Semantic Understanding of Human Mobility Lifestyle to Support Crisis Management

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

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

ABSTRACT

In this paper, we propose a method for understanding the semantics of mobility (mainly related to lifestyle) patterns based on stay point detection from tracking data. The method identifies the context (trip purpose and visited point of interest) of tracking data by using large-scale data collection infrastructure. We evaluate our method with a tracking dataset in Birmingham (European project SETA) generated by 534 users from September 2017 to September 2018. To this end, we compare insights from the tracking data with check-in mobility in social media. The results show that both data capture rich human lifestyle features related to the visited point of interest. Our study provides solid evidence that lifestyle patterns from tracking and social media data can indeed be useful for understanding and gauging the level of disruption after a crisis, as it is possible to check the deviation of habits from normal conditions and post-crisis.

Keywords

lifestyle patterns, mobility patterns, semantic annotations, semantic mobility.

INTRODUCTION

Disaster severity and frequency trends have risen significantly in recent decades. Understanding human mobility, specifically lifestyle patterns is crucial for crisis responders and policymakers in determining effective strategies for disaster response, especially recovery or evacuation plans (Wang and Taylor, 2016). These lifestyle patterns are related to spatial and temporal regularity. Spatial regularity means a person is going to similar places or point of interest (POI); temporal regularity can be considered as when a person visits the same POI at a similar time of the day. Natural disasters such as earthquakes, storms, etc. are serious threats to urban areas affecting daily lifestyles because such events can devastate public infrastructure usually visited by people. Adjusting lifestyles to deal with crisis events can be problematic. Therefore, understanding human daily life pattern changes such as trip purpose and use of public infrastructure (shopping malls, food places, stations, etc.) will enable decision makers to monitor and analyse the impact of any disaster in the urban area, providing advantages for risk assessment, preparedness, emergency management and plans for recovery or reconstruction actions primarily related to the needs of citizens, such as where to obtain supplies of food and clothing.

Recently, technologies have been developed to capture human mobility data. The movement of people may be tracked and recorded in real time easily using wireless and sensor technology. Apps created by big companies such as Google, Apple, and Facebook collect our daily patterns of regularity related to place and the time spent there. Furthermore, with the increased popularity of social media, a large amount of check-in data can also be used to capture human mobility. Learning and understanding this data can help to analyse lifestyle patterns related to a crisis event. However, the accuracy, sparsity and semantic representation of tracking data pose challenges in understanding lifestyle patterns. Raw tracking data in the form of a time sequence of locations are not easily understandable without semantic representation.

Human mobility analysis has been applied to support crisis management using crowdsourcing (Ludwig et al., 2017), GPS data (Song et al., 2017), and social media (Xiao, Huang and Wu, 2015). Furthermore, effective models for simulating and predicting human movement after a crisis have been proposed (Song et al., 2015, 2017). However, there is still a lack in previous studies related to understanding human mobility to support crisis management. Most are concerned with specific mobility data. They only use tracking data (GPS trajectory), transportation networks, CDR or social media, but do not integrate these data to give more insight into daily urban life patterns. By incorporating these data, habit deviation from normal and post-event, the changes in crowd mobility, levels of disruption after an event can be assessed accurately to support crisis management. If mobility and distribution patterns can be efficiently identified then post-crisis risk can be minimalised, crisis responders or rescue resources will arrive on time, understand what the population needs, how the different part of a city are connected and provide better evacuation plans. Specifically, the research uses GPS trajectory data only focused on motorised mobility (ex: using taxi trajectory) and does not discuss other types of mobility (walking, running, cycling, etc.). It does not explain detailed semantics of mobility such as the associated behaviour to the mobility, trip purpose or how the population use the POI.

This work aims to identify semantic mobility patterns in the metropolitan area from two datasets: tracking and social media data. The tracking data collected by apps (iOS and Android apps as a part of the European project SETA, http://setamobility.weebly.com/) that track multimodal mobility patterns in three cities (Birmingham, Santander, Turin). Since social media data relationships could explain 10%-30% of all human movement (Cho, Myers and Leskovec, 2011), we compare the patterns in tracking data with check-in patterns found in social media data. 'Check-in' refers to the specific time and location of a particular user shown by use of online social applications.

In this study, we present a method to build semantic mobility patterns from tracking data. We develop a model to discover trip purpose and important places in people's daily life based on stay point. The model will allow the analysis of human daily lifestyle patterns, when or where people use public infrastructure such as shopping malls, food places, stations, etc. Furthermore, knowledge of everyday patterns allows understanding of the citizen's needs (where they get supplies (e.g. food, clothing), where they go at the weekend, how the different parts of the city are connected, etc.). Specifically, it also shows and gauges the level of disruption after an event, as it is possible to check the deviation of habits from a normal situation and post-event. This detailed information is very useful for decision makers or policymakers.

In summary, the contributions of this paper are:

- We study an important and challenging problem of understanding human lifestyle patterns in tracking data records. We build context annotations that help bridge the gap between raw tracking data and surrounding contexts and hugely enrich our understanding of mobility data.
- Different from existing studies on human mobility for crisis management which do not consider detailed semantics of mobility, we propose a semantic model to capture meaningful user activity based on stay points and type of location visited.
- We conduct experiments on large-scale mobility datasets in Birmingham (534 users in tracking data and 116,468 users in social media).

Below, we discuss related work and challenges in understanding human mobility in crisis events. Next, we describe heterogenous datasets used in this research. We present a method for identifying human semantic mobility in the system overview and framework section. The results of our experiment are discussed in the Results and Discussion section. Finally, we conclude the paper and provide some directions for future work.

RELATED WORK

A number of studies have been made on how mobility patterns can be utilised to supporting crisis management by applying techniques such as crowdsourcing, GPS data, and social media. Many studies have employed the content of social media such as tweets to identify irregularities in human movements or communication before, during, and after disasters (Hong, Fu, Torrens and Frias-Martinez, 2017; Kumar and Ukkusuri, 2018; Xiao, Huang and Wu, 2015). Other researchers have proposed a general model to identify human post-crisis movement patterns based on mobile GPS data (Song et al., 2015, 2017). Call Detail Records (CDR) have also been used by researchers to explore communication behaviour changes caused by disasters, especially in particular locations where resources for evaluating mobility behaviours are rarely available (Hong, Mashhadi and Frias-Martinez, 2018). While these studies present in-depth investigations of large-scale human mobility, they often focus on specific types of disaster without considering other potential considerations in the aftermath of crises such as the dynamics of population lifestyle in the affected area (how the citizens use POI in the affected area).

