SOCIAL MEDIA AND TRANSPORT CHOICES: HOW SOCIAL MEDIA CAN AFFECT TRIPMAKERS' CHOICES

ANASTASIA NIKOLAIDOU & PANAGIOTIS PAPAIOANNOU
Laboratory of Transportation Engineering, Division of Transportation and Project Management,
Department of Civil Engineering, Aristotle University of Thessaloniki, Greece

ABSTRACT

During the last years, the evolution of the World Wide Web from static websites to the Web 2.0, a platform where content and applications are continuously modified by all users, has created new online applications, known as "Social Media", through which users interact with one another, share their experiences, opinions, knowledge, and sometimes their locations, an aspect quite useful for the transport sector. Many attempts have been made recently to implement social media analysis in the domain of transportation and extract useful information from social media. However, despite the efforts made so far, it has not been examined to a large extent what the characteristics of social media users are and how the existence of information on various topics in such platforms can affect the transport sector. In the framework of the present work, an attempt is made to investigate the likely influence of social networking services on the behavior and choices of commuters, through a questionnaire survey composed in the Greek language. This research aims to identify, at first, information on how often and in which way social media are being used and, secondly, to examine whether the existence of real-time information on traffic conditions from social networks or through interaction with other users of such networks can affect the way users travel and make their choices. The preliminary results indicate that almost 64% of the survey's participants receive information from social media about their trip, especially when they use public transport. They also reveal that this information may alter, in a substantial percentage, their route or the period they travel. Understanding the influence of social media on the behavior and choices of commuters can be a valuable tool for mobility managers to alter the trip characteristics for various activities such as work, but mainly trips for entertainment, shopping, etc.
Keywords: social media, questionnaire survey, travel behaviour, behavioural characteristics.

1 INTRODUCTION

The term "social media" is used to describe a range of web applications that allow uses to interact with each other and encourage them to share experiences, opinions, knowledge and sometimes their location [1]. Additionally, through social media sites users can create, modify and publish on the internet their own user-generated content [2]. To this end, social media contain a large amount of information and data produced by users with different interests and socio-economic characteristics such as gender, age, occupation, etc. [3] Social media are therefore an important potential tool for the promotion of research in various fields and researchers are able, through social media data mining, to study and analyse human attitudes and behaviour.

In the domain of transportation, mining and analysing data derived from social media, and in particular from users' publications, should allow the analysis of issues that were difficult to be identified with precision so far. Such issues can be understanding of mobility and activity patterns [4], [5], measuring rider satisfaction [6], [7], event and incident detection related to system disruptions [8], [9], etc.

To exploit information from social media, in addition to the advanced data and text mining required techniques, it is necessary to examine and explore further the sample from which this information is derived in an attempt to measure and correct any sampling error that may

WIT Transactions on Ecology and The Environment, Vol 226, © 2017 WIT Press
www.witpress.com, ISSN 1746-448X (on-line)
doi:10.2495/SDP170531

exist. Sampling errors may be due to the lack of representativeness of the total population as most of social media users are younger and more familiar with technology and thus a selective bias may evolve. Furthermore, errors may be present because many people publish fake or misleading information and at the same time they do not to reveal their real identity in social media.

In the context of this paper, an attempt is made to identify, through a questionnaire survey, the profile of social media users. This profile, except from socio-economic characteristics, is shaped from information regarding the use of social media both for general and transport purposes and information about the effect social media information may have on travel behavior and commuters' choices. Finally, based on the results of the questionnaire survey, logistic regression models are constructed to study and analyze the possibility of the trip makers to use or post transport-related information in social media, based on their personal characteristics.

The paper is structured as follows: After this introductory section, the next section is devoted to the description of the undertaken research. Section three presents the most important results of the statistical analysis of the questionnaire survey. The development of binary models is described in section four whereas the model results are given in section five. Finally, the conclusions drawn are given in the last section of the paper.

