CrossMark

Exploring the potential of phone call data to characterize the relationship between social network and travel behavior

Miguel Picornell[1] · Tomás Ruiz[2] · Maxime Lenormand[3] ·
José J. Ramasco[3] · Thibaut Dubernet[4] · Enrique Frías-Martínez[5]

Abstract Social network contacts have significant influence on individual travel behavior. However, transport models rarely consider social interaction. One of the reasons is the difficulty to properly model social influence based on the limited data available. Non-conventional, passively collected data sources, such as Twitter, Facebook or mobile phones, provide large amounts of data containing both social interaction and spatiotemporal information. The analysis of such data opens an opportunity to better understand the influence of social networks on travel behavior. The main objective of this paper is to examine the relationship between travel behavior and social networks using mobile phone data. A huge dataset containing billions of registers has been used for this study. The paper analyzes the nature of co-location events and frequent locations shared by social network contacts, aiming not only to provide understanding on why users share certain locations, but also to quantify the degree in which the different types of locations are shared. Locations have been classified as frequent (*home*, *work* and *other*) and non-frequent. A novel approach to identify co-location events based on the intersection of users' mobility models has been proposed. Results show that *other* locations different from *home* and *work* are frequently associated to social interaction. Additionally, the importance of non-frequent locations in co-location events is shown. Finally, the potential application of the data analysis results to improve activity-based transport models and assess transport policies is discussed.

✉ Miguel Picornell
  miguel.picornell@nommon.es

1 Nommon Solutions and Technologies, Madrid, Spain

2 Universitat Politècnica de València, València, Spain

3 Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB),
   07122 Palma de Mallorca, Spain

4 Institute for Transport Planning and Systems (IVT), ETH Zurich, 8093 Zurich, Switzerland

5 Telefonica Research, Madrid, Spain

🙋 Springer

Introduction

Travel behavior is nearly always modelled as a set of independent decisions across travelers. This approach provides satisfactory results for regularly-scheduled or very inelastic activities, like work trips, but ignores the fact that intra- and extra-household interactions play a key role in many other trips and activities (e.g., leisure trips) that are planned jointly and/or depend on the trips and activities of the social contacts. The concept of a 'full individual daily pattern', which constitutes the core of the original activity-based approach, needs to be expanded to account for the influence of the social network. A key issue is incorporating realistic geographic social networks into agent-based models, which makes it necessary to characterize the form and statistical properties of the underlying social structures and the strengths of their influences. The analysis of new data sources, such as online social networks or mobile phone data, can help improve the understanding of the interdependencies and co-evolution of the social networks and the activity-travel patterns. In recent years, there has been an increasing interest in studies related to human mobility patterns (e.g., Brockmann et al. 2006; Gonzalez et al. 2008; Song et al. 2010a; Gould 2013) and social networks (e.g. Onnela et al. 2007; Lazer et al. 2009; Carrasco et al. 2008a; Clifton 2013), some of them using different spatiotemporal information from non-conventional, passively collected data sources (e.g., GPS, mobile phones, Twitter, etc.). However, only a few studies have analyzed both aspects at the same time using mobile phone data records (Calabrese et al. 2011a; Cho et al. 2011; Phithakkitnukoon et al. 2012; Chen and Mei 2014). The main objective of this paper is to examine the relationship between travel behavior and social networks using mobile phone data. The paper focuses on the analysis of the characteristics of the locations shared by social contacts, aiming to understand and quantify why and in which degree those locations are shared.

The structure of the paper is as follows: first, a review of previous work related to the interaction between social network, travel behavior and the use of mobile phones is presented. Secondly, the scope and contributions of the paper are shown. In the third place, the characteristics of the dataset used in this study are described. Fourthly, the methodology followed to obtain users' social network and travel behavior from mobile phone data and to analyze the relationship between them is explained. Then, the results and main findings are presented, and their application to inform activity-based models and assess mobility policies such as carpooling is explained. Finally, the main conclusions of the paper and further research avenues are discussed.

Literature review

Social networks and travel behavior

It has been recognized that the characteristics of people's social network influence social activities and related travels (Axhausen 2005; Arentze and Timmermans 2006; Carrasco and Miller 2006). There is an increasing number of transport studies that are including

social networks as an important factor to improve travel demand models. Earlier applications of social networks in transport planning and travel behavior studies date from the beginning of the present millennium. Dugundji and Walker (2005) derived a mode choice model using various static associative social networks that group individuals by several statistics. Paez and Scott (2007) presented a similar approach to estimate the share of telecommuting at a firm in consideration of peer pressure to appear at one's desk. Carrasco and Miller (2006) explicitly included social networks in a conceptual model of social activity-travel behavior. Marchal and Nagel (2006) allowed cooperative agents in a microsimulation to share information with each other about activity locations and about other agents, in order to optimize trip chains. Arentze and Timmermans (2006) presented a framework for a multi-agent microsimulation that produces a dynamic social network which evolves together with activity-travel patterns. Hackney et al. (2006) also studied interdependencies between social networks and travel behavior. Silvis et al. (2006) found relations between number of trips and locations visited, and the social network size and number of repeated contacts. Molin et al. (2007) analyzed the influence of the size and composition of the social network on travel demand. Arentze and Timmermans (2008) focused on direct effects of social networks on activity-travel interactions. Carrasco et al. (2008a) studied the spatial distribution of social activities, focusing on the home distance of individuals. Carrasco et al. (2008b) explored the relationships between travel behavior, ICT use and social networks. Carrasco and Miller (2009) studied the effects of characteristics of individuals' personal networks and interactions on activity frequency. Hackney and Marchal (2009) developed a microsimulation model which incorporated a social network on top of a daily activity scheduler. More recently, Hackney and Marchal (2011) and Ronald et al. (2012a, 2012b) have taken into consideration the role of social networks in travel behavior using an agent-based approach. Habib and Carrasco (2011) analyze the effects of social networks on the timing and duration of activities. Van den Berg et al. (2013) studied the effects of social networks and telecommunications on activity-travel patterns. Moore et al. (2013) studied links between personal networks, time use and geographical location of people. Sharmeen et al. (2013, 2014) analyzed face-to-face social interaction and geographic accessibility.

Although significant theoretical advances have been made in understanding how the social network influences travel behavior, data availability is still a significant limitation for this kind of studies. As remarked by Van den Berg et al. (2013), only a few data collection efforts have been made so far in order to incorporate social networks in models of travel demand. Furthermore, data is usually collected through personal surveys which are limited in sample size (hundreds of users) and period of time (few days). For instance, Carrasco et al. (2008c) carried out a survey of 350 people and in-depth interviews of a subsample of 84, and Van den Berg et al. (2013) performed a survey combining a questionnaire and a 2-days social interaction diary, obtaining 747 responses (response rate 20 %). On the other hand, non-conventional passively collected data from Twitter, Facebook or mobile phones, which provide relevant information on social relations and user's location data, can open an opportunity to deal with data limitation problems. In contrast to surveys, these new data sources provide location information as well as social interaction of millions of users during long periods of time. In terms of mobility, mobile phone data is one of the best sources to obtain spatiotemporal information for a long period of time covering a big percentage of the entire population (Lane et al. 2010). Additionally, when analyzing social networks, mobile phone data have the advantage of providing more relevant face-to-face personal relationship information compared to other data sources such as Twitter or Facebook (Phithakkitnukoon et al. 2012). Therefore, mobile phones

seem to be one of the most appropriate data sources to simultaneously analyze social network and travel behavior. At this point, it is worth mentioning that, as well as data from surveys, mobile phone data have their own limitations and drawbacks (e.g. limited socio-demographic information available due to privacy issues), which will be discussed at the end of this paper.

