EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

CrossMark

Countrywide arrhythmia: emergency event detection using mobile phone data

Didem Gundogdu[1,2*], Ozlem D Incel[3], Albert A Salah[4] and Bruno Lepri[1]

*Correspondence:
gundogdu@fbk.eu
[1] Fondazione Bruno Kessler, via Sommarive, 18, Trento, Italy
[2] DISI, University of Trento, via Sommarive, 9, Trento, Italy
Full list of author information is available at the end of the article

Abstract

Large scale social events that involve violence may have dramatic political, economic and social consequences. These events may result in higher crime rates, spreading of infectious diseases, economic crises, and even in migration phenomena (e.g., refugees across borders or internally displaced people). Hence, researchers have started using mobile phone data for developing tools to identify such emergency events in real time. In our paper, we apply a stochastic model, namely a Markov modulated Poisson process, for spatio-temporal detection of hourly and daily behavioral anomalies. We use the call volumes collected from an entire geographic region. Our work is based on the assumption that people tend to make calls when extraordinary events take place. We validate our methodology using a dataset of mobile phone records and events (emergency and non-emergency) from the Republic of Côte d'Ivoire. Our results show that we can successfully capture anomalous calling patterns associated with violent events, riots, as well as social non-emergency events such as holidays, sports events. Moreover, call volume changes also show significant temporal and spatial differences depending on the type of an event. Our results provide insights for the long-term goal of developing a real-time event detection system based on mobile phone data.

Keywords: behavioral anomaly detection; emergency events; mobile phone data; Markov modulated Poisson process

1 Introduction

Large scale social events can happen anytime, anywhere, and without warning. Examples are clashes between ethnic communities, violence among supporters of political groups or sports clubs, demonstrations and celebrations. Some of these events cause migration phenomena (e.g., refugees and internally displaced people) [1, 2], higher crime rates and spreading of infectious diseases [3, 4], and result in economic crises [5]. Subsequently, researchers have recently started to automatically identify emergency events by using new sources of data, such as geo-referenced social media and mobile phone data [6–8]. In particular, the almost universal adoption of mobile phones is generating an enormous amount of data about human behavior with a breadth and depth that were previously inconceivable. In 2013, there were 6.8 billion mobile phone subscriber accounts worldwide, with millions of new subscribers every day, corresponding to a penetration of 128% in the developed world and 90% in developing countries [9].

In our paper, we investigate the usage of mobile phone data to detect and characterize emergency and non-emergency events. Specifically, we adopt the Markov modulated Poisson process framework [10, 11] (MMPP) for the spatio-temporal detection of hourly and daily behavioral anomalies in call volume, and we discuss the relationship between these anomalies and the actual emergency and non-emergency events that might have caused them. Compared to previous studies [8, 12–17], we do not start with the location and the time of an already known event, but we use an unsupervised approach that spatio-temporally identifies unusual calling behavior. Moreover, our approach detects not only daily anomalies, but also hourly anomalies. Hence, we are able to capture behavioral responses occurring within hours of an event.

To validate our approach, we use mobile phone records from Ivory Coast (officially Republic of Côte d'Ivoire) in Africa. The data, collected from December 1, 2011 to April 28, 2012 during the post-election crisis, were released for Orange's 'Data for Development Challenge' (D4D) [18, 19] and contain calls between 5 million customers.

Our results show that we can successfully capture anomalous calling patterns associated with violent events, riots, as well as social non-emergency events such as holidays, sports events on an hourly and daily basis. Moreover, we illustrate that call volume changes also show significant temporal and spatial differences depending on the type of an event. Unlike previous work on classification of social events [20], we find that the coverage of the spatial impact is more significant than the duration.

In summary, the main contributions of this study are:

- We construct a detailed database of emergency and non-emergency events in the Ivory Coast using multiple sources of data and merge them with the geographical locations of the cell towers. This database is used as ground truth in our analysis;
- We adopt a Markov modulated Poisson process (MMPP) to spatio-temporally detect hourly and daily behavioral anomalies in call volume;
- We test our methodology using the Call Detail Records (CDRs), aggregated to the cell tower level, of an entire country;
- We discuss the correspondence between the anomalies found and the actual emergency and non-emergency events that might have caused them;
- We highlight and discuss the different spatial and temporal signatures of the discovered events.

The rest of the paper is structured as follows. Section 2 discusses related work on using mobile phone data for measuring human behavior and previous approaches to similar event detection problems. In Section 3, we present the proposed Markov modulated Poisson process for identifying anomalies. We describe the 'Ivory Coast Dataset', the Call Detail Records (CDRs) and the event records, used to validate our approach in Section 4. In Section 5, we evaluate the correspondence between anomalies found and actual emergency and non-emergency events, reporting comparative experimental results with two other approaches from the literature. Before concluding the paper, we elaborate on what we have learned from matched and unmatched events, and we discuss some limitations of our study (Section 6).

2 Related work

In this section we review the literature on understanding human behaviors from Call Detail Records (CDRs), and event detection methodologies.

2.1 Understanding human behaviors from call detail records data

Mobile phone operators can analyze the behavior of a large number of people from their aggregated mobile phone usage [21]. The Call Detail Records (CDRs) stored by operators (typically for billing purposes) can be exploited to extract mobility patterns [17, 22, 23], to model social interactions [24, 25], to analyze the dynamics of a city [26, 27], to understand epidemics [28, 29], to estimate population densities [30], and to predict energy consumption patterns [31], and socio-economic indicators and outcomes of territorial disputes [32–34].

In the last few years, several studies have shown that natural and man-made emergency events (e.g. earthquakes, floods, bombings, riots) can be reflected by dramatic increases in calling and mobility behaviors [8, 12–17, 20, 35].

The assumption behind these works is that significant changes in behavior, captured by mobile phone data, will indicate the occurrence of extreme events. Indeed, people tend to share and inform each other about an emergency event typically right after it is realized [36]. However, planned non-emergency events also occur and may provoke significant changes in mobile phone behavior. Bagrow *et al.* found easily detected changes in call frequency during festivals, concerts, sport events, and it is likely that also other events such as holidays may produce similar changes [13].

The call data can be analyzed at the individual level, to detect the changes from the expected call and mobility behavior [8, 37, 38]. This approach requires some restrictions for data privacy, and incurs high computational cost. These problems can be overcome by computing on the aggregated call volume of each cell tower [20].

Typically, the approaches that analyze CDR data start with the time and the location of an already known event and then look for anomalous calling behavior at that time and location. The most common way of anomaly detection in time series is to define a baseline. With a supervised approach, very good results can be achieved. In this paper, we take an unsupervised approach to build a model without having prior information.

Recently, Dobra *et al.* proposed an unsupervised behavioral anomaly detection system that identifies days and locations with unusual calling or mobility behavior without knowing the event [39]. The authors used mobile phone records from Rwanda in order to connect the identified anomalous days and locations with extensive records of violent and political events (e.g., protests, violence against civilians) and natural disasters (e.g., earthquakes). Specifically, they computed the aggregated daily mobility and calling patterns of each site, starting from the individual behavior of all assigned subscribers for the corresponding site. This methodology suffers from high computational costs and works only on a daily basis. Our approach instead takes as input the calling volume of a cell tower without accessing individual data. Hence, we are also able to detect hourly variations and the behavioral responses occurring within hours of an event.

There are two main approaches for classification of emergency and non-emergency events from mobile phone data. The first approach is the analysis of mobility patterns to model individuals and crowds, which has been used to identify concerts and sport events [8, 40]. The second approach is the analysis of temporal patterns, which can be used to differentiate between event types according to the duration of behaviors [20]. We propose a third approach by observing spatio-temporal characteristics of anomalous behaviors.