2 METHODOLOGICAL APPROACH

For the purposes of this research, i.e. to investigate the influence of social media on travel behaviour and choices made by the commuters, a questionnaire survey was designed and carried out. The aim of the specific questionnaire has been to collect data regarding the use of social media sites and information about how these sites are used in matters related to transportation. Furthermore, the questionnaire attempts to investigate the degree of impact that real-time information, from different sources, about traffic or other issues can have on commuters' choices.

2.1 Organization and implementation of questionnaire survey

The first step, following the design of the questionnaire survey, was a pilot survey, through personal interviews on a small sample, with the aim to identify potential weaknesses and gaps. The pilot phase was successfully conducted prior to the implementation of the full survey. A commercial web platform (SurveyMonkey) was used for the implementation of both the pilot phase and the full the questionnaire survey. This platform relies on a specialized software for designing questionnaires and simplifies the data collection process, as participants' responses are automatically stored in an online database.

The questionnaire was available only in the Greek language and since there was no clearly defined study area, it was addressed to all those who speak Greek. The only limitation in the survey sample is that it was addressed only to internet users. The questionnaire collection period was approximately 3 months, from the end of December 2016 till March 2017, while the promotion of the survey was carried out through various channels such as: a) posts on social media sites (Facebook, Twitter, LinkedIn), b) by e-mail to all members of Aristotle University of Thessaloniki (students, graduates and employees) and c) publications in internet magazines and in newsletters of associations (e.g. Hellenic Institute of Transportation Engineers).

Initially, the minimum threshold for correctly-completed questionnaires was set to 300. However, a total number of 648 valid questionnaires was finally reached.

2.2 The questionnaire survey

The questionnaire form consists of 5 parts, each of which plays a specific role in this research effort. The five parts of the questionnaire are briefly described below.

In the first part, there is an introductory text that informs participants about the research purposes and, the privacy policy; it also provides contact details for communication purposes.

The second part aims at collecting information regarding the use of social media in general, such as the extent to which different social media sites are used, the purposes they are mainly used for, the activities performed by the users and the information users publish about their profile. There are also additional questions about the use of internet in general, and about their familiarity level with technology. Participants who do not use any social media site were not obliged to answer the above questions and instead they answered a set of questions about their decision not to use these sites.

The fourth section of the questionnaire was designed to investigate whether and how the information that a user gets from social media about transportation can influence his/her choices. To investigate the influence of social media, there are at the beginning of the section some questions about the mode of transport commonly used by each user as well as the usual trip purpose and frequency of travel with that mode. In the end, there were questions aiming at exploring how social media can influence commuters, with the help of Likert scale.

The fifth and final part of the questionnaire includes questions related to the socio-economic characteristics of the sample.

The full questionnaire can be found online by using the following link: https://www.surveymonkey.com/r/H5887DB.

3 DESCRIPTIVE STATISTICS OF THE SAMPLE

This section presents the main descriptive statistics and the data analysis derived from the questionnaire survey. The summary of the most important results of the analysis is presented along the questionnaire structure as presented above.

3.1 Social media use

From all the respondents in the sample, 85% use a social media site and 74% of them have an active account in at least one of these media for time periods longer than four years. Most of the respondents use social media to communicate with other people but also for collecting information. When using these sites, most users usually perform activities related to the participation in groups of common interests, monitoring and commenting on interesting content posted by others (articles, images, videos).

Participants use Facebook, Instagram, Twitter and LinkedIn more often compared to other social media sites. 46% of the sample uses the popular social platform of Facebook for more than 2 hours a day. Despite the systematic use of the above sites, most of the sample subjects state that they do not rely heavily on the information that they receive through them.

In addition to using social media, participants are quite familiar with the internet as they use it daily, and about 50% of the sample is connected online for more than 2 hours daily for work or other purposes. Most of the respondents in the sample also uses to a significant extent their mobile phone for getting real-time information through the internet.

3.2 Social media use for transportation purposes

Regarding the use of real-time information/updates from social media for transportation, the survey findings indicate that 64% of the sample receives such information. For those who do not belong into this percentage, 62% state that they did not know that such information could

be provided; on the other hand, those who were aware of the existence of such information stated that they were not interested in this information and that they consider it inadequate. Information received by commuters from social media sites mainly derives from the official profiles maintained by public transport operators such as OASA (Athens Urban Transport Organisation) and OASTH (Organization of Urban Transportation of Thessaloniki).

