XII Conference on Transport Engineering, CIT 2016, 7-9 June 2016, Valencia, Spain

Social Networks, Big Data and Transport Planning

Tomás Ruiz[a*], Lidón Mars[b], Rosa Arroyo[a] Ainhoa Serna[c]

[a]*Transport Department, School of Civil Egineering, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain*
[b]*Social psychology Department, Faculty of Psychology, Universitat de València, , Spain*
[c]*Faculty of Engineering, Information Systems Department, Mondragon Unibertsitatea, Loramendi, 4, Arrasate, 20500, Spain*

Abstract

The characteristics of people who are related or tied to each individual affects her activity-travel behavior. That influence is especially associated to social and recreational activities, which are increasingly important. Collecting high quality data from those social networks is very difficult using traditional travel surveys, because respondents are asked about their general social life, which is most demanding to remember that specific facts.

On the other hand, currently there are different potential sources of transport data, which is characterized by the huge amount of information available, the velocity with it is obtained and the variety of format in which is presented. This sort of information is commonly known as Big Data. To use this data on Transport Planning application is a challenge, which require employing complex data mining techniques.

In this paper, we identify potential sources of social network related big data that can be used in Transport Planning, discussing their advantages and limitations. Then, a review of current applications in Transport Planning is presented. Finally, some future prospects of using social network related big data that are included in the MINERVA project are highlighted.

Keywords: Social Networks; Big Data; Transport Planning; MINERVA project

1. Introduction

The prediction of individual's travel behavior is an essential component of transport planning and policy analysis. The ultimate mission of travel demand forecasting research is to develop the capability to predict how individuals respond to changes in their travel environment (Kitamura, 1988). The activity-based approach to study travel behavior recognizes the fact that travel demand is derived from the activities that have to be performed in different places (Hensher and Stopher, 1979; Heidemann, 1988). The main focus is on decisions for each activity of type, here, when, for how long and with whom to participate (Axhausen and Gärling, 1992).

Non-compulsory activities, in particular leisure activities, are increasing in importance. Leisure activities in the area of Valencia represented 6.3% in 1991, and 9.3% in 2009 (COPUT, 1994; Generalitat Valenciana, 2010). Other non-recurrent travels represented 14.4% in 1991 and 32.0% in 2009. Leisure travel is more complex and variable that work-related travel, being less rigid in temporal and spatial patterns and more influenced by external factors such as social contacts or weather conditions (Kowald and Axhausen, 2015). Social interactions are an important explicative variable of leisure travel. Therefore, there is a need to adequately characterize social interactions through Social Networks if we want to study better leisure and social travel.

A Social Network is defined as a set of persons who are linked pairwise, so that each person can reach any other through an active tie (Axhausen, 2006). It is argued that the geography of the members of the travelers' social networks together with the geography of the relevant activity locations determines the amount and style of travel. This is especially true for leisure and social activities. It is important to mention that Social Networks are commonly related to on-line tools that facilitates people sharing information. In this paper, we adopt the first definition for Social Network.

Social network data is being collected mainly in person, using ego approaches and combining qualitative or quantitative techniques. They are long surveys (two hours long), and the respondents are asked about the characteristics of their social networks (alter's demographics, spatial location, frequency of interactions, etc.) (Ohnmacht, 2015; Carrasco et al., 2008; Van den Berg et al., 2009; van den Berg and Timmermans, 2015) and information on habitual mobility, preferences, biographical questions (Axhausen and Frei, 2007). The amount and characteristics of Social network data requires new sources of information that could complement survey data.

2. Social Network Big Data

Big Data represents the information assets characterized by such a High Volume, Velocity and Variety (the three Vs) to require specific technology and analytical methods for its transformation into value (de Mauro et al., 2016). Big Data sources for Transport Planning applications include: GPS traces (smart phones, etc.), cell phones traces, transit smart card transactions, and other count and sensor data (including on-board vehicle sensors), online Social Media/Networks, credit card transactions (travel-related activities), RFID traces (NFC…).

Besides the three Vs, Social Network related Big Data usually have the following characteristics: individual data, spatio-temporal data, socioeconomics and demographics, information shared, messy/disorganized data, multiparametric data, event-oriented data, type of social link/tie (depends on the data source).

Considering those characteristics, Social Network related big data could be inferred from Internet-based Social Media/Network tools. A Social Media is a system that disseminates information to others. Internet-based are currently the main social media data source. The terms on-line social media and social network sites are used in this domain interchangeably. Examples of online social networks are Twitter, Facebook, Linkedin, etc. Data from these and other similar online social networks have been used to characterized social interactions: Davis et al. (2011), Luo and Zhong (2015), Gilbert and Karahalios (2009), Aiello et al (2012).