Travel behavior and mobile phone data

Recent studies from the human and social research area have demonstrated the usefulness of mobile phone data to study travel behavior. González et al. (2008), Song et al. (2010a, b), and Bagrow and Lin (2012) have demonstrated that human mobility is highly structured and governed by certain patterns. Slim and Ahas (2010) used mobile phone positioning data to identify individuals' residential locations in Estonia. Ahas et al. (2010) also monitored the movements of suburban commuters in the city of Tallinn, Estonia. Mobile phone positioning data has been used to study how people move during social events (Calabrese et al. 2010). Song et al. (2010a) studied the predictability of human mobility from location data of GSM tower IDs. Becker et al. (2011) identified residential location of daily workers and the late-night revelers in the city of Morristown, New Jersey, USA, in order to understand daily flows of people in and out of city. Isaacman et al. (2011) used call detail records (CDRs) to identify locations where people spend most of their time. They validated the algorithms used, derived via logistic regressions, by comparing their results to ground truth data provided by a group of volunteers. The algorithms identified home and work sites with median errors under one mile. Do and Gatica-Pérez (2012) developed algorithms to predict user mobility using various types of data collected from mobile phones of 153 volunteers during 17 months (GPS, WiFi APs, calling logs, etc.). In the transport field, research interest in relation to mobile phone-based data has been concentrated on using mobile phones as probes for estimation of aggregate level traffic parameters, such as travel time and travel speed (Bar-Gera 2007), mode share (Wang et al. 2010; Doyle et al. 2011), origin–destination matrices (White and Wells 2002; Cáceres et al. 2007; Sohn and Kim 2008; Calabrese et al. 2011b) and traffic flows (Cáceres et al. 2012). Reviews of current practices using mobile phone as traffic probes can be found in Yim (2003), Rose (2006), Cáceres et al. (2008) and Steenbruggen et al. (2011).

Social networks, travel behavior and mobile phone call data

Communication information from mobile phone data can be used to infer social network structures. For example, Eagle et al. (2009) used call logs, Bluetooth devices in proximity, cell tower IDs, application usage and phone status collected from mobile phones of 94 volunteers to study friendship behaviors. Mobile phones were equipped with software applications that recorded and sent the data to a central server. The analysis of the mobile phone data was compared with self-reported data. They found that friendship is related to in-role communication and proximity (those interactions likely to be associated with work, e.g. proximity at work), as well as with extra-role communication and proximity (those interactions that are unlikely to be associated with work, such as Saturday night proximity). Using just the extra-role communication factor from that analysis, it was possible to accurately predict 96 % of symmetric non-friends (subjects who work together but neither considers the other a friend) and 95 % of symmetric friends; in-role communication produced a similar accuracy. Thus they could accurately predict self-reported friendships based only on objective measurements of behavior. Landline and mobile phone data were

used by Sobolevsky et al. (2013) as a proxy for interactions to identify community regions. They detected coherent areas, and most of their boundaries closely follow existing political or socio-economic borders.

A considerable number of studies have used mobile phone data to either analyze social network or travel behavior. However, studies using mobile phone data to jointly analyze social networks and travel behavior are scarce. Phithakkitnukoon et al. (2011) identified residential locations of individuals with mobile phone positioning data and quantified the strength of social ties based on call duration. They found that residential migration can affect the strength of social ties over time: strong ties persist through a migration, while weak ties tend to disappear. In a subsequent study (Phithakkitnukoon et al. 2012), the authors found that 80 % of individuals' mobile phone traces were within the 20 km proximity of their nearest social ties' residential locations. Calabrese et al. (2011a, b) used a subset of mobile phone data from 1 million users in Portugal to study the relationship between their telecommunications patterns and physical locations. They found that there was a strong positive correlation between the call frequency between two individuals and the frequency of co-location occurrences. Cho et al. (2011) studied social travel using cell phone location data (estimated from the nearest cell phone tower of both the persons making and receiving the call), and data from two online location-based social networks. Ythier et al. (2013) used data from phone calls, sms logs and GPS of 111 people to investigate the influence of communication and social contacts on travel behavior. They found that people tend to travel in a similar manner as those they are socially connected to (consistently with the social network and travel literature) and that communication use is a complement to physical travel (consistently with the telecommunication and travel literature). Chen and Mei (2014) identified social ties and characterized basic mobility patterns using a mobile phone dataset of around 425,000 users with both location and calling information for a large urbanized city in China.

Scope and paper contribution

The use of mobile phone data to analyze social network and travel behavior interaction is gaining interest due to its potential to identify social and travel behavior patterns based on a large sample of individuals. This source of data has the advantage of being collected passively, with no human errors, no non-response and no fatigue/attrition. The main purpose of this paper is to contribute to efforts in this area by focusing on two main aspects: (1) the relationship between social network and frequent locations visited by social network individuals, and (2) the analysis of co-location, defined as the events in which two individuals of the same social network are in the same place at the same time. Note that when analyzing frequent locations of the social network, co-location is not strictly required.

With respect to frequent locations, research on social networks and travel behavior has mainly focused on home locations, taking the spatial proximity of residential locations as a proxy of social interaction, although some studies have also addressed the distribution of the work location of the social network (e.g., Phithakkitnukoon et al. 2012). Also, mobile phone data have been analyzed to infer home and work locations, defined as the most frequent locations in a certain period of time. In this paper, in addition to home and work locations, a methodology is proposed to identify other frequent locations. Additionally, instead of analyzing the spatial distribution of frequent locations of the social network, the

focus is on the analysis of the nature (i.e. home, work, other location) of the locations shared by the individuals of the social network, aiming to understand if the reason why the user is at that location is influenced by its social contacts.

Regarding co-location, few studies have faced this problem through the use of mobile phone data. Calabrese et al. (2011a, b) identified a co-location event when two individuals who are in the same area (different from users' home and work) call each other, which is seen as coordination between them to meet in a nearby area. Chen and Mei (2014) concluded that other attributes related to mobility patterns such as co-location are equally or even more strongly related to social interaction than the spatial distribution of residential locations of the social network; co-location is assumed to occur when two individuals are in the same place during a time frame, dividing every day into two time frames, daytime (8 a.m.–8 p.m.) and night-time (8:01 p.m.–7:59 p.m.). In the present paper, a novel methodology to analyze co-location is proposed. For each individual, a mobility model identifying locations visited along the different days of the sample is defined. By crossing the mobility models of the individuals, a co-location event is identified if two users of the same social network are in the same place at the same time. The methodology proposed allows the identification of co-location events even if there is not phone communication between the individuals and with a high temporal resolution. Similarly to frequent locations analysis, one of the main objectives is to analyze the nature of co-location events.

To the best of our knowledge, the dataset used for this study (described in detail below) is the largest one considered so far to analyze the interaction between the social network and travel behavior.

Dataset

The mobile phone data used for this study consists of a set of CDRs. CDRs are generated when a mobile phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS, etc.). For invoicing purposes, the information regarding the time and the base transceiver station (BTS) tower to which the user was connected when the call was initiated and ended is logged, providing an indication of the geographical position of the user at certain moments. No information about the exact position of a user in the area of coverage of a BTS is known. Also, no information about the location of the cell phone is known or stored if no interaction is taking place. The CDRs used in this study were collected for Spain, comprising anonymous call information for around 24 million users, accounting for more than 50 % of the 2009 Spanish population. The CDRs cover a period of time from September to November 2009 consisting of 53 days (including weekdays and weekends) which provide more than 10 billion spatiotemporal registers. From the information contained in each CDR, the following call information was extracted: caller's anonymous ID, callee's anonymous ID, day of the call, time when the call starts, duration of the call, caller's connected tower when the call starts and caller's connected tower when the call ends. Users' positions are collected from BTS towers around Spain, leading to a location accuracy of few hundreds of meters in urban areas and several kilometers in rural areas due to the different density of towers. In order to preserve privacy, original records were encrypted. Additionally, all the information presented in this paper is aggregated. No contract or demographic data were available for this study. None of the authors of this study participated in the encryption or extraction of the CDRs.

