Environment and Planning B: Planning and Design 2008, volume 35, pages 961 – 980

DOI:10.1068/b3317t

Collecting social network data to study social activity-travel behavior: an egocentric approach

Juan Antonio Carrasco¶
Department of Civil Engineering, University of Toronto, 35 St George St, Toronto, ON M5S 1A4, Canada; e-mail: j.carrasco@udec.cl

Bernie Hogan
Department of Sociology, University of Toronto, 725 Spadina Avenue, Toronto, ON M5S 2J4, Canada; e-mail: bernie.hogan@utoronto.ca

Barry Wellman
Centre for Urban and Community Studies, University of Toronto, 455 Spadina Avenue, 4th Floor, Toronto, ON M5S 2G8, Canada; e-mail: wellman@chass.utoronto.ca

Eric J Miller
Department of Civil Engineering, University of Toronto, 35 St George St, Toronto, ON M5S 1A4, Canada; e-mail: miller@ecf.utoronto.ca
Received 10 February 2006; in revised form 24 August 2006; published online 1 October 2007

Abstract. This paper presents a data collection effort designed to incorporate the social dimension in social activity-travel behavior by explicitly studying the link between individuals' social activities and their social networks. The main hypothesis of the data collection effort is that individuals' travel behavior is conditional upon their social networks; that is, a key *cause* of travel behavior is the social dimension represented by social networks. With this hypothesis in mind, and using survey and interview instruments, the respondents' social networks are collected using an egocentric approach that is constituted by the interplay between their individual social structures and their social activity behavior. More explicitly, individuals' networks are a context within which to elicit social activity-travel generation, spatial distribution, and information communication and technology use. The resultant dataset links aspects, in novel ways, that have been rarely studied together, and provides a sound base of theory and method to study and potentially give new insights about social activity-travel behavior.

1 Introduction

Within the study of travel demand using the activity-based approach (Axhausen and Gärling, 1992), there is an increasing interest in the influence of social interactions on activity-travel decisions (Bhat and Lawton, 2000). Concomitantly, there is a growing interest in the study of social activities, which recognizes both their importance in the overall travel patterns and that they differ behaviorally from better-studied trip purposes, such as work and shopping (Bhat and Gossen, 2004). Nevertheless, despite this interest, the lack of data collection effort has inhibited an explicit and reliable link between social travel behavior and social interactions. Moreover, although data collection and modeling techniques have gone far towards understanding individual activity-travel decision-making processes in time and space (Doherty and Miller, 2000), little is known about the linkages between social and spatial interactions.

Axhausen (2005, page 100) explicitly discusses this need to incorporate the 'social dimension' in travel behavior:

> "Transport planning and even more so transport modeling has ignored the social dimension of travel in the past. There is therefore no empirical literature to fall back on. The general lack of detailed address geocoding of previous travel diary

¶ Correspondence address: Department of Civil Engineering, Universidad de Concepcion, PO Box 160 C, Concepcion, Chile.

data makes these large data sets less useful than they could be, as they cannot be used to trace the development of the spatial visiting and meeting patterns in detail."

In order to incorporate formally the social dimension, researchers should take account of the existing structure of social relations represented in the *individual's social networks*, for it is within this structure that social interactions and interaction decisions are made. In fact, although 'with whom are activities performed?' now constitutes a standard question in transportation data collection [especially in time-budget studies, see, for example, Arentze et al (1997), Doherty et al (2004)] and questions 'for whom activities are performed' (Goulias and Kim, 2005) have been made, more information is needed to capture the overall importance of the social dimension in activity-travel patterns. In fact, asking *with whom* the activity was performed does not necessarily capture the relevance of those individuals in the overall activity-travel behavior, since interactions with specific persons are only collected if they occurred within the specific time frame covered by the instrument. This issue is particularly relevant considering that surveys, in general, collect information for short periods of time. Consequently, very few social activities and almost no social repetitive patterns are captured because social activities are in general less frequent than work and shopping, and because activity-variety seeking is present over long time spans, especially for social activities (Schlich et al, 2004). In sum, time–space fixity and recurrence in social activity travel are in general very difficult to study.

