

Social Media and Travel Behaviour

Abstract

Social media has emerged as a trend that greatly influences transportation and travel behaviour. This influence is identified both on the way that travellers make decisions concerning transportation related matters and the fact that social media allow tracing back the way that these decisions were made and allow for the collection of data, related to understanding these decisions. This chapter introduces these interdependences between Social Media (SM) and travel behaviour by presenting a collection of use cases and methods. Firstly, an introduction to the existing dominant SM is presented, including current availability of data and a discussion of helpful frameworks that could facilitate transportation related research on the subject. The pertinent literature on the User Generated Content (UGC) generation process and users' personality characteristics is reviewed, in order to gain understanding on the characteristics of the users, who generate the content. Secondly, the relation of SM to travel behaviour is established by investigating the impact of UGC to the transportation system and vice versa. The capabilities of SM to allow for behavioural interventions is discussed towards the direction of inducing more socially responsible behaviour. Finally, related case studies are presented.

Introduction

Social media (SM) has emerged as a prominent trend in social communication. Online platforms, such as Facebook and Twitter, conquer the Internet with millions of visitors and users per day. For the past few years the rankings of most visited webpages (Alexa.com) illustrate the steady positioning of Social Media and their overall use. Specifically, in Feb 2018, Facebook was ranked as the third, Twitter as the thirteenth and Instagram as the fifteenth most visited websites. These ranks are widely indicative of the overall trend, especially when we consider that most of the remaining Top-15 websites are search engines.

Although the history of Social Media is short, with their wide use to emerge in the 2000s, the idea of opinion exchange and in general communication platforms is not at all new. The first attempt was recorded in 1979, with Tom Truscott and Jim Ellis creating Usenet, a virtual space

for users to post public messages (Kaplan and Haenlein, 2010). Following up on that, Internet Relay Chat (IRC) debuted in 1988 as a way for friends and strangers to communicate with major success, and paving the road for modern forum platforms, such as bulletin board systems. Towards this direction, Social Media started as a space for communication, user profiles, and updates with SixDegrees.com, often considered as the first Social Media Platform, launched in 1997 (Boyd and Ellison, 2007).

These initiatives, and developed platforms, were providing the means for sharing (or exchanging) personal information about a variety of topics, and, in turn, facilitating the collection of the participants' behaviour (in terms of posts, responses, or just reading) in a very organized manner. Focusing on topics of 'everyday life', travel behaviour stands for a significant activity consuming a valuable share of peoples' daily time-budget, where decisions are made based on the available information that travellers have and thus users' to have a strong incentive to search for updated and reliable information. Moreover, based on the fact that SM provide a two-way location- and person-specific communication structure (users are contributing information to the 'networked community'), it gradually becomes a rich 'bank' of disaggregate behavioural information about users' characteristics, information collection activities, preferences, reactions, choices and other elements that ultimately constitute behaviour.

In particular, these days SM usage generates an astonishing amount of information (and consequently, data). Interestingly, most users use their smartphones for accessing SM applications -88% of Facebook, 83% of Twitter users on mobile- (<http://www.pewinternet.org/fact-sheet/social-media>). This vast 'information bank', contains records about travel activities, which extends typical information that users have access to (network/route conditions, traffic information, etc.), toward a user-centric, both fetch- and push-type, updated information that effect travellers' decisions.

On the other hand, the collection of transportation-related data conventionally relies on either travel surveys or traffic counts. However, both methods have limitations and the validity of the data collected is frequently questioned. Transportation surveys' execution is costly and there is a long-standing debate on their ability to capture human behaviour. On the other hand, traffic counts do not offer the data proximity, as well as the spatial coverage, required, especially in urban settings. These limitations have imposed the development of voluminous methods and techniques aiming to accurately represent the transportation system and shape transport policies, using the available means. Some of the limitation of the conventional types of data

and data collection methods can be overcome with the evolution of pervasive systems, such as GPS handsets and cellular networks, and particularly of Social Media (SM), such as Facebook, Twitter, Flickr and Google+. Data originating both directly from pervasive systems and from SM have shifted the interest of the scientific community from the focus on overcoming limitations towards utilizing the use of the increasingly available data, namely Big Data. Specifically, the utilization of User Generated Content (UGC), originating from SM, have been examined in the literature. However, the validity of the data as well as the context of usage has received little attention.

