The influence of social contacts and communication use on travel behavior: A smartphone-based study

Jeanne Ythier, RGR Robert-Grandpierre et Rapp SA
Joan L. Walker, Institute of Transportation Studies - University of California, Berkeley
Michel Bierlaire, Transport and Mobility Laboratory - EPFL

Conference paper STRC 2013

STRC | **13th Swiss Transport Research Conference**
Monte Verità / Ascona, April 24-26,2013

The influence of social contacts and communication use on travel behavior: a smartphone-based study

Jeanne Ythier
RGR Robert-Grandpierre et Rapp SA
Ingénieurs Conseils
Genève, 1227

Phone: +4122.308. 98. 00
email: ythier@rgr-sa.ch

Joan L. Walker
University of California, Berkeley
Institute of Transportation Studies
111 McLaughlin Hall
Berkeley, CA 94720-1720

Phone: +1510.642.6897
email: joanwalker@berkeley.edu

Michel Bierlaire
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne

Phone: +4121.693.25.37
email: michel.bierlaire@epfl.ch

April 2013

Abstract

In this paper we will explore the potential of a smartphone database to investigate influences on travel behavior. Our aim is to exploit the rich individual-level data available from the smartphone to study the influence of communication and social contacts (collected via phone call and text message logs) on spatial movement (collected via GPS). The advantage of smartphone data is the ability to collect such rich data without user input over a long period of time, and the disadvantage is the difficulty associated with processing the data. We will work with three months of data from 111 people collected via a snowball sample. In studying travel behavior, we focus on high level measures of mobility as represented by the size of activity space and travel intensity (our dependent variables). We use as explanatory variables sociodemographics, spatial relationship between home and work, use of communication (number of phone calls and text messages), and the travel behavior of those in the sample who are connected to the respondent (where connectivity is measured by phone and text message contact). We will describe how these variables were processed from the smartphone data and present estimation results from the regression analysis. We find that people tend to travel in a similar manner as those whom they are socially connected to (consistent with the social network and travel literature) and that the use of communication is a complement to physical travel (consistent with the telecommunication and travel literature). The results, although preliminary, illustrate how smartphone data can be exploited to reveal complex features of travel behavior, even when they are not collected for travel behavior purposes.

Keywords

Smartphone data – Travel behavior– Social network – Communication

1. Introduction

The availability of smartphone data opens new opportunities to analyze travel behavior. This paper is an exploratory study of the potential there is using a smartphone dataset to evaluate how the social contacts of a traveler, together with his profile as a user of communication services, are related to travel behavior.

Compared to traditional surveys, such as those based on travel diaries, smartphone data are not biased by interpretation, judgment or omission from the travelers. The various sensors available in the current generation of smartphones reveal rich information about the location, the movements, the contacts and the usage of the phone, in particular the communication profile via phone calls and text messages. An important point is that the dataset has not been collected for travel behavior purpose. Our objective in this paper is to explore how this information can be used to quantify the impact of various measured quantities on travel behavior.

The paper is organized as follows: first a literature review is provided, then, the methodology and model are presented, followed by a case study and conclusions.

2. Literature review

The literature on travel behavior is vast. Most articles focus on measuring travel habits and activity patterns are based on travel diaries or GPS data. For instance, Buliung and Kanaroglou (2006) analyze how households and individuals are using space to conduct their activities. Schlich and Axhausen (2003), Pas (1988) and Gonzalez et al (2008) measure habitual travel behavior and Hanson and Huff (1988) study the variability in individual travel patterns. Global Positioning System data is used to measure person travel. Schönfeleder et al (2002) use GPS data to analyze the long term mobility patterns and Hood et al (2011) estimate a bicycle route choice model collected with GPS data.
On one hand travel diary provide information about social characteristics or social network but it is limited for long term travel behavior studies, whereas GPS data allows long term studies but do not provide any information about user's characteristics.

The use of smartphone data to analyze human behavior has recently gained a great deal of attention. Several articles focus on data collection and algorithm development. For instance, Laurila et al. (2012) summarize the research initiatives for generating innovation around smartphone-based research, Nitsche et al (2012) introduce an approach to use smartphone data for a large-scale mobility survey and Chen and Bierlaire (2012) proposed a probabilistic method by matching transport mode and physical path with smartphone sensors. Others focus on traditional travel demand models or apps with smartphones. Vautin and Walker (2010) studied the influence of smartphones on transportation behavior, operations and planning. Li et al (2011) developed a trip analyzer that identifies travel mode and purpose. Do and Gatica-Perez (2012) create models for smartphone-based human mobility. Also, Mulder et al. (2005) measure social phenomena. Jariyasunant et al (2011) use personal travel data to promote sustainable behavior.

In this research we are using smartphone data to explore the opportunity of studying the influence of social contacts as well as communication patterns on travel behavior. Both of which have a rich literature.

On the social network side, there is growing research in the link between social interactions and travel behavior. For example, Silvis et al. (2006) find two different socio-mobility styles: the first one consists in performing many shorter trips to visit a large number of people individually, and the second one consists in doing fewer longer trips to visit many people simultaneously. Their results show that social interaction is an important predictor of trips. However, the validity of a self-estimated social network size is questioned by the authors. The objectivity of smartphone data may circumvent these limitations.

