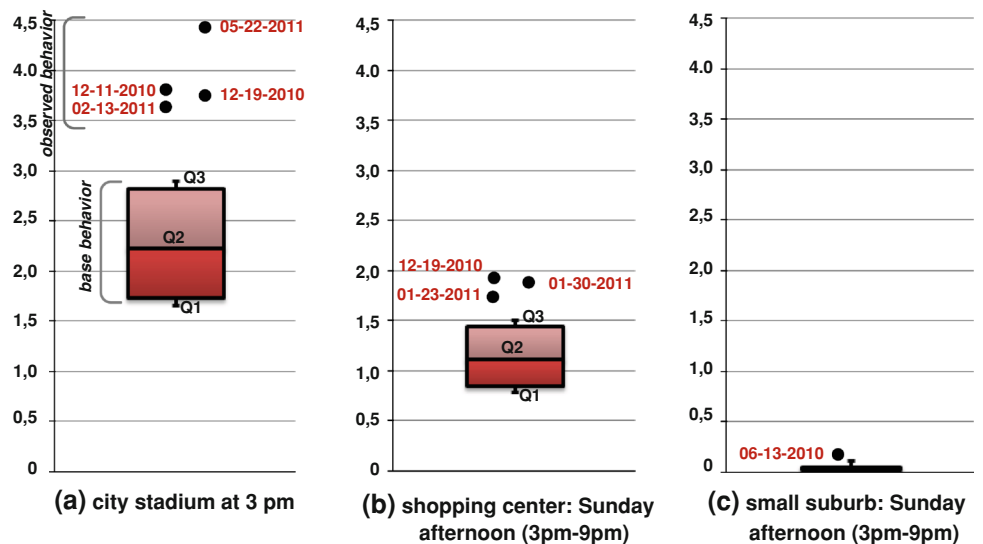


Fig. 7 *Boxplots* of Erlang data for three different areas of interest in Torino, Italy: **a** a city stadium; **b** a shopping mall and **c** a small residential suburb



approach best performs in large cities in which events tend to attract more people (see for example the different performances with regard to stadiums).

The above results illustrate that our approach is able to identify events happening in the city under different circumstances. However, there are two main limitations in the



Fig. 8 Number of false positive and false negative events detected by our algorithms. In particular, in (a) we considered 3 shopping malls, 2 stadiums and 1 small residential suburb while in (b) we considered 10 shopping mall, 1 football stadium and 1 residential suburb

current approach. On the one hand we verified that our algorithm is effective only for places that are well populated (thus producing enough data) and spatially separated from other main venues (so that GSM localization is accurate). Our performance degrades in scarcely populated areas in which we do not have enough data, or in dense areas in which multiple (event-producing) venues are covered by the same network cells, and thus it is difficult to discriminate around which venue people are located. In fact, in our analysis we experimented how the city of Turin shows better results than Modena, which is less populated (see Fig. 8). Moreover, we noticed how shopping malls located a little outside the city center show better results than the others, where multiple venues might contribute to generate the network usage. These two factors have a higher impact than the size of the cell network. In fact, as already mentioned, cells in Modena have a size of 150×150 m while cells in Turin have a size of 250×250 m. Even though the size of the cells in Turin is less refined, the recognition of events in that city is easier thanks to the higher number of population that produces a wider audience during the events. On the other hand, it is fair to point out that identifying accurate ground truth information (against which our approach is tested) is still an open challenge:

1. It is possible that we included in the ground truth a “damp squib” event that did not actually attract a lot of people (e.g., a football match of little importance), or that attracted a lot of people who did not use the phone. This may result either in an incorrectly classified true positive (if our approach found the event), or in an additional false negative (if our approach did not find the event).
2. it is possible that—despite our search efforts—we did not include in the ground truth an event that actually happened. This may result either in an incorrectly classified true negative (if our approach did not find the

event), or in an additional false positive (if our approach found the event).

With regard to this latter point, we speculate that the high number of false positives in stadiums (see Fig. 7a) is actually due other events in the stadium’s proximity that we did not record, and so our approach *correctly* classify them as events.