In this chapter we critically review various components of the use of social media in transport. We approach this topic by first providing a short summary of crucial aspects of travel behaviour that are related to social media use. We continue with the review of the use of social media in transport, discussing how the field has emerged and the directions it has taken, as well as some privacy and data related implications with regards to the use of social media data. Finally, aiming at showcasing its potential, three use cases that illustrate possible uses of Social Media are summarized.

Travel Behaviour

In order to understand the effect that SM have on transportation systems operations (mainly), it is valuable to recall some fundamental elements of the system organization and integration. For the purposes of the current chapter, among various approaches that could be used, the description of transportation systems organization will follow the socio-economic approach, where social systems are described based on demand-supply interrelations. Focusing on the term travel demand (especially regarding personal travel), this corresponds to manifestation of choices made by all (potential and actual) users of the available transportation supply system (infrastructure), or in other words, the manifestation of people's behaviour that reflects their characteristics, their needs, their opportunities and their available information. The developments in the information, communications and marketing scientific and commercial fields, provided (or promoted) additional opportunities for *ad-hoc* connections among people interested in sharing or exchanging information about particular topics, an element of interaction that—as earlier described—corresponds to a particular type of social behaviour. Also, by considering the fundamental axiom that behaviour is 'governed' (beyond personal characteristics) by the available information, it is evident that the (almost sudden) blossom of the additional channel of information that SM offered, inevitably they are playing an increasingly vital role in peoples' choices (and thus behaviour). Finally, as much as travel

activities are having an important role in everyday life, these are affected by information recorded in SM that is relevant, shared by the connected community and can be reached (and appraised as valuable) by potential users/travellers.

On the other hand, the development of a communication network of interested users, provided an additional mean from the side of system operators, in order to improve systems' performance, in an intervening scheme. In particular, the fact that information provision may influence the users' choices (and thus travel demand), the typical scheme used in transportation systems of 'guiding' users' choices may be employed using the opportunities of SM. As so, suitably selected/composed 'messages' could be broadcasted such as to encourage (discourage) users of making particular decisions, resulting to a targeted, timely and effective additional mechanism for treating travel demand and optimize system performance. In order to achieve such objectives, a thorough multidisciplinary analysis is needed. Possibly new methodological tools should be employed or developed and additional scientific disciplines should be engaged in order first understand the mechanisms of this new type of communication channel (and datasets available) and then optimal intervention strategies can be formed. In the following paragraphs, a brief description of the use of SM data in transportation is provided.

Social Media in Transport

Social Media characteristics have been the subject of multiple studies. Their particular focus in terms of functionalities that each social media covers can be organised in seven blocks (Kietzmann et al., 2011).

Social Media Functionalities

Chaniotakis, Antoniou and Pereira (2016) have adjusted these seven blocks to reflect the impact that Social Media has to transportation. In particular: a) **Presence** describes when someone is accessible and/or located. This functionality, and especially its reflection on the real world, is interesting in transportation research, because it lets researchers analyse traces from individuals to study mobility and activity patterns; b) **Sharing** refers to content that can be shared. It is interesting to mention that some platforms are built around a particular type of content, whereas others embed content within other functionalities. Social Media content sharing exploration can provide valuable information on the individual identity and also on some of the personality characteristics. However, such an endeavour would require the use of advanced image and video processing algorithms and would raise privacy issues; c) **Relationships** refers to the

inclusion of social networks (defined as a network of social interactions and personal relationships). The exploration of online social network structures and interactions can contribute to the exploration of the impact that social networks have to mobility patterns (Axhausen, 2008); d) **Identity** describes the extent to which SM allow and require users to reveal their true identity by including information and can be perceived as the act of disclosing thoughts and feelings that describe an individual's preferences (Kaplan and Haenlein, 2010). Identity characteristics are particularly interesting for transportation research on the identification of samples for modelling purposes; e) **Interactions** describes the communication (via messages, pokes, and posts) with people in their SM social network and with strangers (in some cases, from groups defined within each SM platform). SM conversations can be performed in public or with private messages. The interactions that take place among individuals in SM platforms and their tone, form, and content can reveal each individual's social network and, consequently, can allow transportation researchers to get insights in the social network and mobility relation. They also allow for the identification of identity and personality characteristics that have been found to be factors affecting mobility; f) **Groups** refer to the formation of communities and groups; g) **Reputation** refers to the individual and the content posted. Reputation is particularly important in transport, as it describes the scheme of internet influencers, people who are being followed in order to provide advices for in many cases transport-related activities.