A number of works used the D4D Ivory Coast dataset to validate event detection approaches [19]. Paraskevopoulos *et al.* [41] used numerical analysis to detect anomalies.

Their model is based on call duration statistics, which is calculated by dividing the cumulative call duration by the number of calls, for each day and for each cell tower. During an emergency, people may indeed tend to make more calls, but these can be with shorter durations. Hence, looking at the total duration may not be sufficiently discriminative. Dong *et al.* used the D4D dataset for modeling the movements of flocks of people [38]. Their assumption is that people move in groups when an extraordinary event happens. This assumption can provide valuable insights to detect and characterize protests. However, this mobility pattern is not observed during attacks against civilians that take place in urban areas. In Section 5.1 we will provide comparative results with the two recent methods described briefly here.

2.2 Event detection methodologies

In the literature, anomalous event detection, which corresponds to the task of detecting anomalies from time series, is strongly related to *outlier detection* [42], and *change point detection* [43, 44].

In our application scenario, the normal behavior is observed by the temporal (hourly) changes of the aggregated call volumes, collected from different cell towers. Similar discrete, count-based analysis problems have been tackled in statistics, econometrics, psychology, and ecology [45–48].

In particular, the change point methodology is a well-studied topic in statistics [43]. Previously, Zhang *et al.* used this approach to detect anomalies in mobile phone data [49]. However, the problem of this method is that it does not preserve the periodicity of data.

Other researchers proposed the usage of Hidden Markov Models (HMM) [50, 51] for the anomaly detection task. As highlighted by [52], the MMPP approach that we propose to use in this paper is a special case of HMM and Markov Chain Monte Carlo (MCMC) approaches. MMPP has also similarities with the change point methodology [53] but it has the advantage of preserving periodicity.

3 Methodology

Our goal is to investigate the usage of mobile phone data to detect and characterize anomalous events. Poisson processes are commonly used to model count data (data in which the observations can take only the non-negative integer values) [11, 54] in many domains, such as modeling rare incidents in psychiatric hospitals [45], and traffic analysis [52, 55].

3.1 Markov modulated Poisson process

In our setting, the count data are represented by the number of total calls in 1-hour time intervals, denoted as $N(t)$, for each cell tower, denoted as k.

The model is represented as a two state Markov chain, with a normal call behavior state $z(t) = 0$ and an abnormal call behavior state $z(t) = 1$ (see Equation (1)). The transitions between the states are defined with a transition probability matrix M_z in Equation (2)

$$z(t) = \begin{cases} 1, & \text{if there is an event in time } t, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

$$M_z = \begin{pmatrix} 1 - z_0 & z_1 \\ z_0 & 1 - z_1 \end{pmatrix}. \tag{2}$$

Our observations are denoted as $N(t)$, and hidden variables are the amount of calls in a normal call pattern $N_0(t)$, and the amount of calls initiated because of an anomalous event $N_E(t)$. The definition of $N^k(t)$ for a cell tower k is given by the summation of hidden variables $N_0^k(t)$ and $N_E^k(t)$, as shown in Equation (3)

$$N^k(t) = N_0^k(t) + N_E^k(t).$$
(3)

Hence, we assume that the number of calls are generated through a heterogeneous Poisson distribution $\text{Pois}(N, \lambda(t))$, with the rate value that is a function of time $\lambda(t)$, as shown in Equation (4)

$$\text{Pois}(N; \lambda(t)) = e^{-\lambda(t)}(\lambda(t)^N / N!).$$
(4)

The normal call pattern of a cell tower k is dependent on the day (i) of the week (δ_i^k) and the hour (j) of the day $\eta_{j,i}^k$, and λ_0^k represents the average rate of cell tower (k) in one week. We can formulate the rate function of the Poisson distribution $\lambda^k(t)$ as the product of the initial rate λ_0^k, the daily effect δ_i^k, and the hourly effects $\eta_{j,i}^k$. The detailed derivations can be found in Appendix A.1.

To evaluate event detection models, it is necessary to annotate existing data by cross-referencing it with news sources. We provide a list of important events in Table 3 and Table 4 (details explained in Section 4.2). For each important event, we have grouped the cell towers close to the location provided in the ground truth, and associated events to cell towers probabilistically. Each cell tower's probability distribution is normalized before calculating the average probability of the region. If the event probability is higher than a defined threshold τ (e.g. 0.15), we classified this event as detected in the region.

3.2 Baseline model

To assess the MMPP performance, we have implemented a baseline model, which we describe in this section. For each cell tower, the hourly average call volume is calculated as shown in Equation (5). In our notation, each cell tower is denoted with k, days are indexed with i, hours with j, $N(t)$ is the observation for time t, and D is the total number of days for the data collection period. Once the averages are calculated, we subtract the average value, Φ, from the observed call volume for time t. If the obtained value is higher than a defined threshold (τ), the event is labeled as anomalous.

$$\Phi_{i,j}^k = \frac{\sum_{i=1,j}^{D} \sum_{i,j=1}^{24} N(i,j)^k}{D}, \quad \forall k,$$
$$\left| (N(i,j)^k - \Phi_{i,j}^k) \right| > \tau.$$
(5)

4 The Ivory Coast dataset

In this section, we introduce the datasets that have been used to evaluate our proposed methodology: (i) a mobile phone records dataset, obtained by Orange for the 'D4D Challenge,'[a] and (ii) an event records dataset obtained from a variety of public sources (e.g., Armed Conflict Location and Event Data Project, United Nations Council and International Crisis Group reports, local and international news).

Table 1 Summary statistics of the Ivory Coast dataset and the country

Total area of the country (hectares)	32,246,300
Population of the country	around 20 million
Number of subscribers	5 million
Type of the data	Cell tower to cell tower
Total number of calls	190,638,909
Time period	7 weeks / 20 weeks

4.1 Call details records

The Orange dataset contains anonymized and aggregated calls between 5 million customers from December 1, 2011 to April 28, 2012 (referred to as Set 1 in the D4D Challenge). Specifically, the dataset contains the total volume (number of calls and SMSs) and duration of calls between each pair of cell towers over the entire period. The total number of cell towers is 1,238; however, in the pre-processing phase, we have eliminated the cell towers, which are not present during the whole period, ending up with 970 cell towers to analyze. The exact locations of the towers are not provided by Orange Telecom, due to the company's operational confidentiality. Finally, it is worth noticing that the data consist of the total traffic between cell towers, and hence no individual data are ever accessed.

Mobile phone penetration is high enough (95%) [56] in Ivory Coast to make such a dataset sufficiently representative of the population. Moreover, the network operator holds a dominant position in Ivory Coast (48% of the market share). Table 1 provides summary statistics for the country and the dataset.

Specifically, the call data consist of (i) date, (ii) hour, (iii) initiating cell tower, (iv) destination cell tower, (v) aggregated number of calls, and (vi) aggregated duration of calls. The CDR data with an unassigned cell tower are deleted. From December 14, 2011 to January 19, 2012, half of the country had a shortage of energy, confirmed by the Orange Telecom authorities. The data collected during this period are deleted. Additionally, we have missing data from 22nd of January to 30th of January. To preserve the weekly periodicity without being affected by data collection issues, we analyze the data from December 5 to December 11, and from January 30 to March 11 (49 days in total).

To provide some geographical context, Figure 1 shows the rough locations of cell towers in relation to the regional boundaries of Ivory Coast (255 subprefectures). It shows the dense concentration of cell towers in and around the largest cities of each region. Abidjan, the economic capital of Ivory Coast, has a significantly higher concentration of cell towers compared to the rest of the country.