Even more importantly, collecting data only about with whom the activity was performed implies a conception of the social dimension as a mere *attribute* of the social activity (at the same level as destination or time of the day, for example), rather than as the *cause* of the social activity. This approach may hide behavioral processes, such as the propensity to interact with some people rather than with others, and, more importantly, it can hide the potential importance in frequency, spatial location, and other activity-travel attributes that the 'with whom' dimension implies. For example, we visit an elderly lady once a week, traveling one hour not only because she is a nice lady and our costs make this travel feasible, but *because* she is our grandmother and she lives at an hour's distance. Hence, travel behavior is *caused* by the social dimension (our social network).

This paper presents a data collection effort designed to address some of these challenges, linking social activity-travel behavior and social interactions. The instruments designed explicitly collect the individuals' social networks using an egocentric scheme, constituted by the social structure of specific individuals (egos), and the interplay between their social activities and social networks. More explicitly, the objective of the data collection is to provide a dataset that can help to study the effect of social networks in the following aspects of social activities: (i) their generation, (ii) their spatial distribution, and (iii) their relationship with information communication technology (ICT) use.

Regarding the generation and spatial distribution of social activities, the major interest of the social network approach resides in the explanation that it can provide an understanding of individuals' social patterns, testing whether the 'with whom' dimension constitutes a relevant *cause* of individuals' activity-travel patterns. In the case of the relationship between social activity travel and ICT, social networks constitute a privileged way of studying the effect of different media in social interactions and activity travel, considering that new technologies are increasingly embedded in ordinary life (Haythornthwaite and Wellman, 2002).

Also, as a background motivation, the interest of this work is in studying the general usefulness of social networks for understanding travel behavior, taking into account Axhausen's (2002, page 3) remarks about "the need to underpin our travel models with a better understanding of the social structures of daily life ... as we implicitly forecast/speculate about them when we predict travel behavior over long

time horizons, anyway." This study of social networks and travel behavior is in its early stages, and research has concentrated mainly on social influence aspects (Dugundji and Walker, 2005; Páez and Scott, 2007). An overall interest in this work is informing, and enriching with the social network perspective, the behavioral components of operational activity-travel demand models' such as TASHA (Miller and Roorda, 2003; Miller et al, 2005), and integrated transportation and land-use models, such as ILUTE (Salvini and Miller, 2005). The motivation for studying social networks is also based on a general interest in studying not only outcomes but also behavioral processes (Doherty and Miller, 2000), echoing the long-discussed need to incorporate complementary explanations into the dominant microeconomic paradigm (Gärling, 1998). From a data collection perspective the method employed here also responds to the interest in exploring mixed quantitative and qualitative data collection in travel behavior research (Clifton and Handy, 2003).

The rest of the paper is divided as follows. Section 2 reviews the key background concepts; section 3 describes the survey and interview instruments of the data collection; and section 4 presents some discussion and conclusions.

2 Key concepts

This section presents the conceptual background that motivated and guided the design of the data collection. After a brief review of the social networks approach, the key hypotheses and concepts of the interplay between social networks and activity-travel behavior are described. Finally, the main characteristics, issues, and challenges in social network data collection are discussed, setting the context of the specific design chosen for this study.

2.1 The social networks approach

The social networks approach of this work is more than metaphorical, incorporating network analytic theory and methods, and four decades of substantive findings. It draws from a long tradition in sociology and, to a lesser extent, other disciplines, such as anthropology, graph theory, and management science [for further reviews of the paradigm and techniques, see, for example, Carrington et al (2005), Scott (1991), Wasserman and Faust (1994), and Wellman and Berkowitz (1988)]. Tindall and Wellman (2001, pages 265 – 266) define the social network approach in the following way:

> "Social network analysis is the study of social structure and its effects. It conceives social structure as a social network, that is, a set of actors (nodes) and a set of relationships connecting pairs of these actors."

Thus, two key components define this paradigm: *actors*, who represent different entities, such as groups, organizations, nations, as well as persons; and relationships or *ties*, which represent flows of resources that can be related with aspects such as control, dependence, cooperation, information interchange, and competition.

The core concern of the social network paradigm is to understand how social structures facilitate and constrain opportunities, behaviors, and cognitions. Social network analysis conceives the overall behavior as being more than the sum of individual behaviors, and contrasts with explanations that treat individuals as independent units of analysis, as traditionally used in travel behavior research. Thus, behavior is explained not only through personal attributes but also through social structural attributes that incorporate the interaction among the different social network members. This approach assumes that *the whole is more than the sum of its parts*—that is, social phenomena cannot be understood solely by individual characteristics (such as socioeconomic attributes) but must also consider the social structure emerging from the interaction among individuals.