Social Media Use in Transport

Data from social media can be categorized as a passively collecting data source forming a particularly useful user generated data source. Efforts in working with this yet growing amount of data have been directed towards all aspects of the Big Data Life Cycle (data acquisition, information extraction and cleaning, data integration, aggregation and representation, modelling analysis and interpretation; see Jagadish et al. 2014) constituting a rather multidisciplinary research topic. In transportation, these efforts have been mainly focusing on the aspects of data acquisition –mostly in terms of data collection–, information extraction and cleaning and modelling analysis.

The literature on Social Media and Travel Behaviour is vast. Rashidi et al. (2017) presented a discussion on the evolution of the literature related to Social Media and transport. They performed a Scopus search only for title and abstract, which resulted in 935 papers until late 2015 performing the following query:

("Social media" OR Twitter OR foursquare OR facebook OR yelp OR instagram) AND ("travel" OR "transport" OR "mobility" OR "geo")

Performing the same search in December 2018 (undertaken by the authors of this chapter), yields 2442 documents, with approximately 400 papers being published every year. To get a better idea of the various focus areas of social media use in transport, we have performed a number of different queries that illustrate the various research trends that emerge (Table 1). For comparison purposes we have used the same query structure for social media (("Social media" OR Twitter OR foursquare OR facebook OR yelp OR instagram)) and we have been changing the transport-related part.

Table 1: Number of papers resulting from Scopus search based on keywords used.

Keyword*	Total Number of Papers
Traffic	1096
Travel	838
Tourism	791
Mobility	685
Transportation	403
Trip	227
Travel behaviour	44
Travel demand	21

* *Query performed as a combination of ("Social media" OR Twitter OR foursquare OR facebook OR yelp OR instagram)) and the keyword mentioned, only for title and abstract.*

The commonly exploited SM-originated information is based on the use of the spatial information accompanying posts (geotag) and in some cases the language processing of posted content. This is primarily performed in terms of historic data analysis and corresponds to the investigation of research areas, such as travel demand (Lee et al., 2016b; Maheswaran et al., 2007), the exploration of activity modelling (e.g. Chen et al., 2017; Lee et al., 2016a), the identification of urban settings (e.g. Jiang and Miao, 2015), mobility patterns (e.g. Hawelka et al., 2013) and social networks (e.g. Cho et al., 2011). Also interesting is the use of SM continuous streaming of information for identifying disruptions or special events and for forecasting (e.g. Gu et al., 2016; Kumar et al., 2014; Marcus et al., 2011). Not limited to the use of data, SM exploitation for transportation purposes is reflected upon the direct communication of transport service providers with users and information sharing by SM platforms (Gal-Tzur et al., 2014). For more details on the different areas of using social media for transportation studies the reader can refer to Chaniotakis et al. (2016). In the next few

paragraphs we shortly explore the use of social media for the categories identified as most pertinent to travel behaviour.

Interestingly, different trends emerge with the establishment of Social Media research in Transport. As presented in Figure 1, although in the first few years the keyword “traffic” was dominant, there has been a gradual but steady shift towards “tourism” and “travel”.