Pedestrians and public transport users use at a great extent information from social media for transportation, in comparison with users of other transport modes, both before and during their journey. Moreover, most respondents are looking more for transport-related information before a trip aiming to better plan it.

Participants receive mainly information on topics related to weather and traffic conditions as well as to public transport (route delays, line modifications, systems disruptions). The dominant transport modes of these respondents are public transport and walking.

In addition to receiving information from social media for transportation, users are also able to post information about their trips. Some 74% of the sample never publish such information as they do not consider it useful and is a waste of time, while from the remaining 26% of the participants, only 19% publish information about this subject a few times a month. The main reasons users publish information about their trip are to inform other users, express their dissatisfaction with an issue related to transportation and to share their opinion and experiences with other users. Respondents publish more often publish information regarding their location by using the service "check-in", which is available in many social media sites, and information about public transport.

Based on the survey findings, it appears that most users are reluctant to publish information about their journey on social media. This may be because in Greece people are not quite familiar with such issues; reluctance may also be due to concerns about privacy issues, as most users link this type of information with check-in and location information.

3.3 Effect of social media information on transportation

The purpose of this section is to determine whether information that a user can get from social media sites about transportation related matters, may affect travel behavior and choices made by the commuters. Based on the findings of the questionnaire survey, users rely more about trip information on the official sites of public transport operators, on existing applications for transportation (such as Google maps) and on comments and posts from users that they personally know.

Concerning the possible influence of social media to travel behavior, the majority of the respondents consider that the existence of trip information on social media sites may affect their choices. Information from social media can possibly change the route chosen by commuters and may change the departure time of a trip; moreover, the availability of information from official sites of public transport operators can affect users in choosing this particular mode of transport.

To a smaller extent, users feel that having information about transportation from social media can make them change the preferable mode of transport. This may be related to the fact that a large part of the sample usually uses for their trips car and public transport. In both cases, users are usually "captive" of their choices, as they choose car for comfort and minimum travel time and public transport mainly for economic reasons and due to lack of alternatives.

3.4 Socio-economic characteristics of the sample

Regarding the socio-economic characteristics of the sample, 60% of the respondents are women and most of them are between 18 and 35 years old; this was something expected as

the questionnaire survey was carried out through the internet, in which younger people have better and more frequent access. From the total sample, 43% hold a postgraduate/PhD diploma; this can be attributed to the fact that the survey was promoted via e-mail to all members of the Aristotle University of Thessaloniki. For the same reason, almost 70% of the sample are student or employees. Regarding monthly income, most respondents stated incomes in the range of 800-2400€, whereas 72% of them have a car or motorbike license. Finally, 92% have a smart phone showing that most persons in the sample are quite familiar with technology.

3.5 Inferential statistics and hypothesis testing

The main objective of a statistical analysis is to investigate a phenomenon based on sample data and extract conclusions from the sample to the population under study. This type of investigation and analysis which is known as inferential statistics is presented in this section [10].

A series of appropriate statistical tests were carried out between two variables which represent the first the probability to use transport-related information from social media and the second the probability to post such information on social media. These two represent the dependent variables of the models that will be created subsequently, and all the other variables, constitute the independent (explanatory) variables of the model. Table 1 presents the dependent and independent variables that were examined.

Table 1: Encoding and description of examined variables.