Another stream of work developed a model to predict human mobility during or after a crisis. Song et al (Song et al., 2015) proposed an effective model based on a Hidden Markov Model to predict human mobility following disasters, the Great East Japan Earthquake and Fukushima nuclear accident, based on GPS records, news, and transportation networks. Recently, a deep learning framework was developed to predict regular and irregular human mobility during rare events in the city from historical GPS data (Fan et al., 2018). These prediction models, however, usually lack semantic understanding of mobility because they did not consider transportation mode - stay, walk, bicycle, car, train - and were not associated with other context data such as POI, weather, or event information which could provide detailed knowledge to boost prediction accuracy.

The most similar work to this paper examined a process to annotate raw tracking data with some contextual sources. Adding semantic annotations is a key for human mobility analysis, especially finding behaviour patterns from tracking data. Point-of-Interest Extraction (PIE) Algorithm annotated stay points in indoor activities by defining parcel polygons and polygons-of-interest (de Graaff, de By and van Keulen, 2016). Furthermore, a framework which combined trajectory segmentation (move or stop) with contextual information (OpenStreetMap[1] and Ordnance Survey UK[2]) examined the type of place visited by the volunteered GPS trajectories (Siła-Nowicka et al., 2016). Other works also studied the function of each region (zones) in a city using a topic modelling approach based on taxi trajectories and POI data (Yuan, Zheng and Xie, 2017). The method annotates functional zones with semantic terms by calculating an average POI feature vector (rank POI category in a zone) and frequently observed mobility pattern in each zone. In addition, research related to mobility based on check-in location in social media has also been conducted. Constructing a tourist travel diary in Hongkong based on POI check-in obtained detailed trip and contextual information reflecting tourist activities (Vu et al., 2017). Chao et al analysed college community patterns in Wuhan based on geotagged Weibo messages (Yan et al., 2018).

HETEROGENEOUS DATA SOURCES

In this work, we employ various data sources to build semantic understanding of human mobility to support crisis management; they can be summarized as follows:

- **Tracking Data:** We collected tracking data from thousands of users in Birmingham, Santander, and Turin from September 2017 - September 2018. As a case, we use tracking data from 534 users in Birmingham. The SETA apps have been distributed to the citizen with some social economic category (a specific type of population). Users are voluntary and entirely aware that they are being tracked for research purpose. Data were collected and managed by large scale architecture that provide query processing, integrated processing, Reliability, Data Privacy, and Security. The framework and architecture for data collection were discussed in the previous research (Ciravegna et al., 2018)
- **Social media data:** We extracted check-in data in Birmingham, UK from the geo-located tweets (Twitter) of 116,468 users from September 2017 to September 2018.
- **POI database:** We built a POI database based on data from Foursquare[3] and google places[4]. The data contain information about POI, such as name, location, POI category, opening times.
- **OpenStreetMap Database:** We used the UK OpenStreetMap database to map tracking data with potential stay points.
- **Postcode database:** We used the UK postcode database to map stay points with the nearest postcode.

SYSTEM ARCHITECTURE AND FRAMEWORK

In this section, we describe the system architecture for developing a semantic understanding of human mobility from tracking data to support crisis management. As explained above, the framework and data collection infrastructure has been discussed in previous research (Ciravegna et al., 2018). Therefore, the following system architecture (Figure 1) focuses on the data processing method.

[1] https://www.openstreetmap.org/
[2] https://www.ordnancesurvey.co.uk/
[3] https://foursquare.com/
[4] https://cloud.google.com/maps-platform/places/

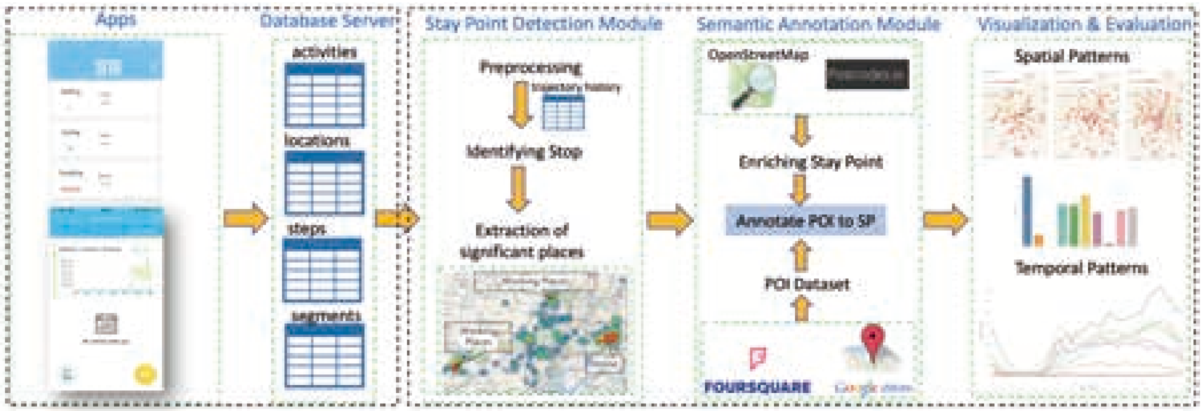


Figure 1 System Architecture & Framework

Our framework comprised four components: Data Collection Tool; Stay Point Detection Module; Semantic Annotation Module; and Visualization and Evaluation module. The detailed description of each part is explained in the following sub section.

Data Collection Tool

The SETA app is a mobile application that monitors a person's movement and presents a summary and breakdown of their day so that a user can track their level of mobility. The app tracks vehicular and active mobility to allow the next analysis related to lifestyle patterns such as stay point or trip purpose that refer to the points of interest. This application captures data from mobile phone sensors, such as accelerometer, GPS (location), and pedometer (steps). Data from multiple sensors are analysed and combined to produce segments of activity such as walking, cycling, and running, including step counts, time of activity started and ended, and individual locations (latitude, longitude).