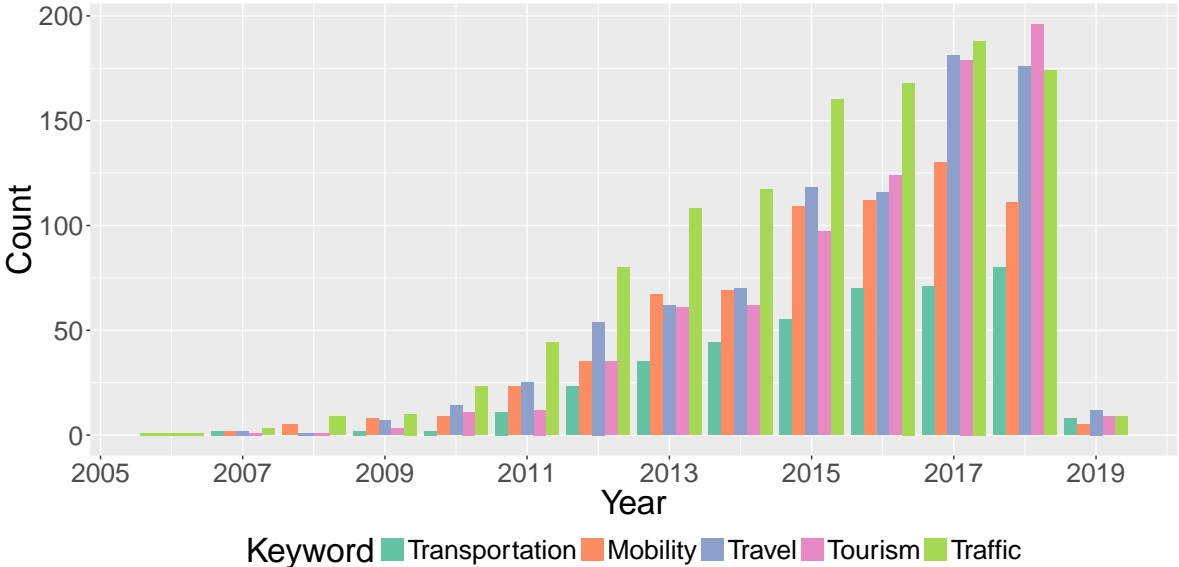


Figure 1: Evolution of Literature for different keywords explored, based on scopus search

To summarize, the main interest -especially in the first few years of the exploration of social media use- in transport was around the data properties itself and the resulting travel characteristics. This was explored on the basis of what mobility patterns can be extracted from social media, both on a local (Cheng et al., 2011) and a global level (Hawelka et al., 2013), focusing in many cases to model locations visited (Cho et al., 2014; Yuan et al., 2013). Accordingly, the investigation of the potential to substitute or complement existing data collection methods was explored. While initial efforts aimed to substitute conventional travel survey data, this was quickly abandoned as discrepancies were observed, such as high representation of specific fractions of activities (e.g. leisure) or differences in direct comparison emerged (Lee et al., 2016b; Rashidi et al., 2017). Lately, efforts on the use of social media are slowly directed towards the combination of different data sources employing data fusion methods (Akbar et al., 2018).

Privacy and Data Implications of Social Media use

Privacy implications of Social Media have long been examined by different angles. Two main streams can be identified: a) potential of privacy infringements through social media and b) changes of privacy perceptions through Social Media. Both are widely connected to transportation research and the use of social media in transport.

Starting from the latter, it has been evidenced that Social Media has changed the culture of sharing information, especially for younger generations (Madden et al., 2013). Users are more likely to reveal their real name, talk about their interests and be open in revealing their thoughts and experiences. The reason why this is important for transport is that it relates to the difficulties associated with data collection. Users revealing true information could lower the value of privacy (Antoniou and Polydoropoulou, 2014); increasing the response rate on transportation related surveys and allow for the combination of Social Media (which provide an inexpensive stream of information) with other sources of data.

On the potential of privacy infringements through social media, the case of Cambridge Analytica (Greenfield, 2018) was one of the cases that changed the landscape on Social Media privacy awareness and the need of safeguarding user privacy. One could argue that this was only the tip of the iceberg. Users providing geolocated information or information concerning activities could have adverse impact on their safety and their ability to safeguard their interests e.g. insurance premiums and employment (Sánchez Abril et al., 2012). Although numerous measures have been taken from Social Media providers to safeguard user privacy, the possibility of privacy infringement is more relevant than ever, as new methods of user identification collection of user data emerge (Smith et al., 2012; Zheleva and Getoor, 2009). Users of social media in many cases stand helpless against identification of possible treats and issues, reducing trust and overall creating problems, even to a healthy use of Social Media.

With regards to the use of data from Social Media, there is a long-lasting discussion on the premises of data provision and inherent biases. Data provision is tightly connected to issues of privacy and the code of conduct of using sensitive data. As extensively discussed in Chaniotakis, Antoniou and Pereira (2016), data availability relies on the Social Media platform providers' policies and one of the main disadvantages of its use, is the uncertainty with regards to the continuation of its provision. Already in the last few years, and due to privacy concerns, many providers changed the policies concerning the use of APIs and limited the data provided (e.g. provision of number of visits from Foursquare, events visited data from Facebook) in ways that could restrict its use in Transport. With regards to biases, the main point raised with the

sampling bias that related to the socio-demographics of Social Media users. With the establishment of Social Media platforms this seems to be alleviated (Rashidi et al., 2017); however, there is still the issue of why and when people use social media, which poses the question of the capabilities that Social Media data have to increase the observability of the transportation system.

