Mapping Mobility to Support Crisis Management

Fabio Ciravegna
University of Sheffield
f.ciravegna@sheffield.ac.uk

Jie Gao
University of Sheffield
j.gao@sheffield.ac.uk

Chris Ingram
K-Now Ltd
chris@k-now.co.uk

Neil Ireson
University of Sheffield
n.ireson@sheffield.ac.uk

Vita Lanfranchi
University of Sheffield
v.lanfranchi@sheffield.ac.uk

Humasak Simanjuntak
University of Sheffield
htasimanjuntak1@sheffield.ac.uk

ABSTRACT

In this paper we describe a method and an infrastructure for rapid mapping of mobility patterns, based on a combination of a mobile mobility tracker, a large-scale data collection infrastructure, and a data and visual analytics tool. The combination of the three enables mapping everyday mobility patterns for decision makers, e.g. city council, motorways authorities, etc. and can support emergency responders in improving their preparedness and the recovery in the aftermath of a crisis. The technology is currently employed over very large scale: (i) in England it is used by a public body to incentivise physical mobility (400,000 app downloads and hundreds of millions of data point since September 2017); (ii) in Sheffield UK, through the MoveMore initiative, tracking active mobility of users (5,000 downloads); and (iii) the European project SETA, to track multimodal mobility patterns in three cities (Birmingham, Santander and Turin).

Keywords

GIS, mapping, mobility tracking, large-scale data and visual analytics.

INTRODUCTION

A detailed understanding mobility pattern is a major need for emergency responders and in general for decision makers such as city councils, governments, etc. Apart from the importance for the day to day management of the territory, large environmental emergencies (e.g. storms, earthquakes, flooding) affect mobility first and foremost which in turn affects life and the economy of the area, as well as the ability to bring emergency relief where needed. People tend to have very regular mobility patterns and altering adjusting mobility behaviour to deal with crisis events can be problematic, particularly when the crisis can disrupt the usual mobility infrastructures. Rapid mapping of the changes in mobility allows a range of benefits: from planning changes to provision of services, to planning of remedial actions, etc. Mobility issues can have a major impact on the effectiveness of crisis management, as it can impact on the effectiveness of response and relief operations, and in particular evacuations (Wang and Taylor, 2016; Pan et al; 2007).

There is however a lack of studies on how mobility patterns can be studied to support crisis management. Most studies concerning human mobility patterns and crisis management are focused on prediction models (Yan et al; 2014; Ye and Chen 2013) and tend to focus on a specific type of disaster or disease (Cova, 2005; Eubank et al, 2004). The collection methodologies for human pattern analysis tend to vary between household surveys, census data and more efficient and powerful techniques such as crowdsourcing, mobile phone GPS data (Yan et al; 2014) and social network data (Song et al, 2014). Mobile phone billing data has been used by Bagrow et al (2011) to track communications between individuals during emergency events. Crowdsourcing techniques have

been employed to analyse group activities and behaviours in emergency situations using aggregated mobility patterns (Cardone et al, 2013). Song et al (2014) have worked on analysing individual movement data extracted from Twitter to study population movements after the Great East Japan Earthquake to predict population movements in urban areas impacted by the earthquake.

Previous studies have focused on discovering motorised mobility trends and do not discuss different types of mobility and their associated behaviours (e.g. walking vs driving). Studies on pedestrian behaviours have been carried out in the past, mostly using manual survey techniques (e.g. direct observations, photographs, surveys, etc.) for simulation purposes.

Standard world-wide emergency protocols are based on predefined response and evacuation plans to provide response and relief. Although these plans are carefully constructed, taking into account all potential risks and mitigation strategies and working towards optimization of resources, they may fail when dealing with unpredictable human behaviour (Wang and Taylor, 2016). Previous studies analysed the impact of mobility conditions on the success of evacuations from wildfires, demonstrating how traffic queues and road blockages caused people to lose their lives (Church and Sexton, 2002; Cova, 2005).

Human movements can be efficiently tracked on a very large-scale using low-cost sensors integrated in smartphones – analysing this data can enable the detection of patterns and situations that can influence mobility (Zhang et al. 2014) and therefore can impact crisis management, as they can collect high granularity mobility data from hundreds of thousands of citizens. This data can be qualified and quantified, i.e. it is possible to collect detailed geolocated vehicular and active mobility (walking, cycling, running, etc.) so to allow insights well beyond simple location patterns, and to extend tracking to multimodal mobility and life patterns, inclusive of analysis of stay points (where people carry out activities), purpose of travelling (shopping, commuting, free time, etc.), usage of infrastructures such as shopping centres, etc. This would allow a very detailed assessment of the everyday life patterns as well as their changes in the face of crises.

This work aims to identify patterns of human movements in rural and metropolitan area by collecting data from apps that track movements for different purposes (e.g. health and wellbeing and sustainable mobility, with explicit acknowledgement of data reuse in the Terms and Conditions of the apps) and by using large scale data and visual analytics techniques to support decision makers, in particular to both support the next generation emergency plans, and to assess the effect of a crisis. Collected data could also be used as input to simulation algorithms.

In this paper we describe and discuss a large-scale framework for mobility patterns tracking. The framework is composed of a mobile tracker (Android and iOS) that has been distributed to citizens and volunteers as an add-on to existing apps, a distributed large-scale data collection architecture and a data and visual analytics tool. The framework allows mapping mobility in normal times, but has the potential to become key to support emergency response and reconstruction after a crisis, providing (i) quasi real time situation awareness on the ground; (ii) qualifying and quantifying the effects of the crisis by comparing patterns pre and post crisis (iii) gauging the effectiveness of the recovery action by quantifying and qualifying the level of return to normality.