2.2 Social networks and activity-travel behavior

The main hypothesis in this work is that communication and activity-travel patterns *emerge* from the individuals' social networks, or, in other words, they can be inferred in part from knowing individuals' social network characteristics. This hypothesis has consequences in the understanding of the generation and spatial distribution of social activities, and communication media behavior among individuals.

The *generation* of (or decision to perform) social episodes can be explained by the individual's propensity and opportunity to engage in a social activity (Chapin, 1974; Hägerstrand, 1970). Propensity not only depends on the individual's socioeconomic and lifestyle attributes (Lu and Pas, 1999), but also on *with whom* individuals perform social activities, who constitute the individual's social network characteristics, and also on their related communication patterns (eg frequency and type of media used). As a consequence, ties or links between the individual and other people in the network represent a flow of potential activity travel *generated* by the interaction between those 'nodes' of individuals. Opportunities to engage in social activities are represented by individuals' time and space prisms (Hägerstrand, 1970), which, in part, depend on the spatial distribution of individuals' social networks, and are fixed in the short time horizon.

In the case of the *spatial distribution*, the above hypothesis implies studying activity destinations from another, complementary, perspective, which views the observed individuals' activity patterns as a direct consequence, not only of their preferences and restrictions but also of the *spatial location* of their social network. The concept of *social anchor points* can be used, which describes the main places where the individuals 'move around' when they interact with other network members. Several anchor points are defined by the individual's social network, either directly—for example, homes of the social network—or indirectly—for example, pubs or restaurants, which are themselves defined in part by the network members' home or workplace. These anchor points are hypothesized as key pivotal places defining the social *activity space* (Horton and Reynolds, 1971).

Finally, regarding communication patterns, the previous hypothesis conceives interpersonal relationships as the focal element from which different communication interaction media emerge, incorporating in the same framework face-to-face, telephone, and Internet-based socializing episodes. This focus of communication subjected to individuals' social networks sets an intuitive and consistent conceptual framework about the way in which individuals satisfy their interaction needs, thus recasting social activity travel as only one of the possible means of interaction.

2.3 Social networks data collection

There is a long tradition concerning the techniques and issues of collecting social network data; for an in-depth review, see Marsden (1990; 2005). Overall, some key challenges in this kind of data collection are:

- Network boundaries are difficult to define.
- People do not easily recall their network members, and need appropriate 'prompts' to elicit them; in addition, networks are very large in general, and different social network members may have different importance, depending on the phenomenon studied.
- Information about the network members needs to balance detail and the interviewee's burden.

Each of these three challenges was considered in specific social network data collection techniques, as described in more detail below.

2.3.1 *Whole versus egocentric networks*

Social network data collection differs when studying 'whole' or 'egocentric' networks. Whole-network studies examine actors "that are regarded for analytical purposes as bounded social collectives" (Marsden, 2005, page 8). Actors in these studies are named in closed lists, usually predefined, and are known a priori. Sine these boundaries are very difficult to define in urban settings with large populations (lists of the population are not known in advance or are too long), whole-network studies are often impractical, making egocentric data collection a more feasible method.

Egocentric-network studies concentrate on specific actors or egos and those who have relations with them, called *alters*. That is, from the respondent's perspective, egocentric networks constitute a 'network of me' or a network of actors (alters) with whom the respondent has some relationship. Egocentric-network data are thus composed of two levels: (i) an ego – network level constituted by the ego's characteristics and overall network features; and (ii) an ego – alter level, constituted by the characteristics of each alter and alter – ego ties.

2.3.2 *Name generators*

As discussed before, defining the network's boundary is a crucial challenge. For egocentric networks the problem is twofold: choosing appropriate egos and selecting appropriate network members. In the first case, egos must be representative of the context studied (eg specific urban settings). In the second case, eliciting 'appropriate' network members is difficult, due to the large size of networks and the need to sample adequate network members for the phenomenon of interest. In egocentric methods the most used technique to elicit network members is the *name generator*, which consists of free recall questions that elicit alters from an ego's network (Burt, 1984; Marsden, 2005).