Carrasco and Miller (2006) and Axhausen (2008) collect social network and context information by capturing ego-centric or personal networks. Also, Carrasco et al (2008) incorporate the social dimension in social activity-travel behavior. They explicitly study the link between individuals’ social activities and their social networks using an egocentric approach. The main hypothesis is that communication and activity-travel patterns emerge from the individuals' social networks. This hypothesis has consequences on the generation and spatial distribution of social activities, and the usage of communication among individuals.

Axhausen (2003) shows the interactions between spatial structure of social network and travel patterns, especially for leisure trips. Besides, leisure travelling is mostly socially employed to meet friends, relatives and contacts. The distribution of these friends, relatives and contacts across space is crucial in leisure travel generation. This hypothesis is also confirmed by Ohnmacht (2009). Finally, the spatial spread of social network has increased, explaining the observed increase in leisure travelling. Axhausen (2005) and Marsden and Campbell (1984), measure social interaction and social network structure. Giuliano (1997) shows the relationship between societal change and transport and Goetzke (2008) illustrates the network effects in transit use. Finally, many articles focus on capturing influences from social networking among household members (e.g. Arentze and Timmermans (2009), Bradley and Vovsha (2005), Gliebe and Koppelman (2002), Kang and Scott (2010), Scott and Kanaroglou (2002), Srinivasan and Bhat (2005) and Timmermans and Zhang (2009)).

There is also a large variety of literature covering the interaction between telecommunications and travel behavior. Golob (2001) study the effect of information technology on personal travel behavior. Senbil and Kitamura (2003) study the relationship between telecommunications and activities and Handy and Yantis (1997) show the impact of telecommunication technologies on nonwork related travel behavior. Choo et al 2002, 2005 and 2007 study the impact of telecommunications on travel demand, supply and telecommuting. Mokhtarian (2002) provide a comprehensive survey. They identify four types of cross-mode relationships from the literature (e.g. Claisse, 1983, Mokhtarian and Salomon, 2002, Mokhtarian (1990), Niles, 1994, Salomon, 1985 and 1986):

Substitution refers to the replacement of trips by the use of telecommunication.
Complementarity refers to the growth of the number of trips as a consequence of the increased use of telecommunication.
Modification refers to the influence of the use of telecommunication on the types of trips (for example, the transportation mode or the destination).

> **Neutrality** refers to instances where usage of telecommunications has no influence on travel behavior. A typical example is a trip to the grocery store.

Their analysis uses data collected by trade organizations, government agencies or public agencies (National time series data spanning 1950-2000). Mokhtarian (2002) conclude that "*impact focusing on a single application (such as telecommuting) have often found substitution effects, such studies are incomplete and likely to miss the more subtle, indirect, and longer-term complementarity effects that are typically observed in more comprehensive studies. From the comprehensive perspective, substitution, complementarity, modification, and neutrality within and across communication modes are all happening simultaneously. The net outcome of these partially counteracting effects, if current trends continue, is likely to be faster growth in telecommunications than in travel, resulting in an increasing share of interactions falling to telecommunications, but with continued growth in travel in absolute terms*."

In summary, the influence of social networking and the use of communication services on travel behavior are well acknowledged in the literature. However, the effects of social networks and telecommunications are for the most part studied separately (see Páez and Scott (2007) and Kwan (2008) for an exception). Furthermore, the derivation of quantitative models capturing this relationship, based on smartphone data, has not yet been proposed. This is the objective of this paper. Moreover, the literature focuses mainly on teleworking aspects of communication, whereas we investigate how the patterns of communication usage are related to travel behavior.

3. The variables

The Nokia dataset is not collected for transportation study purposes. In our study, we define variables of interest and propose how to process the data.
In order to derive quantitative models, we first characterize the main concepts by variables that can be observed. The three concepts in our analysis are (i) travel behavior, (ii) social contacts and (iii) usage of communication. For each of these general concepts we define key variables that, hypothesized, will reveal the relationships we are investigating.

One of the advantages of smartphone data is the ability to collect data over longer periods of time without burdening the respondent. Having such individual-level data over a longer period of time provides more insight into the general mobility methods of people than in a one- or two-day survey. We choose to focus on such higher-order mobility styles for our analysis, and investigate *travel behavior* as described by (i) the travel intensity, and (ii) the size of one's activity space. The travel intensity is characterized by the total number of different activity locations visited. The activity space is defined as the area where most of the activities of the traveler are located. The location of an activity is a place where the traveler spends time. It includes home location, work location, leisure locations, etc.