The framework has been used in real world applications, collecting hundreds of millions of data points from hundreds of thousands of citizens as part of three mobility tracking programs: (i) a public body in England (which we cannot name for legal reasons) is using it to support better physical mobility in the population. The resulting app has been downloaded over 400,000 times and the infrastructure has collected over hundreds of millions of data point since September 2017. (ii) MoveMore, in Sheffield, UK, has released it to track the physical mobility of the citizens of Sheffield (5,000 downloads). (iii) as part of the EU project SETA, the infrastructure is being used to track multimodal mobility patterns in three cities: Birmingham, Santander and Turin, where its potential for planning has been tested and successfully verified by local decision makers.

In this paper we describe the methodology, the architecture, the release to hundreds of thousands of users and the results of the preliminary release as part of the SETA project.

REQUIREMENTS

Awareness of transport network conditions and mobility patterns are of fundamental importance for all phases of emergency management, from preparation to response to recovery. In particular government emergency organisations have a duty to consider transport in terms of:

- Devising strategies for evacuations
- Facilitating fast, flexible and effective response and recovery
- Gauging the level of recovery and return to normality

To understand requirements for the infrastructure, we have consulted with decision-makers from three European cities (Birmingham, UK; Turin, Italy; Santander, Spain) during stakeholders workshops for the EU project SETA[1] and we have derived the following requirements:

- High granularity, widespread coverage of mobility patterns: while most of the knowledge about mobility patterns tend to be concentrated on major roads, knowledge of minor roads is less common. It is however fundamental to be able to collect rapidly detailed mobility patterns at very high granularity (i.e. both major and minor roads, both central urban and peripheral rural areas, etc.) so to have complete picture of the situation and to plan urgent and planned remedial actions.
- Tracking different types of vehicular mobility as well as active mobility (walking, cycling, etc.): understanding the modes of transport used in different areas and moments could prove of fundamental importance during and after a crisis, when availability of certain means of transport may be restricted (e.g. major roads are closed, railways are unavailable, etc.). Being able to track the way people move (e.g. building rapidly typed (by vehicle) Origin/Destination matrices and trajectories) would allow rapid appreciation of the effects of the crises, understanding changes and needs of the population, etc. This will allow analysis of the impact of any disruption, well beyond the effect of simple transportation, to inform a number of stakeholders on e.g. effects on everyday life patterns, usage of infrastructures such as station and shopping centres, etc.
- Resilience: to be useful in an emergency a tracking solution must be robust to communication loss in the affected areas, providing caching and a distributed data collection architecture to be deployed locally and/or in the cloud;
- Energy and data conscious: any technology running on citizens' phones must be low impact with respect to the user's phone availability, in terms of battery usage and data allowance. In this way, any emergency app can be kept on the user phone indefinitely, available to collect data when needed.

FRAMEWORK

Our solution is composed of:

(i) A mobile data collection tool (iOS and Android apps) that has been distributed to thousands of citizens and able to track their mobility in terms of geo-located patterns (walking, cycling, use of vehicles, etc.). Movements are also qualified in terms of intensity (intensity of walking, speed of cycling and vehicles, etc.). The tracker works in the background and comes as a self-contained library that can be included in any existing application; there are three apps currently available for download from the iOS and Android app stores that include the tracking library: SETA[2] and MoveMore[3].

(ii) The data collection architecture achieves a combination of scalability, efficiency and flexibility for mobility Big Data management. The resulting infrastructure created is distributed, scalable, highly robust, able to handle huge number of sensors (**>500,000** users), processes multiple high velocity data streams (avg. **500,000** daily requests), and provides scalable and reliable backend for mobile applications (with **0.247ms/12.6ms** minimum/maximum response time, serves **>24 million/day** query and **>1 million/hour** query, with **>30K/hour** write throughput[4]). The Service-Oriented Architecture (SOA) is developed and further scaled out to provide real-time read and write access to structured and semi-structured big data. The leading-edge computing technologies developed allow the balancing of data collection and distributed processing performing on both local servers (near the source of the data or locally) and remote cloud servers.

(iii) A data and visual analytic platform that enables to identify mobility patterns from large scale GPS data traces; in terms or path usage, stay points, area-to-area mobility (i.e. Origin/Destination matrices).

In the next sections we will go into the details of each module.

Mobile Tracker

The tracking technology is implemented as a low power library that can be added to different types of

[1] setamobility.eu
[2] https://itunes.apple.com/us/app/seta/id1195607470 / https://play.google.com/store/apps/details?id=oak.shef.ac.uk.abstractcityactivity

[3] https://itunes.apple.com/gb/app/movemore-sheffield/id1093481731 / https://play.google.com/store/apps/details?id=uk.ac.shef.oak.ActivityRecognition.WeSenseIt.MoveMore

[4] Numbers are from our AWS master database nodes based on the statistics in December, 2017

applications, such as mobility apps, health and wellbeing apps, routing apps, etc. It runs in the background and senses mobility features through a range of sensors (e.g. step counters, activity recognition, accelerometer, gyro, etc.) as well as from location services (GPS, network, etc.). It fuses the data to create geolocated multimodal mobility segments (user trips), which are then sent to the central server. The tracker is designed to identify both vehicular and active mobility (walking, cycling, running, etc.) so to allow insights well beyond simple traffic patterns, and to extend tracking to multimodal mobility and life patterns, including the analysis of stay points, purpose of travelling, usage of infrastructures such as shopping centres, etc.