4.2 Event records

Our data on violent and political events, major holidays and major events (e.g., elections, the African Football Cup) come from a various structured and unstructured public data sources such as Armed Conflict Location and Event Data Project (ACLED) [57], United Nations Council and International Crisis Group reports, local and international news. ACLED and UN Security reports include extensive data on conflict-related events including riots, protests, killings, and battles. The information is obtained by local or international newspapers (e.g. Notre Voie, Le Patriote, France24, BBC) and radio sources, and it includes details such as the date and location of each event, the type of event, the groups involved, and the fatalities. Ivory Coast has been suffering unstable political conditions during the data collection period. According to the United Nations Human Rights De-

Figure 1 Cell tower locations within subprefectures in Ivory Coast.
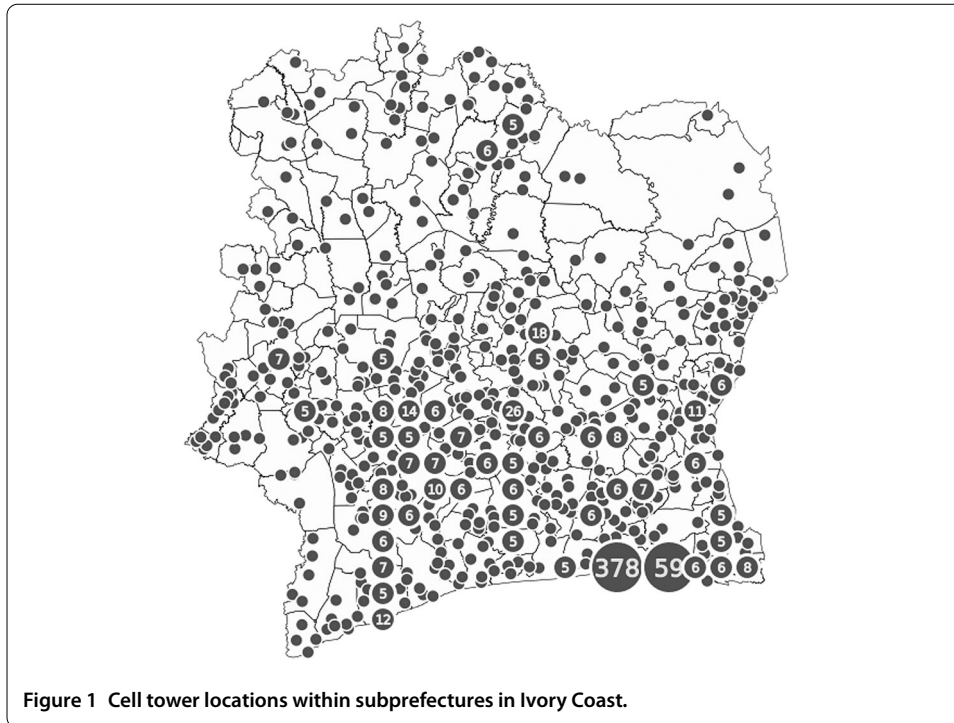
partment, more than 600,000 Ivorians were displaced in the country and around 200,000 Ivorians migrated to neighboring countries in order to be in secure living conditions [58].

In total, we have gathered 19 emergency events (e.g., confrontations between groups, protests) and 11 non-emergency events (e.g., national and regional holidays, African Football Cup games). The complete list of events (emergency and social non-emergency events) is shown in Table 3 and Table 4.
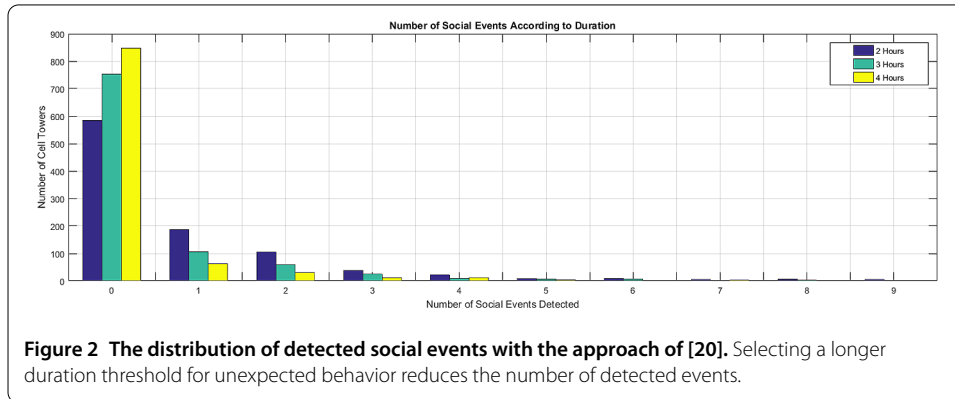
5 Experimental results

Our approach identifies many hours and days with unusual increase in calling volume. As in [39], these anomalies are found sometimes in a specific site and sometimes across multiple sites. Specifically, our approach provides the probability of having an anomalous event in a specific location.

As we mention in Section 1, our goal is to identify emergency events. However, non-emergency events might also produce behavioral changes in calling volume. Hence, it is relevant to highlight and discuss the specific behavioral signatures of the different events in order to detect the emergencies effectively. In this regard, from the 19 annotated emergency events, our method automatically detected 15, while for the social non-emergency events it was able to identify 8 out of 11 events. Similar to [39] and [11], we find some hourly and daily anomalies that we were not able to match to any of the recorded events (i.e. possible false positives).

We compare our MMPP approach with a baseline model, which has been explained in Section 3.2, and show that our approach outperforms the baseline. Indeed, from the 19 annotated emergency events, the baseline approach is able to identify 8, while for the social non-emergency events, it is able to identify 7 out of 11. While the nature of the problem dictates a very large dataset with very sparse true positives, the results are quite promising. Before describing our results in detail, we compare our method with the event

Table 2 Comparative results for the MMPP model and for Dong *et al.*'s approach [38]

	Precision	**Recall**	***F*-score**
MMPP	0.31786	0.9219	0.4727
Dong *et al.*	0.0676	0.92	0.125

Figure 2 The distribution of detected social events with the approach of [20]. Selecting a longer duration threshold for unexpected behavior reduces the number of detected events.

identification approach by Dong *et al.* [38], as well as with the event type classification approach described by Young *et al.* [20] in the next subsection.