Category	Name	Description
Dependent variables	Social media use for transport	Probability to use transport-related information from social media
	Social media post for transport	Probability to post transport-related information on social media
Independent variables	Sex	Gender of the respondents
	Age	Age of the respondents
	Occupation	Occupation of the respondents
	Education	Education level of respondents
	Income	Monthly family income of respondents
	Car license	Possession of car or motorbike license
	Smartphone own	Possession of smart phone
	Mode	Dominant mean of transport used
	Trip purpose	The purpose for which a trip is usually made with the selected mean of transport
	Trip frequency	How often a trip is made with the selected mean of transport
	Information from phone	The extent to which respondents receive real-time information from the internet through their phone
	Social media trust	The extent to which respondents trust the information they receive from social media
	Utility of social media transport information	The extent to which respondents consider useful the information they receive about transportation from social media
	Facebook use	Degree of use of Facebook
	Twitter use	Degree of use of Twitter

The statistical tests conducted examined the hypotheses that a change in the independent variable can affect the dependent variables and have a significant effect on it. Based on the findings regarding the correlation between dependent and independent variables Table 2 presents both the variables the change of which have an influence on the dependent variable and also the specific categories of these variables which relate to specific characteristics of users who use information from social media for transportation purposes or post transport-related information accordingly more often. The independent variables are divided into 3 categories: socio-economic characteristics, social media/internet use and trip characteristics.

4 DEVELOPMENT OF BINARY MODELS

For the purposes of the present research, a series of binary logistic logit models were developed to examine the intention (in terms of probability) of respondents to use real-time information from social media for transportation purposes or post transport-related information on social media.

Logistic regression is a technique designed to perform data analysis concerning the study and forecasting of values of a categorical dependent variable by using both quantitative and qualitative independent variables. In addition to predicting, a logistic regression model can estimate the effect of each independent variable on the configuration of the dependent variable values [11].

More specifically, 2 sets of binary models were formed, one for each of the examined dependent variables. In the statistical analysis carried out for the development of the models, all independent variables presented in detail in previous section were examined. To examine how and at what extent specific characteristics of individuals improve the explanatory power

Table 2: Overview of variables having a significant effect on the dependent variables.

	Social media use of transport	**Social media post of transport**
Socio-economic characteristics		
Sex	Women	-
Age	18-35	18-35
Occupation	Students and employees	Students and employees
Education	High school and MSc/PhD	University and Msc/PhD
Income	800-1600€	800-1600€
Car license	Car/motorbike license holder	-
Smartphone own	Smartphone owner	Smartphone owner
Social media/internet use		
Information from phone	Great extent	Great extent
Social media trust	Moderate degree	Moderate degree
Utility of social media transport information	Great extent	Great extent
Facebook use	More than 2 hours per day	More than 2 hours per day
Twitter use	Sometimes per month	Sometimes per month
Trip characteristics		
Mode	Public transport and car users	Public transport and car users
Trip purpose	From and to work/school	From and to work/school
Trip frequency	More than once per day	More than once per day

of the models, three binary models were set up for each dependent variable. The first model (Model 1) included only variables related to social media / internet use like social media trust and Facebook use, where the second model (Model 2) included additional variables that are used to describe trip characteristics, such as mode and trip purpose. Finally, the third and last model (Model 3) included in addition to those of Model 2, variables of socio-economic nature such as age, gender, income etc.

The models were calibrated using SPSS software and the parameters where estimated through the Maximum Likelihood Estimation (MLE) method [12].

MLE approach assumes that a given sample could be generated by different populations and is more likely to come from one population rather than another. The likelihood function L* for N observation and K parameters can be defined as:

$$L^* (\beta_1, \beta_2, \dots \dots \dots \beta_\kappa) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}} , \quad (1)$$

and the logarithm of L*, defined as L, will be written as:

$$L (\beta_1, \beta_2, \dots \dots \dots \beta_\kappa) = \sum_{n=1}^{N} [y_{in} \log P_n(i) + y_{jn} \log P_n(j) \quad (2)$$

For the comparison of the three models constructed for each dependent variable, a Likelihood Ratio test was performed each time. The Likelihood Ratio test usually is used to measure the performance of one model relative to other. The mathematical formula for the two models is written as follows [12]:

$$-2\log L = - (2\log_{(model1)} - 2\log_{(model2)}) \quad (3)$$

The result of eqn (3) is compared against the critical tabulated Chi-Squared value at a specific level of confidence for given degrees of freedom. If the estimated L* value exceeds the critical chi-squared value, the null hypothesis is rejected, meaning that the extended model has a better model fit (e.g. the inclusion of the individual specific variables does significantly improve the model fit).