The underlying algorithms include sensor data collection, data fusion, as well as segment identification, classification and fusion. The library is generic and provides a set of combinable logics to track mobility according to needs: for tracking vehicular mobility, for tracking walking and running, for tracking cycling, etc. They are composable, so applications covering different types of mobility are possible, while allowing the creation of strategies for specialised trackers (for example we recently created an application optimised for cycling detection). On top of these, sit the data processing methods for specific applications, i.e. user services based on the detected patterns, such as stats and motivational messages for health and wellbeing apps, routing, mapping for mobility apps, etc.

The tracker is designed to work 24/7 with limited impact on the battery so not to affect the phone's availability to the user, which is crucial during emergencies but is equally essential for everyday life. The app works unobtrusively in the background tracking mobility while turning itself off when the user is inactive. The sensors accessed are low power and their activation is based on consumption levels: for example only one type of sensor is active when the phone is still and these have virtually no power consumption. An example of these is the Significant Movement Detection Sensor, which is centrally managed by Android and hence low cost battery-wise. Location tracking, a major source of battery drain, is optimised in terms of frequency of polling, as well as via use-only-when-strictly-needed strategies. For example, dynamic geofencing (Pongpaichet, 2013) is used to restart geolocation only if specific conditions are met (e.g. the user is leaving the area where they stopped): geofencing is a technique that allows to set a trigger when the phone moves inside or outside a geographic area of fixed radius (e.g. 500m), it uses low power strategies based on network information rather than full location services. Geofences are set dynamically every time specific conditions are met, e.g. the phone has been still for a period of time. We also make reuse of already detected locations: the phone regularly polls locations for its own use and we always reuse any location when it becomes available through other apps, avoiding turning off the GPS. We can also vary the accuracy of location detection according to needs: for example if a user is following a previously recorded frequent pattern, there is no need for precise geolocation and tracking if the timing and the way of travelling is compatible with the previous patterns, as there is no sensible change. The exact location can be predicted from previous data. This brings high battery savings.

Data sending is also a major source of power drain in mobile phones. This is optimised by the tracker: sending of large chunks of data is preferred over frequent sending of small packets; so during normal periods and when following usual patterns, data is sent infrequently (e.g. every six hours), hence reducing considerably power usage. During specific periods (e.g. emergencies) or when moving through dynamically defined geofences, data can be sent in real time, if necessary.

The algorithms interpreting the sensor data are based on a cascade of weak classifiers fusing and cleaning sensor data polled at regular intervals (depending on application but typically a few times a minute) to create larger aggregates of coherent single mode activities (e.g. vehicular segments) that are then combined into multimodal segments (e.g. walking, followed by vehicle, followed by walking). Methods used are low and high pass filters[5] to clean sensor data, maximum entropy modelling and weighted moving average for aggregation, as well as Kalman filters for trajectories. These methods are designed to provide reasonable results to the user in real time, while using limited battery power: data is then sent to the central server where it is analysed with more powerful algorithms to provide full scale accuracy for decision makers (se below).

The tracker has a local database where data is stored when communication to base is impossible, which is a frequent event in everyday life (especially in rural areas) and of course during crises. Network availability polling is performed strategically to avoid sending data (and using both power and data allowance) when connection is unavailable or unreliable.

Precision of tracking has been verified by an independent testing company, as well as by the usage by hundreds of thousands of users.

Data Collection Architecture

[5] https://www.w3.org/TR/motion-sensors/

Figure 1 shows the general architecture of data collection infrastructure. Because the user mobility data are geographically distributed and produce multimodal and multidimensional data at high rates, a distributed data management infrastructure is required. In emergency situation, with this architecture, data can be stored near its source, and data processing and filtering can be performed locally, and synchronised to the central server at a later time. Such an architecture reduces bandwidth requirements and enables parallel processing of high volume streams. Although queries can be posed from anywhere in the world, in practice the requester and the data of interest will be in a proximate geographical region (for example, local city council will query data about citizens in their vicinity).

Meanwhile, Cloud computing technologies play a pivotal role in the infrastructure and offer a solution to coordinate, integrate and manage large volume of data stream. The major part of infrastructure is hosted in a public cloud (Amazon's *Elastic Compute Cloud*, EC2), where a virtual data centre/cluster is created. This is based on the consideration that multi-dimensional spatial indexing, data/request load balancing, and data clustering can be quickly combined and deployed so that the performance can be optimised effectively across the clusters with low overhead cost. Specifically, *Amazon Route 53* is adopted as DNS hosting solution, which provides reliable and cost effective way to route web and API requests to our RESTful based application layer. Then, Amazon' *Elastic Load Balancing* automatically distributes incoming traffic across multiple EC2 web servers. Multiple web servers are set up and take request assigned from load balancer. Every EC2 server runs a Node.js application managed by PM2[6], and provides users with secure access to the application through an Nginx reverse proxy. To maximise requests throughput capacity, the data processing method is optimised by a asynchronous pipeline. The continuous stream of raw data is collected and immediately stored into a database (the "master" node in the figure). This process is to minimise the data collection latency and transmission errors (packet corruption or loss) due to various network and memory issue. The original dataset collected are extracted, validated, analysed and filtered asynchronously by a set of I/O and computation intensive extract-transform-load (ETL) tools (parallelized with multiple "processing nodes") in order to allow sophisticated data mining and pattern recognition. The solution of separating data collection and data ingestion splits two problems: 1) processing high velocity of data; 2) handling the value and validity of data. The formal problem concerns the speed of handling data stream and latter one focuses on the usefulness, quality and accessibility of the data. This becomes more important when data streams from multiple heterogeneous platforms and sources.