The apps fuse individual tracking data, sent to the server and immediately stored in a database. The final dataset collected is extracted, validated, analysed and filtered to form tables which store data about activity, locations, segments, and steps.

Stay Point Detection Module

Real tracking data are often not accurate, due to noise, missing, or incomplete data. Processing this low-quality data will lead to poor results. The pre-processing step is performed to provide clean and valid data to obtain useful features. Any duplicate record is removed, the timestamp locations checked with excluded locations with the same timestamp. The pre-processing step also does temporal alignment by aligning raw data (activities, locations, steps, and segments) captured from different phone sensors into one history table. The alignment process classifies locations into activity segments based on the location timestamp. All locations between activity segment start time and end time are classified as one location segments and then the activity is attached to the location. The segments categorised in 6 activities are Vehicle, Cycling, Stationary, Other, Walking, and Running.

Stay point is the set of geographical points where a person stays for a while within a certain distance threshold. In general, stay point can be represented as a set of consecutive GPS locations $L=\{l_m, l_{m+1}, l_{m+2}, \ldots, l_n\}$ *where* $\forall i,\ m< i\leq n,\ distance(l_m, l_n)\leq dthreshold$ *and* $|duration(l_m, l_n)|\geq tthreshold$. The distance between two geographical locations is computed by using the Haversine formula and the time span between two location points is defined as stay duration. Stay point plays a major role in adding a context or semantic meaning to trajectories. Identifying the stay points allow us to segment the trajectory to be a stop or move, determine mode transition e.g. walking to bus, and infer trip purposes by associating it with POI information.

We develop an algorithm based on the PIE algorithm (de Graaff, de By and van Keulen, 2016) and the algorithm proposed by Li et al (Li et al., 2008) to obtain stay points from tracking data. The procedure to acquire stay points from tracking data can be seen in algorithm 1 (see 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.  SP = {sp1, sp2, ....., spn} the list of stay points
2.  Clustering of Stay Point
3.  Home Place

begin
    for each user do
        //Remove inaccurate locations that greater than the average of accuracy
        listLocationsPoints = {lp ∈ UT | lp.a ≤ avg(a)}

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

        //Cluster candidate Stay Point by applying DBSCAN and generate centroid
        listStayPointCluster = getClusterDBSCAN (listCandidateStayPoints)

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

        //define some significant places: home, SP1, SP2, SP3, etc
        defineHomePlace(stayPoints, summarySP)
end
```

Figure 2 Algorithm to extract stay points

The algorithm works by retrieving a sequence of locations for each user trajectory. The inaccurate locations are removed based on the average of location accuracy in the data. The stay point candidates for each user are generated by considering the distance and duration threshold parameters. We also consider the type of location segment in deciding stay point candidates. The '*stationary segment*' means that users stop in certain locations for a certain time based on segment timestamp. We focus only on the stay point where the user stays for equal to or more than to 5 minutes and not give much attention to shorter stay points such as traffic light stops, traffic at roundabouts, etc.

The stay point candidates are location clusters which consist of one or more stop locations. Even though the most inaccurate locations are already previously removed, the stop locations could contain locations outside of the POI. Therefore, we utilise DBSCAN algorithm (Ester et al., 2008) with parameter *eps=30 meters* and *minPts=5* to get more accurate locations in the stay point and get a better centroid of the cluster based on remaining stay point locations.

For spatial and time regularity, we record features such as frequency, day, arrival time, leaving time, and stay duration to produce a summary of user stay points. We also create Origin-Destination matrices based on stay point to get detail journey of the user. We use this summary to detect Home place and extract other significant places. Home is considered the place where a person spends most of his/her time. Therefore, the rules for defining Home are:

i. The most frequently visited place with the most total duration of visits.
ii. The start and end location for every day.
iii. There should be at least 5 stop locations per week in the stay point.

We aggregated and clustered stay points for each user by temporal features (time, day) and spatial features (POI category, postcode) to understand human lifestyle patterns. Knowing daily patterns of citizens is critical to allow us in measuring the level of disruption after a crisis, as it is possible to monitor the deviation of the habit before and after the event. Figure 3 shows the distribution of stay points (on weekdays and weekends), with the sizes of the circle indicate number of stay points at the POI.

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

Semantic Annotation Module

Semantic annotation is a process of adding context to the stay point so we can classify it to the certain POI. We want to analyse which POI motivated stops in the stay point during the trip. There are two subprocesses to semantically annotate the individual trajectories: enriching stay points with additional data from OpenStreetMap and postcode.io[5]; and finally, annotating POI to the stay point.

To get more meaningful knowledge about activity label, we enriched stay point with related contextual information, such as OpenStreetMap and postcode. We mapped the stay point to the OpenStreetMap data by exploiting the information around the stopping place to get POIs nearby, land use, or amenity information. We utilized spatial relationship features supported by the PostGIS[6] database to determine if a stay point was entirely inside another geometry (polygon) or within the specified distance (maximum 100 m) of points or polygons representing the POIs. If this condition was fulfilled, then we assigned a name of POI nearby, land use, and amenity information (the nearest POI within 100m is assigned to the stay point which is not inside of another polygon). We also assigned postcode information to the stay point based on the nearest postcode.

The last process in building semantic annotation to the stay point is classifying stay point to the POI.

The POI dataset enables us to model the functionality of each area in the city. POI dataset gives some information about the venues, such as name, category, coordinate, opening times, and a bounding box. We built POI dataset based on data from Foursquare, OpenStreetMap, and google places. The POI dataset comprises 10 categories: *arts & entertainment, college & education, event, food, nightlife, outdoors & recreation, professional, residence, shops travel and transport*.

We queried foursquare Places API to collect POI information in Birmingham city with bounding box [{'south west': '52.381054, -2.033649', 'north east': '52.608707, -1.728858'}]. To cope with a limited request in the API, we used the technique introduced by Fei et all (Fei Wu, 2016). For each search query, the API returns at most 50 POIs. We performed recursive steps by partitioning the Birmingham map into smaller grids (four partitions for each iteration) if a query on the map or current grid returned 50 POIs. As a result, we obtained 19122 POIs. Figure 4 shows the distribution of POI categories in Birmingham City.

[5] http://postcodes.io/

[6] https://postgis.net/

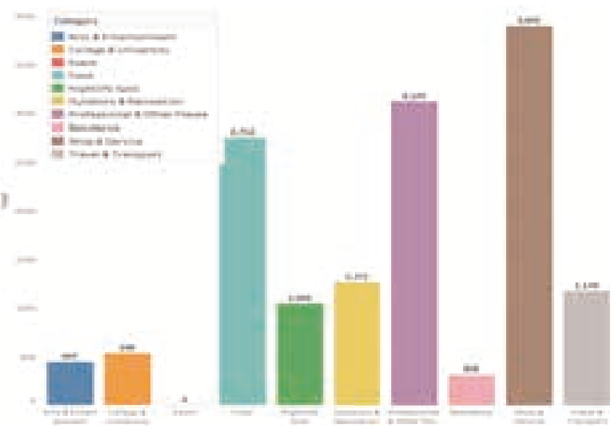


Figure 4 POI distribution in Birmingham, UK

Based on the defined POI dataset, the stay point is annotated with POI category. Home stay points located in residential areas are excluded from further analysis. The semantic annotation rule is applied by combining the distance-based method and contextual information previously found from OpenStreetMap. Other stay points are annotated with the POI based on the nearest distance, highest similarity POI name to the stay point and the arriving time in the range of POI opening times.

The algorithm to build semantic annotation to the stay point can be seen in the following algorithm (Figure 5).