Ordinal and nominal variables of frequency, age, income, etc., were coded as dummy variables to be inserted into the binary models. For the selection of the final binary models an assessment of the goodness of fit of the model is required. The tests carried out for examining the statistical significance of the models were as follows:

- The “Nagelkerke R Square” index that gives an indication regarding the size of the sample variance eventually interpreted by the regression. Increase of the index’s value is an indication that each imported variable adds information to the equation. The closer to 1 the value of the index is, the better the model fits the sampling data.
- For the control of the goodness of fit of the sample data “Hosmer and Lemeshow” test is also used. Values of sig. > 0,05 at a confidential level of 95% indicate that the logistic regression model is well-adapted to the data,
- Another measure of evaluating the fit of a given logistic regression model is the “Classification Table”, which compares the observed probabilities with those observed by the model. The larger the percentage of dependent variables correctly predicted, the better the fit of the model is.

4.1 Model results

This section presents the results of the binary logit models that have been developed within the framework of the study for each of the examined dependent variables. The following Tables 3, and 4 illustrate the parameter estimates of the binary models for each dependent

variable respectively as well as they are presenting the results of the Likelihood Ratio tests for the comparison of the simplified Model 1 and the extended models (Model 2 and Model 3).

Based on the statistical tests' results for the dependent variable "social media use for transport", all models fit the sample data. The addition of variables related to trip characteristics are not of statistical significance whereas the addition of variables of socio-economic nature improves the adaptability of the model.

In the three models that were constructed for this dependent variable, it was found that for the two of them, the constant term was statistically significant and thus users have a prior preference to use transport-related information from social media, where in Model 3 this is not the case.

Variables related to social media/internet use are the variable related to posting information about transport on social media and four dummy variables for the degree that users get real-time information from their phone and the utility of transport-related information from social media. Obviously, people who post information regarding transportation are more likely to use such information from social media and people who get to a small degree real-time information from their phone are less likely to use social media for this purpose. The increase of the utility of social media transport-related information increases the possibility to get information from social media about transportation.

Table 3: Parameter estimates of the binary models for "social media use for transport".

Variables	**Model 1**		**Model 2**		**Model 3**	
	B	sig.	B	sig.	B	sig.
constant	0.509	0.000	0.509	0.000		
Social media/internet use						
Social media post for transport	1.235	0.000	1.235	0.000	1.256	0.000
Information from phone (small degree)	-0.811	0.005	-0.811	0.005	-0.884	0.002
Utility of social media transport information (none)	-1.693	0.001	-1.693	0.001	-1.594	0.002
Utility of social media transport information (small)	-1.141	0.001	-1.141	0.001	-1.008	0,002
Utility of social media transport information (medium)	-0.919	0.000	-0.919	0.000	-0.905	0.000
Trip characteristics						
Socio-economic characteristics						
Sex					0.540	0.001
Education (high school)					0.679	0.002
Nagelkerke R Square	0.178		0.178		0.229	
Hosmer and Lemeshow Test	0.974		0.974		0.136	
Classification	67.4		67.4		68.4	
-2 Log likelihoods	653.254		653.254		617.972	
-2logL	0			35.282		
LR test	significant difference at 95% confidential level			significant difference at 95% confidential level		

The variables related to the socio-economic characteristics of the users suggest that women and high school graduates are more likely to be informed from social media about transportation purposes than men and people of different educational level.

For the dependent variable, "social media post for transport" the statistical tests undertaken show better fir of Model 2 compared to the others. Moreover, the introduction in the model of variables related to trip characteristics of the sample improved the explanatory power of the simplified model whereas the introduction of socio-economic variables improved the simplified model too, but not Model 2.

In this case, it was found that the constant term of the models is statistically significant, meaning that users have a prior preference not to post transport-related information on social media.