Concretely, the data collection infrastructure addresses four key challenges:

Query processing - For large-scale data collection, the "**master**" server must be write-intensive, high volume, and often time critical. The data that support emergency management is often multimodal (from spatio-temporal location readings to rich multimedia data), and the system itself will be heterogeneous in its mix of data processing platform, storage platform, communication capabilities, and administrative domains. The design of architecture enables low-latency continuous processing of data streams from geographically distributed sources. The engines on EC2 cloud provides sophisticated fault-tolerance, load-management, and federated-operation features for distributed data streams. "**Slave"** nodes serves as highly available read replicas and can be easily scaled out horizontally to more commodity servers on-demand (e.g., standard db.t2.large RDS instance used in our current infrastructure[7]).

Integrated processing - Traditional databases only support one-time queries over stored data. The system is designed to recognize the need to support integrated queries over live and historical data (see the "**Backup Node**" in the figure). Such query support is critical to help users understand new events in the context of pass observations, while ensuring high availability of live database. To enable queries over live and historical data, the infrastructure has additional ETL components to decide where to store data, how much data to store, what data to summarize, and when to discard data. Additional challenges are the data archive size and the high aggregate data rate, which can be mitigated by vertical and horizontal scaling, and the use of big data technologies depending on needs (e.g., replace the "Backup Node" with Apache HBase[8])

Reliability - It is well known in big data analysis that a large-scale architecture must have built in failure mode, as a subset of the nodes will encounter components and communication failures. Suddenly increasing the number of users can increase the failure rate of the server. Thus, the infrastructure is designed to quickly react to large-scale load variations and it provides redundancy so to avoid Single Point of Failures.

Data privacy and security - The sensor data must be highly private (for example, location information). A set of mechanisms has been developed to control access to this data, keep it secure, and regulate its use. First, no user identity information (e.g., phone number, name, home address) will be collected. Additionally, a security

[6] http://pm2.keymetrics.io/
[7] https://aws.amazon.com/rds/instance-types/
[8] https://aws.amazon.com/emr/details/hbase/

layer is built based on Secure Socket Layer (SSL) and Transport Layer Security (TLS) 3.0 protocol to the data with scalable and efficient encryption algorithms. SSL/TLS certificate is issued and used to establish identity and trust between server and client apps, ensuring privacy and security whenever communicating sensitive data. GPS data is never sent to the client, preventing any possibility of 3rd party requests for location information.

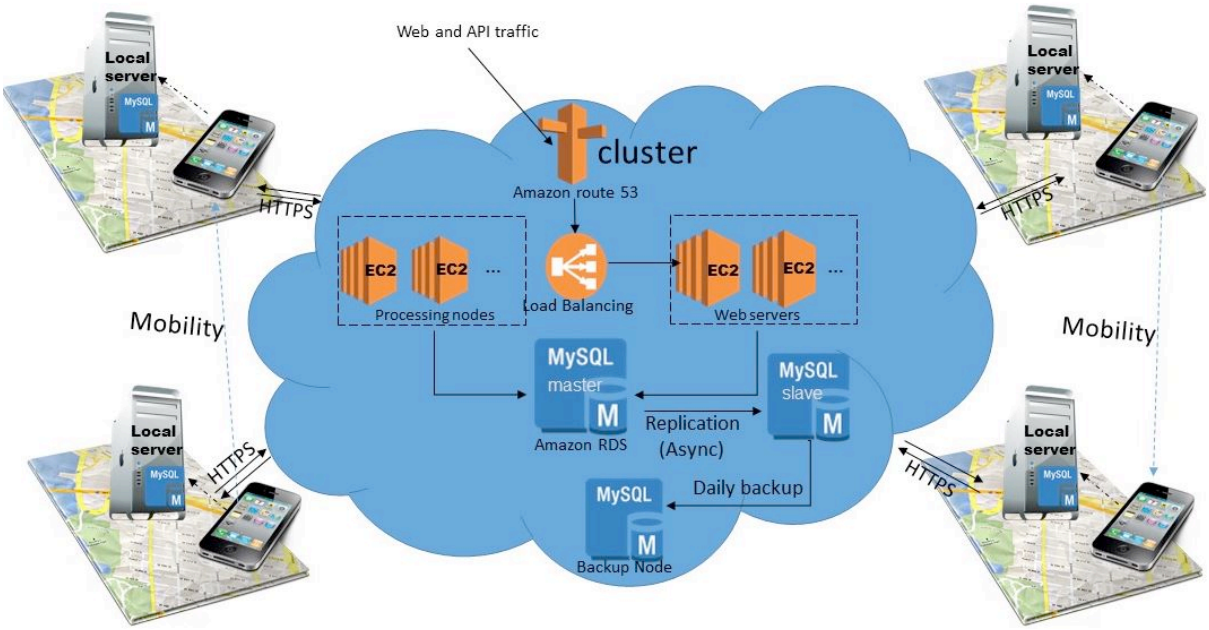

Figure 1: General Architecture of Data Collection Infrastructure

Data Processing

The server-side data processing mirrors a number of processes which are performed on the mobile device, however on the server there are fewer constraints in terms of processing power and time, and historic and population data can be considered. The server can be used to experiment with the algorithms for noise-reduction, activity recognition, etc. and these can then be transferred to the phone during the regular updates. The data analysis framework has four interacting sub-processes:

Noise detection and removal

The precision of a GPS signal is primarily affected by the quality and location of the GPS unit and the number and position of satellites, which produces both random and structural noise. The noise detection process involves identifying GPS traces, which are both feasible and expected, i.e. removing outliers, determining common paths, etc.