More specifically, we used five variables to characterize travel behavior (three to capture travel intensity, two to capture size of activity space). These variables are the dependent variables in our analysis. The variables as well as the form of the regression model are as follows:

Measures of travel intensity

1. Total number of trips: it is the number of visits to activity locations performed by the traveler during the period of analysis. It is captured by generalized linear model with a negative binomial error term. Such a model is designed for over-dispersed count data.
2. Total number of activity locations: it is the number of places visited by the traveler during the period of analysis (three months in our case), irrespectively of the number of times each location is visited. It is also represented by a negative binomial linear model.
3. Number of occasional activities**:** an occasional activity is defined as an activity performed fewer times over the period of analysis. In our case study, occasional activities are performed less than 5 times over the 3 month analysis. This variable is designed to distinguish routine and non-routine travel behavior. It is also represented

by a negative binomial linear model.

Measures of size of activity space

4. Maximum distance traveled [kilometers] it is the greatest distance traveled between home and an activity location. It is captured by generalized linear model with a log normal error term.
5. Average distance per trip [kilometers] it is the total number of kilometers traveled divided by the total number of trips. It is also represented by a log normal linear model.

The explanatory variables we use are classified into three categories: (i) socio-economic characteristics of the traveler (variables 1 to 6 below), (ii) variables describing aspects of the social contacts (variable 7 below) and (iii) variables describing the usage of communication (variables 8 to 12 below).

Socio-economic characteristics

1. Housemate is a dummy variable equal to 1 if the traveler has a housemate, 0 otherwise.
2. Male is a dummy variable equal to 1 if the person is male, 0 otherwise.
3. Dummy variables Age: levels: under 16, over 33. The reference model corresponds to ages between 21 and 32 years old.
4. Dummy variables work**:** levels: part time work, not working, full time studying, work other. The reference model corresponds to full time work.
5. The distance between home and work, in kilometers.
6. The number of visits to work is the number of times the user goes to workplace. A high number of visits to work corresponds to a person who also goes to many places during the day.

Characteristics of social contacts

7. Travel profile of contacts**:** the travel behavior of the contacts (that is, persons who have been in communication with the target traveler) is considered. Again, this behavior is characterized by the variables 1 to 5 above. In order to account for the strength of social connection, these variables are weighted by the total number of communications that have occurred between the target traveler and each of his contacts, that is the number of phone calls (including missed calls) and text messages sent and received. If *i* is the identifier of the contact, and *Y_contact_i* is the value of the travel behavior variable Y for contact *i*, the corresponding travel profile variable is defined as:

$$Travel\_profile\_contacts = \frac{\sum number\_of\_communications\_i \cdot Y\_contact\_i}{\sum number\_of\_communications\_i}$$

Usage of communication

8. The total number of calls: missed, sent and received calls.

9. The total number of text messages: sent and received text messages.

10. The number of occasional contacts: contacts that have been called once. This variable is designed to capture the heterogeneity of the social network. The total number of contacts could be also a variable. It is designed as a proxy for the size of the social network. However, this variable is strongly correlated with the number of occasional contacts.

11. The proportion of long calls: proportion of long calls (that is, any call longer than 3 minutes) over the total number of calls (except missed calls):

$$Proportion\_of\_long\_calls = \frac{Nb\_of\_long\_calls}{Nb\_of\_long\_calls + Nb\_of\_short\_calls} \cdot 100$$

12. Dummy variables Who pays: levels: phone bill paid by the traveler and by others. The reference level is “paid by the traveler”.

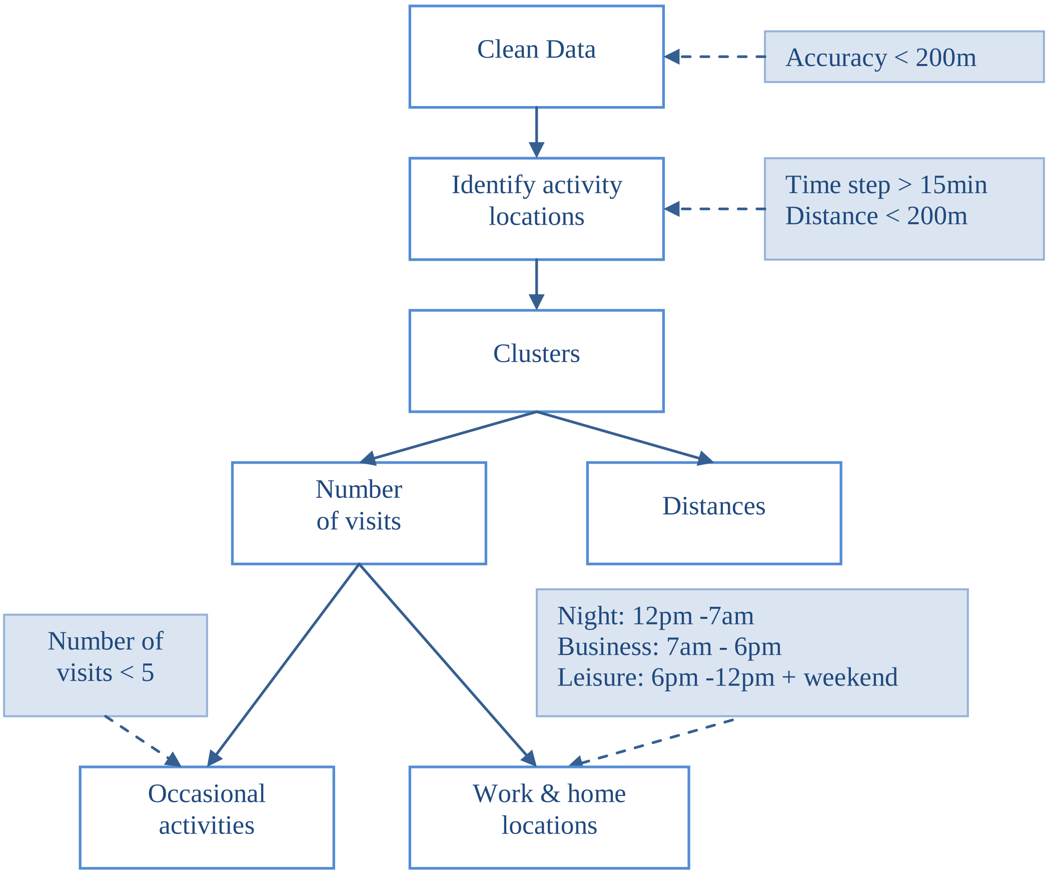
Each of the five models is potentially explained by all independent variables, although some may be insignificant in the model results.