Variables related to social media/internet use are "social media use for transport" and six dummies for the degree that users get real-time information from their phone and the degree that they use Facebook and Twitter, the two selected social media platforms for the analysis. As expected, people who tend to use transport information from social media are more likely to also post such information on these sites and people who get to a small degree real-time information from their phone are less likely to do so. Regarding the use of specific social media sites, people who do not use at all Facebook are less likely to post transport information on social media where an increase in the use of Twitter increases also the likelihood that users will take advantage of this possibility.

Table 4: Parameter estimates of the binary models for "social media post for transport".

Variables	**Model 1**		**Model 2**		**Model 3**	
	B	sig.	B	sig.	B	sig.
constant	-2.316	0.000	-2.230	0.000	-2.382	0.000
Social media/internet use						
Social media use for transport	1.005	0.000	0.913	0.002	1.005	0.000
Information from phone (great degree)	0.635	0.007	1.117	0.000	0.952	0.000
Facebook use (none)	-1.269	0.002	-1.332	0.009	-1.027	0.019
Twitter use (sometimes a month)	0.705	0.002	1.049	0.003	0.886	0.007
Twitter use (sometimes a week)	1.024	0.002	1.307	0.001	1.287	0.000
Twitter use (one hour per day)	1.455	0.003	1.577	0.003	1.596	0.003
Twitter use (more than 2 hours per day)	1.990	0.001				
Trip characteristics						
Mode (taxi)			2.592	0.039		
Trip purpose (from and to school)			-0.835	0.015		
Socio-economic characteristics						
Education (University graduate)					0.620	0.013
Income (1600-2400€)					-0.833	0.007
Nagelkerke R Square	0.223		0.260		0.235	
Hosmer and Lemeshow Test	0.168		0.240		0.825	
Classification	80.9		80.9		79.1	
-2 Log likelihoods	546.054		375.574		451.097	
-2logL	170.48		-75.523			
LR test	significant difference at 95% confidential level		nonsignificant difference at 95% confidential level			

Variables of socio-economic interest indicate that university graduates are more likely to post transport-related information on social media. Additionally, people with an income of 1600–2400€ are less likely to post such information compared to people who have higher or lower income. Taxi users and users travelling from and to school are more likely to post on social media.

5 CONCLUSIONS AND FURTHER RESEARCH

The research presented in this paper was focused on analyzing and studying the profile of users who receive real-time information about transportation from social media and furthermore of users who publish transport-related information on such sites. The investigation is based on data collected by an online questionnaire survey conducted in Greece. Additionally, binary models were developed to examine the probability of using or posting transport-related information on social media, based on users' personal characteristics.

Based on the results of the descriptive statistics of the sample data, it is shown that social media data mining is vulnerable to errors as the representation of elderly and people who have small or no relationship with technology is considered unsatisfactory. Although the rapid development of technology will enable in the future more people and with different backgrounds to gain access in the use of internet in general, social media cannot replace traditional methods of data collection, such as traffic counts and household surveys, but can be a supplementary source of information and knowledge.

Information that users can access through social media may affect to some extent the way they travel. Particularly, public transport users stated that the existence of information from the official sites of the operators could increase the attractiveness of the specific mode of transport. In this context, authorities and public transport operators could exploit the power they have through social networking pages to attract more users and to inform passengers in a more timely and valid manner about issues related to the provision of transport services. Public authorities and policy makers could exploit social media for encouraging citizens to use more sustainable means of transport, such as buses and fixed track modes.

Findings of the questionnaire survey indicate that users do not post at a great extent data about their trips on social media which means that extracting information from social networking services in Greece will provide weak information about habits and commuter choices. However, data from location sharing services (check-in), can be used for drawing useful conclusions about origin and destination of commuters.

The parameter estimates for the dependent variable "social media use for transport" indicate that the inclusion of socio-economic variables to the simplified model which was only consisted of variables related to social media/internet use, improved the explanatory power of the model. The increased utility of social media transport-related information increases the value (probability) of the dependent variable. Additionally, women are more likely to use social media for transport information which may be because women tend to use more social media in general compared to men. The increased probability of high school graduates may be linked to the fact that people belonging to this education category are usually students which are more active on social media compared to the entire population.