Case Studies of Social Media use in Transport

Aiming at illustrating the way that Social Media interact with travel behaviour, in the next few sections the exploration of three related case studies takes place. The first case study aims at illustrating how Social Media can influence travel behaviour by the exploration of the activities shared within them, the second case study aims at illustrating the focus of social media usage in a spatio-temporal pattern, while the third explores the differences of social media use in different areas around the world. Naturally, there are many other applications, focusing on various modes of transport. Arguably, of particular interest are applications related to public transport (e.g. Steiger et al., 2014 and Cottrill et al., 2017).

Extracting Activities from Social Media

In this case study the implementation of the framework derived by Chaniotakis et al. (2017a) has been used to combine data from Foursquare and Twitter in order to perform text modelling for the identification of users' activities (Figure 2). An efficient method is implemented that considers the locations that a user visits multiple times in order to extract texts that refer to the same activity. User data is classified using a density-based spatial classification (DBSCAN). This allows the distinction of locations visited frequently and enriches the text available for modelling using Foursquare posts, in a robust and fast way. In terms of data collection, Twitter data was collected for a period of one year for the greater London area, resulting in 482,883 unique users and approx. 4.5 million tweets. 90,000 users were selected as a sub-sample to collect 200 tweets (from their timeline). In total, the database included 11,060,814 tweets. From those, 8,141,996 were tweeted in the greater London area. From the geotagged tweets in the greater London area, 3,764,230 (46.2%) included a link that could be parsed, 220,118 of which originated from Foursquare (2.7%).

The Foursquare links were used to extract information concerning the activities that relate to the tweets. As presented in Figure 3, a tendency was observed towards leisure activities for all days of a week. Activities such as education and work were clearly represented more highly during week days and less represented during weekend days. The activities with the highest

representation were found to be the “Bar – Pub” and “Restaurant” activities. Classification methods were applied on the resulting dataset, after removing stop words and punctuation, with the highest accuracy to be observed with the use of Generalized Linear Model via Penalized Maximum Likelihood (GLM) which had an overall classification accuracy of 83%.

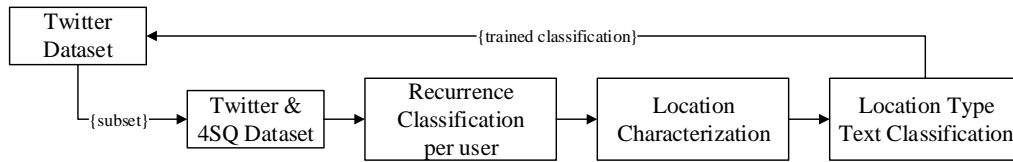


Figure 2: User-Centric Activity Enrichment high level methodology for extraction of activities (adapted from Chaniotakis et al., 2017)

The particular case study provided an understanding of the direct merits in decoding social media data and particularly the exploration of users’ activities commonly under-represented in conventional travel survey. User activity patterns can also be derived and modelling of long-term activity-chains can be realized, as Social Media users share information in a much larger span of time in comparison to conventional surveys.

Spatio-temporal patterns of social media usage

The second case study presents the analysis performed in order to compare the characteristics of social media use in comparison to conventional measurements (Chaniotakis et al., 2016a) on a temporal, spatial and activity level. The main driver of this study is the exploration of the possibilities of using social media in transportation modelling with regards to travel behaviour.

The study was based on the collection of publicly available data from three social media platforms (Facebook, Foursquare and Twitter), which was later compared to a conventionally collected travel survey performed in the same period of time for the city of Thessaloniki, Greece. Within this case study, the activity characteristics are explored, aiming at identifying which activities are commonly represented by each data source. Then, the temporal characteristics are investigated, including the derivation of correlations among the various data sources. Finally, the spatial distribution of the data is examined in the Thessaloniki city centre using spatial heat-maps that extract the density of locations visited from each data source.