4. Methodology

4.1 Data processing

The data has been processed to obtain the value of the dependent and explanatory variables defined above. The process is summarized in figure 1 and further described below.

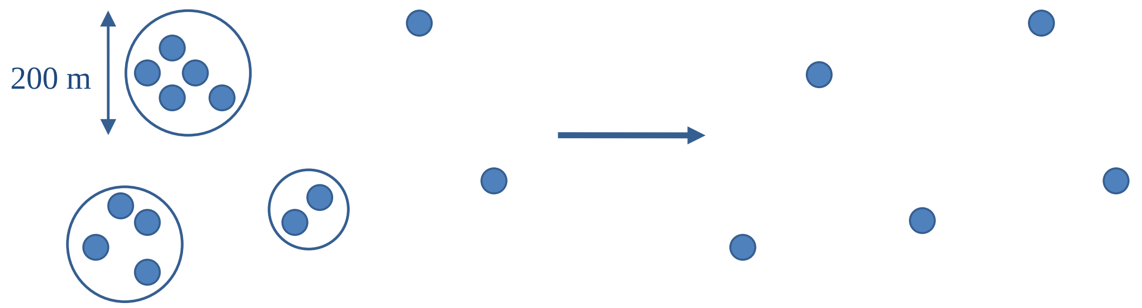
Figure 1 Data processing



1. Data cleaning: GPS data with poor accuracy has first been dropped. Each point with a confidence interval larger than 200m has been considered of poor accuracy.
2. Identification of the activity locations: We consider that an activity occurs when a user stays in an area of a radius less than 200m during more than 15 minutes. (A sensitivity analysis testing cutoffs from 10 to 20 minutes did not lead to significant differences in the results). If GPS measurements *i* and *i+1* meet this criteria, measurement *i* is associated with an activity.

3. Spatial clustering: In order to identify the locations of the various activities, we need to group the measurements selected in the previous step. We group together measurements that are less than 200 meters apart, and associate an activity location to each of the groups, as illustrated in figure 2. Some ambiguities had to be processed manually.

Figure 2 Clustering to specific activity locations



4. Time clustering: in order to identify the number of visits to each location, we regroup GPS measurements that are less than 15 minutes apart. The number of visits to a location is defined as the sum of the number of such clusters that are within a radius of 200m of the location, and the number of measurements in the spatial cluster associated with the location at step 3. With this procedure, we capture both the instances where the GPS was turned on and the instances where it was off during the activity.
5. Once we detected all activities we divide time in three periods. Night (from midnight to 7am), business hours (from 7am to 6pm) and leisure time (from 6pm to midnight plus the weekend). We use this partition to identify the location of home and work for each person in the sample. The assumption is that the place with the largest number of visits during leisure time and night is home and the place with the largest number of visits during business hours is the work place. Activity locations with less than 5 visits are defined as occasional activities.

4.2 Regressions

The independent variables related to the size of activity space are continuous data whereas the independent variables related to travel intensity are count data.

The log normal linear regression model is a classic model to evaluate continuous positive data. The negative binomial regression model is used to count over-dispersed data. Over-dispersed means that the conditional variance is not equal to the conditional mean. In both cases, a log-linear model is considered. Model equations are presented below:

$$\ln(Y) = \beta_0 + \sum_i \beta_i \cdot x_i + \varepsilon_i$$

5. Case study

5.1 Data description

The Nokia Research Center in Lausanne organized a data collection campaign involving 200 users from September 2009 to October 2010 in Switzerland. Each user carried a N95 smartphone equipped with an application that continuously collected and uploaded data from the sensors of the phone. The data was collected without any intervention from the traveler. The details concerning the data collection campaign are described in a technical report (Kiukkonen (2009)).