Methodology

In this section we explain the methodology followed to: (1) determine the social network of the users, (2) identify the frequent locations visited by each user, (3) develop user mobility models, (4) analyze the interactions between social network and frequent locations and (5) analyze co-location events.

Social network

The determination of the social network of each user is based on an egocentric network approach, leading to a network of users (alters) with whom the main user (ego) has some relation. It has been considered, as in other similar studies (Onnela et al. 2007; Phithakkitnukoon et al. 2012; Chen and Mei 2014), that a relationship between two different users only exists if the phone communication between them is reciprocal. Therefore, the social network of the ego is defined as a set of nodes (one node per user) and undirected connections or links between them representing reciprocal calls. In order to measure the strength of these relations, links have been weighted by the total numbers of calls between users.

Frequent locations

User frequent locations are defined as those places repetitively visited by the user along a certain period of time. Previous studies identifying frequent locations based on mobile phone data have mainly focused on home and work locations (e.g., Isaacman et al. 2011; Phithakkitnukoon et al. 2012; Chen and Mei 2014). In this paper, other relevant locations are additionally considered. Some frequent locations are hard to identify due to their particular spatiotemporal characteristics: for instance, it seems that a place where a person goes swimming all Mondays should be considered as a frequent location; however, if the frequency is measured on a weekly or monthly basis (i.e., 7 or 30 days) this location would probably be wrongly discarded. To give response to this problem and maximize the identification of relevant frequent locations, different criteria have been defined. A location is considered frequent if the user appears at that location a minimum number of days on a single day basis; on a working day basis, considering working days from Monday to Thursday; or on a weekend basis, considering weekend from Saturday to Sunday. Fridays have intentionally not been classified neither as working days nor weekend due to their particular mixed characteristics. The minimum number of days (minimum frequency) to consider a location as a frequent location is determined by the following expression:

$$Minimum\_frequency = \alpha \cdot total\_sample\_days$$

where 'α' is a reduction coefficient and 'total_sample_days' is the total number of days of a certain type present in the sample (e.g., total number of Mondays, total number of working days, etc.). The alpha coefficient determines the ratio between the minimum number of appearances on days of a certain type (single day, working day, etc.) and the total number of days of that type present in the sample. It is important to note that the frequency of appearance of a user at a certain location is in most cases underestimated, since the user will only appear in that position if he/she makes or receives a call. Therefore, these considerations about the nature of the data have to be taken into account when

selecting the value of the alpha coefficient. As a first approach, an alpha coefficient of 0.35 has been considered adequate to estimate frequent locations.

Additionally, frequent locations have been classified into three different groups: *home*, *work* and *other*. *Home* and *work* locations are estimated considering only working days. A frequent location is classified as *home* if it is the most frequent location between 8 p.m. and 7 a.m. Similarly, *work* location is considered the most frequent location between 8:00 a.m. and 5 p.m. Finally, all other frequent locations different from *home* and *work* are classified as *other* locations. Note that a single location can be classified simultaneously as *home* and *work* location. In contrast to *other* frequent locations, as *home* and *work* locations present time restrictions, it seems reasonable that lower alpha coefficients should be considered in these cases. Moreover, as the effective hours (hours when there exists a high probability of making or receiving a call) considered for *home* locations are lower than those considered for *work* locations, the alpha coefficient considered for *home* locations should be lower. Under these considerations, alpha coefficients of 0.2 and 0.3 have been considered appropriate to estimate *home* and *work* locations respectively.

Mobility model for co-location analysis

CDRs provide, on average, spatiotemporal information of each user every several hours. This level of detail could be useful for analyzing individual daily mobility partners such as *home* and *work* trips; however, when analyzing co-location events between an individual and its social network, more detailed information is needed. To respond to this limitation, a mobility model which expands the spatiotemporal information present in the CDRs providing an estimation of the position of the user along the day has been developed (see Fig. 1). The mobility model for each user is defined as follows:

1. *CDR of the user = User's location and time information ($L0$, $t0$)*
2. *Next CDR of the user = User's location and time information ($L1$, $t1$)*
3. *if ($t1$–$t0$) > $T\_threshold$ ⤳ Location information missed from $t0$ to $t1$*
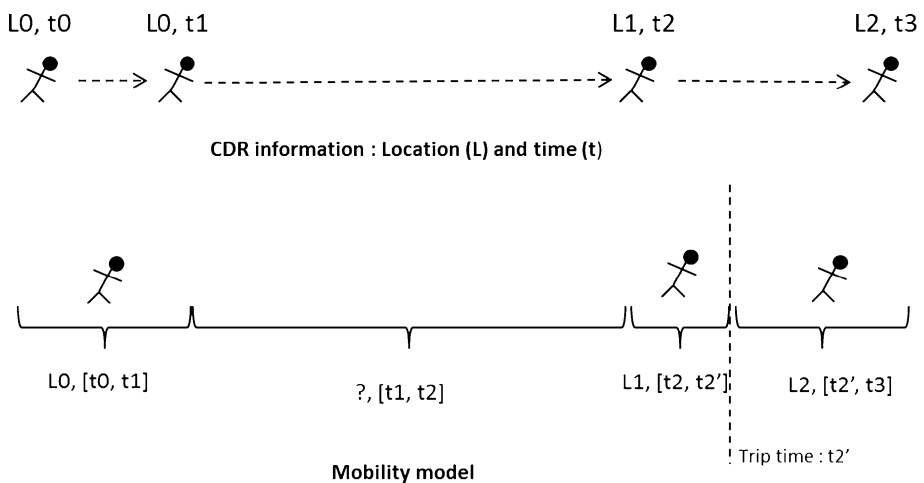

Fig. 1 Comparison between the information provided by CDRs and the information provided by the mobility model

4. *else:*

- *if L0 = L1 = L ⇢ User location between [t0, t1] is L*
- *else ⇢ t' = f(t0, t1)/User location is L0 between [t0, t'] and L1 between [t', t1].*

   T_threshold represents the maximum time distance between 2 instances (t0, t1) to consider that no relevant intermediate locations exist between those instances; and $f(t_i, t_j)$ is a probability function that determines the time when a trip is performed.

Social network and frequent locations analysis

The main objective of this analysis is to explore the relation between the frequent locations visited by the ego and those visited by its social network. Only users whose *home* and *work* locations have been identified are considered for the analysis (around 2,300,000 users). For each egocentric network, the frequent locations of the ego are compared with the frequent locations of the alters. The common frequent locations are identified and the type of relation between those locations is classified according to the characteristics of the locations shared. There are 9 possible types of relations derived from the combination of the three possible types of frequent locations [*home, work, other*]. As mentioned before, *home* and *work* could correspond to the same location; in these cases, the type of relation is proportionally assigned (e.g., user 'A' and user 'B' share a common location 'L', for 'A' location 'L' is simultaneously *home* and *work*, and for 'B' it's *other* location; the relation type will be classified as 50 % home-other and 50 % work-other).

Co-location analysis

The mobility models of the different users belonging to the same social network are compared to identify co-location events. In contrast to the previous analysis, both frequent and non-frequent locations are considered. The different locations are classified as *home*, *work*, *other* and *non-frequent*. For each egocentric network, the different locations of the ego and the alters are compared along the different days of the sample. Co-location is identified and classified according to the characteristics of the locations shared. There are 16 possible types of co-locations derived from the 4 types of locations. The co-location analysis has been performed using a subset of the whole dataset covering the users living in the metropolitan area of Barcelona and their social network (independently of the place of residence), leading to a sample of around 250,000 users. Note that, as in the previous analysis, the mentioned sample only considers users whose *home* and *work* locations have been identified.