Binary models constructed for the dependent variable "social media post for transport" showed that, for all three models, users have a negative prior preference to post transport-related information on social media which is also in line with the results of the questionnaire survey. Trip characteristics variables improved the goodness of fit of the model in contrary to the socio-economic variables. The "increase use of Twitter" increases the possibility of posting transport information on social media, as Twitter, due to the limitation on the length

of posts (only 149 characters), is ideal for short conversations such as comments regarding traffic conditions or the quality of a transport service. People who use taxi and people who travel from and to school - mainly students - are also more likely to post on social media. In terms of socio-economic characteristics, university graduates tend to post more and this may be explained by their higher education level, whereas people of high income are less likely to post. This could be linked to the fact that people of high income usually travel by car, whereas people of lower income use mainly public transport and they are also more likely to post on social media information regarding their trip and express their opinion about transport services.

Exploring and understanding the profile of social media users can provide an interesting insight when mining and analyzing social media data for transportation purposes. On the one hand, knowledge of the personal characteristics of social media users, such as socio-economic characteristics, information regarding the use of social media as well as information about trip characteristics, can help researchers better understand and calibrate mobility, activity patterns and travel behavior. On the other hand, based on the results of the binary models presented in this study, if the probability to use or post transport-related information on social media is known, it may enable both researchers and transport stakeholders better understand which are the specific user characteristics that affect this probability. To this end, when social media data is mined it can be known in advance which are the characteristics of the sample from which data would be collected; similarly, when public transport operators publish transport-information on social media they will be able to know to whom they are addressed to, and thus they will be able to better plan their strategy for attracting new users, based again on the sample characteristics.

ACKNOWLEDGEMENT

The current paper is based on the research conducted in the framework of a research project funded by the programme "RESEARCH PROJECTS FOR EXCELLENCE IKY/SIEMENS".

REFERENCES

[1] Kaplan, A.M. & Haenlein, M., Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, **53**(1), pp. 59–68, 2010.

[2] Sterne, J., *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*, John Wiley & Sons: Chichester UK, 2010.

[3] Efthymiou, D. & Antoniou, C., use of social for transport data collection. *Proceedings Social and Behavioral Sciences,* **48,** pp. 775–875, 2012.

[4] Gao, H., Tang, J., Hu, X. & Liu, H., Exploring temporal effects for location recommendation on location-based social networks. *Proceedings of the 7th ACM conference on Recommender systems*, pp. 93–100, 2013.

[5] Hasan, S., Zhan, X. & Ukkusuri, S.V., Understanding urban human activity and mobility patterns using large-scale location-based data from online social media, *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, (6), 2013.

[6] Schweitzer, L., How are we doing? Opinion Mining Customer Sentiment in US Transit Agencies and Airlines Via Twitter. *Proceedings of the 91th Transportation Research Board Annual Meeting*, 2012.

[7] Collins, C., Hasan, S. & Ukkusuri, S.V., A novel transit rider satisfaction metric: rider sentiments measured from online social media data. *Journal of Public Transport*, **16** (2), pp. 21–45, 2013.

[8] Fu, K., Rakesh, N. & Jason, T.X., Social Media Data Analysis for Traffic Incident Detection and Management. *Proceedings of the 94th Transportation Research Board Annual Meeting*, 2015.
[9] Abel, F., Hauff, C., Houben, C.J., Stronkman, R. & Tao, K., Twitcident: Fighting fire with information from social web streams. *Proceedings of the 21st International ACM Conference Companion on World Wide Web*, pp. 305–308, 2012.
[10] Dyer, C., *Research in Psychology: A practical guide to methods and statistics*, Blackwell Publishing, UK, 2006.
[11] Ben-Akiva, M.E. & Lerman, S.R., Discrete choice analysis: theory and application to travel demand. *MIT press*, UK, 1985.
[12] Shumacher, R.E. & Lomax, R.G., *A beginner's guide to structural equation modelling*, **2**, Lawrence Erlbaum Associates, UK, 2004.