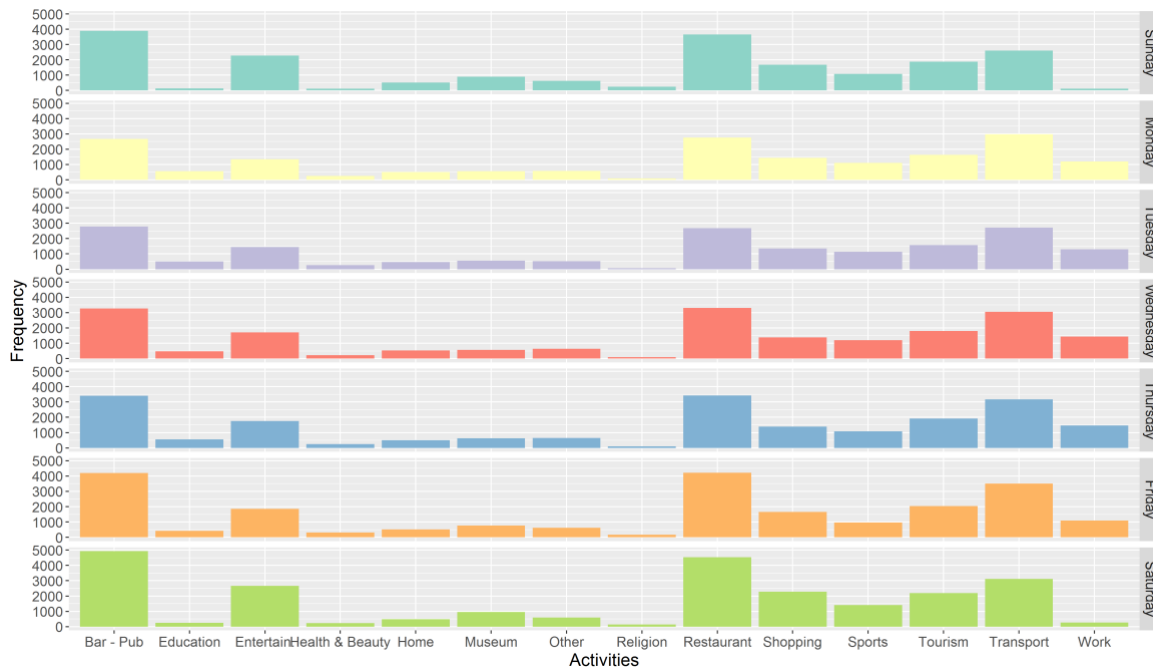


Figure 3: Temporal distribution of Foursquare labelled activities (adapted from Chaniotakis et al., 2017)

On a spatial level it was clearly evidenced that the distribution of the locations from social media were found to be distributed in areas with high concentration of recreational land uses which was not observable for the case of the conventional travel surveys, data of which, illustrates a more evenly distributed concentration of attractions with little fluctuations in the city centre. This was observable also from the correlations of the observed temporal distributions, where the check-ins for Facebook and Twitter with the Travel Survey illustrated very low, and in many cases negative, correlations.

Social Media use – comparability and transferability

In this third case study, data collected from a selection of 10 cities around Europe and the USA are analysed for the extraction of information to compare different social media use around the world and transferability (Chaniotakis et al., 2017b). The different descriptive statistics that can be identified from the REST API data collection of Twitter are first explored and the timeline (twitter history) of a sample of users per city is collected. The analysis performed includes the classification of users' posts and the use of distributions to compare different cities' results, following a user-centric analysis of the posting activity and the connection with other Social Media Platforms. Additionally, aspects of activity space have been analysed and compared.

The data used was collected for 4 cities in the United States of America (Los Angeles - California, New York, Orlando, Seattle) and 6 cities in Europe (Amsterdam, Athens,

Copenhagen, London, Munich, Paris). Based on the collected dataset, a random sample of at least 1000 users were selected for each city, in order to collect their twitter timeline. Based on a spatial, temporal and activity space analysis, we can conclude that there are clear differences in the use of social media for different areas around the world. Starting from the posting characteristics, the percentages of the number of geotagged tweets posted in each case clearly differ, with cities in the USA having a range of geotagged tweets that is higher than that observed in European cities (32.9% to 48.4% in the USA vs. 11.7% to 29.2% in Europe). Specifically, the highest percentage of geotagged tweets (to the total number of tweets) is found in New York (73.9%), while the lowest is found in Copenhagen (24.3%). Differences have also been observed on the spatial distribution of tweets, corresponding to the different urban structures observed between the two continents. Finally, the exploration of the activity space of different individuals has also confirmed the differences in the use of Social Media in Europe and in the USA, while it is worth noting that all European cases have a significantly larger activity space in comparison to the activity space of the American cases.

This case study has illustrated that the posting process and in general the use of social media is not the same between different areas around the world, even for somewhat similar regions. Consequently, the transferability or generalization of solutions has to be directed through the exploration of why and how people use social media and how this is translated in terms of transport demand.