4.3 Results stability

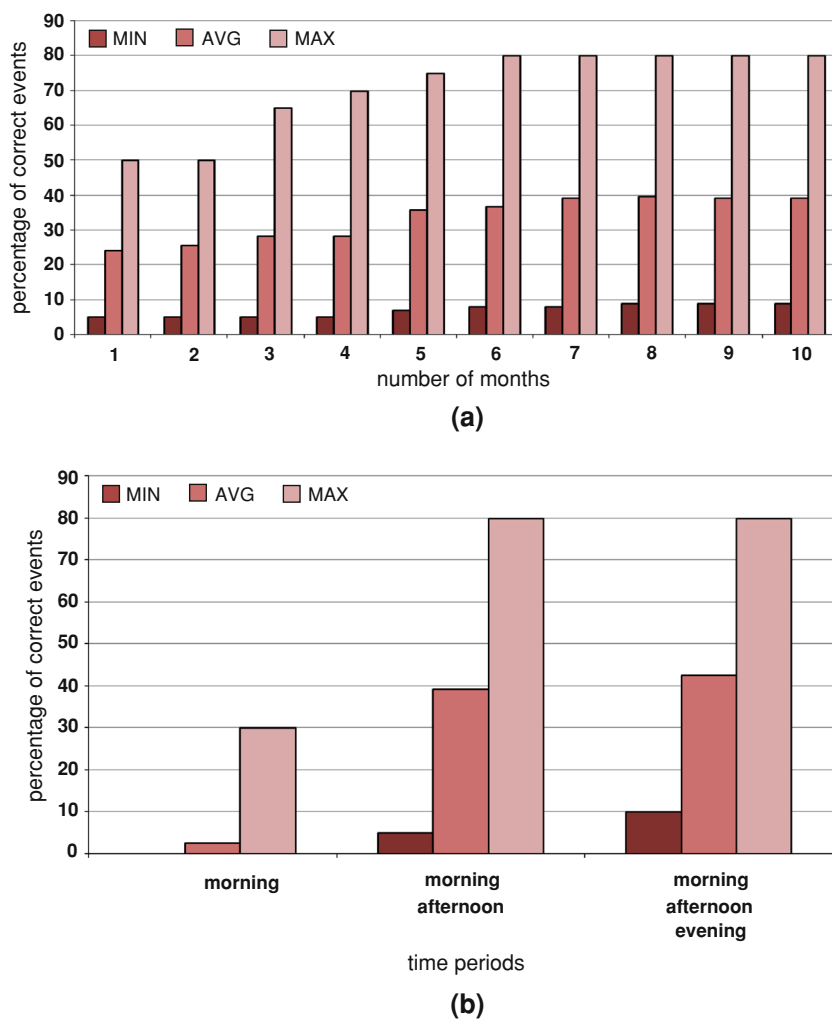
In this set of experiments we tried to validate the stability of our results with regard to different subsets of our data. In particular our goal is twofold:

1. We want to understand the amount of data required to compute reliable statistics (i.e., *box plots*) associated to a given cell for outliers identification. This is very important to assess how much “training” data has to be collected, before running event detection.
2. We want to understand the number of samples required to compute the observed behavior of a single day before analyzing whether that value is an outlier or not. This point is very relevant for the actual applicability of our approach: if all the samples of a given day are required to compute a stable observed behavior, then our approach can identify events only after they actually happened. Viceversa, if events can be identified only after few hours, then our system could work in real-time and detect events that are underway.

In this set of experiments we focused on the detection of shopping centers’ special openings during the Sundays of December in the cities under study. We focused on these events as they are represented by a rather reliable ground truth and they are temporally confined within a specific time period.

Following our methodology, for each shopping center, we construct the base statistics (i.e., box plot) of a normal

Fig. 9 All these results have been averaged over all the shopping malls of the two cities under investigation. **a** Detection accuracy of shopping malls' events as a function of the number of months of data used for computing the base statistics. **b** Detection accuracy of shopping malls' events as a function of the time periods used to estimate the observed behavior



Sunday. Then we test whether a given Sunday of December is an outlier with respect to that distribution. Results have been averaged over all the shopping malls of the two cities under investigation.

For the former analysis we computed the base statistics by considering only x months of data before the time period in which events happened. So with $x = 1$ we computed the statistics using only the Sundays from November, with $x = 2$ we computed the statistics using the Sundays from October and November, and so on.

The plot in Fig. 9a shows the minimum, average and maximum accuracy obtained with such data. It is worth noticing that results stabilize after 6 months of data. Such a large number can be explained in that, since we are building the box plot for a normal Sunday, each month of data produces only 4 days of useful data.