Name-generating questions elicit "a fraction of respondents' social contacts" (Marsden, 2005, page 12). The key decision then is choosing the appropriate specific question(s) that will elicit the network members relevant to a specific phenomenon of interest, constrained by the available time, and the desired level of complexity of the data collection instrument. Also, the number of alters elicited can be limited by a specific number (Marsden, 1987) or unlimited (as here). There is an extensive literature that compares different name generators, discussing aspects such as their influence on network size, the number of 'core' and 'extended' network members that each elicits, the importance of the instrument's context, the relevance of the order and wording of questions, and the forgetting phenomena [for further review, see Marsden (2005) and the references therein]. Finally, name generators are also important to measure *tie strength* between egos and each alter, and between alter – alter pairs. As before, the chosen indicator will affect which tie strength measure is collected, with emotional closeness being the most usual and accepted (Marsden and Campbell, 1984). In the case of egocentric networks, tie strength is, in general, measured only *from the ego perspective*, for both ego – alter pairs and alter – alter pairs. This is because it is prohibitively expensive to interview both egos and alters in large sample surveys.

2.3.3 *Name interpreters*

After eliciting network members, a second set of questions is usually performed to obtain more information about the characteristics of each alter (eg socioeconomic status, relationship with ego) and ego – alter relationship (eg frequency and characteristics of interaction). From a practical point of view, a key challenge here is gathering an adequate amount of information in a nontedious, relatively short, and reliable way. These issues are especially critical when the number of alters is not defined beforehand, as in this study. In general, a *sampling* strategy is usually performed (Marsden, 2005), although no firm guidelines can be found from the literature about this issue.

3 The Connected Lives Study

This section describes the overall study and specific instruments used in this data collection effort, as well as some insights as to how these data can help researchers to capture the link between social networks and social activity-travel behavior. After a brief presentation of the overall study, the survey and interview instruments are described in detail.

3.1 The study

The data were gathered in the East York area of Toronto, Canada, between July 2004 and April 2005 as part of the 'Connected Lives Study', a broader study composed of surveys, interviews, and observations about people's communication patterns. The first author conducted the study in conjunction with sociologists (among them, the second author) and a social worker at the NetLab—a part of the Centre of Urban and Community Studies at the University of Toronto, and led by the third author. The multidisciplinary setting allowed for a rich cross-fertilization in data collection techniques, and a broad set of information was collected. The East York area is located on the east side of downtown Toronto, and is fairly representative of the overall central city characteristics regarding sociodemographics and general transportation level of service. We randomly sampled English-speaking nonfrail East York adults (>18 years old), and completed 350 surveys; the sampling frame yielded 621 valid names, and we obtained a response rate of 56%. After the survey stage we conducted in-home interviews and observations with a quarter of the survey participants, leading to a subsample of eighty-seven people. For a further overview of the study, including further discussions about the validity of the data, see Kayahara and Wellman (2005) and Wellman et al (2006).

3.2 Name generator

The name generator in both the survey and interview instruments concentrated on the individual's *affective network* or a network of people the respondent defines as *close*—this is an approach that seems to be useful for understanding communication and social activity-travel patterns. Specifically, respondents were asked to name the people who live outside their household with whom they felt *very close* and *somewhat close*.

Very-close people consist of 'people whom you discuss important matters with, *or* regularly keep in touch with, *or* are there for you if you need help'. Somewhat-close people consist of 'more than just casual acquaintances, but not very close'. The approach of eliciting only outside household members is made to simplify the name-generator process. Although some aspects of household dynamics and social interactions can be lost, information about household characteristics from other parts of the study partially helps to understand these dynamics.

This 'closeness' approach defines two aspects. First, closeness becomes an operational measure of tie strength: strong (very close), and weak (somewhat close).[1] Second, closeness defines the social network 'boundary'—and thus the sociable activity-travel patterns captured in the data—that excludes people who are only casual acquaintances.

3.3 Survey

The survey instrument covered a variety of aspects about people's communication patterns (Wellman et al, 2006); the focus here is on the social network composition section. The specific items prompted are shown in the appendix. The 'summation method'

[1] Note that this definition of tie strength is very consistent with the original Granovetter (1973, page 1361) definition: "... the strength of a tie is a ... combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie".

was used (McCarty et al, 2000), which consists of asking respondents *how many* strong-tie and weak-tie people live outside their households in each of the following roles: immediate family, other family, neighbors, work or student mates, known only online, from organizations, other friends, and others not previously included. As an aid, respondents were provided with a sheet that enabled them to write the names of the people in each category, and to ensure that the set of individuals counted in each role was mutually exclusive. After prompting the number of network members in each role, further questions included the number of strong-tie and weak-tie network members of each gender, of different ethnic heritage, living outside Canada, and living in Canada at a distance greater than one hour's travel.