For this analysis, we used the location and communication data. When the GPS is turned on, we have access to the longitude, the latitude and the altitude with a time step of 10 seconds. We have also access to the entire list (caller/recipient) and duration of incoming, outgoing, and missed calls, as well as the list (sender/recipient) of incoming and outgoing text messages.

Note that the travel profiles of contacts (variable 7 above) are available only for contacts who participated in the survey. As the participants of the data collection campaign have been recruited based on a snowball sampling strategy, the average number of contacts who participated in the survey is 4.81. ($\sigma = 4.36$).

For our study, we considered a period of 3 months, from March 1, 2010 to May 31, 2010 (3 months), selected to be a period free of major holidays. We extracted the data of 111 users that are usable for this study.

5.2 Strengths and weaknesses

The Nokia smartphone dataset has a large and comprehensive amount of information on movement and smartphone use for the individuals in the sample, which provides access to an abundance of information concerning each participant. Moreover, the data is “objective”, as it is collected without the intervention of the users. The possible biases are only due to technological reasons. The negative side of this is that the data is difficult to analyze and process. The main difficulty with GPS coordinates is the time of tracking is not continuous. Indeed, the GPS is regularly turned on and off to save battery life. In addition, it is difficult to precisely evaluate the amount of time users spend in their activity locations, which motivated the procedure described above to use the number of visits to each activity location. Finally, our knowledge of the social network is limited to the portion that participated in the survey, which is not the complete social network.

Besides, Nokia uses a snowball sampling method to enroll users. On the one hand, this method could bias the results, but on the other hand, it is the best way to get a social network in the sampling.

5.3 Descriptive statistics

After processing, we obtained the value of the variables for 111 users over the 3 month period. We report some statistics in Table 1 (data processed from the smartphone data) and 2 (data obtained via a supplemental survey).

Table 1 Travel and communication statistics for the 111 survey participants (processed from smartphone data)

Variables	Units	Mean	Variance	Minimum	Maximum
Number of trips	-	376.9	85859.2	4	1540
Number of activities	-	20.9	140.7	1	67
Number of occasional activities	-	8.55	31.4	0	33
Average distance per trip	km	18.4	546.9	0.2	160.7
Max distance traveled	km	75.3	3803	0.2	239.8
Distance home-work	km	15.8	1075.9	0	224.9
Number of visits to work	-	63.3	3793.7	0	451
Number of calls	-	1650.2	1141948.4	89	4946
Number of text messages	-	1203.8	1201807.1	90	5495
Number of occasional contacts	-	12.9	54.4	2	40
Part of long phone call	-	18.9	89.7	3.43	49.3
Number of trips	-	376.9	85859.2	4	1540
Number of activities	-	20.9	140.7	1	67
Number of occasional activities	-	8.55	31.4	0	33

Table 2 Socioeconomic characteristics of the sample (obtained via survey questions)

	Observations (of 111 total)
Housemate	80
Male	40
Female	62
Age < 21	9
Age 22 - 32	69
Age > 33	24
Working full time	49
Working part time	8
Not working	8
Studying full time	35
Other employment status	2
Bill paid by self	87
Bill paid by other	15
No survey data	9

5.4 Results

The estimated parameters of the two models related to the size of the activity space (log normal linear regression models) are reported in the Table 3 and the models related to travel intensity (negative binomial regression models) are reported in the Table 4. Several specifications were tested to get to these final models, including different explanatory variable combinations and residual analysis to verify the appropriateness of the model forms selected.

In examining the estimation results, a first general comment is that the signs of the coefficients are consistent with expectation. The socio-demographics were on the whole not particularly significant. One's work status was never significant, gender only influenced average distance per trip (men travel farther on average), having a housemate only influenced the number of occasional activities (having a housemate leads to more), and having your bill paid by someone else other than an employer (e.g., parents) lead to smaller average distances per trip. In these models, the most interesting variables we have included are those related to ones communication behavior (how much does one use the smartphone to make calls and send texts) and the travel behavior of one's contacts. Recall that the "contact behavior" is simply a weighted average of the dependent variable for other people in the sample whom the traveler has contacted by smartphone call or text, where the weight is a function of the number of smartphone contacts. In all cases, this variable is statistically significant with a positive sign. This indicates that one tends to have similar travel behavior characteristics as those one is socially connected to. In terms of the communication use influence, while the number of calls was not significant, the number of texts was in all three measures of travel intensity but was not significant in explaining the size of the activity space. This suggests communication as a compliment to travel and activity. The number of occasional contacts (also a proxy for the number of contacts as these are highly correlated) also significantly and positively increases the measures of travel intensity. This also makes sense: the more contacts the more activities.

In addition to these general results, each model is discussed briefly below, with emphasis placed on those factors that are statistically significant.

Max distance

If the maximum distance of the contacts increases, the maximum distance of the target traveler also increases. The maximum distance is high for people between 22 and 32 and shorter for young people. If the number of visits to work and the home-to-work distance increase, the maximum distance increases too.

Average distance per trip

If the average distance per trip of the contacts increases, it does for the target traveler too. Men have a bigger average distance per trip than women. If the home-to-work distance increases, the average distance per trip increases too. In terms of communication, if the phone bill is paid by someone else, the average distance appears to be smaller.