Conclusions and Discussion

Social media has emerged as a trend that greatly influences mobility and travel behaviour. This influence is identified both on the way that travellers make decisions concerning mobility-related matters and the fact that social media allow tracing back the way that these decisions were made and allow for the collection of data that related to understanding these decisions. This chapter offers a comprehensive ‘guided tour’ on the added-value that SM is expected to offer in the organization transportation systems, starting from pointing out the specific elements that this new communication paradigm extends the information that traditionally was used in analysing the phenomenon of travel.

The literature review presented highlights advances and methods that have guided the use of Social Media in Transportation Research. The most significant outcome is the shift from the exploration of the data properties towards the investigation of its potential to substitute or

complement conventional transport data. In this regard, particular interest is paid on identifying interdependences between Social Media (SM) and travel behaviour by presenting a collection of different use cases and methods. The indicative research outcomes presented were selected mainly in order highlight the degree of detail that SM datasets offer and that never before were available for the understanding the transportation system operations, especially that of travel of people. The differences of social media use in the USA and Europe is first discussed, aiming at a better understanding of aspects of social media use and transferability of solutions. Then, the potential of using the posted text is investigated for the exploration of activities; this is believed to be one of the most valuable information that social media data can provide in this context. Finally, the direct comparison between different sources of data highlights the need for using heterogeneous data in transport related applications.

The collection of social media data is a process that is amenable to a great degree of automation. Kuflik et al. (2017) describe the potential and challenges of automating this process. It should be noted, however, that besides the merits of using Social Media in transportation studies, the highlighted issues concerning privacy and data availability can force the use of poorer -in terms of information- datasets that may rely in extremely expensive data collection processes, of small sample sizes. The same is evidenced with other (big) data owners that -although collecting data of high value for transportation- refuse to open it for research.

References

- Akbar, A., Kousiouris, G., Pervaiz, H., Sancho, J., Ta-Shma, P., Carrez, F., Moessner, K., 2018. Real-time probabilistic data fusion for large-scale IoT applications. *IEEE Access* 6, 10015–10027. <https://doi.org/10.1109/ACCESS.2018.2804623>
- Antoniou, C., Polydoropoulou, A., 2014. The Value of Privacy: Evidence From the Use of Mobile Devices for Traveler Information Systems. *J. Intell. Transp. Syst.* 0, 0. <https://doi.org/10.1080/15472450.2014.936284>
- Axhausen, K.W., 2008. Social networks, mobility biographies, and travel: survey challenges. *Environ. Plann. B. Plann. Des.* 35, 981.
- Boyd, D.M., Ellison, N.B., 2007. Social network sites: Definition, history, and scholarship. *J. Comput. Commun.* 13, 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Chaniotakis, E., Antoniou, C., Aifadopoulou, G., Dimitriou, L., 2017a. Inferring Activities from Social Media Data. *Transp. Res. Rec. J. Transp. Res. Board* 2666, 29–37.

<https://doi.org/10.3141/2666-04>

Chaniotakis, E., Antoniou, C., Goulias, K., 2017b. Transferability and Sample Specification for Social Media data: a Comparative Analysis, in: International Conference on Intelligent Transport Systems in Theory and Practice, Mobil.TUM. Munich, Germany.

Chaniotakis, E., Antoniou, C., Grau, J.M.S., Dimitriou, L., 2016a. Can social media data augment travel demand survey data?, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. <https://doi.org/10.1109/ITSC.2016.7795778>

Chaniotakis, E., Antoniou, C., Pereira, F.C., 2016. Mapping Social Media for Transportation Studies. IEEE Intell. Syst. <https://doi.org/10.1109/MIS.2016.98>

Chen, Y., Mahmassani, H.S., Frei, A., 2017. Incorporating social media in travel and activity choice models: conceptual framework and exploratory analysis. Int. J. Urban Sci. 1–21. <https://doi.org/10.1080/12265934.2017.1331749>

Cheng, Z., Caverlee, J., Lee, K., Sui, D.Z., 2011. Exploring Millions of Footprints in Location Sharing Services. ICWSM 2011, 81–88.

Cho, E., Myers, S.A., Leskovec, J., 2011. Friendship and Mobility: User Movement in Location-based Social Networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11. ACM, New York, NY, USA, pp. 1082–1090. <https://doi.org/10.1145/2020408.2020579>

Cho, Y.-S., Ver Steeg, G., Galstyan, A., 2014. Where and Why Users "Check In".