Stay-point detection and clustering

Identifying the journey start and end points, as well as intermediate stops and mode transitions is a crucial process in determining mobility patterns. In general, a stay-point is defined in spatio-temporal terms, i.e. a GPS signal staying in a given location (e.g. within a circle) for a given about a time using experimentally fine tuned threshold values for location size and duration of stay. The stay-points are then clustered together, currently clusters are defined by postal regions, in the UK these have four levels of granularity: (~1.75 million) units, (11,199) sectors, (2,983) district and (124) areas. The clustering of stay-points allows for the creation of Origin-Destination matrices, which, given the various levels of granularity, allows for the analysis of intra- and inter-urban movements.

We adopted the PIE algorithm (de Graaff et al, 2016) and the algorithm proposed by Li et al (2008) to extract stay points from user trajectories. The steps to extract stay points from user trajectories can be seen in the algorithm 1 below (Figure 2).

Algorithm 1 Stay Point Detection

INPUT:
1. *UT* = list of user trajectories {<lat,lon,t,seg>, <lat,lon,t,seg>, ...<lat,lon,t,seg>}
2. *eps*, the stay point distance threshold in meters
3. *t*, the stay point duration threshold in millisecond

OUTPUT:
1. SSP = {sp_1, sp_2, ..., sp_m} the list of semantic stay points
2. Clustering of Stay Point

ALGORITHM:
for each user **do**

```
    //Remove inaccurate locations that greater than the average of accuracy
    listLocationsPoints = {l_p ∈ UT | l_p.a ≤ avg(a)}

    //Get candidate stay points for each day
    listCandidateStayPoints = getCandidateStayPoints (listLocationsPoints, eps, t)

    //Get centroid of stay point by computing mean of stay points
    listStayPoints = computeMean (listCandidateStayPoints)

//Generate summary of stay point (occurrence, days, duration)
listSummarySP = generateSummarySP (listStayPoints)

//Add semantic to the stay points by considering spatial relationship
//with Point of Interest and Postcode
SSP = annotateStayPoint(listSummarySP)
```

$$\left\{ sspI \in SSP \mid distance(sp, polygon) = 0 \vee distance(sp, polygon) < eps \vee distance(sp, point) < eps \right\}$$

findClusters(SSP)

Figure 2: Algorithm to extract stay points

In this algorithm, the sequence of locations is retrieved from all user trajectories. Noise points are filtered based on the most frequently occurring (or an average) of location accuracy in the data. The daily stay point candidates are identified based on two parameters, distance and duration threshold (i.e. dist=30 metres and dur=5 minutes). The distance between two location points is calculated by using the Haversine formula. The centroid is calculated to get only one point in each stay point. Regarding spatial and time regularity, we keep the frequency, day, arrival and leaving time features. Finally, the algorithm produces a summary of stay points that consists of the significant places, duration, day and frequency. Furthermore, stay point is enriched with information such as land-use type and postcode.

All user stay point are aggregated and clustered by time, day, or certain period to understand what people tend to do. The changing of stay point patterns identified before the disaster and those recorded after the disaster is critical to know the people go to other places during the crisis/disaster. Figure 3 shows the heatmap of stay points (on weekdays and weekends), with the location's colour intensity determined by the number of times that location is allocated as a stay point.

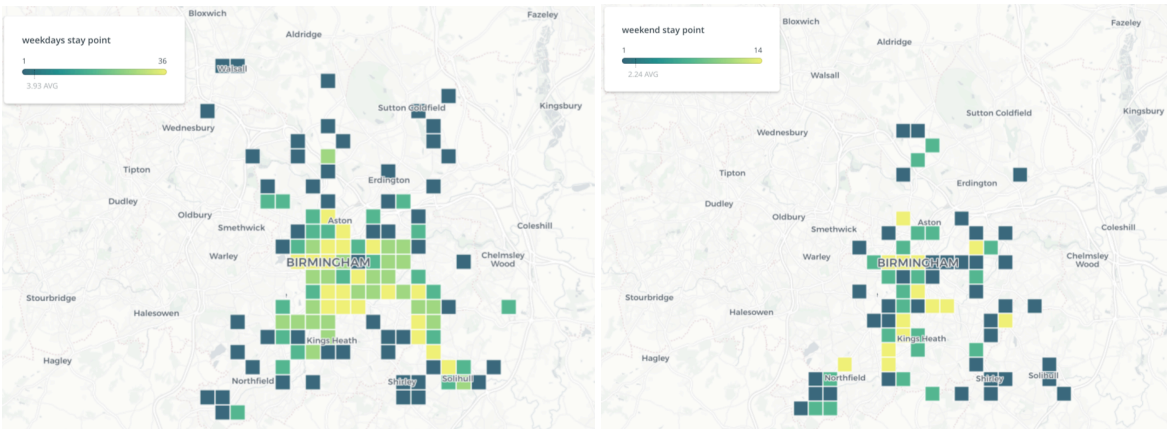

Figure 3: Distribution of Birmingham stay point for weekdays (left) and weekend (right)

Mode detection

Identifying whether, when and where people use the various modes of mobility, i.e. walking, running, cycling or using private of public vehicles is fundamental to understanding how populations can respond to emergency situations. For example, is the population reliant on public transport and the road network; activity modes are slower but less affected by disruption and congestion, as shown by the increasing use of Cycle Response Units in urban areas. The server mode detection using the user's history sensor data (GPS, steps, etc.) and the combined data from all users to determine the most likely mode of travel.