Number of trips

If the number of trips performed by the contacts increases, it does so for the target traveler too. The number of trips is the largest for people older than 33 years old and the smallest for people under 21. The larger the number of visits to work, higher the number of trips. In terms of communication usage, the more texts are sent and received, the larger the number of trips.

Number of activities

If the number of activities performed by the contacts increases, it does so for the target traveler too. The number of activities is largest for people over 33 years old and the smallest for people under 21. If the number of visits to work and the home-to-work distance increase, the number of activities increases too. In terms of communication usage, the more texts are sent and received, the larger is the number of activities. Also, the larger is the number of occasional contacts, the higher the number of activity locations.

Number of occasional activities

If the number of occasional activities performed by the contacts increase, it does so for the target traveler too. The number of occasional activities is larger when the traveler has a housemate. The number of occasional activities is the largest for people over 33 years old and the smallest for people under 21. If the home-to-work distance increases, the number of occasional activities increases too. In terms of communication usage, the more texts are sent and received, the larger is the number of occasional activities. Also, the larger the number of occasional contacts is, the higher is the number of occasional activity locations.

Table 3 Models related to size of activity space (log-normal model)

			Maximum distance		Average distance per trip	
		units	β	p-value	β	p-value
	Constant	[ ]	3.56	0.000	1.83	7.86e-08
Characteristics of social contacts	Contact behavior	[ ]	**0.00332**	**0.0143**	**0.00677**	**0.0628**
Socio-economic characteristics	Housemate	dummy	-0.0807	0.701	0.346	0.212
	Male	dummy	0.0160	0.922	**0.375**	**0.0235**
	Age < 21	dummy	**-0.806**	**0.0788**	0.0306	0.918
	Age > 33	dummy	-0.325	0.133	0.0549	0.815
	Work part time	dummy	0.146	0.563	-0.279	0.389
	Not current work	dummy	-0.0209	0.960	-0.304	0.647
	Study full time	dummy	-0.266	0.147	-0.232	0.247
	Work Other	dummy	-0.413	0.646	-0.149	0.847
	Distance home work	[km]	**0.00438**	**0.0162**	**0.0109**	**5.3e-14**
	Number of visits to work	[ ]	**0.00274**	**0.0203**	0.000667	0.439
Usage of communication	Number of call	[ ]	6.62e-05	0.298	-8.34e-06	0.890
	Number of text	[ ]	8.02e-05	0.262	0.000103	0.127
	Number of occasional contacts	[ ]	0.0138	0.127	0.00474	0.578
	Percentage of long call	[%]	0.000455	0.955	0.00819	0.291
	Bill paid by other	dummy	0.328	0.327	**-0.773**	**0.056**
	No data	dummy	-0.358	0.312	0.411	0.271

Table 4 Models related to travel intensity (Negative binomial model)

			Number of activities		Number of trips		Number of occasional activities	
		units	β	p-value	β	p-value	β	p-value
	Constant	[ ]	2.16	2e-16	4.51	2e-16	1.06	1.69e-06
Characteristics of social contacts	Contact behavior	[ ]	**0.00854**	**0.0672**	**0.0129**	**0.0534**	**0.0101**	**0.0526**
Socio-economic characteristics	Housemate	dummy	0.0712	0.556	-0.00612	0.971	**0.238**	**0.0923**
	Male	dummy	0.0867	0.421	0.193	0.212	0.0683	0.574
	Age < 21	dummy	**-0.622**	**0.00287**	**-0.701**	**0.0150**	**-0.682**	**0.00447**
	Age > 33	dummy	**0.258**	**0.0409**	**0.443**	**0.0156**	**0.242**	**0.0864**
	Work part time	dummy	0.0150	0.934	0.0922	0.726	0.0101	0.960
	Not current work	dummy	0.110	0.618	0.296	0.350	0.0233	0.926
	Study full time	dummy	-0.0171	0.884	0.139	0.403	0.0251	0.848
	Work Other	dummy	0.0161	0.964	0.372	0.471	0.0749	0.848
	Distance home work	[km]	**0.00481**	**0.00197**	0.00316	0.175	**0.00557**	**6.35e-04**
	Number of visits to work	[ ]	**0.00298**	**8.72e-05**	**0.00971**	**2e-16**	0.00116	0.177
Usage of communication	Number of call	[ ]	-1.39e-05	0.763	-4.93e-05	0.461	-4.25e-06	0.933
	Number of text	[ ]	**1.76e-04**	**4.23e-05**	**1.76e-04**	**0.00521**	**1.85e-04**	**8.41e-05**
	Number of occasional contacts	[ ]	**0.0109**	**0.0929**	0.00287	0.763	**0.0131**	**0.0623**
	Percentage of long call	[%]	-0.00495	0.343	-0.00348	0.643	-2.405e-05	0.997
	Bill paid by other	dummy	0.0338	0.843	0.134	0.580	0.0358	0.852
	No data	dummy	0.0409	0.839	0.0352	0.902	0.215	0.346

Table 5 Goodness of fit

	Maximum distance	Average distance per trip	Number of activities	Number of trips	Number of occasional activities
R2	0.32	0.74	0.41	0.43	0.38

To report the goodness of fit we compute a R^2 for both models. For log normal models

$$R^2 = 1 - \frac{\sum resuduals^2}{\sum (Y - mean(Y))^2}$$

And for negative binomial regression we used the log likelihood

$$R^2 = 1 - \frac{final\_Likelihood}{initial\_Likelihood}$$

Remark: R^2 is high for average distance per trip showing the model fit well the data. The others are less significant which can be explained by the small size of the sample.