For the latter analysis we computed the observed behavior by considering an increasing fraction of the day. In particular, relying on the results of the Fourier analysis described in Fig. 2, we focused on 3 eight-hours time intervals associated to morning, afternoon and evening

time periods. The plot in Fig. 9b shows the minimum, average and maximum accuracy obtained by computing the observed behavior over such portions of the day. Given the low difference by using 2 or 3 eight-hours time intervals, our result shows that our approach could be applied in real time and provide events detection as events are underway.

In both these experiments (Fig. 9) it is possible to see a large variation between minimum, maximum and average accuracies. This is because our approach works best (80 % accuracy) for places that are well populated and spatially separated from other main venues, while performance degrades (10 % accuracy) in scarcely populated areas, or in dense areas in which multiple (event-producing) venues are covered by the same network cells.

4.4 Comparisons with other events detection methodologies

In this last set of experiments we compared the obtained results with other two methodologies for outliers detection. Similarly to the proposed approach, each methodology

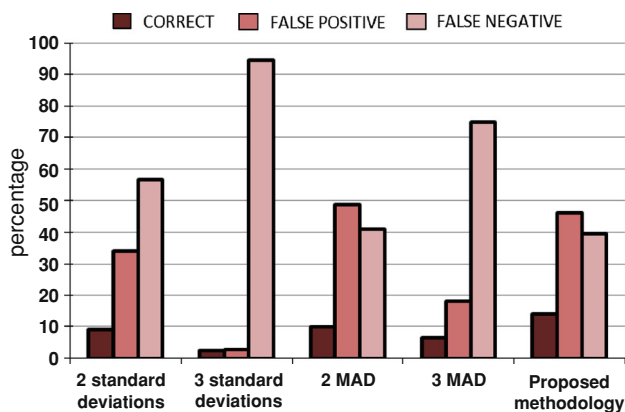


Fig. 10 Percentage of correct, false positive and false negative events detected with each methodology considering all the shopping malls and football stadiums taken into consideration

firstly consists of finding a threshold, and secondly marking as outlier each point that falls outside such a threshold. The main difference basically consists in the choice of the thresholds.

In particular, we focused our attention on the following methodologies (for a more detailed description see (Chandola et al. 2009; Wilcox 2012)):

- **MAD median rule** In statistics, the median absolute deviation (MAD) for a data set X_1, X_2, \dots, X_n , is defined as the median of the absolute deviations from the data's median:

$$MAD = 1.483 \times \text{median}_i(|X_i - \text{median}_j(X_j)|)$$

that is, starting with the residuals (deviations) from the data's median, the MAD is the median of their absolute values. The constant 1.483 is a correction factor which makes the MAD unbiased at the normal distribution. Then, a point in the data is considered an outlier (in our case an event) if it falls outside $\text{median} \pm k \times MAD$. Following the literature of this area (Wilcox, 2012), we considered as thresholds $k = \pm 2$ and $k = \pm 3$.

- **Standard deviation** Another common method for outlier detection consists of finding the mean and the standard deviation of the data set and then call anything that falls more than k standard deviations away from the mean as an outlier. That is, x is an outlier if

$$\frac{|(x - \mu)|}{\sigma} > k$$

As with the previous methodology, we considered as thresholds $k = \pm 2$ and $k = \pm 3$.

Figure 10 shows the results obtained for all the above mentioned methodologies considering all the shopping malls and football stadiums taken into consideration. Looking at the figure, it is clear that the choice of $k = \pm 3$ produces the worst results by using a threshold that is too

high and thus the methodologies do not recognize any events (high number of false negative events). The methodology proposed in this paper, by considering two different thresholds for overcrowded and underpopulated events, and by fine-tuning the threshold values outperforms the other approaches. The most similar results are produced by the 2 MAD technique, which has a slightly higher number of false positive events and a lower number of correct events.

It is important to remark that while these results are significant for a comparison perspective, the absolute percentages of correct, false positive and false negative results are less meaningful in that we average together very different areas of the city. While the proposed approach is effective in identifying events in populated and spatially separated areas (see previous section), it is not able to identify events in scarcely populated areas, or in dense areas in which multiple event-producing venues are covered by the same network cells. The overall average among such different places leads to low absolute values for our approach.