Map Matching

The assigning of the given (inaccurate) GPS location to the most likely path requires considering the mode (i.e. vehicles are largely limited to roadways unlike cyclist or pedestrians) and the most likely route given all the GPS locations allocated to a single journey. The A* graph traversal algorithm is used to find the minimum cost route between the journey start and end points considering all the intermediate locations on the journey. The path does not have to pass directly through the intermediate locations, but instead within the circle around the locations with the radius given by the GPS accuracy (provided by the GPS unit). The mode of travel determines the allowable paths. The cost of traversal is a function of the edge distance (of the path between intersections) and the frequency a path is travelled both by the user in question and the population as a whole.

Visual Analytics

The processed data, described above, is indexed in a high performance cloud-based faceted search engine[9], and visualised via faceted browsing interface, which allows an analyst to constrain and relax filters over the data to explore multiple dimensions of the data. To maintain security the data is stored on a separate server to the, Node.js, interface server. The interface server acts as a secure gateway to the data, and access is regulated by the rights granted to the visual analytics user profiles. For example, some users may only be able to view summary data, whilst other can drill-down to individual users. Figure 4 shows the interface focused on the city of Turin, with the map displaying the frequency of the paths taken, and the facets allowing the filtering of the data according to the journey mode, duration, date, day of week and hour of day. The Origin/Destination matrix can be used to filter the journey to those to/from given areas. The interface also animates the flow of the population, providing an understanding of the movement over time.

The map also provides visualisation of the location data as a heatmap. The locations are indexed using geohashing, which divides the Earth into a hierarchical grid structure, by selecting the appropriate hierarchy level (given the map zoom level) it is possible to efficiently display 100s millions of location points. Figure 5 shows the comparison of common locations in Sheffield during the day versus the night. Such information is useful to know the distribution of the population during an emergency and therefore the number of people affected in each area.

[9] http://lucene.apache.org/solr/

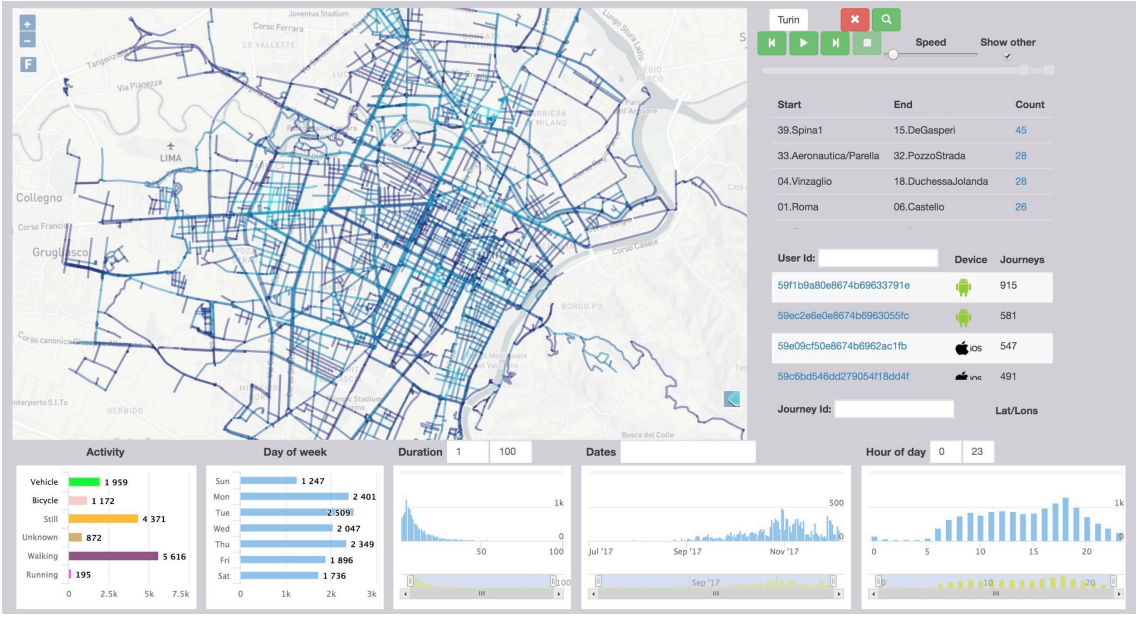

Figure 4: The visual analytics interface

Figure 5: The Birmingham locations heatmap for weekdays (left) and weekend (right)

Evaluation

The framework was evaluated in terms of:

- Usability and user acceptance of tracking application
- Usefulness and potential insights derived from the data collected and processed for emergency managers

The framework was formally tested as part of the EU Project SETA, focusing on assessing the app usability and the activity recognition capabilities. The app was distributed to a group of 59 users in Birmingham for use during October 2017. All the participants were invited to use the app for a prolonged period of time and ask to report any issues using the online reporting tool. The participants were also invited to complete an online questionnaire to report on the usability and precision of activity recognition.

There have been 315 users of the SETA app who have provided data for at least one journey in October 2017. Note that, as users are anonymous, it is possible that a single person can represent multiple users of the app. Of these 315, 59 (18.7%) have provided journey data in the Birmingham area, i.e. within the boundary box (52.38 >= latitude <= 52.61 AND -2.04 >= longitude <= -1.79). Those users produced 7,100 journeys; Figure 6 below shows the distribution of journeys per user (min=1, max=914, mean=120).