The advantages of using Social Media information for travel behavior studies are: the ability to capture the whole trip; preserve elements of the associated context and/or the individual socio-characteristics; bring in qualitative data on large scale; cost effectiveness. On the other hand, they have some limitations: lack of accompanying context in the data; informal nature of textual content; scarce location related information; need of periodic monitoring and enhancement of the system.

Characteristics of social network Big Data can also be inferred from wi-fi connections (Sapiezynski et al., 2015), and cell phones traces like Call Detail Records (CDR): Eagle et al. (2009), Zhang and Dantu (2010), Calabrese et al. (2011), Cho et al. (2011), Phithakkitnukoon et al. (2012), Chen and Mei (2014), Toole et al. (2015), Picornell et al. (2015).

3. Current Applications on Transport Planning

Current applications of Social Network big data in Transport Planning are still very scarce. We found in the literature references related to improving sustainable mobility, enhancing road safety, promoting public involvement, and facilitating management and operation. Only one work is related to characterize (air) travel behavior.

Rovniak et al. (2013) designed and evaluated of a social network intervention for promoting physical activity. Bartle et al. (2013) explored, using a qualitative approach, the social processes that occurred when a group of 23 commuter cyclists interacted with one another through a specially designed, map-based website over six weeks, sharing their routes and other cycling-related information.

Fire et al (2012) used a collaborative community –based navigation application to improve road safety by informing police departments on locations that need more attention. Elevant (2013) combined a personalized service of weather information to drivers and the introduction of a social network component, including user-generated local weather observations. Both methods used together proved efficient as the amount of changed decisions was of considerable amplitude. A correlation between time of exposure and changed decisions implies that social components and interactivity may be a powerful tool in traffic weather services and ITS. Luan et al (2015) developed a vehicular social network to enable social communications and interactions among users on the road during their highway travels.

Nash (2012) describes an Internet application designed to test the use of ITS techniques in improving transport planning. It consists of a game, best practices library and social network. The game and best practices are designed to educate residents, while the social network allows them to suggest ideas for improving transport. The application was developed and tested in Vienna during 2011. The application was successful technically, but failed to attract a critical mass of users. Bakht and El-Diraby studied the underlying social networks formed in the background of public relations processes on online social media public involvement during planning of four light rail transit projects in various cities in the United States and Canada. Such networks are referred to as infrastructure discussion networks (IDN) and can be rich sources of information for decision makers.

Ai et al. (2012) proposed a novel social network architecture for rail traffic transportation system is proposed, including the SNS Management Center, a dedicated social network for internal information exchanges among the drivers, the dispatchers, and the train control center, the SNS for the information exchanges among the train marshals, crews, and passengers. The proposed framework attempts to exploit social networks in an active manner in the field of rail traffic transportation systems to make the system safer, more secure, reliable, comfortable, and humane.

Wall et al. (2014) explored the use of online social network data in characterizing individuals' air travel behavior. Data were collected through a web-based survey including information about individuals' air travel history and online social network information, specifically participants' Facebook networks. They found a positive relationship between the size and distribution of individuals' Facebook social networks and both their engagement in air travel and the odds that their air travel will be leisure related or will include a leisure component.

4. Future Prospects: the MINERVA Project

As described earlier, current applications of data on social interactions collected from online social networks or mobile phones on transport planning are very scarce. We describe below an ongoing project that aims to contribute on the improvement of travel demand forecasting methods.

Recently, the Minister for Economy and Competiveness of Spain has funded MINERVA, a research project to enhance data collection methods for travel demand forecasting. The aim of MINERVA is to evaluate the usefulness of data from on-line social media and mobile phones to obtain information related to activities and trips, perceptions and attitudes related to travel, and social network interactions.
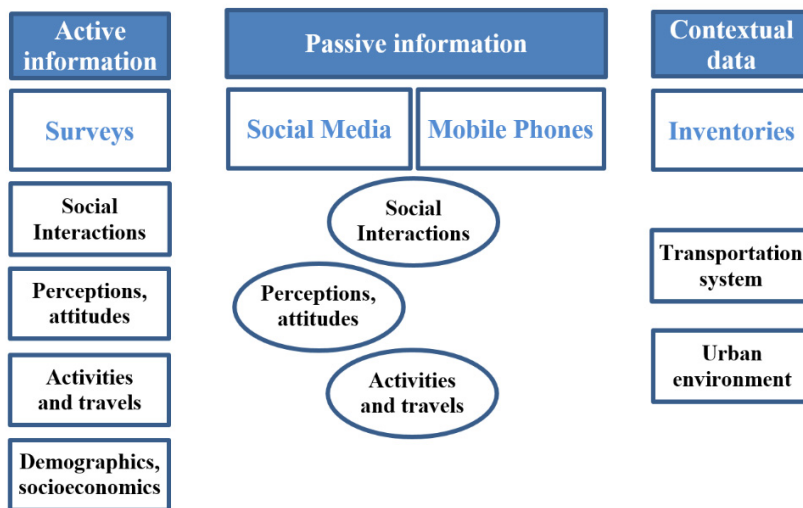