6. Discussion

In this article, we explored the potential of using dataset from smartphones for transportation purposes. We extracted information on travel, major activity locations, personal characteristics, telecommunication behaviors and social contacts. The first main conclusion is that the behavior of the social contact influences the traveler, who has a tendency to adopt similar travel behaviors. This confirms the work by Axhausen (2003, 2005) about the importance of social networking in trip generation. Secondly, usage of communication does not influence the size of the activity space but it influences the travel intensity. Indeed, it seems that there is a complementarity between the number of text messages and the travel intensity. This result is consistent with the findings of Mokhtarian (2002). In addition, the diversity of contacts in the address book influences also the travel intensity.

We can conclude from the analysis above that the social network, the socio economic characteristics and the use of communication indeed influences travel behavior, characterized by the size of activity space and the travel intensity.

Although consistent with the literature and intuition, these results should be taken with a grain of salt. Indeed, the causality of some variables may be questioned. For instance, is the number of trips explained by the number of text messages sent and received, or the other way around? Both hypotheses should be tested. The same can be said for the relationship between one’s travel behavior and of one’s contacts.

Additional improvements of the model include the usage of e-mails, as well as the inclusion of land use characteristics, such as population density and accessibility of home location. Finally we could include some variables related to the spatial relation between the different users such as common activities locations and distance between the users home. And the work could be extended to other measures of mobility, such as mode usage. To do so, we need further research on data processing and additional data collected campaign designed for travel behavior topics.

The preliminary analysis presented in this paper demonstrates the potential smartphone data, collected without the user’s intervention, can indeed be exploited to analyze in detail travel behavior.

7. Acknowledgment

We wish to thank various people for their contribution to this project, Mr. Akshay Vij and Mr David Gaker, for their valuable technical support on this project, Mr Sebastian E. Guerrero and Mr Haotian Liu, for their help in programming and Mr. Edouard Bulteau, Mr. Christopher Lalau Keraly and Mr. Romain Leblanc for their helpful advice.

8. References

Arentze T.A., Timmermans H.J.P., A need-based model of multi-day, multi-person activity generation. Transportation Research Part B 43 (2), 2009, p.251–265.

Axhausen K.W., Social networks and travel: Some hypotheses, STELLA, 2003.

Axhausen K.W., Social networks and travel: some hypotheses. In Social Aspects of Sustainable Transport: Transatlantic Perspectives, 2005.

Axhausen K.W., Social networks, mobility biographies, and travel: survey challenges. Environment and Planning B: Planning and Design 35, 2008, p.981-996.

Bradley M., Vovsha P., A model for joint choice of daily activity pattern types of household members. Transportation 32 (5), 2005, p.545–571.

Buliung RN. and Kanaroglou PS., A GIS toolkit for exploring geographies of household activity/travel behavior. In Journal of Transport Geography, 2006, No. 14 p.35-51.

Carrasco J-A., and Miller E., Exploring the propensity to perform social activities: asocial network approach. Transportation 33, 2006, p.463-480.

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

Chen J. and Bierlaire M., Probabilistic multimodal map matching with rich smartphone data, 2012.

Choo S. and Mokhtarian P.L.., Telecommunications and travel demand and supply: Aggregate structural equation model for the US. In Transportation Research Part A 41, 2007, p.4–18.

Choo S., Mokhtarian P.L., and Salomon I., Impact of home-based telecommuting on vehicle-miles traveled: a nationwide time series analysis. California Energy Commission Research Report UCD-ITS-RR-02-05, 2002.

Choo S., Mokhtarian P.L.., and Salomon I., Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the US. In Transportation No. 32, 2005, p.37-64.

Claisse G., Transport and telecommunications, ECMT round table 59. Paris, European Conference of Ministers of Transport, 1983.

Do T.M.T. and Gatica-Perez D., Contextual conditional models for smartphone-based human mobility prediction, at 14th ACM International Conference on Ubiquitous Computing, Pittsburgh, Pennsylvania, United States, 2012

Gliebe J.P., Koppelman F.S., A model of joint activity participation between household members. Transportation 29 (1), 2002, p.49–72.

Goetzke, F., Network effects in public transit use: evidence from a spatially autoregressive mode choice model. Urban Studies 45, 2008, p407-417.