Results and discussion

Social network statistics

From the whole sample of CDRs, around 24 million of egocentric networks have been identified. The average number of alters per ego is 9.31 with a standard deviation of 17.19. The average number of phone calls between two users (considered as a proxy of the strength of the social relation) is 21. Ninety percent of the egos have less than a phone call per day with each alter.

Frequent locations statistics

For each user of the sample, the frequency of appearance of his/her locations has been calculated. The average number of frequent locations per user is 3.47 with a standard deviation of 2.83. Considering that the minimum number of frequent locations is two (*home* and *work*), on average every user has 1.5 *other* locations which could be associated to social activities. This result shows that there are other frequent locations, apart from the commonly considered *home* and *work*, whose importance (measured as the number of locations) is similar to the non-social ones and in some cases even more important (based on the standard deviation results). This result supports the idea that considering other activities different from *home* and *work* is essential to properly capture users' travel behavior. Figure 2 shows the distribution of *other* locations according to the day of the week. Most of the *other* frequent locations have been identified on Tuesdays, Thursdays and Fridays. Moreover, there are some locations that are at the same time frequent on a working day basis and on a weekend basis.

To validate the methodology used to estimate *home* locations, a correlation analysis comparing the results and the 2009 population distribution of Spain has been performed. The results are compared at province level (52 provinces), showing a high correlation, with $R^2 = 0.93$ (see Fig. 3). Similarly, to validate the relation between *home* and *work* locations (commuting trips) a correlation analysis comparing the results obtained and the 2011 census for the metropolitan area of Barcelona has been carried out. The results are compared at municipal level (36 municipalities) providing an $R^2 = 0.99$ (see Fig. 4). According to correlation results, the alpha coefficients used to estimate *home* and *work* locations seem to be adequate. The alpha coefficient for *other* locations is more difficult to validate because of the fact that there are no relevant statistics available. However, since *home* and *work* locations coefficients seem to be adequate, a value of 0.35 for *other* locations seems reasonable. It is important to highlight that the alpha
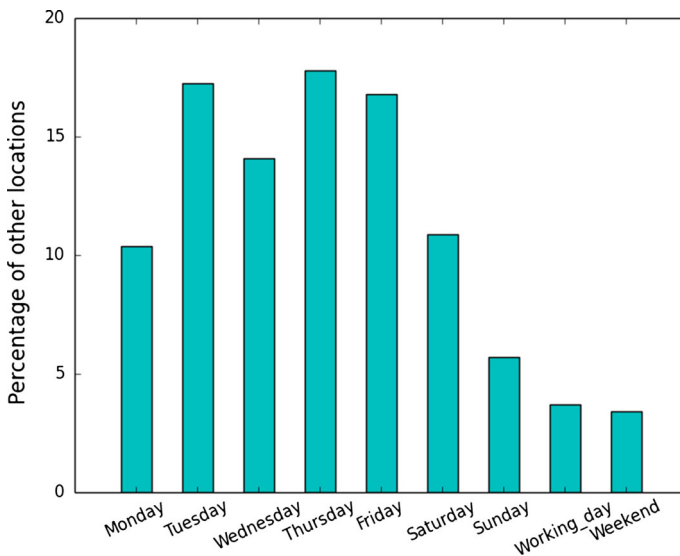

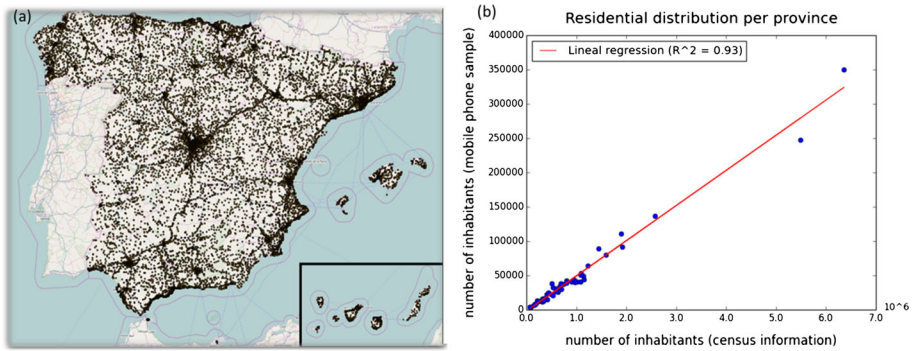**Fig. 2** Distribution of other frequent locations according to the day of the week

Fig. 3 **a** Home distribution based on mobile phone data analysis **b** Correlation analysis between 2009 census population information and mobile phone results
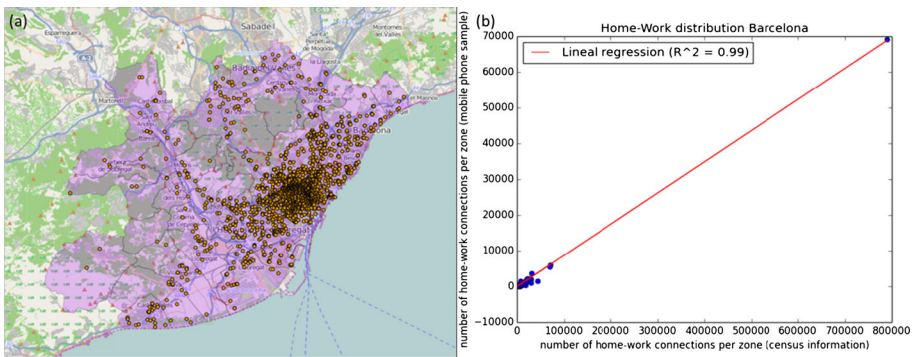

Fig. 4 **a** Home and work locations of the metropolitan area of Barcelona obtained from mobile phone data analysis **b** Correlation analysis between 2011 census Barcelona information and mobile phone results

coefficients proposed are appropriate for the temporal and spatial resolution of this dataset; and that other similar alpha coefficients applied to this dataset may also lead to good results. To determine which range of alpha values is appropriate for each type of location, a sensitivity or robustness analysis would be needed, being this analysis out of the scope of the present paper.

Mobility model

Mobility models have been defined for the residents of the metropolitan area of Barcelona and their social networks (250,000 users). The probability function $f(t_i, t_j)$ used to determine the time when the trip is performed has been extracted from the trip statistics presented in the 2009 mobility survey of the metropolitan area of Barcelona ('Enquesta de Mobilitat en Día Feiner', EMEF 2009). A *T_threshold* of 4 h has been considered appropriate for the analysis. User mobility models provide on average 4.6 h of location information on weekdays and 2 h on weekends (2.5 h on Saturdays and 1.5 h on Sundays), distributed as shown in Fig. 5.
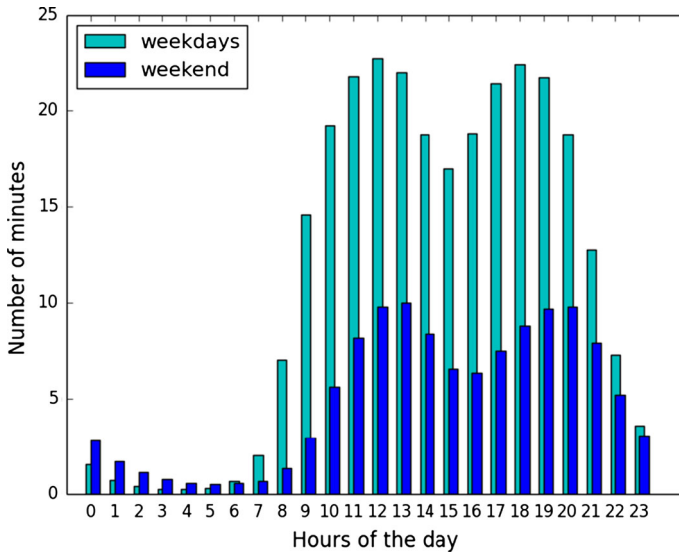
Fig. 5 Time coverage of mobility models on weekdays and weekend

Social network and frequent locations interaction results

From the 24 million egocentric networks in the sample, only those in which the ego has *home* and *work* location information are considered for the analysis (around 2.3 million). Similarly, only alters with information about their frequent locations are considered (for each egocentric network, around 17 % of the alters provide that information, with a standard deviation of 13 %). Results show that each ego shares, on average, at least one frequent location with 61.23 % of the alters (standard deviation of 36.88 %), suggesting a significant relationship between the social network and the frequent locations visited by the users. Egos share 1.36 frequent locations with each of the alters, with a standard deviation of 0.94. Considering that users have 3.47 frequent locations, each ego shares 40 % of those locations with each of the alters of the network. This result shows that not only the number of alters who share common locations with the ego is significant but also the number of common locations between the ego and each alter. Moreover, it is observed a strong positive correlation ($R^2 = 0.97$) between the average number of phone calls and the number of frequent locations in common; suggesting that as the strength of the relation increases (number of phone calls between the ego and the alters), so does the number of common locations (see Fig. 6).