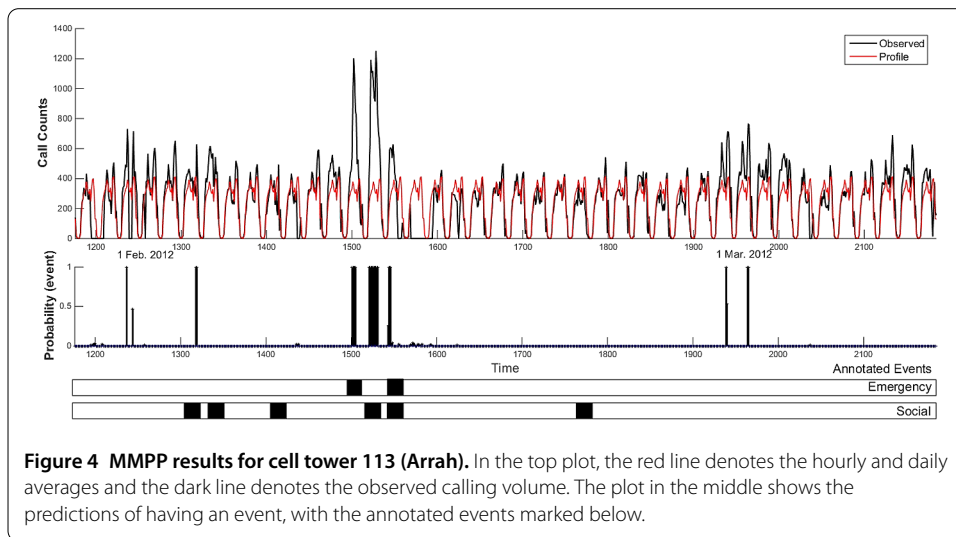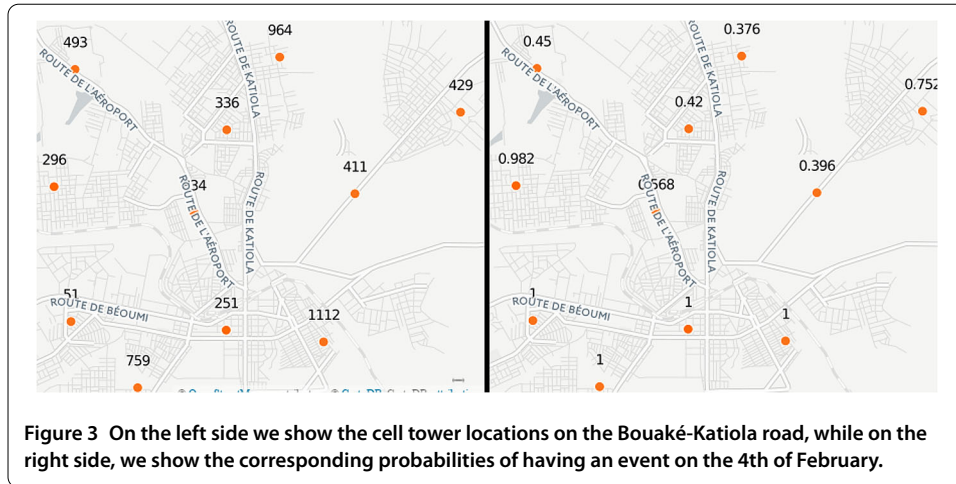## 5.1 Comparison with recent approaches

Dong *et al.* proposed a methodology in 2015 for event identification based on modeling individuals' mobility as flocks through the city, and they tested this methodology on the D4D dataset [38]. They reported a *precision* value of 0.0676 and a *recall* value of 0.9200. In order to provide a fair comparison, we tested the MMPP approach on the same experimental setting proposed in their paper, with 140 days of data and 25 ground truth events. The results show that MMPP outperforms this approach in all the measures (see Table 2).

We also compare our approach with the one proposed by Young *et al.* in 2014 to classify social non-emergency events [20]. This approach posits that social events have longer temporal duration than emergency events. Since the duration of deviation is a major parameter, we compared the methods for observation periods of two, three, and four consecutive hours. Setting this to two hours means that an event is detected if the observed behavior deviates from the expected for two hours. For each of the three settings, Figure 2 shows the number of social events detected by a given number of cell towers. The best performance is obtained by using two consecutive hours, where the average number of detected social events is equal to 1.34. Using MMPP, on the other hand, we are able to detect 8 social events. We now describe the findings of our method in detail.

6 Discussion and limitations

In this section, we match some of the anomalies identified along with key events that occurred in Ivory Coast and we discuss what we have learned from matched and unmatched events. Then, we discuss some limitations of our study.

In 2010 and 2011, Ivory Coast suffered from a *post election crisis.* The election results initiated a civil war, and effects of the war were still present during the data collection period. Therefore, our emergency events are mainly due to these unstable political conditions.

Figure 3 On the left side we show the cell tower locations on the Bouaké-Katiola road, while on the right side, we show the corresponding probabilities of having an event on the 4th of February.

Figure 4 MMPP results for cell tower 113 (Arrah). In the top plot, the red line denotes the hourly and daily averages and the dark line denotes the observed calling volume. The plot in the middle shows the predictions of having an event, with the annotated events marked below.

Violence against civilians - February 3, February 19, and March 8, 2012 On February the 3rd, the proposed system did not detect any significant changes in call volume near the location of the event, close to Bouaké. Instead, we see a clear anomaly in this area on February the 4th (see Figure 3).

Similarly, no anomalies were detected on February the 19th, while on March the 8th we detect an anomaly close to Bouaké (cell tower 964) around 11am. Interestingly, this finding is not detected by the baseline model.

Violence against civilians - February 11 and February 13, 2012 In Arrah, it was reported that the confrontation of two groups ended with three people killed and at least 19 injured. The visualization of our model's output is shown in Figure 4 for cell tower 113. In the top plot, the red line denotes the hourly and daily averages and the dark line denotes the observed calling volume. The plot in the middle shows the predictions of an event occurring, with the annotated events marked below.

Protests against the government - February 18, 2012 MMPP detects calling volume increases from a single cell tower at 3pm and 5pm on February 18, 2012. The cell tower is
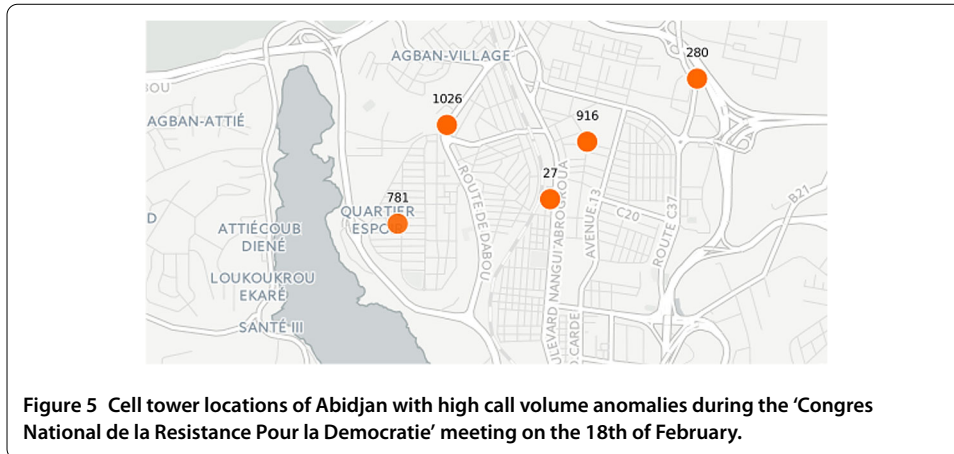
Figure 5 Cell tower locations of Abidjan with high call volume anomalies during the 'Congres National de la Resistance Pour la Democratie' meeting on the 18th of February.
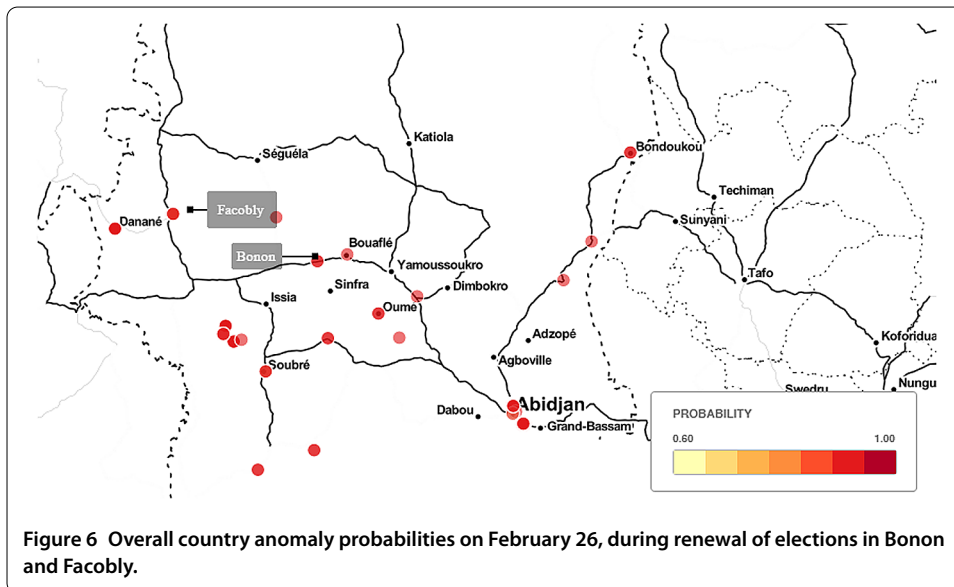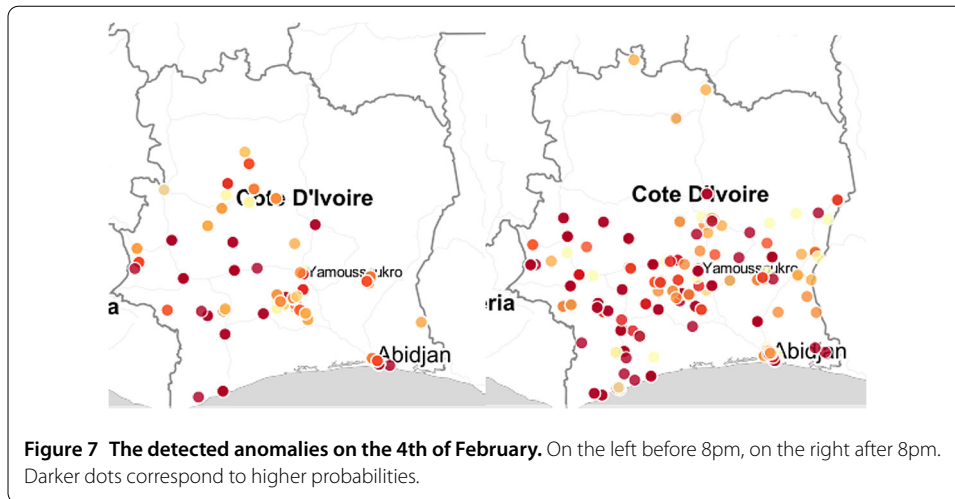