```
Algorithm 2 Annotate Semantic SP

INPUT:
1. SP: list of significant places [<list stay point>,...], SSP:summary stay point [<summary SP>]
2. OpenStreetMap Data
3. POI database

OUTPUT:
1. SSP = [sp1, sp2, ......, spn] the list of semantic stay points

begin
    for each SP do
        //enrich the stay points with postcode
        SSP=getPostcode(stayPoint centroid)

        enrich the stay points with OSM by considering spatial relationship
        SSP = enrichStayPoint(listSummarySP)
            [sspi ∈ SSP | distance(sp, polygon) = 0 ∨ distance(sp, polygon) < eps ∨
                              distance(sp, point) < eps]

        //Add semantic to the stay points by considering same name and nearest POI
        If not Home and not Land-use residential
            SSP=getSemanticSP(SPname, arrivalTime, POI Database)
end
```

Figure 5 Semantic annotation algorithm

Visualization and Evaluation Module

Semantic POI finds in the tracking data are visualized via a faceted browsing interface. We show the statistic on how the population in Birmingham used POI in the city and visualized the spatial distribution in the map. We also use this module to visualize POI check-in from social media. Such information is very useful for identifying the distribution of population during an emergency and therefore the number of people affected in each area.

EXTRACTING POI CHECK-IN FROM SOCIAL MEDIA

Location-based services in social media applications facilitate people sharing their check-in places (Cho, Myers and Leskovec, 2011). The check-in information contains geotagged venue and its geolocation attribute which are very useful in identifying human movement behaviour. Mobility pattern can be understood from social media data by extracting check-in tweets to get information about activity purpose. The activity purpose is inferred based on the category of the visited POI appearing in the tweet. Furthermore, the spatial distribution of popular places is represented by considering the total number of check-ins for different POI category and the popularity over time for those POI categories is considered as temporal distribution.

In this paper, we use 2,876,645 geotagged tweets from 116,468 users in Birmingham, UK from September 2017 to September 2018. We recorded geo-located tweets using Twitter search and streaming API[7]. After collecting this dataset, we selected the subset of tweets containing check-in word and Instagram. As a result, we obtained 43,971 check-in tweets for our analysis. Based on our observation, we extracted the check-in tweets by defining the strategy below:

i. Check-in tweet contains check-in words: ["I'm @", "I'm at", "I am @", "I am at", "arrived at", "arrived @", "I am in front of", "I'm in front of", "I'm going to", "I am going to"] followed by POI name.
ii. POI name in the check-in tweet is matched with POI name in POI database if:
    a. the distance between the geotagged tweet and POI coordinate is within 100 meters and
    b. the similarity between POI name is large or equal than 0.8. We use Lavenshtein distance[8] to compute the similarity between POI name.

Another rule applied to obtain check-in tweets was to utilize Instagram links found in the tweets. People publish their status on Instagram when they check-in to one place and this status links to their Twitter account. The Instagram content includes POI name without check-in word pattern. The text in the tweet will contain the Instagram link. Therefore, we extract the Instagram link in the Twitter and crawl the link to obtain a POI name where the user checks in. Then the matching rule in point ii is also applied to define the POI name. We also consider the POI opening time to validate the POI name. The example content of Instagram link can be seen in Figure 6 below.

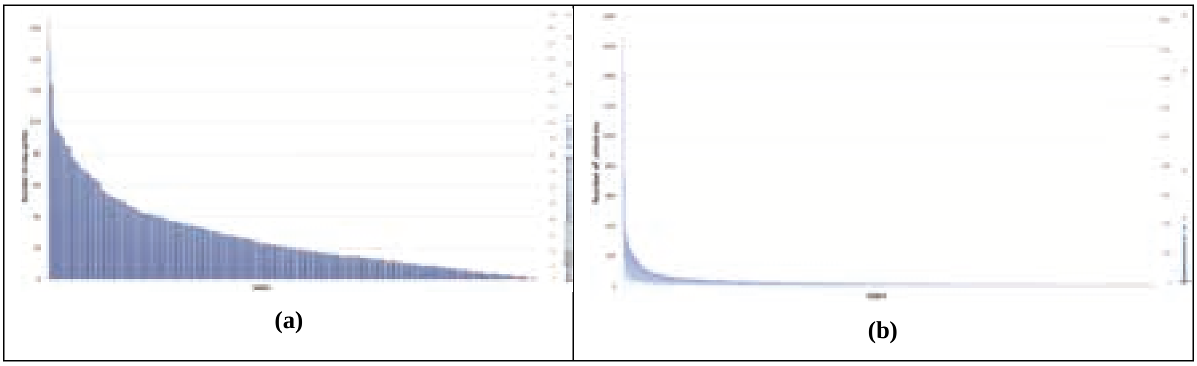
Figure 6 Example Instagram content link in a tweet

RESULTS AND DISCUSSION

In the following section, we present key findings derived from tracking and social media data. We demonstrate the capability of our method to capture rich information about human lifestyle patterns. Approaches were applied to analyse the behavioural characteristics of users in Birmingham from two perspectives: temporal pattern and spatial pattern analysis.