Finally, a set of questions about frequency and media of interaction were asked. Specifically, respondents were asked the number of strong-tie and weak-tie network members they typically: (i) call by cell phone, (ii) call by regular phone, (iii) send an e-mail to, (iv) send an instant message to,[(2)] (v) talk with face-to-face, (vi) meet at restaurants or bars, and (vii) visit or host as visitors. Each of these questions differentiated between two time horizons: (1) at least once a week, and (2) between once a week and once a month. Figure 1 shows the overall number of network members by ego in selected categories, illustrating that the method captures a broad set of social networks, and social activity-travel and communication patterns.

The approach adopted in the survey could be defined as a 'meso' approach, since it captures structural features, such as size of the network, approximate density, and aggregated composition according to aspects such as role and gender, and the ego's characteristics. This approach contrasts with a 'micro' approach (adopted in the interview), which also captures more disaggregated aspects, such as each alter's characteristics (eg gender, spatial location), and each interaction between alters and the respondent.

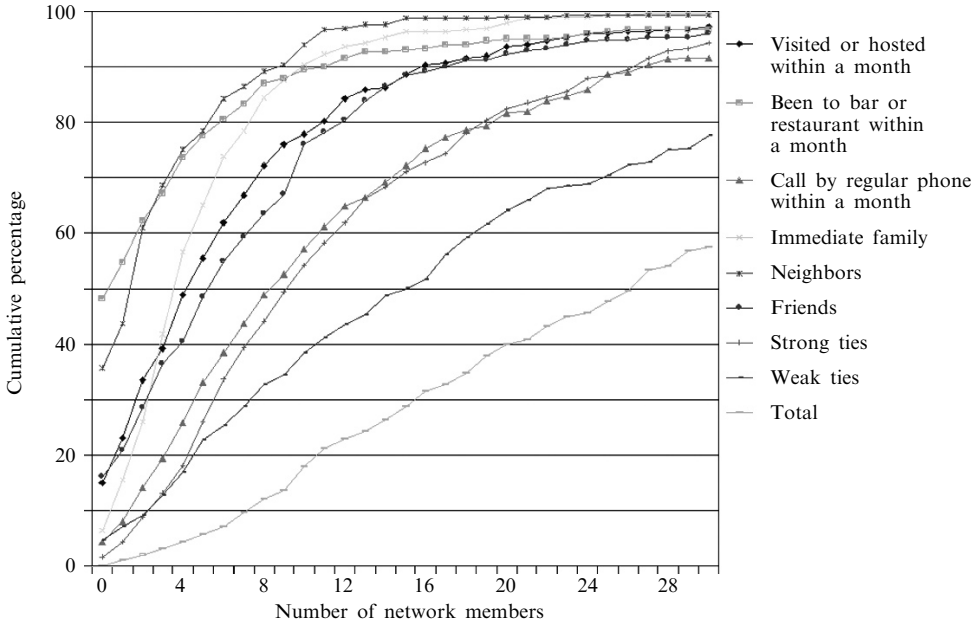


Figure 1. Cumulative percentage of network members by selected categories.

[(2)] Instant message refers to the interaction through short messages via the Internet (using providers such as Microsoft®, Yahoo®, and AOL®), which is a different technology from the interaction through cell phones using short message services, SMS, which are still not common in North America.

Yet, the 'meso' approach constitutes a quick way, in the time-limited survey, of eliciting structural features without the use of more complex instruments, such as those used in the interviews.

This 'meso' approach can be used to study the link between the individuals' network composition and their propensity to socialize. For example, using these data, Carrasco and Miller (2006) analyze how the number of network members—grouped in the different categories mentioned before—is related to the propensity to socialize, measured by the number of alters with whom the ego socializes in different time spans (weekly, monthly, yearly) and activity types (household versus nonhousehold social activities). The data allowed Carrasco and Miller to explore the effect of aspects such as network distance composition (eg neighbors versus people living at a distance greater than one hour's travel) and ICT use (eg e-mail frequency with weak versus strong ties) on the individual's propensity to socialize.