Figure 6 Overall country anomaly probabilities on February 26, during renewal of elections in Bonon and Facobly.

located in the city center of Abidjan, the economic capital of the Ivory Coast. A non-violent protest against the government was reported by the United Nations on the same day in Abidjan, in front of the 'Congres National de la Resistance Pour la Democratie' headquarters located in the city center. We assume those protests took place in the western part of the map, shown in Figure 5. Each orange dot represents cell towers with an anomalous call volume. Unfortunately, the reports of United Nations do not specify the duration of the protest.

Elections and violent clashes in Bonon and Facobly - February 26, 2012   After 5pm, we observe anomalous calling patterns in two regions of the mid-west, Bonon and Facobly, for the cell towers 239, 1154, 374, 181. On the same day, a couple of events were recorded for these two regions, political elections and consequent violent clashes between political opponents, as shown in Figure 6. The violent events ended with the death of five people. Interestingly, we observe that these violent events seem to have effects not only in the regions involved, but also in Abidjan. Indeed, we detect an anomaly in Abidjan after 5pm.

Figure 7 The detected anomalies on the 4th of February. On the left before 8pm, on the right after 8pm. Darker dots correspond to higher probabilities.

Although Abidjan is not the administrative capital of the Ivory Coast, the headquarters of the two opposition parties are located in this city.

Violence against civilians - February 29, 2012   The event records report that FRCI (Forces Republicaines de Côte d'Ivoire) shot civilians in Séguéla, and this violent repression resulted in the death of two people. On the same day, we observe an unusual calling activity in the northern part of Séguéla, close to the border with Guinea.

As previously mentioned, we also collected 11 non-emergency events (social events and holidays). The main characteristic of those types of events is that they affect almost the entire country. In the following, we discuss a couple of relevant non-emergency events.

African football cup, January 21 - February 12, 2012   In 2012, the African Football Cup was held in Equatorial Guinea and Gabon. Ivory Coast played in the final match and dramatically lost the championship on penalties. It is worth noting that football is one of the most important social activities in Ivory Coast and that the day after the finals was proclaimed a national holiday. This enthusiasm is spread across the entire country. Interestingly, African football cup matches have a highly specific call signature, causing a very high call volume for one or two hours after the match on cell towers across the country. Figure 7 contrasts the anomalies detected before and after the match. Specifically, the left side of Figure 7 shows the anomalies detected before 8pm and the right side shows the ones detected after 8pm.

Mavlid an Nabi (celebration of the birth of the prophet), February 4 - February 5, 2012 The Ivory Coast is composed of different ethnic groups and religions. The north is mainly Muslim and the south is mainly Christian. During these two days we found calling anomalies in the northern part of the country.

Ash Wednesday Christian festival, February 22, 2012   We do not observe any significant difference in the calling patterns of the Christian south and the Muslim north. Only in the capital city, Yamoussoukro, during the whole day and night we register some unusual calling activity. Interestingly, Yamoussoukro has the biggest church in the world, the Basilica

of Our Lady of Peace. This monument usually sees large number of Christians gathering during the festival.

Our results do not confirm the patterns observed by [59] and [60] that found a positive correlation between the mobile phone coverage and the probability of having a conflict. In our setting, the densest population and the highest coverage is in the Abidjan area, while the violent events tend to happen in the mid-west of Ivory Coast, where many conflicts between different ethnic groups exist.

Another finding is an irregularly high call volume on the first day of each month. One possible reason is the high penetration of mobile payments in Ivory Coast. People typically pay their utilities, rents, etc. on this day. However, we do not have any formal way of confirming this, and we assume this behavior is normal. Note that our work suffers from a number of limitations. First of all, we focused our attention only on anomalous increases in calling patterns. However, emergency and non-emergency events may also cause changes in mobility patterns, as shown by [39]. Second, the annotated event records may be incomplete given the scarcity of information and the chaotic socio-political situation during the data collection period. Moreover, the exact event locations are often not provided and the dates can fluctuate a bit in the reports. For example, UN records report events in Arrah from February 11 to February 13, while *Le Figaro* newspaper [61] reports the same events from February 12 to February 13 (our results seem to confirm the dates provided by *Le Figaro*). Third, we observed daily data losses in several cell towers, e.g. *cell tower 49* on the 5th of December. This data loss tends to generate a high number of false positives when the average call volume of a cell tower is high, e.g. greater than 1500. Finally, the MMPP performance is sensitive to prior parameters, even though we have empirically shown that cell tower call volumes can be easily and robustly estimated from a few days worth data [11]. Yet, in our case, the model's performance may be sensitive to the cell tower selected for tuning the parameters.

7 Conclusions

In this paper, we have proposed an approach based on Markov modulated Poisson processes for spatio-temporal detection of hourly and daily behavioral anomalies in call volume. Our work is based on the assumption that people tend to make calls when extraordinary events take place. We validate our methodology using a dataset of mobile phone records, together with emergency and non-emergency events from Ivory Coast. Our results show that we can capture anomalous calling patterns associated with violent events, protests, holidays and major sport events (e.g. African Football Cup games). One of our findings is that the impact area is a better feature than the temporal duration for identifying social non-emergency events from mobile phone data, which runs counter to the general opinion held in the literature. Another valuable finding is that for event detection, analysing mobile phone activity as a time series gives better performance compared to tracing movements of the masses. It is worth noticing that our methodology uses only aggregated call volumes of cell towers and no data can be traced back to individuals. Hence, there are minimal - if any - privacy concerns.