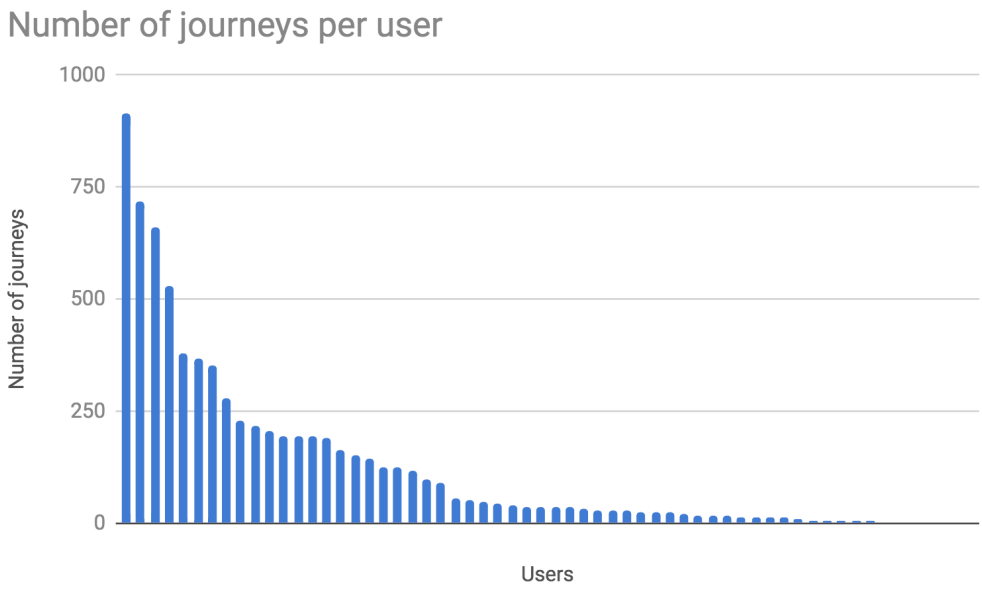

Figure 6: Number of journeys per users

Table 1 below shows the number and type of journeys taken by the 59 users. As can be seen although almost 25% of users had been running, they only had a mean of 4 runs per user, and therefore running was by far the least frequent activity. The most frequent activity was walking, with a mean of over 47 walks per user.

Table 1: Number of users, journeys and journeys per user for each activity type

	Users	Journeys	Journeys / user
Total	59	7100	120.34
IN_VEHICLE	56	1638	29.25
ON_BICYCLE	36	598	16.61
STILL	27	2186	80.96
UNKNOWN	21	253	12.05
WALKING	50	2369	47.38
RUNNING	14	56	4

The visual analytics interface was evaluated by a cohort of 11 stakeholders from specific teams of Birmingham City Council, including Data Analysis, Transport, and Health & Wellbeing teams.

In the following section we will present key findings derived from users' replies to the questionnaire. The questions were based on a Likert Scale from 1 to 5 where 1 = disagree/difficult, 5 = agree/easy. In general the system was well received and seen as "easy to use" and "well integrated". Aside from the general questions concerning the usability of the interface, there was a general agreement that the system provided insight into the understanding of activities in the city (mean=3.3), Figure 7 below.

During the focus groups the users were asked to evaluate the potential for the interface to offer insights on the mobility of the city and to highlight new requirements that would improve the usefulness of the interface. The users particularly focused on the importance of transparency, suggesting that the system should show the degree of accuracy of the map matching, journey segmentation and activity recognition. Whilst the system works by processing and analysing the data over large scale to present a visual interpretation, the users expressed the desire of seeing the "raw" latitude/longitude data. Another requests from the users was to use a different zoning/area systems, e.g. wards and electoral divisions, which requires the provision of area shape data or centroids, this is public data in the UK. This would also enable analysis of mobility data in terms of census ward data, e.g. demographic information. These functionalities have now been added to the interface and will be evaluated in the next phase on the project.

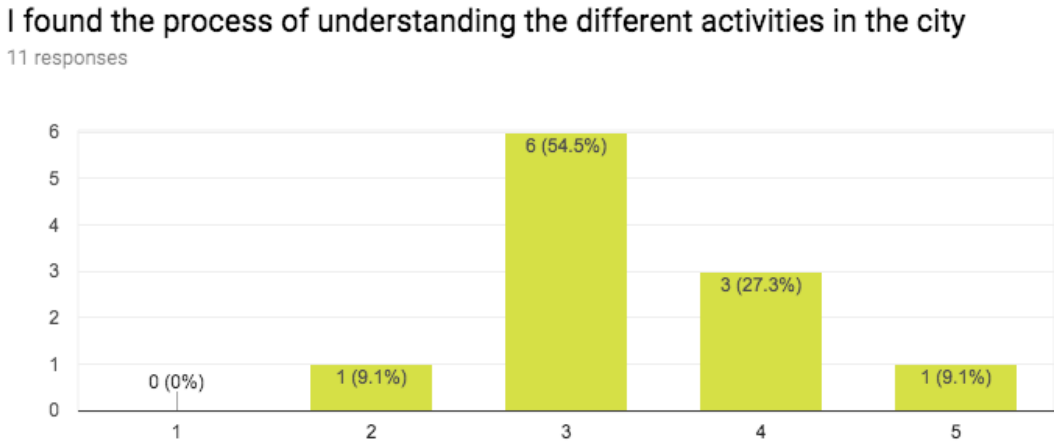

Figure 7: Understanding of different activities in the city

CONCLUSION

In this paper we have described a method and an infrastructure to map mobility patterns at high granularity over large areas. The method is based on a highly efficient tracker running on citizens' mobile phones and a large-scale data collection and analytics platform.

Although the infrastructure has not yet been used during crises, it has been used to track mobility of hundreds of thousands of users, studying their patterns of mobility in terms of Origin/Destination matrices, geolocated patterns, mobility means, etc. This allows decision makers to understand mobility during normal periods and to capture mobility patterns before, during and after crises, so to analyse details of any change in patterns, land use, use of infrastructure, etc. as a proxy to identification of changes in habits, unavailability of resources and infrastructures, road blocks, etc. which are very relevant to decision makers.