FIGURE 4.1 MINERVA framework for data collection of travel demand

It is planned to collect passive information (without the intervention of individuals) to inferred characteristics of social network interactions and activities and travels from Social Media and mobile phones. Perceptions and attitudes related to travel will be only possible to be collected from Social Media. The estimation and calibration of algorithms to extract that passive information requires a secondary source of information, which will be collected through surveys. Additionally, contextual data related to characteristics of the transportation system and urban environment will be collected.

The development of MINERVA includes some challenges concerning the extraction of valid information from Social Media and mobile phones, estimation and validation of Big Data algorithms, and development of data fusion techniques to take advantages of the benefits of each type of information.

4.1. Data extraction from Social Media

Qualitative and quantitative content analysis methods will be used to get data on perceived sentiments on travel modes from online social networks. The methodology to be used consists on four main phases: data acquisition, data preparation for analysis, data curation and storage (Serna et al, 2016).

4.2. Data collection from mobile phones

Smartphones have multiple sensors than can be utilized to infer characteristics of the activities and travels performed by their users: GPS, accelerometer, orientation, gyroscope, etc. Additionally, Call Detail Records (CDR) are being used to extract information to infer characteristics of trips and activities and of social networks interactions (González et al., 2008; Picornell et al., 2015, Toole, 2015).

4.3. Estimation and calibration of data mining algorithms

During the process of data extraction from Social Media, several program applications interfaces (API's) and algorithms will be used. Tasks to be carried out include: gathering, filtering and cleaning the unstructured data; correcting texts using spell checkers customized with localisms and abbreviations; analyzing morphosyntactically all words; identifying categories and carrying out a sentiment analysis process; and storing and managing data (Serna et al, 2016).

The information to be collected from mobile phones require appropriate algorithms to eliminate noise data and identify valid information related to timing, duration and travel mode of trips, and type and location of activities (Ferrer and Ruiz, 2014). Similarly to infer social ties and their characteristics.

4.4. Data fusion techniques

The use of Social Media and mobile phones has increased the amount of data available. But that data often does not contain enough information to carry out travel behaviour studies. Similarly, data from travel surveys has deficits of sample size, lack of representation, etc. Integration and data fusion techniques allow enriching a database to meet the study requirements.

Data from Social Media, mobile phones and surveys will be fused at two levels (Varshney, 1997). Level 1 fusion is oriented towards processing the raw data and the estimation of travel indicators. The objective of Level 2 fusion is to derivate distinctive characteristics and behavioural patterns from the estimation of the previous level.

5. Conclusions

This paper presents a review of the use of data from Social Media and mobile phones to infer social interactions characteristics, and how those results are being used in the field of Transport Planning. The aim of MINERVA project is to contribute to enhance the Transport Planning process through improving data collection methods for travel demand forecasting. MINERVA's aim is to complete travel survey data with Social Network and travel behaviour information inferred from the use of Social Media and mobile phones.

References

Ai, B., Cheng, X., Yang, L., Zhong, Z., Ding, J-W., and Song, H. (2012). Social Network Services for Rail Traffic Applications. Intelligent Systems, 29 (6), 63-69.

Aiello et al. (2012) Friendship prediction and homophily in Social Media. ACM Transactions of the Web, 6, 2:9.

Axhausen, K.W. (2006). Social factors in future travel: A qualitative assessment. IEE Proceedings Intelligent Transport Systems, 153 (2): 156–166.

Axhausen, K.W. and Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models and research problems. Transport Reviews, 12, 4, 323-341.

Axhausen, K. W. and Frei, A. (2007). Size and structure of social network geographies, Arbeitsbericht Verkehrs und Raumplanung, No. 439 (Zurich: Institute for Transport Planning and Systems, ETH Zurich).

Bakht, M. and El-Diraby, T. (2014). Hidden Social Networks That Drive Online Public Involvement in Infrastructure Construction. Case Study of Light Rail Transit Projects in North America. Transportation Research Record: Journal of the Transportation Research Board, 2453, 84-93.

Bartle, C. Avineri, E. and Chatterjee, K. (2013). Online information-sharing: A qualitative analysis of community, trust and social influence amongst commuter cyclists in the UK. Transportation Research Part F: Traffic Psychology and Behavior, 16, 60-72.

Calabrese F, Smoreda Z, Blondel VD, Ratti C (2011) Interplay between Telecommunications and Face-to-Face Interactions: A Study Using Mobile Phone. Data. PLoS ONE 6(7): e20814.