For each of the common locations between the ego and the alters, the type of interaction between them is identified and classified according to the types of the shared locations. For example, if an ego lives in location A, and the alter visits the ego frequently (so position A is an *other* frequent location for the alter), the type of interaction between them will be classified as *home-other*. A total of 9.6 million interactions ego-alter were identified and analyzed. Table 1 shows the distribution of the types of interaction between the ego and the alters. From the ego's point of view, most of the interactions with the alters occur in *other* frequent locations (54 %) and to a lesser extent in *home* and *work* locations (21 and

Fig. 6 Correlation between the average number of phone calls between two users and the number of frequent positions in common
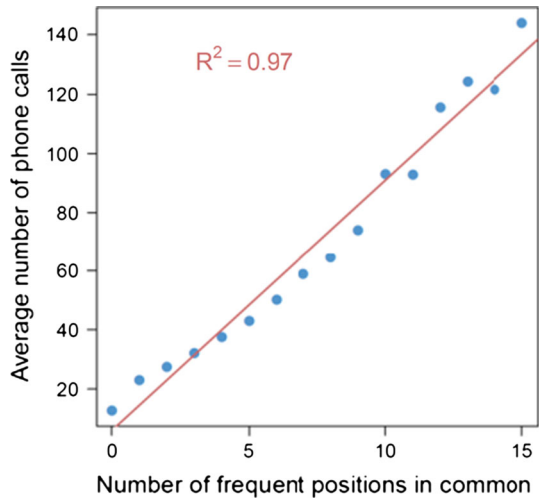

Table 1 Distribution of the ego-alter frequent locations interaction types

Ego/Alters	Home (%)	Work (%)	Other (%)	Total (%)
Home	7.41	6.42	7.62	21.45
Work	6.42	9.07	8.58	24.07
Other	7.62	8.58	38.28	54.48

24 %, respectively). The most common type of interaction (38.28 %) between the ego and the alters is one in which they share an *other* location different from their home and work.

Social interactions can be associated to those where are least one of the locations is classified as *other*. Results show that at least 27 % (standard deviation 32 %) of the ego's *other* locations are shared with alters, being 14, 16 and 70 % the probabilities that the location corresponds to the *home*, *work* and *other* location of the alter respectively. It is said 'at least' because only a percentage of the alters provide location information and therefore some shared frequent locations may be missing, and consequently the percentage of *other* locations shared is probably underestimated.

Co-location analysis results

Co-location events have been analyzed for the metropolitan area of Barcelona during a period of 53 days from September to October 2009. The average number of appearances per user along the 53 days is 58.3 with a standard deviation of 51.29. Those locations are not necessarily places where the user performs an activity, but they could also be locations along a certain trip. From an ego's perspective, the number of locations shared per alter is 8.75 (standard deviation of 3.94), corresponding to 15 % of the ego's locations. From those 8.75 common locations, 1.22 locations (standard deviation 0.74) are co-located locations. Each of those co-located locations has produced on average 4.36 co-location events along the sample, with a standard deviation of 4.17. It is important to note that, as it happened

when analyzing common frequent locations, co-location events are probably underestimated since mobility models do not cover the 24 h of the day.

For each co-located location, the type of the interaction between the ego and the alter has been classified according to the types of the location shared (*home, work, other* and *non-frequent* location). Around 1.4 million interactions ego-alter were identified and analyzed. Table 2 shows the distribution of the types of interaction between the ego and the alters. From the ego's point of view, a significant number of co-located locations are *non-frequent* locations (43 %) and *other* frequent locations (30 %), and to a lesser extent *home* (11 %) and *work* (16 %) locations.

Most of the places where egos co-locate are ego's frequent locations (57 %), being 19.5, 28.5 and 52 % the probability of co-location at ego's *home, work* and *other* locations respectively when co-location occurs at a frequent location. These percentages are quite similar to those obtained when analyzing the types of interaction of ego's frequent locations (21.5, 24 and 54.5 % for *home, work* and *other* locations respectively). At first glance, this result might seem obvious since the more the frequent location interactions of a specific type, the higher the probability of co-locating in this type of interaction. However since the determination of *other* frequent locations does not consider the variable time, it could be the case that an ego and an alter sharing several *other* positions do not co-locate in any of them. This result also supports the hypothesis that *other* frequent locations could be associated to places where individuals of the same social network interact (co-locate), which suggests that the distribution of the type of interaction ego-alter when analyzing frequent locations could be used as a proxy of the probability of an ego to co-locate in its different types of frequent locations when co-location occurs in a frequent location. This is of considerable importance since the calculation of the type of interaction ego-alter considering frequent locations is simpler and less time consuming than the co-location analysis.

On the other hand, although co-location in frequent locations is majority, *non-frequent* locations are almost equally important (43 %). When an ego co-locates at a *non-frequent* location, there is a probability of 8.1, 11.3, 22.8 and 57.8 % that this location corresponds to the alter's *home, work, other* and *non-frequent* location respectively. *Non-frequent* locations can be seen as destinations rarely visited by the users. These locations will be hardly explained by transport models which only consider generalized travel costs (time, economic cost, etc.) and omit the influence of the social network. In almost half of the cases (42 %), when the ego co-locates at a *non-frequent* location, it is because that location corresponds to a frequent location of the alter. In the rest of the cases, both (the ego and the alter) shared a *non-frequent* location. The most common type of ego-alter interaction (24.81 %) is one in which they share a *non-frequent* location. This type of interaction (*non-frequent-non-frequent*) can be seen as a joint decision between the ego and the alter to decide a place to meet different from their frequent locations.

Table 2 Distribution of the ego-alter co-location interaction types

Ego/Alter	Home (%)	Work (%)	Other (%)	Non-frequent (%)	Total (%)
Home	2.51	2.14	2.92	3.48	11.05
Work	2.14	5.15	4.11	4.86	16.26
Other	2.92	4.11	12.94	9.78	29.75
Non-frequent	3.48	4.86	9.78	24.81	42.94

Mobile phone data discussion: characteristics and limitations

Although it has been shown that mobile phone call data have the potential to provide rich information about the interaction between social networks and travel behavior, they are not free of drawbacks and limitations.

First, it is important to remark the limitations associated to the fundamental nature of mobile phone data. The use of mobile phone call data to identify social relationships inevitably misses other kind of interactions conducted through other communication channels, such as face to face or e-mail interactions. Moreover, it may misidentify certain interactions as social, such as sporadic work relationships. These intrinsic limitations may lead to errors in the estimation of some social network variables (e.g., number of social contacts) and may introduce bias in some analysis (e.g., when non-social interactions are considered in the analysis).