Golob T.F. Travelbehaviour.com: Activity Approaches to Modeling the Effects of Information Technology on Personal Travel Behaviour, in Travel Behavior Research, The Leading Edge, D. Hensher, Editor. Elsevier Scince/Pergamon: Kidlington, Oxford, 2001, p.145- 184

Gonzalez M.C., Hidalgo C.A., and Barabasi A.L., Understanding individual human mobility patterns. In Nature, No. 453, 2008, p.779-782.

Giuliano G., Gillespie A., Research issues regarding societal change and transport. In Journal of Transport Geography, vol. 5, No. 3, 1997, p.165-176.

Handy S.L. and Yantis T., The impact of telecommunications technologies on nonwork travel behavior, 1997.

Hanson S. and Huff J.O., Repitition and day-to-day variability in individual travel patterns: Implication for classification. In Behavioral Modelling in Geography and Planning, R. Golledge and H. Timmermans, 1988, p.368-398.

Hood J., Sall E. and Charlton B., A GPS-based bicycle route choice model for san Francisco, California, Transportation letter: The international Journal of Transportation Research, 2011, p.63-75

Jariyasunant J., Carrel A., Ekambaram V., Gaker D., Kote T., Sengupta R. and Walker J., The Quantified Traveler: Using personal travel data to promote sustainable transport behavior, 2011.

Kang H., Scott D.M., Exploring day-to-day variability in time use for household members. Transportation Research Part A 44 (8), 2010, p.609–619.

Kiukkonen N., Technical Report: Data collection campaign, 2009

Kwan M.P., Mobile Communications, Social Networks, and Urban Travel: Hypertext as a New Metaphor for Conceptualizing Spatial Interaction, 2008.

Laurila J.K., Gatica-Perez D., Aad I., Blom J., Bornet O., Do T.M.T., Dousse O., Eberle J. and Miettinen M., The mobile Data Challenge:Big Data for Mobile Computing Research, 2012.

Li M., Dai J., Sahu S. and Naphade M., Trip analyzer through smartphone apps, at 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, United States, 2011, p.537-540.

Marsden P.V., Campbell K., Measuring tie strength' Social Forces. In Social Forces, No. 63, 1984, p.482-501.

Mokhtarian P.L., Telecommunications and travel: The Case for Complementarity. In Journal of Industrial Ecology Vol. 6 No. 2, 2002, p.43-57.

Mokhtarian P.L. and Salomon I., Emerging travel patterns: Do telecommunications make a difference? Perpetual motion: Travel behavior research opportunities and application challenges, 2002.

Mokhtarian, P.L., A typology of relationships between telecommunications and transportation. In Transportation Research A, 1990, p.231-242.

Niles J., Beyond telecommuting: A new paradigm for the effect of telecommunications on travel, National Technical Information Service (NTIS), 2002.

Nitsche P., Widhalm P., Breuss S. and Maurer P., A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys, Procedia Social and behavioral sciences 48, 2012, p1033-1046.

Ohnmacht, T., Social-activity travel: do the ‘strong-tie relationships’ of a person exist in the same community? The case of Switzerland. Environment and Planning A 41 (12), 2009, p.3003–3022.

Páez A., Scott D., Social influence on travel behavior: a simulation example of the decision to telecommute'. In Environment and Planning A, vol. 39, No. 3, 2007, p.647-665.

Pas E., Weekly travel-activity behaviour. In Transportation, Vol. 15, issue 1-2, 1988, p.89-109.

Salomon I., Telecommunications and travel: Substitution or modified mobility? Journal of transport Economics and Policy, 1985, p.219-238.

Salomon I., Telecommunications and travel relationships: A review, Transportation Research A 20A, 1986, p.223-238.

Schlich R. and Axhausen K.W., Habitual travel behaviour: Evidence from a six-week travel diary. In Transportation, Vol. 30, issue 1, 2003, p.13-36.

Schönfeleder S., Axhausen K.W., Antille N. and Bierlaire M. Exploring the potentials of automatically collected GPS data for travel behaviour analysis- A Swedish data source, GITechnologien für Verkehr und Logistik 13, 2002, pp.155-179.

Scott D.M., Kanaroglou P.S., An activity-episode generation model that captures interactions between household heads: development and empirical analysis. Transportation Research Part B 36 (10), 2002, p.875–896.

Senbil M. and Kitamura R., Simultaneous Relationships Between Telecommunications and Activities, at 10th International Conference on travel behavior Research Lucerne, 2003.

Silvis J, Niemeier D, D'Souza R., Social Networks and Travel Behavior: Report from an Integrated Travel Diary. At the 11th International Conference on travel Behavior Research, Kyoto, 2006.

Srinivasan S., Bhat C.R., Modeling household interactions in daily in-home and out-of-home maintenance activity participation. Transportation 32 (5), 2005, p.523-544.

Timmermans H.J.P., Zhang J.Y., 2009. Modeling household activity travel behavior: examples of state of the art modeling approaches and research agenda. Transportation Research Part B 43 (2), 2009, p.187–190.

Vautin D.A. and Walker J., Transportation Impacts of Information Provision and Data Collection via Smartphones, Revised paper submitted for presentation at the 90th Transportation Research Board Annuel Meeting, 2010.