The tracking information was collected from 534 users in Birmingham during September 2017 to September 2018. The users were part of the population with certain characteristics related to a social-economic indicator such as income, ethnic community, medical condition or mental well-being. All tracking users were anonymous, and a single person could represent multiple users of the SETA app. Based on the framework implementation on tracking data, we found 72% (389) of users had more than 10 stay points and 50% (275) of users had between 10 and 38 stay points. Meanwhile, geotagged social media data were collected from Twitter users in Birmingham in the same period. We obtained 44,225 check-in locations in Birmingham with average 4 check-in locations per user (minimum=1, maximum=1680). Figure 7 shows the distribution of stay point or check in locations per user.

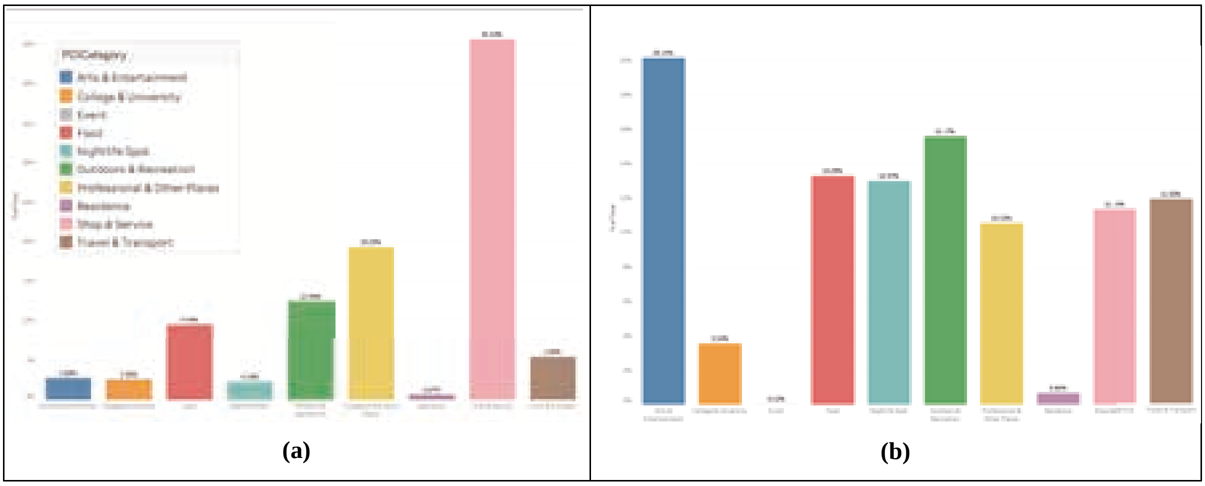
[7] https://developer.twitter.com/
[8] https://en.wikipedia.org/wiki/Levenshtein_distance



(a) (b)

Figure 7 Number of stay point or check-in location per user. (a) tracking data (b) social media

Figure 8 shows the POI distribution for tracking and social media data. Overall, trend activity based on POI category visited by the user was very different. As can be seen, Shop & Service-related activity was the highest among user in tracking data. The remaining activities (Professional & Other places, Outdoor & Recreations, and Food related activity) follow with big different gaps. In contrast, activity patterns in social media were more uniform. Arts & Entertainment related activity was the most frequent and other activities (Food, Nightlife, Outdoor & Recreations, Professional & Other Places, Shop & Service, Travel & Transport related activity) had similar levels of popularity.



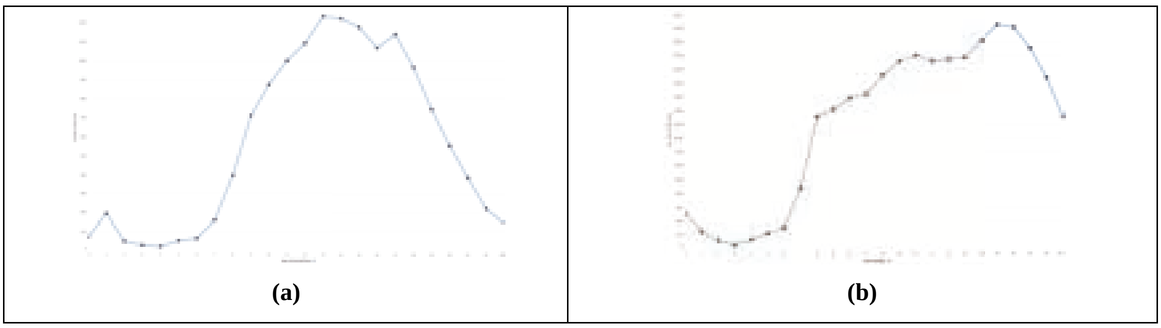
(a) (b)

Figure 8 Statistic POI distribution (a) tracking data (b) social media

The following subsections discuss the temporal and spatial pattern results obtained from tracking and social media data.

Temporal Mobility Patterns

We extracted time (arrival time in tracking data and check-in time in social media) when users followed the activity. Figure 9 shows the number of stay points and check-ins per day. As can be seen, this pattern presents a glimpse of the active time of the user in tracking and social media data. Clearly, both users tended to follow their activities in the morning at 7 am. There is a major peak between 12 and 17 in tracking data that indicates most users followed their activity in the stay point at this time. In social media, the pattern reached a peak between 17 and 21 pm; this means that most users tended to post their activity information after office hours.



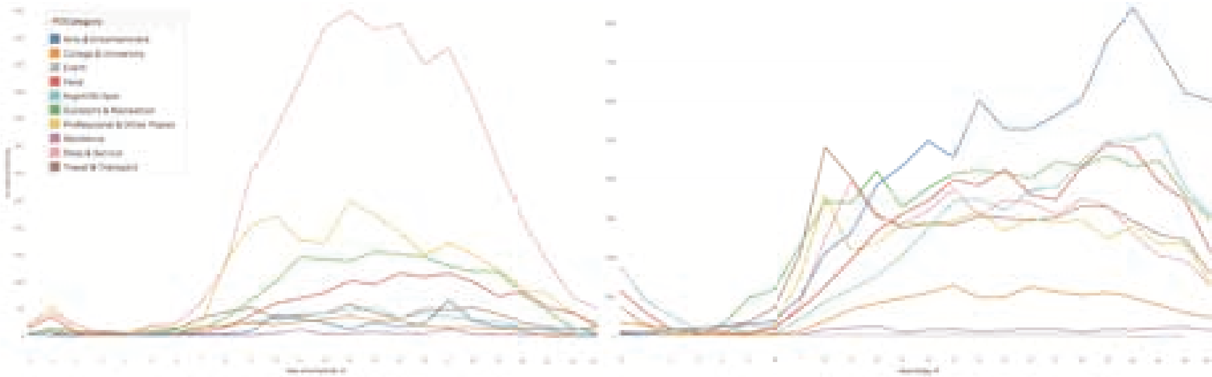
(a) (b)

Figure 9 Time distribution of (a) tracking data and (b) social media

The temporal pattern indicates time-varying popularity where a place may be of interest to people. To reveal detailed temporal mobility regularity, we investigated the distribution of visits for POI category at different hours of the day. We aggregated the number of stay points or check-ins for each POI category in each hour of the day (see Figure 10). The weekly pattern of these visits was also examined and can be seen in Figure 11.

We observed that activity purpose had a pronounced impact on the time of activities. For instance, we can see that food-related activities in these two datasets had three distinct peaks around noon (12-13 pm), evening (15 pm) and early night (17-19 pm). Nightlife activities were more popular during afternoons and nights and peaked around night (21 pm) as these visits mostly constituted going to bars and nightclubs (even though, nightlife activity was not common in tracking data). Furthermore, we can see that activity related to professional places was increasing significantly in the morning (after 7 am) and remained steady between 8 am - 17 pm. This shows that this activity was mostly in a working place or school. Similarly, outdoor and recreation-related activity also remained constant between 6 am - 19 pm.

As pointed out above, shop and service-related activities were more dominant for tracking users (Figure 10(a)), increasing significantly in the morning (9 am), peaking around afternoon (13 pm), and dropping steadily until 17 pm. This pattern decreased dramatically after 17 pm. Using semantic stay point, we identified most of those places were the Shopping centre, Shopping Park, or supermarkets. This indicates that most users bought their supplies, such as food, clothing, etc. at this time. Most of the users also went to eat at food outlets or fast food restaurants between 13 and 17 pm. Professional & other place related activities rose sharply in the morning - at 8 am (going to work), reached a peak at 13 pm (back to work places after rest breaks) and fell gradually around 18 pm. Furthermore, the pattern shows that more people visited parks, leisure or sports centres around 11 am - 17 pm. Nightlife and Art & Entertainment related activities were less popular in tracking data at only 4.8% of all activity.