In sum, we believe that our work contributes to the process of creating an effective emergency detection system that may be used by governments, policy makers, and international organizations to significantly increase human well-being and the wealth and security of countries.

As future work, we are planning to target anomalies in mobility patterns and apply our approach to other datasets, both from developing and developed countries.

Appendix

A.1 Derivation of Markov modulated Poisson process

In this section we present how we implement the Markov modulated Poisson process in order to detect anomalous events.

The model is represented as a two state Markov chain, with a normal call behavior state $z(t) = 0$ and an abnormal call behavior state $z(t) = 1$, as shown in Figure 8. The transitions between the states are defined with a transition probability matrix M_z in Equation (6), which is time independent but dependent on previous transition probabilities

$$z(t) = \begin{cases} 1, & \text{if there is an event in time } t, \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

$$M_z = \begin{pmatrix} 1 - z_0 & z_1 \\ z_0 & 1 - z_1 \end{pmatrix}. \tag{7}$$

Our observations are denoted as $N(t)$, and hidden variables are as follows: the amount of calls in a normal call pattern $N_0(t)$, the amount of calls initiated because of an anomalous event $N_E(t)$ and the transition probabilities of events $z(t)$ as in Equation (7). The definition of $N^k(t)$ for a cell tower k is given by the summation of hidden variables $N_0^k(t)$ and $N_E^k(t)$, as shown in Equation (8)

$$N^k(t) = N_0^k(t) + N_E^k(t). \tag{8}$$

We assume that the number of calls are generated through a heterogeneous Poisson distribution $\text{Pois}(N, \lambda(t))$, with the rate value that is a function of time $\lambda(t)$, as shown in Equation (9). Changes in the count data are modeled taking into account the periodicity depending on the day and on the hour of the day. In Appendix A.1.1, this property is used to calculate the posterior distributions

$$\text{Pois}\big(N; \lambda(t)\big) = e^{-\lambda(t)}\big(\lambda(t)^N / N!\big). \tag{9}$$

The normal call pattern of a cell tower k is dependent on the day (i) of the week (δ_i^k) and the hour (j) of the day $\eta_{j,i}^k$, and λ_0^k represents the average rate of cell towers in one week.
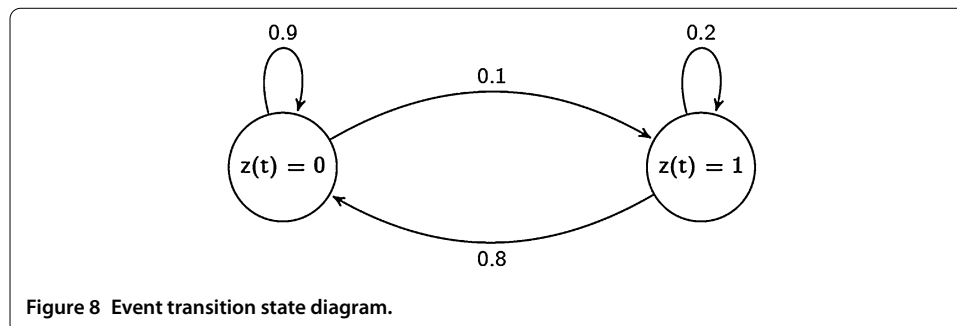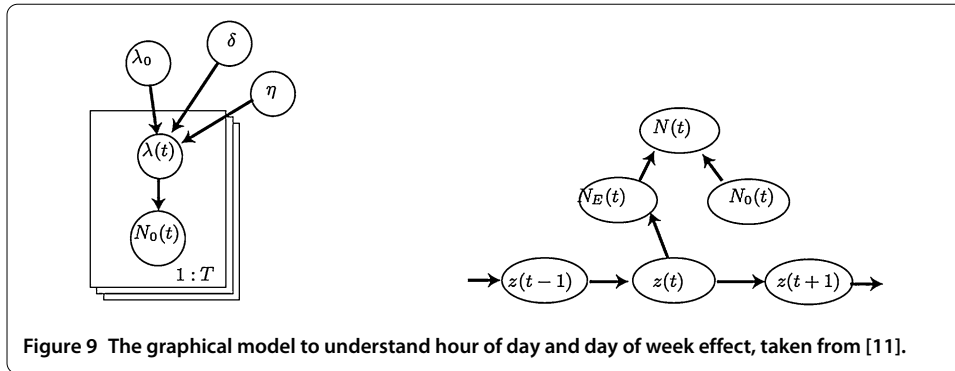

Figure 8 Event transition state diagram.

Figure 9 The graphical model to understand hour of day and day of week effect, taken from [11].

We can formulate the rate function of the Poisson distribution $\lambda^k(t)$ as the product of the initial rate λ_0^k, the daily effect δ_i^k, and the hourly effects $\eta_{j,i}^k$, Equation (10). Plate notation is used to depict the repeating form of the data, as shown in Figure 9

$$\lambda^k(t) = \lambda_0^k \delta_{d(t)}^k \eta_{d(t),h(t)}^k. \tag{10}$$

Conjugate prior ensures that the posterior distribution is coming from the same family of the prior distribution. Hence, our model becomes tractable and it can be written in closed form. Random variables δ, η, denoting the day of the week and the hour of the day, are coming from a multinomial distribution as shown in Equation (11) and the conjugate prior is selected as Dirichlet distribution in Equation (12)

$$\sum_{i=1}^{7} \delta_i = 7,$$
$$\sum_{i=1}^{D} v_{j,i} = D \quad \forall j, \tag{11}$$

$$\lambda_0 \sim \Gamma\left(\lambda; a^L, b^L\right),$$
$$\frac{1}{7}[\delta_1, \ldots, \delta_7] \sim Dir\left(\alpha_1^d \cdots \alpha_7^d\right), \tag{12}$$
$$\frac{1}{D}[\eta_{j,i}, \ldots, \eta_{j,D}] \sim Dir\left(\alpha_1^h \cdots \alpha_D^h\right).$$

In the right side of Figure 9, $z_{(t-1)}$, $z_{(t)}$ and $z_{(t+1)}$ show the time series properties of the event transitions. $N_0(t)$ and $N_E(t)$ are hidden variables, and $N(t)$ is our observation

$$z_0 \sim \beta\left(z; a_0^Z, 0^Z\right), \tag{13}$$

$$z_1 \sim \beta\left(z; a_1^Z, b_1^Z\right), \tag{14}$$

$$N_E(t) = \begin{cases} 0, & z(t) = 0, \\ P(N; \lambda(t)), & z(t) = 1, \end{cases} \tag{15}$$

$$\lambda(t) \sim \Gamma\left(\lambda; a^E, b^E\right). \tag{16}$$

Table 3 List of emergency and social non emergency events. MMPP and baseline results columns show the detected events with 'X' and the undetected ones with '-'. Specifically, we use 'E' for emergency events and 'S' for social non emergency events. For the emergency events not gathered from UN reports [62] we cite the source

Date		Results MMPP		Results baseline		Cell T.	Emergency events	Social events
		E	S	E	S			
Dec. 11	7		X		X			Anniversary of death of Felix Houphouet Bougny
	11		X		X			New Parliament Selection
	15						Yopougon [63]	
	16						Yale, violent incident	
	17						Yale and Vavoua [63] violent incident	
	18						Vavoua [63]	
	25						Sikensi [63]	Christmas holiday
	26						Sikensi [63]	
Jan. 12	1							New Years Day
	4						Peite Guiglio	
	4						Boumi near Bouaké	
	5						Dobia	
	6						Toa Zeo near Dukou	
	6						Daloa	
	8							Baptisim of Lord Jesus fest
	9						Ayami	
	14							Arbeen Iman Huseyin fest
	15						Gagnoa	
	17							Annan and Clinton Visit
	18							Annan and Clinton Visit
	20						Touba	
	21						Abidjan Meeting	
	22						Konsou, Gnabra and Zedekan	CUP CIV & Sudan 17-18
	30	X		-				CUP CIV & Angola 19-20

Event probabilities can be sampled through the densities of a normal event $P(N; \gamma)$ times the probability of having an anomalous event $\gamma(t)$ as shown in Equation (17)

$$\int P(N; \gamma)\Gamma\left(\gamma; a^E, b^E\right) = Nbin\left(N; a^E, \frac{b^E}{1 + b^E}\right). \tag{17}$$

The hyper parameters of Gamma Distribution Γ, a^E and b^E in Equation (16) set the distribution's sharpness or smoothness between the transition states Z_0 and Z_1. In traumatic events, the transition generates a sharp peak. Estimation of these two parameters are shown in Appendix A.1.1.