Apart from these intrinsic limitations, the temporal and the spatial resolution of the data highly influence the results:

- The temporal resolution of the data can be defined as the quantity of data available per unit of time. For this study, information is only available when the mobile phone user makes/receives a call. Therefore, it could be the case that a frequent location is missed or misclassified as non-frequent location if the number of calls made or received at that location is not a good proxy of the time spent by the user at that location. Likewise, some social interactions may be missed if information is not available when that social interaction is occurring.
- The spatial resolution of the data determines the accuracy in the estimation of the user position. In this study, the spatial resolution corresponds to the size of the Voronoi area associated to each BTS, which varies from few hundreds of meters in urban areas to kilometers in less populated areas. This research assumes that two users co-locate if they are in the same Voronoi area, which may lead to overestimate co-location, especially in less populated areas. Additionally, from a social perspective, it is important to remark that co-location does not ensure social interaction. Just because two users are in the same area at the same time, it is not possible to know with certainty that a social interaction between them is taking place.

The improvement of the temporal and spatial resolution of the data will lead to more accurate results. Temporal resolution can be improved by recording other type of registers apart from calls, such as text messages or Internet connections, while spatial resolution can be enhanced by means of signal triangulation, WiFi or GPS information.

Finally, apart from data characteristics (intrinsic and extrinsic), the representativity of the mobile phone data sample is a key question. The sample has to be of enough size and homogeneously distributed among the population in order to minimize bias. The mobile phone data analyzed for this study accounts for more than 50 % of the 2009 Spanish population and it is homogeneously distributed across the territory, as the comparison with census information confirms. However, as socio-demographic information is missing, some population profiles may not properly be represented in the sample. For future studies, this problem could be mitigated if the dataset provided by the mobile phone operator included some basic sociodemographic parameters available to the mobile phone company, such as age, gender, etc.

In summary, mobile phone data open an opportunity to better understand the relationship between social network and travel behavior. However, the characteristics and

limitations mentioned above have to be considered when analyzing and interpreting the results in order to devise how to effectively use mobile phone data to complement the insights gained from traditional surveys.

Applications to the transport sector

Activity-based models: improvement of travel behavior modelling

The results obtained from the joint analysis of users' social network and travel behavior provide relevant information to enrich activity-based models. Indeed, one important challenge for operational daily mobility models is the prediction of location choice for *discretionary* (as opposed to mandatory) activities: while home and work locations can typically be obtained from reliable sources, such as census, the high flexibility of discretionary types makes them much more difficult to handle, the tendency being to *underestimate* traveled distances for those purposes.

In the past, various approaches have been proposed to tackle this problem. The first one, building on the classical *random utility framework*, proposes to account for *unobserved heterogeneity* in location characteristics and individual tastes using *random error terms* for each agent-location pair (see e.g. Horni 2013). This approach yields pretty good results, but has two main drawbacks:

- it substitutes an explanation of *why* individuals travel further than expected by random noise,
- the choice of the location being independent across agents, it is unable to represent *joint traveling to a joint activity*. Not only does this represent a substantial part of travel, but it is of prime interest for forecasting the impact of policies aiming at effecting *car occupancy*.

For those reasons, a second approach to discretionary activity location choice has been proposed, which takes into account the *willingness to pass time with social contacts* in the utility an agent derives from its daily plan (Axhausen 2005). The basic idea is the following: the choices of an agent result from a tradeoff between the benefits it derives from performing activities and the generalized cost (money, time, etc.) of the associated trips. For instance, the MATSim software platform (www.matsim.org) considers utility-maximizing agents, trying to get the most of their day given travel times (influenced by others via congestion). The basic utility function simply separately scores activity performance and travel, and sums the resulting values:

$$V = \sum_i V_i perf + \sum_j V_j leg$$

where $V_i perf$ is the reward (normally positive) to perform an activity, and $V_j leg$ the penalty (normally negative) of traveling. As long as the marginal utility of travel time is lower than the marginal utility of performing an activity, agents have an incentive to perform shorter trips. If the utility derived from an activity is allowed to vary depending on *who* participates, however, agents may get an incentive to travel further to meet social contacts—possibly reproducing the tendency elicited by the analysis in the previous section. This *joint* location choice is a first, necessary step, to include *joint travel to joint activities* in a simulation framework.

This comes however at high cost: the universal destinations choice set is enormous, and the multi-objective aspect of the problem requires the usage of non-traditional *solution concepts* to represent joint decisions (see Dubernet and Axhausen 2014, for a comparison of two solution concepts to simulate household mobility, or Ronald et al. Ronald et al. 2012a, 2012b and Ma et al. (2011, 2012) for rule-based simulated bargaining approaches). The interaction patterns between the social network and travel behavior obtained from the mobile phone data analysis may however help to make this kind of simulations tractable, for instance using the following steps:

- Consider as possible destinations the frequent locations of an agent's social contacts, since results show that co-location frequently occurs in those locations.
- Look into the intersection between isotims of the social network (line of equal transport cost), in order to find possible destinations (probably non-frequent destinations) that maximize users utility. This proposal is based on the fact that co-location events also take place at non-frequent locations.
- Use the kind of patterns obtained from the mobile phone data analysis to calibrate/validate the model.

Figure 7 shows an example of how the introduction of social interaction in activity-based models could influence the results. As results show, there is a significant probability that individuals of the same social network share *other* frequent locations. If no social interaction is considered, agents will take their own decisions and select the *other* location as a function of the benefits and costs they get. However, if social interaction is considered, there is a probability that an agent chooses the same *other* location as a contact of its social network even when that decision implies more generalized travel costs.

Transport policy applications

There are several policy applications where social interaction together with travel behavior information can be useful for policy planning and assessment purposes. A better understanding of the influence of the social network on travel behavior and the availability of transport modelling tools taking into account these considerations are important for evaluating transport policies where social interaction is relevant. Such transport policies are usually related with services in which transport resources (vehicles) are shared. Some
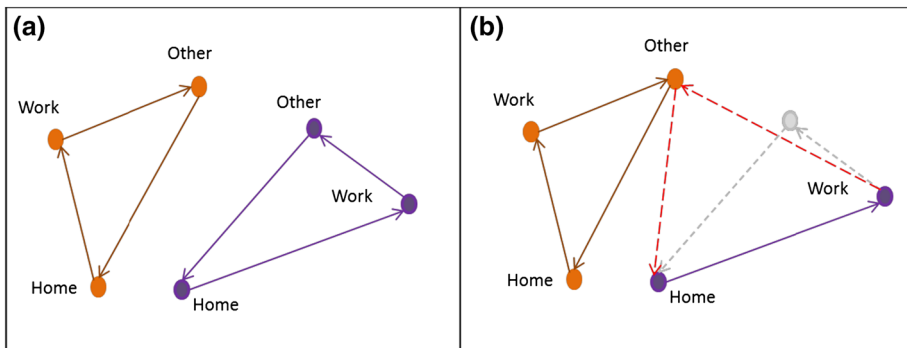

Fig. 7 Diary of activities and trips of 2 agents of the same social network during a standard day: **a** Model results without considering social network influence; **b** Model results considering social network influence

examples of policies where the approach proposed in this paper can be useful are shown below:

- Transport on demand: transport on demand aims to minimize the underutilization of the transport services by dynamically adjusting supply to demand. For example, it is possible to identify frequent locations where people go out on Saturday night and determine their places of residence. Depending of the distribution of homes and the time variance of return trips, the impact (congestion, safety, etc.) of a new collective transport which maximizes the usage (number of passengers) and minimizes the cost (minimum route) could be assessed.
- Carpooling: it is usually recognized that people belonging to the same social network are more conducive to sharing transport resources. In the case of carpooling, social interaction is especially important for several reasons: the driver will probably not be a professional driver; people may feel uncomfortable sharing a car with people they don't know, etc. From the information of the home locations of the social network and the possible destinations (frequent or not), it is possible to evaluate if sharing a car could be beneficial for them.