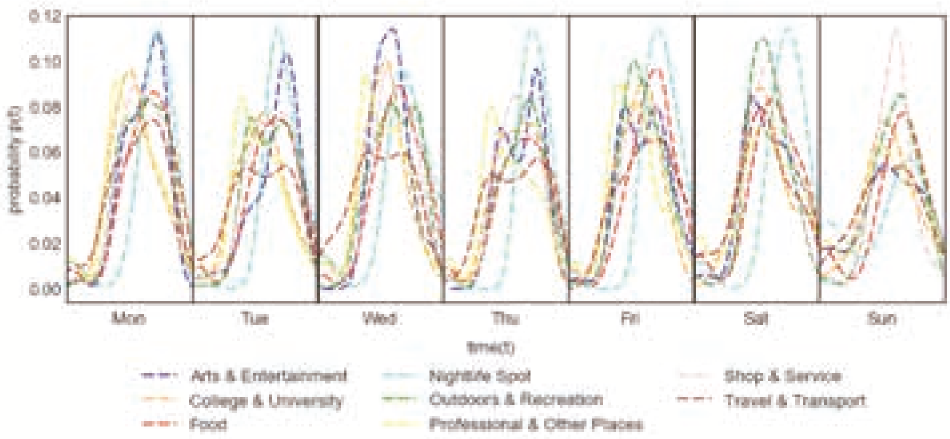


(a) (b)

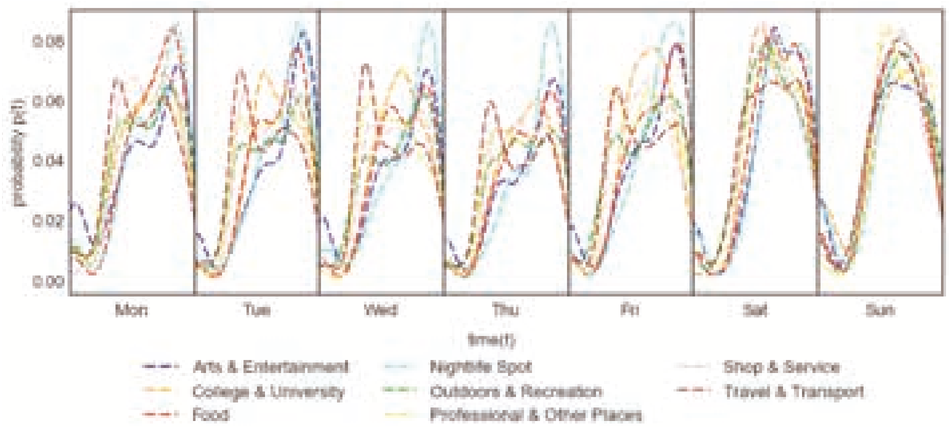
Figure 10 POI popularity over 24 hours in a day. (a) tracking data (b) social media

Figure 10(b) shows the activity patterns in social media were more comprehensive than tracking data. More people visited Art and Entertainment places around 14 pm and between 17 - 22 pm (after office hours) as these activities mostly involved attending an event in such places (Birmingham arena, The Genting Arena, O2 Academy, etc.). Shop and service-related activities fluctuated between 8 am - 19 pm and peaked around morning (9 am), and noon (13 pm). This activity was mainly conducted in food markets, shopping plaza, or shopping mall and some service places. Travel and transport activity reached a peak at 8 am when people were going to work, remained steady between 11 am and 18 pm, and decreased gradually after 18 pm when people finished working. Furthermore, most people bought food (eat and drink) in a restaurant or cafe at lunch (13 pm) and dinner time (19 pm - 20 pm).

Figure 11 shows the stay point or POI check-in density for every day in tracking and social media data. As can be seen, art and entertainment, and nightlife related activity was more popular on weekdays, and shopping activity was more dominant at the weekend. In social media, weekly patterns suggest that art and entertainment and nightlife related activity was dominant on weekdays after 12 pm, whereas shopping and recreation trips were predominant at the weekends. Users in social media also tended to post their activities from travel and transport places in the morning on weekdays. Similarly, tracking users visited art and entertainment and nightlife places on weekdays. Most users followed shopping & service, outdoors and recreation related activities at the weekend.



(a)



(b)

Figure 11 Temporal POI Check-in densities every day for a) tracking data b) social media data

Based on the temporal pattern results explained above, we can see our method has the capacity to capture detailed lifestyle patterns from tracking data and social media. Different patterns in tracking and social media data were obtained related to when users followed the activity and the type of POI visited. Clearly, most of the activity in tracking data was carried out between 9 am - 17 pm. During this period, the trend of activities (except shopping and service-related activity) showed small changes. In contrast, most users in social media tended to post their activity at a particular time: in the morning when they were starting an activity, during their lunch break and after working hours. Related to type of POI, most users in tracking data followed a routine activity by visiting the same place of POI several times (shopping in a specific kind of supermarket, etc.). Social media users posted their activity when attending an entertainment event, shopping in the plaza or mall, or eating in a restaurant. The POIs visited by user in social media were sparser than in tracking data.

Spatial Distribution of Popular Places

The current section demonstrates that our method can be used to visualise popular POIs in the city. The popular places visited by the citizen from tracking and social media data are generated by counting the total number of stay points or check-ins for each POI. The density for each region corresponds to the number of stay points or check-ins to that POI. Figure 12 and 13 show a visualisation of the spatial distribution of stay points and check-ins in the Birmingham area. The sizes of the circle indicate the number of stay points or check-ins at the POI.

Figure 12 Spatial POI Check-in densities every day for a) tracking data b) social media data

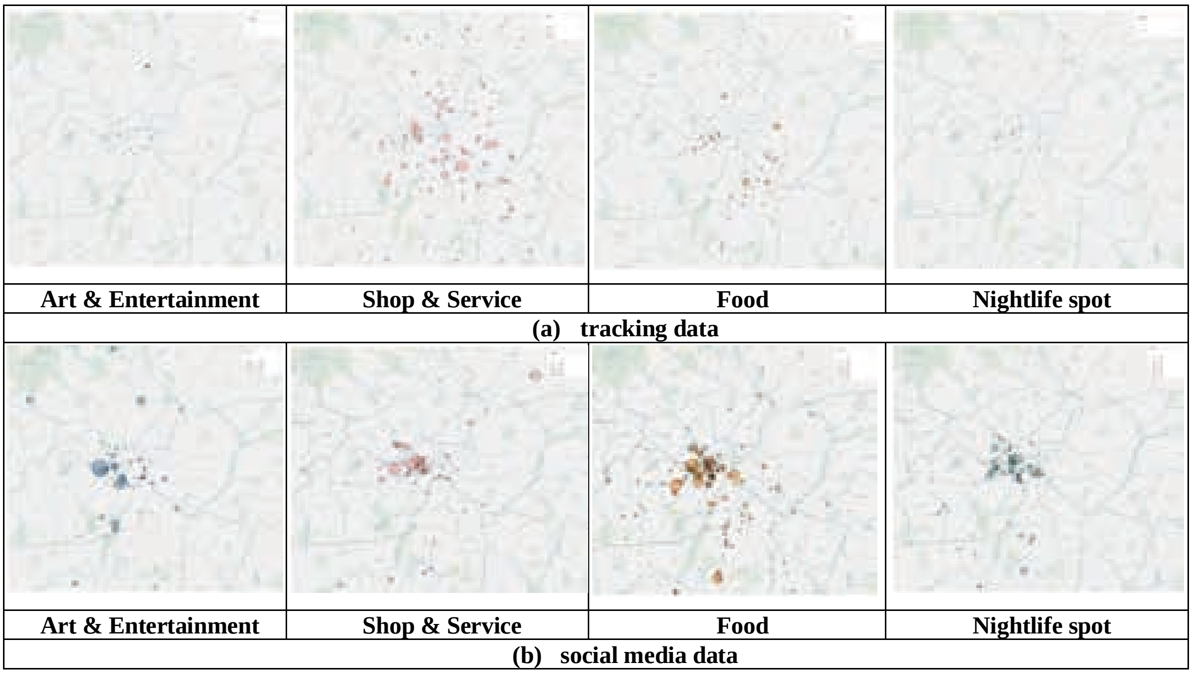


Figure 13 Spatial POI Check-in density for some POI Categories