Cottrill, C., P. Gault, G. Yeboah, J. D. Nelson, J. Anable and T. Budd (2017). Tweeting Transit: An examination of social media strategies for transport information management during a large event. Transportation Research Part C: Emerging Technologies, Volume 77, April 2017, Pages 421-432, <https://doi.org/10.1016/j.trc.2017.02.008>

Gal-Tzur, A., Grant-Muller, S.M., Minkov, E., Nocera, S., 2014. The Impact of Social Media Usage on Transport Policy: Issues, Challenges and Recommendations. Procedia - Soc. Behav. Sci. 111, 937–946. <https://doi.org/10.1016/j.sbspro.2014.01.128>

Greenfield, P., 2018. The Cambridge Analytica files: the story so far [WWW Document].

Gu, Y., Qian, Z. (Sean), Chen, F., 2016. From Twitter to detector: Real-time traffic incident detection using social media data. Transp. Res. Part C Emerg. Technol. 67, 321–342. <https://doi.org/http://dx.doi.org/10.1016/j.trc.2016.02.011>

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2013. Geo-

- located Twitter as the proxy for global mobility patterns. *CoRR*.
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C., 2014. Big data and its technical challenges. *Commun. ACM* 57, 86–94.
- Jiang, B., Miao, Y., 2015. The evolution of natural cities from the perspective of location-based social media. *Prof. Geogr.* 67, 295–306.
- Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Bus. Horiz.* 53, 59–68. <https://doi.org/http://dx.doi.org/10.1016/j.bushor.2009.09.003>
- Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horiz.* 54, 241–251. <https://doi.org/10.1016/j.bushor.2011.01.005>
- Kuflik, T, Minkov, E, Nocera, S., S. Grant-Muller, A. Gal-Tzur and I. Shoor (2017) Automating a framework to extract and analyse transport related social media content: The potential and the challenges. *Transportation Research Part C: Emerging Technologies*, 77. pp. 275-291. ISSN 0968-090X, <https://doi.org/10.1016/j.trc.2017.02.003>
- Kumar, A., Jiang, M., Fang, Y., 2014. Where not to go?: detecting road hazards using twitter. *Proc. 37th Int. ACM ...* 2609550, 1223–1226. <https://doi.org/10.1145/2600428.2609550>
- Lee, J.H., Davis, A.W., Goulias, K.G., 2016a. Activity Space Estimation with Longitudinal Observations of Social Media Data, in: *Transportation Research Board 95th Annual Meeting*. Washington, DC.
- Lee, J.H., Gao, S., Goulias, K.G., 2016b. Comparing the Origin-Destination Matrices from Travel Demand Model and Social Media Data, in: *Transportation Research Board 95th Annual Meeting*.
- Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., Beaton, M., 2013. Teens, social media, and privacy. *Pew Res. Cent.* 21, 2–86.
- Maheswaran, M., Tang, H.C., Ghunaim, A., 2007. Towards a gravity-based trust model for social networking systems, in: *Distributed Computing Systems Workshops, 2007. ICDCSW'07. 27th International Conference On. IEEE*, p. 24.
- Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C., 2011. *TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration*. *Proc. 2011 Annu. Conf. Hum. Factors Comput. Syst.* 227–236. <https://doi.org/10.1145/1978942.1978975>

- Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T.S., 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transp. Res. Part C Emerg. Technol.* 75, 197–211.
- Sánchez Abril, P., Levin, A., Del Riego, A., 2012. Blurred boundaries: Social media privacy and the twenty-first-century employee. *Am. Bus. Law J.* 49, 63–124.
- Smith, M., Szongott, C., Henne, B., Von Voigt, G., 2012. Big data privacy issues in public social media, in: *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference On. IEEE*, pp. 1–6.
- Steiger, E., Ellersiek, T. and Zipf, A., 2014, November. Explorative public transport flow analysis from uncertain social media data. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 1-7). ACM.
- Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M., 2013. Who, where, when and what: discover spatio-temporal topics for twitter users, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM*, pp. 605–613.
- Zheleva, E., Getoor, L., 2009. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles, in: *Proceedings of the 18th International Conference on World Wide Web, WWW '09. ACM, New York, NY, USA*, pp. 531–540. <https://doi.org/10.1145/1526709.1526781>