Carrasco, J.A., Hogan, B., Wellman, B. and Miller, E.J. (2008). Collecting social network data to study social activity-travel behavior: an egocentric approach. Environment and Planning B: Planning and Design, 35 (6), 961-980.

Chen, C. and Mei, Y. (2014). Does distance still matter in facilitating social ties? The roles of mobility patterns and the built environment. TRB 2014 Annual Meeting Compendium.

Cho, E., Myers, S.A. and Leskovec, J. (2011). Friendship and Mobility: User Movement In Location-Based Social Networks. KDD 2011, August 21-24, San Diego, California, USA.

Conselleria d'Obres Públiques, Urbanisme i Transports (1992). El transporte metropolitano en el área de Valencia: encuesta de movilidad metropolitana, 1991.

De Mauro, A., Greco, M., Grimaldi, M. (2016). "A Formal definition of Big Data based on its essential Features". Library Review, 65, 122–135.

Eagle et al. (2009) Inferring friendship network structure by using mobile phone data. Pnas, 106 (36), 15274-15278.

Elevant, K. (2013). Trust-networks for changing driver behaviour during severe weather. IET Intelligent Transport Systems, 7 (4), 415-424.

Ferrer, S. and Ruiz, T. (2014). Using Smartphones to Capture Personal Travel Behavior. In Mobile Technologies for Activity-Travel Data Collection and Analysis (Rasouli, S. and Timmermans, H., Eds.), 171-186.

Fire, M., Kagan, D., Puzis, R., Rokach, L. and Eolivici, Y. (2012). Data mining opportunities in geosocial networks for improving road safety. 2012 IEEE 27th Convention, 14-17 November, 1-4.

Generalitat Valenciana (2010). Estudio de movilidad área metropolitana de València.

Gilbert, E. and Karahalios, K. (2009) Predicting tie strength with Social Media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 211-220.

González, M.C., Hidalgo, C.A., Barabasi, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453, 779-782.

Heidemann, C. (1988). Regional planning methodology: the first and only annotated picture primer on regional planning. Discussion paper 16, Intitut für Regionalwissenschaft, Universität Karlsruhe, Karlsruhe.

Hensher, D.A. and Stopher, P.R. (Editors) (1979). Behavioral travel demand modelling. London: Croom Helm.

Kitamura, R. (1988). An evaluation of activity-based travel analysis. Transportation, 15, 9-34.

Luan, T., Lu, R., Shen, X. and Bai, F. (2015). Social on the road: enabling secure and efficient social networking on highways. Wireless Communications, 22 (1), 44-51.

Nash, A. (2012). GreenCityStreets.com – Using ITS to improve transport planning. 19th ITS World Congress, Vienna.

Phithakkitnukoon S., Smoreda Z., Olivier P. (2012) Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. PLoS ONE 7(6): e39253.

Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J. J., Dubernet, T., & Frías-Martínez, E. (2015). Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. Transportation, 42 (4), 647-668.

Rovniak et al. (2013). Engineering online and in-person social networks to sustain physical activity: application of a conceptual model. BMC Public Health, 13: 753.

Sapiezynski, P., Wind, D.K., Stopczynski, A. and Lehmann, S. (2015). Inferring social ties from WiFi scan results. Proceedings of NetMob 2015, 7-10 April, Madrid.

Serna, A., Gerrikagoitia, J.K. and Bernabé, U. (2016). Discovery and classification of the underlying emotions in the user generated content (UGC). Information and Communication Technologies in Tourism, Proceedings of the International Conference in Bilbao, Spain, February 2-5, 225-237.

Serna, A., Gerrikagoitia, J.K. Bernabé, U. and Ruiz, T. (forthcoming) Sustainability analysis on urban mobility based on Social Media content. Transportation Research Procedia.

Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., & González, M. C. (2015). Coupling human mobility and social ties. Journal of The Royal Society Interface, 12 (105), 20141128.

Van den Berg, P.E.W., T.A. Arentze and H.J.P. Timmermans (2009) Size and composition of ego-centered social networks and their effect on travel distance and contact frequency.

Transportation Research Record: Journal of the Transportation Research Board, 2135, 1-9.

Van den Berg, P. and Timmermans, H. (2015). A multi-level path analysis of social networks and social interaction in the neighborhood. Region, 2 (1), 55-66.

Varshney, P.K. (1997). Multisensor data fusion. Electronics and Communication Engineering Journal, 245-253.

Wall, T. A., Macfarlane, G. S., Watkins, K. E. (2014). Exploring the Use of Egocentric Online Social Network Data to Characterize Individual Air Travel Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2400, 78-86.

Zhang, H. and Dantu, Ram (2010) Predicting Social Ties in Mobile Phone Network. IEEE, 25-30.