A key overall assumption in this section of the data collection is the record of 'usual' communication and activity-travel patterns to capture the *overall* communication and social activity-travel behavior, rather than observed or stated patterns. This approach was adopted due to: (i) the need to capture realized patterns in social activities over long time spans (Schlich et al, 2004), (ii) the reduction in further respondent burden, and (iii) cost restrictions. Further research is needed to assess whether this approach involves biases with respect to the actual respondent's patterns, and whether it constitutes a reasonably adequate proxy of the actual respondent's behavior. However, no technical issues other than these should prevent future studies including more detailed activity-travel aspects in conjunction with social network data. We shall crosscheck survey and interview data in subsequent research.

3.4 Interview

The interviews took, on average, two and a half hours to complete, generally at the homes of the individuals. The interviews were conducted by doctoral students who were members of NetLab and had detailed knowledge of the study. The sections of interest for this paper are: (i) the name generator, (ii) the name interpreter, and (iii) social episodes. The full process of these three sections took between forty minutes and ninety minutes. The time depended mainly on respondent motivation and network size.

3.4.1 *Name generator*

The name generator had three goals: (i) creating a respondent-aided *sociogram*, (ii) maximizing the size and richness of the egocentric network, and (iii) facilitating the recording of the network's connectivity. A sociogram can be intuitively defined as a drawing that permits visualizing the respondents' social network, containing all alter names and the ties among them. The sociogram was represented in the study as a series of four concentric circles, where the ego is at the center, with the alters situated around it (see figure 2).

The importance of the sociogram is threefold. First, it helps to collect tie-level data in an intuitive and easy way for respondents, thereby lowering their burden—especially in the case of elderly or poorly educated people—and facilitating the incorporation of the highest possible number of network members of interest. Second, it makes connectivity recording easier, more reliable, and more complete. Third, it serves as a cognitive aid to prompt follow-up questions about communication and social activity-travel patterns.

The sociogram was built in three steps:

(1) Interview respondents were asked to write their strong-tie and weak-tie network members in a free recall order. They used a 'name template' (see figure 3), and

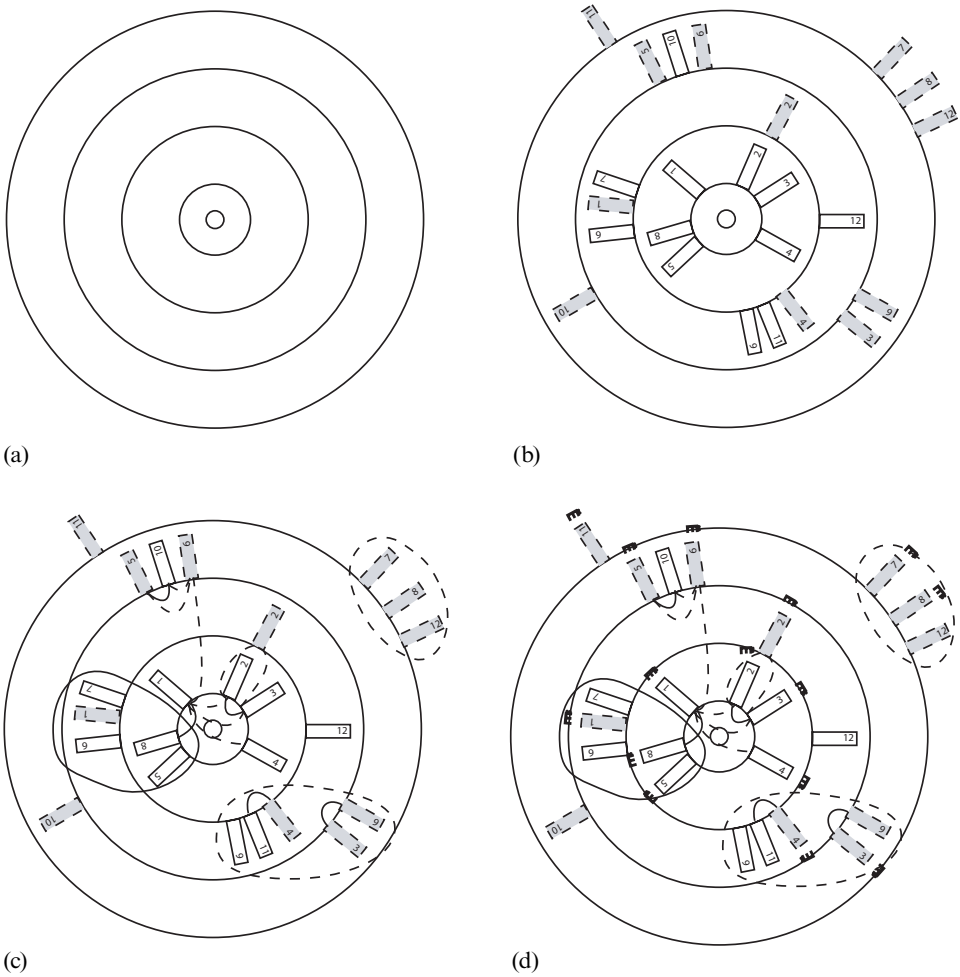
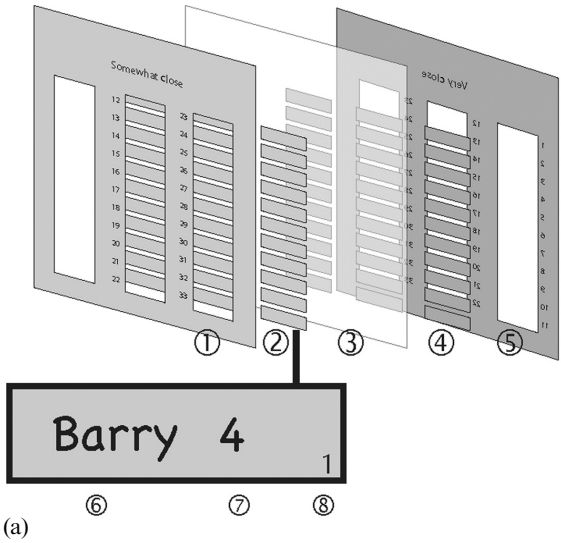