Conclusions

It is widely recognized that social contacts have a significant influence on individual's travel behavior. Most decisions about where to perform an activity are related to the social network. This paper contributes to better understand the way social network influences travel behavior by analyzing the nature (*home, work, other, non-frequent*) of the locations shared by social contacts using mobile phone data, showing the potential of this non-conventional data source to provide relevant information on both social interaction and travel behavior.

From the crossing analysis of social networks with frequent locations and mobility models, relevant statistics about mobility patterns and the nature of locations shared by social contacts have been obtained. The results support the hypothesis that *other* frequent locations of individuals can be considered as potential places where users of the same social network interact. Moreover, it has been shown that most of the co-location inter-actions are those related to ego's *non-frequent* and *other* frequent locations. Indeed, the most common type of ego-alter interaction is one in which they share *non-frequent* lo-cations. Additionally, the potential value of these results to inform activity-based models and assess transport policies in which transport resources are shared has been discussed.

Despite the potential of mobile phone data to provide rich information about the in-teraction between social networks and travel behavior, a number of drawbacks and limitations shall be taken into account, such as the high spatio-temporal heterogeneity of the data or the lack of socio-demographic information. These shortcomings and limitations have been analyzed in depth. Data fusion with other data sources is a promising approach to fill the gaps of information (such as socio-demographic gaps) as well as to validate the results.

This research has thrown up many questions in need of further investigation. An in-teresting future line of research is the analysis of the length distribution of the trips derived from the social activities distinguishing between the different types of interactions ego-alter (e.g. *non-frequent–other*). Especially interesting is the case in which both users share

a *non-frequent* location, aiming to explore if there is any kind of joint decision among them looking for a mutual benefit.

References

Ahas, R., Aasa, A., Silm, S., Tiru, M.: Daily rhythms of suburban commuters' movements in the tallinn metropolitan area: case study with mobile positioning data. Transp. Res. Part C **18**, 45–54 (2010)

Arentze, T.,Timmermans, H. J.: social networks, social interactions and activity-travel behavior: a framework for micro-simulation. Paper presented at the 85th annual meeting of the Transportation Research Board, Washington, D. C., Jan 2006 (2006)

Arentze, T., Timmermans, H.: Social networks, social interactions, and activity-travel behavior: a framework for microsimulation. Environ. Plan. **35**, 1012–1027 (2008)

Axhausen, K.W.: Social networks and travel: some hypotheses. In: Donaghy, K.P., Poppelreuter, S., Rudinger, G. (eds.) Social Aspects of Sustainable Transport: Transatlantic Perspectives, pp. 90–108. Ashgate, Aldershot (2005)

Bagrow, J.P., Lin, Y.-R.: Mesoscopic structure and social aspects of human mobility. PLoS One **7**(5), 1–11 (2012)

Bar-Gera, H.: Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from israel. Transp. Res. Part C **15**(2007), 380–391 (2007)

Becker, R.A., Cáceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., Volinsky, C.: A tale of one city: using cellular network data for urban planning. Pervasive Comput. IEEE **10**(4), 18–26 (2011)

Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature **439**, 462 (2006)

Caceres, N., Wideberg, J.P., Benitez, F.G.: Deriving origin–destination data from a mobile phone network. IET Intell. Transp. Syst. **1**(1), 5–26 (2007)

Caceres, N., Wideberg, J.P., Benitez, F.G.: Review of traffic data estimations extracted from cellular networks. IET Intell. Transp. Syst. **2**(3), 179–192 (2008)

Caceres, N., Romero, L.M., Benitez, F.G., Castillo, J.M.D.: Traffic flow estimation models using cellular phone data. IEEE Trans. Intell. Transp. Syst. **13**(3), 1430–1441 (2012)

Calabrese, F., Pereira, F. C., Lorenzo, G. D., Liu, L., Ratti, C.: The geography of taste: analyzing cell-phone mobility and social events. In: Proceedings of IEEE International Conference on Pervasive Computing (2010)

Calabrese, F., Smoreda, Z., Blondel, V.D., Ratti, C.: Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. PLoS One **6**(7), e20814 (2011a). doi:10.1371/journal.pone.0020814

Calabrese, F., Lorenzo, G.D., Liu, L., Ratti, C.: Estimating origin-destination flows using mobile phone location data. Pervasive Comput. IEEE **10**(4), 36–44 (2011b)

Carrasco, J.A., Miller, E.J.: Exploring the propensity to perform social activities: social networks approach. Transportation **33**, 463–480 (2006)

Carrasco, J.A., Hogan, B., Wellman, B., Miller, E.J.: Collecting social network data to study social activity-travel behaviour: an egocentric approach. Environ. Plan. B **35**(6), 961–980 (2008a)

Carrasco, J.A., Hogan B., Wellman B., Miller E. J.: Agency in social activity and ICT interactions: The role of social networks in time and space, Tijdschrift voor Economische en Sociale Geografie (J. Eco. Soc. Geogr.), **99**(5), 562–583 (2008b)

Carrasco, J.A., Miller, E.J., Wellman, B.: How far and with whom do people socialize? Empirical evidence about the distance between social network members. Transp. Res. Rec. **2076**, 114–122 (2008b)

Carrasco, J.A., Miller, E.J.: The social dimension in action: a multilevel, personal networks model of social activity frequency. Transp. Res. Part A **43**(1), 90–104 (2009)

Chen, C., Mei, Y.: Does distance still matter in facilitating social ties? The roles of mobility patterns and the built environment. Presented at 93rd TRB annual meeting (2014)

Cho E., Myers S.A., Leskovek J.: Friendship and mobility: user movement in location-based social networks. In: KDD '11 Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1082–1090 (2011)

Clifton, K.J.: The social context of travel behavior. In: Zmud, J., et al. (eds.) Transport Survey Methods: Best Practice for Decision Making, pp. 441–448. Emerald Press, London (2013)

Do T., Gatica-Perez D.: Contextual conditional models for smartphone-based human mobility prediction. In: Proceedings ACM International Conference on Ubiquitous Computing, Pittsburgh, Sept (2012)

Doyle, J., Hung, P., Kelly, D., Mcloone, S., Farrell, R.: Utilising mobile phone billing records for travel mode discovery. ISSC 2011, Trinity College Dublin, June (2011)

Dubernet, T., Axhausen K. W.: Solution concepts for the simulation of household-level joint descision making in multi-agent travel simulation tools, paper presented at the 14th Swiss Transport Research Conference (STRC), Ascona (2014)

Dugundji, E., Walker, J.: Discrete choice with social and spatial network interdependencies: an empirical example using mixed GEV models with field and "panel" effects. Transp. Res. Rec. **1921**, 70–78 (2005)

Eagle, N., Pentland, A., Lazer, D.: Inferring social network structure using mobile phone data. Proc. Natl. Acad. Sci. (PNAS) **106**(36), 15274–15278 (2009)

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

Gould, J.: Cell phone enabled travel surveys: the medium moves the message. In: Zmud, J., et al. (eds.) Transport Survey Methods: Best Practice for Decision Making, pp. 51–70. Emerald Press, Bingley (2013)

Habib, K.N., Carrasco, J.A.: Investigating the role of social networks in start time and duration of activities: a trivariate simultaneous econometric model. Transportation Research Record: Journal of the Transportation Research Board **2230**, 1–8 (2011)

Hackney, Jeremy K., Kay W. Axhausen: An agent model of social network and travel behavior interdependence. Paper presented at the 11th international conference on Travel Behaviour Research, Kyoto, Aug (2006)

Hackney, J., Marchal, F.: A model for coupling multi-agent social interactions and traffic simulation, in: TRB 2009 annual meeting (2009)

Hackney, J., Marchal, F.: A coupled multi-agent microsimulation of social interactions and transportation behavior. Transp. Res. Part A **45**, 296–309 (2011)

Horni, A.: Destination choice modeling of discretionary activities in transport microsimulations, Ph.D. Thesis, ETH Zurich, Zurich (2013)