A.1.1 Inference and parameter estimation

In our model, we need to estimate the parameters for the Gamma distribution $\Gamma(\gamma; a^E, b^E)$ in order to model the probability of an anomalous event $p(z(t)^k | N(t)^k)$. Let us assume $\{N_0(t), N_E(t), z(t)\}$ are given. Then, we may estimate the parameters of the distribution by computing the maximum likelihood [43], given that all other variables are conditionally independent. However, as in Figure 9, only $N(t)$ is observed, without separation into $N_0(t)$ and $N_E(t)$. The rest of the parameters, including $z(t)$, are estimated from the observations. Since we do not know when a social or emergency event takes place, we estimate the parameters of the model with the information we obtained. In that respect, we take one cell

Table 4 List of emergency and social non emergency events. MMPP and baseline results columns show the detected events with 'X' and the undetected ones with '-'. Specifically, we use 'E' for emergency events and 'S' for social non emergency events. For the emergency events not gathered from UN reports [62] we cite the source

Date		Results MMPP		Results baseline		Cell Towers	Emergency events	Social events
		E	S	E	S			
Feb. 12	3	-		X		616, 336, 834, 964, 251	Bouaké-Katiola road	
	4		X		X			CUP CIV & Equar. Guina 20-21
	4		X		X			Mavlid an Nabi (Sunni)
	5		-		X			Day after prophet holiday
	8		-		-			CUP CIV & Mali 20-21
	9	X	X	-	-	398, 533, 426, 884, 534	Zibabo Yablo near Duekoue	Mavlid an Nabi (Shia)
	11	X		-		113, 16, 92	Arrah in eastern Cote d_Ivoire	
	12	X	X	X	X	113, 16, 92	Arrah in eastern Cote d_Ivoire	CUP CIV & Zambia 20:30-21:30
	13	X	X	X	X	113, 16, 92	Arrah in eastern Cote d_Ivoire	African cup recovery holiday
	16	X		-		764, 662, 698	Tueho village near Man	
	18	X		X		781, 1026, 27, 198, 436	Abidjan meeting	
	19	-		-		616, 336, 834, 964, 251	Bouaké-Katiola road	
	19	X		-		534, 1071, 645, 677	Westernpart of the country	
	21	-		-		872, 386, 1055	Zriglo in the border area with Liberia	
	22		-		-			Ash Wednesday festival
	26	X		X		239, 1154, 374	Bonon and Facobly election was attacked	
	27	X		-		546, 776, 916, 209, 27	Abidjan	
	29	X		X		285, 1128, 9, 819	Abobo	
	29	X		X		288, 963, 711	Séguéla	
Mar. 12	1	X		X		288, 963, 711	Séguéla	
	2					288, 963, 711	Séguéla	
	3	X		-		152, 980, 414	Daloa	
	5	-		-		42, 39	Agboville near Abidjan	
	8	X		-		616, 336, 834, 964, 251	Bouaké-Katiola road	
	11	X		-		895, 1034, 901, 410, 776, 306, 995	Assasination to Director General Police Academy, in Cocody 2 Plateaux	

tower, which is known for having an associated event, and calculate the deviation from the mean value of the hour of the day $\eta_{j,i}^k$, and day of the week δ_i^k from the observation, as in Equation (3). After tuning these parameters for a single cell tower, we apply them to all $N_E(t)$ in the dataset. We exclude the cell tower that is used for calculating the hyper parameters from the evaluation set. Since the model may be sensitive to the selection of hyper parameters, this selection is important for the evaluation of the model. We recommend the selection of a cell tower that represents the average rate function of the all cell towers. Furthermore, the selected antenna should have a number of observed events, so that the rate change can be quantified. The duration of the experiment is not important

if the data can represent the normal behavior of the cell tower, which corresponds to the mean value of the call volume in our case.

A.2 Event lists

In Table 3 and 4 we list emergency and social non-emergency events, gathered from a variety of public sources (e.g., Armed Conflict Location and Event Data Project, United Nations Council and International Crisis Group reports, local and international news). The cell towers IDs in the event region are given for the analyzed data period.

Author details
[1]Fondazione Bruno Kessler, via Sommarive, 18, Trento, Italy. [2]DISI, University of Trento, via Sommarive, 9, Trento, Italy. [3]Galatasaray University, Istanbul, Turkey. [4]Bogazici University, Istanbul, Turkey.

References
 1. Morrow-Jones HA, Morrow-Jones CR (1991) Mobility due to natural disaster: theoretical considerations and preliminary analyses. Disasters 15(2):126-132
 2. Myers K (2008) Remembering refugees: then and now by Tony Kushner. Cult Soc Hist 5(3):379-382
 3. Bissell RA (1983) Delayed-impact infectious disease after a natural disaster. J Emerg Med 1(1):59-66
 4. Watson JT, Gayer M, Connolly MA (2007) Epidemics after natural disasters. Emerg Infect Dis 13(1):1
 5. Boyle C, Mudd G, Mihelcic JR, Anastas P, Collins T, Culligan P, Edwards M, Gabe J, Gallagher P, Handy S et al (2010) Delivering sustainable infrastructure that supports the urban built environment. Environ Sci Technol 44(13):4836-4840
 6. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proc. 19th Int. Conf. on WWW, pp 851-860
 7. Becker H, Naaman M, Gravano L (2011) Beyond trending topics: real-world event identification on Twitter. In: ICWSM '11, pp 438-441
 8. Traag VA, Browet A, Calabrese F, Morlot F (2011) Social event detection in massive mobile phone data using probabilistic location inference. In: IEEE third international conference on social computing, pp 625-628
 9. The World in 2013, ICT Fact and Figures. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf. Accessed 24 Mar. 2016
10. Cox DR (1955) Some statistical methods connected with series of events. J R Stat Soc, Ser B, Methodol 17:129-164
11. Ihler A, Hutchins J, Smyth P (2006) Adaptive event detection with time-varying Poisson processes. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, pp 207-216
12. Kapoor A, Eagle N, Horvitz E (2010) People, quakes, and communications: inferences from call dynamics about a seismic event and its influences on a population. In: AAAI spring symposium: artificial intelligence for development
13. Bagrow JP, Wang D, Barabási A-L (2011) Collective response of human populations to large-scale emergencies. PLoS ONE 6(3):e17680. doi:10.1371/journal.pone.0017680
14. Bengtsson L, Lu X, Thorson A, Garfield R, Von Schreeb J (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med 8(8):e1001083
15. Gething PW, Tatem AJ (2011) Can mobile phone data improve emergency response to natural disasters? PLoS Med 8(8):e1001085. doi:10.1371/journal.pmed.1001085
16. Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake. Proc Natl Acad Sci 109(29):11576-11581
17. Gao L, Song C, Gao Z, Barabási A-L, Bagrow JP, Wang D (2014) Quantifying information flow during emergencies. Sci Rep 4:3997
18. Data for Development Challenge. http://www.d4d.orange.com. Accessed 24 Mar. 2016
19. Blondel VD, Esch M, Chan C, Clérot F, Deville P, Huens E, Morlot F, Smoreda Z, Ziemlicki C (2012) Data for development: the d4d challenge on mobile phone data. arXiv preprint arXiv:1210.0137