As shown in Figures 12 and 13, popular places differed depend on activity purpose (POI category). The most active POI in tracking data was shopping and service places and few places had very high numbers of people who usually visited for eating, entertainment and nightlife purposes. Whereas in social media, other places such as art and entertainment, food and nightlife spots were also active POIs in the city. Clearly, there were more check-ins for eating, entertainment and nightlife activity in social media than in the tracking data.

Furthermore, stay point and check-in distributions look different for various activities. For example, the distribution of stay points for shopping and service-related activity and other (art and entertainment, food, nightlife) related activity look very different. Distribution of shopping and service-related activity in tracking data spread (scattered) in the city, where those places were shopping centres or supermarkets. Probably, people shopped in these particular places according to the proximity of the POI to their home. While, the distribution of activity from social media was more concentrated in the city centre, it was sparser and spread among different types of POI.

The analysis above highlights the significant differences between patterns in tracking and social media data. Users in social media were random (more variable) since data collected from all active users on Twitter who posted tweets in the Birmingham City area. Furthermore, social media usually contain features of sparsity and uncertainty. Spatial-temporal data provided by geo-located tweets are typically sparse due to the high interval between two consecutive tweets in the spatial dimension. Activity found in social media is often occasional and discontinuous. Users in social media tend to share their geotagged tweets related to an interesting event, exciting moment or express their experience when they are in certain places with/without other people.

In contrast, tracking data provide regular/routine detailed activity conducted in daily life. Therefore, some places were visited several times. Shopping activity was the dominant activity where most places visited by the user were the shopping centre, shopping park, or supermarket. Users in tracking data were not random. As explained above, the apps were used by part of the population in the city with a specific social economy or medical condition characteristics. Therefore, we did not have the largest number of users in the parts of the city where they were the largest part of the population; there does not seem to be a close correlation between tracking user distribution and population size. Nevertheless, the results show that we are able to track citizens with some characteristics, capture the rich context of human lifestyle, and model their lifestyle patterns based on visited stay points. The lifestyle patterns show the power of analysis that clearly explains that these citizens are a part of the population in the city.

Understanding daily lifestyle patterns of the population enables decision makers to understand citizen needs, such as where and when they get supplies (e.g. food), and how the different parts of the city are connected. Monitoring these patterns allows us to understand, learn, and gauge the level of disruption after an event, as it is possible to check habit deviation from normal conditions post-event.

CONCLUSION

In this paper, we have presented a framework to build an understanding of mobility lifestyle patterns from tracking data. Even though the method has not yet been used during a crisis, our result shows that the stay point annotation from tracking data and POI check-in from social media can effectively explain daily life patterns in the city especially related to the usage of public infrastructure, such as shopping places, food, nightlife, etc. This detailed information is very useful for monitoring and learning how the population behaves (deviation of habit) before or after a crisis or gauging the level of disruption after an event.

We note several limitations within our study. The tracking data used are constructed from parts of the population in a specific category in the city and not incorporate data from some representative portions of the population (i.e., people who do not own mobile devices or do not register for a GPS service cannot be incorporated into this study). Therefore, the patterns found are related to certain types of people more likely to have a specific condition (related to social economic indicator). However, we are confident that the tracking data, which offer movement behaviour for the approximately 534 people included in the database, are reflective of part of the population lifestyle patterns. Furthermore, ambiguities in defining home and working place (to distinguish people having fun or working in pubs during the night; people who live and work in the same building (ground and 1st floor) are concerned to be future work to improve our algorithm. The third limitation of our study is the context of mobility only considered the stay points. Therefore, we are concerned to improve our method in future work, especially related to the number of the population examined and analysing the changing patterns of mobility regarding other semantics, such as large events or environmental conditions.