Isaacman, S.,Becker, R., Caceres, R., Kobourov, S., Martonosi, M., Rowland, J., Varshavsky, A.: Identifying important places in people's lives from cellular network data. In: Procedings International Conference on Pervasive Computing, San Francisco, June (2011)

Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. Commun. Mag. IEEE **48**(9), 140–150 (2010)

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational Social Science. Science **323**, 721 (2009)

Ma, H., Ronald, N., Arentze, T.A., Timmermans, H.J.P.: New credit mechanism for semicooperative agent-mediated joint activity-travel scheduling. Transp. Res. Rec. **2230**, 104–110 (2011)

Ma, H., Arentze, T. A., Timmermans, H. J. P.: Incorporating selfishness and altruism into dynamic joint activity-travel scheduling. Paper presented at the 13th international conference on Travel Behaviour Research (IATBR), Toronto, July (2012)

Marchal, F., Nagel, K.: Allowed cooperative agents in a microsimulation to share information with each other about activity locations and about other agents, in order to optimize trip chains (2006)

Molin, E.J.E., Arentze, T.A., Timmermans, H.J.P.: Social activities and travel demands : a model-based analysis of social-network data. Transp. Res. Rec. **2082**, 168–175 (2007)

Moore, J., Carrasco, J.A., Tudela, A.: Exploring the links between personal networks, time use, and the spatial distribution of social contacts. Transportation **40**(4), 773–788 (2013)

Onnela, J.-P., Saramaki, J., Hyvonen, J., Szabo, G., Lazer, D., et al.: Structure and tie strengths in mobile communication networks. Proc. Natl. Acad. Sci. U.S.A. **104**, 7332–7336 (2007)

Páez, A., Scott, D.M.: Social influence on travel behavior: a simulation example of the decision to telecommute. Environ. Plan. A **39**(3), 647–665 (2007)

Phithakkitnukoon, S., Calabrese, F., Smoreda, Z., Ratti, C.: Out of sight out of mind: how our mobile social network changes during migration. Proceedings of the IEEE International Conference on Social Computing, pp. 515–520. Cambridge University Press, Cambridge (2011)

Phithakkitnukoon, S., Smoreda, Z., Olivier, P.: Socio-geography of human mobility: a study using longitudinal mobile phone data. PLoS One **7**(6), e39253 (2012). doi:10.1371/journal.pone.0039253

Ronald, N.A., Arentze, T.A., Timmermans, H.J.P.: Modeling social interactions between individuals for joint activity scheduling. Transp. Res. Part B **46**, 276–290 (2012a)

Ronald, N.A., Dignum, V., Jonker, C., Arentze, T.A., Timmermans, H.J.P.: On the engineering of agent-based simulations of social activities with social networks. Inf. Softw. Technol. **54**(6), 625–638 (2012b)

Rose, G.: Mobile phones as traffic probes: practices, prospects and issues. Transp. Rev. **26**(3), 275–291 (2006)

Sharmeen, F., Arentze, T., Timmermans, H.: A multilevel path analysis of social network dynamics and the mutual interdependencies between face-to-face and ICT modes of social interaction in the context of life-cycle events. In: Roorda, M.J., Miller, E.J. (eds.) Travel Behaviour Research: Current Foundations, Future Prospects, pp. 411–432. Lulu Press, Toronto (2013)

Sharmeen, F., Arentze, T.A., Timmermans, H.J.P.: Dynamics of face-to-face social interaction frequency: role of accessibility, urbanization, changes in geographical distance and path dependence. J. Transp. Geogr. **34**, 211–220 (2014)

Silm, S., Ahas, R.: The seasonal variability of population in estonian municipalities. Environ. Plan. A **42**, 2527–2546 (2010)

Silvis, J., Niemeier, D., D'Souza, R.: Social networks and travel behavior: report from an integrated travel diary. Paper presented at the 11th international conference on Travel Behaviour Research, Kyoto, Aug (2006)

Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., et al.: Delineating geographical regions with networks of human interactions in an extensive set of countries. PLoS One **8**(12), e81707 (2013)

Sohn, K., Kim, D.: Dynamic origin–destination flow estimation using cellular communication system. IEEE Trans. Veh. Technol. **57**(5), 2703–2713 (2008)

Song, C., Koren, T., Wang, P., Barabási, A.-L.: Modelling the scaling properties of human mobility. Nat. Phys. **6**(2010), 818–823 (2010a)

Song, C., Qu, Z., Blumm, N., Barabási, L.-L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010b)

Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P., Scholten, H.: Mobile phone data from gsm networks for traffic parameter and urban spatial pattern assessment: A review of applications and opportunities. GeoJournal **78**, 223–243 (2011). doi:10.1007/s10708-011-9413-y

Van den Berg, P., Arentze, T., Timmermans, H.J.P.: A path analysis of social networks, telecommunication and social activity–travel patterns. Transp. Res. Part C **26**(2013), 256–268 (2013)

Wang, H., Calabrese, F., Lorenzo, G. D., Ratti, C.: Transportation mode inference from anonymized and aggregated mobile phone call detail records. In: 13th international IEEE annual conference on intelligent transportation systems, 318–323 (2010)

White, J. and Wells, I.: Extracting origin destination information from mobile phone data. Road transport information and Control, 19–21 Mar (2002)

Yim, Y.: The state of cellular probes. California PATH Working Paper, UCB-ITS-PRR-2003-25 (2003)

Ythier, J., Walker, J.L., Bierlaire, M.: The influence of social contacts and communication use on travel behavior: a smartphone-based study. In: Transportation Research Board annual meeting (2013)

Miguel Picornell graduated as a Civil Engineer from the Universitat Politècnica de València (UPV) and holds a Master's Degree in Planning and Management in Civil Engineering. He worked as an Assistant Professor at the School of Civil Engineering of the UPV. In October 2012 he joined Nommon, where he is leading different research and consultancy projects related to spatio-temporal data analysis and transport modelling.

Tomás Ruiz is Associate Professor at Transport Department, Universitat Politècnica de València (Spain). His research interests are related to travel behavior analysis and travel survey methods. In particular, he has contributed to better understand the activity scheduling process and sustainable travel habits; to the use of ICT to collect travel data; and to design, implement and evaluate travel behavior change programs.

Maxime Lenormand received a PhD degree in Computer Science from the Blaise Pascal University, Clermont-Ferrand, France in 2012. He is currently working as Postdoctoral Research Associate at the Institute for Cross-Disciplinary Physics and Complex System (IFISC). His research interests focus on big

data analysis and modelling of complex systems in a multidisciplinary context, with particular applications to urban systems and human mobility.

Jose J. Ramasco obtained the PhD in Physics at the University of Cantabria, Santander, Spain in 2002. He then held two postdoctoral positions of 2 years each at University of Oporto (Portugal) and Emory University, Atlanta, USA. Afterwards, he has worked as a researcher at the ISI Foundation in Turin, Italy for 4 years and a half, and now at the IFISC as RyC researcher. His area of expertise is Complex Systems theory with special focus on applications to Transport Networks.

Thibaut Dubernet is member of the Institute for Transport Planning and Systems (IVT) at ETH Zurich since 2011, after graduating in Urban Systems Engineering at the Technical University of Compiègne (UTC). His research focuses on the influence of inter-personal interactions on mobility behavior, with a strong focus on simulation techniques.

Enrique Frias-Martinez is a researcher in Telefonica Research working on Big Data related areas, mainly Urban Computing, Smart Cities and Human Mobility. Before joining Telefonica, Enrique worked for the Department of Biomedical Engineering of the University of California, Los Angeles (UCLA), for the Department of Information Systems and Computing of Brunel University in London, and for the Courant Institute of Mathematical Sciences, New York University. He has a Ph.D. in Computer Science from the Universidad Politécnica de Madrid, Madrid, Spain (2001) where he was the recipient of the Best Ph.D. Thesis Award.