20. Young WC, Blumenstock JE, Fox EB, McCormick TH (2014) Detecting and classifying anomalous behavior in spatiotemporal network data. In: Proceedings of the 2014 KDD workshop on learning about emergencies from social information (KDD-LESI 2014), pp 29-33
21. Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. EPJ Data Sci 4:10
22. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779-782
23. Kung KS, Greco K, Sobolevsky S, Ratti C (2014) Exploring universal patterns in human home-work commuting from mobile phone data. PLoS ONE 9(6):e96180
24. Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. Sci Rep 3:1950
25. Schläpfer M, Bettencourt LM, Grauwin S, Raschke M, Claxton R, Smoreda Z, West GB, Ratti C (2014) The scaling of human interactions with city size. J R Soc Interface 11(98):20130789
26. Louail T, Lenormand M, Cantú OG, Picornell M, Herranz R, Frias-Martinez E, Ramasco JJ, Barthelemy M (2014) From mobile phone data to the spatial structure of cities. Sci Rep 4:5276
27. De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, Lepri B (2016) The death and life of great Italian cities: a mobile phone data perspective. In: Proceedings of the 25th international conference on world wide web. WWW '16, Switzerland, pp 413-423
28. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO (2012) Quantifying the impact of human mobility on malaria. Science 338(6104):267-270
29. Tizzoni M, Bajardi P, Decuyper A, King GKK, Schneider CM, Blondel V, Smoreda Z, González MC, Colizza V (2014) On the use of human mobility proxies for modeling epidemics. PLoS Comput Biol 10(7):e1003716
30. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ (2014) Dynamic population mapping using mobile phone data. Proc Natl Acad Sci 111(45):15888-15893
31. Bogomolov A, Lepri B, Larcher R, Antonelli F, Pianesi F, Pentland A (2016) Energy consumption prediction using people dynamics derived from cellular network data. EPJ Data Sci 5:13
32. Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. Science 328(5981):1029-1031
33. Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F, Pentland A (2014) Once upon a crime: towards crime prediction from demographics and mobile data. In: Proc. 16th ICMI. ACM, New York, pp 427-434
34. Toole JL, Lin Y-R, Muehlegger E, Shoag D, González MC, Lazer D (2015) Tracking employment shocks using mobile phone data. J R Soc Interface 12(107):20150185
35. Altshuler Y, Fire M, Shmueli E, Elovici Y, Bruckstein A, Pentland AS, Lazer D (2013) Detecting anomalous behaviors using structural properties of social networks. In: Social computing, behavioral-cultural modeling and prediction. Springer, Berlin, pp 433-440
36. Gibson M (2006) Order from chaos: responding to traumatic events. The Policy Press, Bristol
37. Akoglu L, Faloutsos C (2010) Event detection in time series of mobile communication graphs. In: Army science conference
38. Dong Y, Pinelli F, Gkoufas Y, Nabi Z, Calabrese F, Chawla NV (2015) Inferring unusual crowd events from mobile phone call detail records. In: Machine learning and knowledge discovery in databases. Springer, Berlin, pp 474-492
39. Dobra A, Williams NE, Eagle N (2015) Spatiotemporal detection of unusual human population behavior using mobile phone data. PLoS ONE 10:0120449
40. Calabrese F, Pereira FC, Di Lorenzo G, Liu L, Ratti C (2010) The geography of taste: analyzing cell-phone mobility and social events. In: Pervasive computing. Springer, Berlin, pp 22-37
41. Paraskevopoulos P, Dinh T, Dashdorj Z, Palpanas T, Serafini L (2013) Identification and characterization of human behavior patterns from mobile phone data. In: International conference the analysis of mobile phone datasets (NetMob 2013). Special session on the data for development (D4D) challenge
42. Zhang Y, Meratnia N, Havinga P (2010) Outlier detection techniques for wireless sensor networks: a survey. IEEE Commun Surv Tutor 12(2):159-170
43. Chib S (1998) Estimation and comparison of multiple change-point models. J Econom 86(2):221-241
44. Raftery A, Akman V (1986) Bayesian analysis of a Poisson process with a change-point. Biometrika 73(1):85-89
45. Gardner W, Mulvey EP, Shaw EC (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychol Bull 118(3):392-404
46. Rodríguez-Avi J, Olmo-Jiménez MJ, Conde-sánchez A, Martínez-Rodríguez AM (2013) A new regression model for overdispersed count data. In: The 29th European meeting of statisticians, p 256
47. Cameron AC, Trivedi PK (2013) Regression analysis of count data, vol 53. Cambridge University Press, Cambridge
48. White GC, Bennetts RE (1996) Analysis of frequency count data using the negative binomial distribution. Ecology 77(8):2549-2557
49. Zhang H, Dantu R, Cangussu JW (2009) Change point detection based on call detail records. In: IEEE international conference on intelligence and security informatics, 2009. ISI '09. IEEE, New York, pp 55-60
50. Luong TM, Perduca V, Nuel G (2012) Hidden markov model applications in change-point analysis. arXiv preprint arXiv:1212.1778
51. Witayangkurn A, Horanont T, Sekimoto Y, Shibasaki R (2013) Anomalous event detection on large-scale gps data from mobile phones using hidden Markov model and cloud platform. In: Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing adjunct publication. ACM, New York, pp 1219-1228
52. Scott SL, Smyth P (2003) The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data. In: Bayesian statistics, vol 7, pp 671-680
53. Chib S, Winkelmann R (2001) Markov chain Monte Carlo analysis of correlated count data. J Bus Econ Stat 19:4
54. Scott SL (1999) Bayesian analysis of a two-state Markov modulated Poisson process. J Comput Graph Stat 8(3):662-670
55. Yoshihara T, Kasahara S, Takahashi Y (2001) Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process. Telecommun Syst 17(1-2):185-211
56. African Mobile Observatory 2011. http://www.gsma.com/spectrum/wp-content/uploads/2011/12/Africa-Mobile-Observatory-2011.pdf. Accessed 24 Mar. 2016

57. Armed Conflict Location and Event Data Project. http://www.acleddata.com. Accessed 24 Mar. 2016
58. United Nations Refugee Agency. http://www.unhcr.org/pages/4d831f586.html
59. Shapiro JN, Weidmann NB (2011) Talking about killing: cell phones, collective action, and insurgent violence in Iraq. Technical report, DTIC Document
60. Pierskalla JH, Hollenbach FM (2013) Technology and collective action: the effect of cell phone coverage on political violence in Africa. Am Polit Sci Rev 107(2):207-224
61. Le Figaro Newspaper. http://www.lefigaro.fr/flash-actu/2012/02/13/97001-20120213FILWWW00689-cote-d-ivoire-3-morts-dans-des-violences.php. Accessed 24 Mar. 2016
62. United Nations Security Council Reports. http://www.securitycouncilreport.org/un-documents/cote-divoire/. Accessed 24 Mar. 2016
63. International Crisis Group Crisis Watch Database. http://www.crisisgroup.org/en/publication-type/crisiswatch/. Accessed 24 Mar. 2016