The tracker has been released to hundreds of thousands of users and the platform has collected hundreds of millions of data points over the past few months. Different incarnation of the data analytics platform have been used by a public body in England to derive patterns of mobility for health and wellbeing studies using the data of 400,000 users, by MoveMore in Sheffield, UK, to track and model active mobility of thousands of citizens over two months, running city wide competitions, including a workplace competition involving dozens of organisations and a derby between the fans of the two local football clubs. We are currently running a pilot city-wide workplace competition in Turin, Italy, involving hundreds of users where the data is used by planners and decision makers to support modelling of multimodal mobility in the city. Our evaluation with decision makers in three cities has highlighted that the data provided by the infrastructure is considered useful to identify mobility patterns and the framework is usable for planning both short terms and long term.

The infrastructure is very flexible and can be ported to different applications, each of them providing data for understanding mobility.

Future work concerns the improvement of the data analytics capabilities in order to further improve the data analysis and the visual analytics capabilities of the tool. Furthermore, we are planning on integration of the tools with a large infrastructure currently developed in the SETA EU project and that is being released to the mobility control rooms of three cities in Europe,

Finally, we expect to verify the usefulness of the framework for emergency management, by analysing the changing patterns of the mobility within a city caused by large-scale events, both planned (e.g. sporting, entertainment and protest events) and any unplanned (large-scale) events that occur within the study cities.

ACKNOWLEDGMENTS

The activity has been generously funded by a number of bodies: from the European Commission as part of the Seta European Project (contract no. 688082), in England through funds from the national health system, as well as funds from the University of Sheffield and the National Centre for Sports and Exercise Medicine (NCSEM)

in Sheffield.

The evaluation described was carried out by partners in the SETA project. We are very grateful to the City Councils of Birmingham (Ravinder Bains), Turin (Giuseppe Estivo, Roberto Cocozza), Santander (Celia Gilsanz Gomez), as well as to Aizoon (Sonia Floris, Daniele Albertin who are running the competition in Turin) and the University of Cantabria (Borja Alonso, Oreña) plus dozens of unnamed users. Collaboration over the years with Hallam University and NCSEM has been fundamental for the development of the tracker and its application to health and wellbeing (Rob Copeland and Ben Heller).

REFERENCES

Agarwal, R. and Karahanna, E. (2000) - Time Flies when You're having Fun: Cognitive Absorption and Beliefs about Information Technology Usage, *MIS Quarterly*, 24, 4, 665-694.

Bagrow JP, Wang D, Barabasi A-L. (2011) Collective response of human populations to large-scale emergencies. PLoS ONE. 6(3):e17680. pmid:21479206

Cardone, G., Foschini, L., Bellavista, P., Corradi, A., Borcea, C., Talasila, M., & Curtmola R. (2013) Fostering participation in smart cities: A geo-social crowdsensing platform. IEEE Communications Magazine, 51(6): 112–119. DOI: http://dx.doi.org/10.1109/MCOM.2013.6525603

Church, R. L., and Sexton, R. M. (2002). "Modeling small area evacuation: Can existing transportation infrastructure impede public safety?" Caltrans Testbed Center for Interoperability Task Order3021, Final Rep.〈http://www.ncgia.ucsb.edu/vital/research/pubs/200204-Evacuation.pdf

Cova, TJ (2005) Public safety in the urban–wildland interface: should fire-prone communities have a maximum occupancy? Natural Hazards Review, 6(3), 99–108.10.1061/(ASCE)1527-6988(2005)6:3(99)

Pan X, Han CS, Dauber K, Law KH. (2007) A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. Ai & Society;22(2):113–32.

Pongpaichet, S., Singh, V.K., Jain, R. and Pentland, A. (2013) "Situation Fencing." Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia - PDM '13, October 22, 2013, Barcelona, Spain, ACM.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W. and Ma, W. (2008) Mining user similarity based on location history. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (GIS '08). ACM, New York, NY, USA, , Article 34 , 10 pages. DOI=http://dx.doi.org/10.1145/1463434.1463477

Song, X.; Zhang, Q.; Sekimoto, Y.; and Shibasaki, R. (2014). Intelligent system for urban emergency management during large-scale disaster. In Twenty-Eighth AAAI Conference on Artificial Intelligence.

de Graaff, V., de By, R.A., and van Keulen, M. (2016) Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16). ACM, New York, NY, USA, 552-559. DOI: https://doi.org/10.1145/2851613.2851709

Wang Q, Taylor JE (2016) Patterns and Limitations of Urban Human Mobility Resilience under the Influence of Multiple Types of Natural Disaster. PLOS ONE 11(1): e0147299. https://doi.org/10.1371/journal.pone.0147299

Yan, X.-Y., Zhao, C., Fan, Y., Di, Z., & Wang, W.-X. (2014). Universal predictability of mobility patterns in cities. Journal of the Royal Society Interface, 11(100), 20140834. http://doi.org/10.1098/rsif.2014.0834

Ye, J. Z. Z., and Cheng, H. (2013) What's your next move: User activity prediction in location-based social networks. Proc. of SIAM International Conference on Data Mining (SDM).

Zhang, D., Wang, L., Xiong, H., & Guo, B. (2014). 4W1H in Mobile Crowd Sensing. Communications Magazine, IEEE, 52(8): 42–48. DOI: http://dx.doi. org/10.1109/MCOM.2014.6871668