From an application perspective, our approach can thus be effectively applied to areas with the above characteristics. While in other situations has to be combined with other sources of information.

5 Conclusions

The recent availability of mobile phone datasets have led to many discoveries on human behavior. In this paper we have presented a methodology to discover events happening in the city from a large set of human mobility traces as recorded by mobile network usage. Using a 10 month mobile network usage dataset over two regions in Italy, we have shown how two types of events (i.e., football matches and shopping malls special openings) can be successfully detected. Our analysis shows that the performances of the methodology strongly depends on the location of the venues under investigation. Moreover, we have shown some insights regarding the amount of data required to compute reliable statistics. These results can have a strong impact on several application scenarios ranging from location-based services (e.g., in which the user gets notified about events nearby) to urban planning (e.g., to dynamically allocate resources to impromptu events).

Finally, many fascinating directions remain open for further research. One is the evaluation of the performances of our approach both using stronger ground truth information, and by comparing our results with events extracted from other data sources (e.g., geo-localized social networks). Using multiple data sources (i.e., observing urban life from multiple perspectives) is a very promising way to cross validate the results and identify events more reliably. A second aspect is simulation: once the event is detected

from a portion of the day, it is natural to investigate how to build on this basis large-scale simulations, capable of predicting realistic evolutions of complex social phenomena. As a final direction, we have observed that mobile network data are huge, so it is important to build methodologies able to deal with this amount of data in a reasonable time.

The final goal would be to create a service to detect the events happening in the city in real time, and to integrate it in a Web application that we are developing to explore city dynamics: Mr.Typ—Mobile and Real-Time Yellow Pages.³

Acknowledgments Work supported by the SAPERE (*Self-Aware Pervasive Service Ecosystems*) project (EU FP7-FET, Contract No. 256873) and by the project Mr.Typ—Mobile and Real-Time Yellow Pages, funded by Telecom Italia.

References

- Andrienko G, Andrienko N, Mladenov M, Mock M, Poelitz C (2012) Identifying place histories from activity traces with an eye to parameter impact. *IEEE Trans Visual Comput Graph* 18(5):675–688
- Calabrese F, Colonna M, Lovisolo P, Parata D, Ratti C (2011) Real-time urban monitoring using cell phones: a case study in Rome. *IEEE Trans Intell Transport Syst* 12:141–151
- Calabrese F, Rades J, Ratti C (2010) Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *IEEE Pervasive Comput* 9(1):78–84
- Candia J, Gonzalez M, Wang P, Schoenharl T, Madey G, Barabasi A (2008) Uncovering individual and collective human dynamics from mobile phone records. *J Phys A Math Theor* 41(22):1–11
- Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. *ACM Comput Surv* 41(3):58
- Ferrari L, Rosi A, Mamei M, Zambonelli F (2011) Extracting urban patterns from location-based social networks. *ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN)*
- Lee R., Wakamiya S., Sumiya K. (2011) Discovery of unusual regional social activities using geo-tagged microblogs. *The World Wide Web J Special Issue Mobile Serv Web* 14(4):321–349
- Mamei M, Rosi A, Zambonelli F (2010) Automatic analysis of geotagged photos for intelligent tourist services. In: *IEEE intelligent environment*, Kuala Lumpur, Malaysia
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybernet* 9:62–66
- Pedersen T, Jensen C (2001) Multidimensional database technology. *IEEE Comput* 34(12):40–46
- Rattenbury T, Good N, Naaman M (2007) Towards automatic extraction of event and place semantics from flickr tags. In: *International ACM SIGIR conference on research and development in information retrieval*, Amsterdam, The Netherlands
- Ratti C, Pulselli RM, Williams S, Frenchman D (2006) Mobile landscapes: using location data from cell phones for urban analysis. *Environ Plann B: Plann Design* 33(5):727–748
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: *International conference on World Wide Web*
- Vieira M, Frias-Martinez V, Oliver N, Frias-Martinez E (2010) Characterizing dense urban areas from mobile phone-call data: discovery and social dynamics. In: *International Conference on Social Computing (SocialCom)*, pp 241–248
- Wilcox R (2012) *Introduction to robust estimation and hypothesis testing, statistical modeling and decision science*, 3rd edn. Elsevier, Amsterdam

³ <http://www.mrtyp.it>