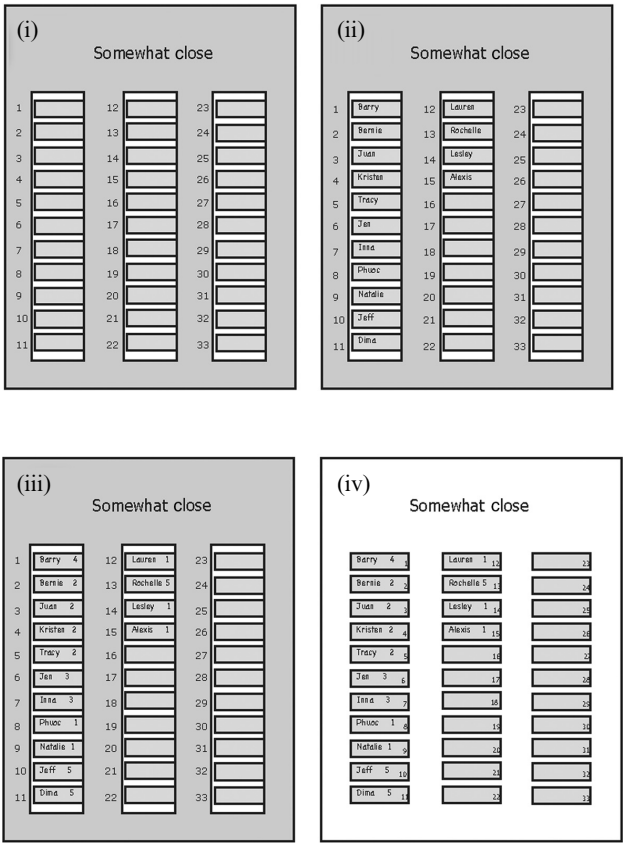
Figure 2. Sociogram in different stages of the interview: (a) a blank sociogram; (b) strong and weak ties in the sociogram; (c) a completed social network; (d) an example of sampling in name-interpreting questions. Each Post-it™ rectangle represents an alter, loops indicate groups of alters with ties among all of them, and lines represent ties between a pair of alters (different shades and line styles denote different tie strengths).

employed the previously mentioned definitions of 'very close' and 'somewhat close'. After the list was completed each individual's role was recorded. Here an alter could have multiple roles (eg a workmate could also be a friend), a phenomenon called *multiplexity*, which is explicitly allowed for and encouraged to be recorded. At this stage, two kinds of numbers were used to record information about each alter: a *rank* number, representing the order in which each alter is elicited (the smaller the number, the higher the rank), and a *role* number(s), representing each alter's role.