ACKNOWLEDGMENTS

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

The authors would like to express their gratitude to the Indonesia Endowment Fund for Education for the research dissertation grant.

REFERENCES

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. *17th ACM SIGKDD international conference on Knowledge discovery and data mining.* San Diego, California.

Ciravegna, F., Gao, J., Ingram, C., Ireson, N., Lafranchi, V., & Simanjuntak, H. (2018). Mapping Mobility to Support Crisis Management. *Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management*, (pp. 305-316). Rochester.

de Graaf, V., de By, R. A., & van Keulen, M. (2016). Automated semantic trajectory annotation with indoor point-of-interest visits in urban areas. *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (pp. 552-559). Pisa, Italy: ACM.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 96*, pp. 226-231. Portland: AAAI Press.

Fan, Z., Song, X., Xia, T., Jiang, R., Shibasaki, R., & Sakuramachi, R. (2018). Online Deep Ensemble Learning for Predicting Citywide Human Mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 105.

Hong, L., Fu, C., Torrens, P., & Frias-Martinez, V. (2017). Understanding Citizens' and Local Governments' Digital Communications During Natural Disasters: The Case of Snowstorms. *Proceedings of the 2017 ACM on Web Science Conference* (pp. 141-150). New York: ACM.

Hong, L., Lee, M., Mashhadi, A., & Frias-Martinez, V. (2018). Towards Understanding Communication Behavior Changes During Floods Using Cell Phone Data. *International Conference on Social Informatics* (pp. 97-107). Springer, Cham.

Kumar, D., & Ukkusuri, S. V. (2018). Utilizing Geo-tagged Tweets to understand Evacuation Dynamics during Emergencies: A case study of Hurricane Sandy. *WWW '18 Companion Proceedings of the The Web Conference,* (pp. 1613-1620). Lyon.

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., & Ma, W.-Y. (2008). Mining user similarity based on location history. *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems,* (p. 34). Irvine, California.

Ludwig, T., Kotthaus, C., Reuter, C., Dongen, S. v., & Pipek, V. (2017). Situated crowdsourcing during disasters: Managing the task s ofspontaneous volunteers through public displays. *International Journal of Human-Computer Studies, 102*, 103-121.

Sila-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science, 30*(5), 881-906.

Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R., Yuan, N. J., & Xie, X. (2015). A Simulator of Human Emergency Mobility following Disasters: Knowledge Transfer from Big Disaster Data. *AAAI*, (pp. 730-736).

Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R., Yuan, N. J., & Xie, X. (2017, January). Prediction and Simulation of Human Mobility Following Natural Disasters. *ACM Transactions on Intelligent Systems and Technology (TIST), 8*(2), 29.

Vu, H. Q., Li, G., Law, R., & Zhang, Y. (2017, August 11). Tourist Activity Analysis by Leveraging Mobile Social Media Data. *Journal of Travel Research, 57*(7), 883-898.

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

Wu, F., & Li, Z. (2016). Where Did You Go: Personalized Annotation of Mobility Records. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 589-598). Indianapolis: ACM.

Xiao, Y., Huang, Q., & Wu, K. (2015). Understanding social media data for disaster management. *Natural Hazards, 79*(3), 1663–1679.

Yan, C., Xiao, M., Ding, X., Tian, W., Zhai, Y., Chen, J., . . . Ye, X. (2018, January 12). Exploring human mobility patterns using geo-tagged social media data at the group level. *Journal of Spatial Science*, 1-18.

Yuan, N. J., Zheng, Y., & Xie, X. (2017). Discovering Functional Zones in a City Using Human Movements and Points of Interest. *Spatial Analysis and Location Modeling in Urban and Regional Systems. Advances in Geographic Information Science,* 33-62.