(2) In the second step respondents were asked to arrange Post-It® notes (one representing each alter) on four concentric circles, according to how 'close' the respondent felt about the represented alters. The closest individuals were on the most central ring, (visually closest to 'ego'), and weaker ties were on the peripheral rings. The closer the respondent felt the alter to be, the closer the ring to the center, starting with the strong-tie people, and followed by weak-tie people. This closeness measure was kept



(a)



(b)

Figure 3. Overview of (a) the name template and (b) the answering process. Name template construction: five layers are placed together as a single 'template' with the use of binder clips; 1: somewhat-close alter plate; 2: somewhat-close alter name tags; 3: divider; 4: very-close alter name tags; 5: very-close alter plate. Also, each tag contains the following information: 6: name; 7: role; 8: rank number.

ambiguous, and is not necessarily linked with tie strength, adding a second and possibly complementary measure of emotional proximity. This step was entirely left to the respondent; the only instruction was to place people who know each other nearby, in order to help the following step.
(3) The third and final step consisted of recording connections among alters. Respondents were asked to draw loops, representing groups of people who they thought were all very close, groups who were all somewhat close among them, and lines among pairs of very-close or somewhat-close alters.

The final result of the name-generator section can be seen in figure 2(c). The respondent has: (i) generated the 'names' in the social network and their roles, differentiating by tie strength, (ii) located each alter in the sociogram, according to a loosely defined emotional closeness, and (iii) recorded ties among all social network members, differentiated by tie strength. The method provides a 'step-by-step' procedure to gather the respondents' social network, which is the base information with which to capture their subsequent communication and activity patterns. More details about the implementation of this section and a comparison of network sizes between the name generator and the summation method are discussed in Hogan et al (2007).

3.4.2 *Name interpreter*
Name-interpreter questions recorded communication and social activity patterns between the respondents (ego) and a selected number of network members (see appendix). Pretests indicated that, in networks of greater than fifteen members, asking detailed travel and communication questions for each alter was highly impractical. This indicated the necessity of sampling network members for whom to retrieve the information. Therefore we introduced a sampling strategy of selecting approximately fifteen alters, representing a compromise between the need for a representative number of people, and keeping the interview to an acceptable duration. Overall, the sampling scheme elicits names from all the rings on the sociogram, from both strong and weak ties, 'covering' the overall network in the best possible way, but at the same time giving higher priority to emotionally closer alters (ie those in the inner rings). From the communication and travel patterns point of view, this sampling scheme balances two objectives: on the one hand, eliciting representative alters of a social network, and thus ego's patterns; on the other hand, capturing a high proportion of the more relevant communication and activity patterns of the respondent, assuming that these correspond to the emotionally closest people (see figure 2).

Two sets of name-interpreting questions were asked for this sample:
(1) *Information about each alter's characteristics, including age, relationship, job, and ethnic heritage*. Two spatial locations were recorded: the alter's home location, and the most frequent place of interaction with the respondent. Both of these were gathered at the level of the intersection to facilitate posterior geocoding.
(2) *Information about the ego's communication and interaction patterns with each alter: 'face-to-face', socializing, and media that potentially could substitute face-to-face interaction, such as telephone, e-mail, and instant message*. 'Face-to-face' and socializing were explicitly separated, since a main interest is differentiating between instrumental interactions (such as those existing in workplaces) and social interactions (visiting, hosting, going to pubs and restaurants)—and also to capture the circumstances in which face-to-face interactions are different from socializing. Thus 'face-to-face' is a superset of socializing, including all contacts made in person. Two additional questions include the direction or 'agency' of the interaction—that is, who starts or triggers the interaction, and the technology used (eg cell phone versus landline).

A paper-and-pencil 'minisurvey' was used to record the information in this section, leaving to the respondents whether they answered the questions on their own or whether the interviewer helped them. The interview conservation was also recorded, a strategy that has proven to be useful, since respondents tended to add extra contextual information about their alters and interaction patterns, providing an interesting mix between quantitative or structured data, and qualitative or contextual information.

3.4.3 *Social episodes*

The last section of interest in the interview involved recording selected social episodes between the respondents and some of the alters previously elicited in the name-generator and name-interpreter sections. 'Social episodes' are defined as those involving visiting, hosting visitors, or meeting in restaurants, pubs, or similar places. The already complex design of the study did not leave room for an extensive collection of all the respondents' social episodes, also considering that a complete account of social activities requires collecting data for long time horizons (Schlich et al, 2004). Instead, the method used in this section involved sampling each respondent's social episodes which potentially serve as a proxy to understand their overall social patterns. Six social episodes were recorded, using a strategy consistent with the main background assumption of the study—that is, communication and social activities emerge from the individuals' social networks.

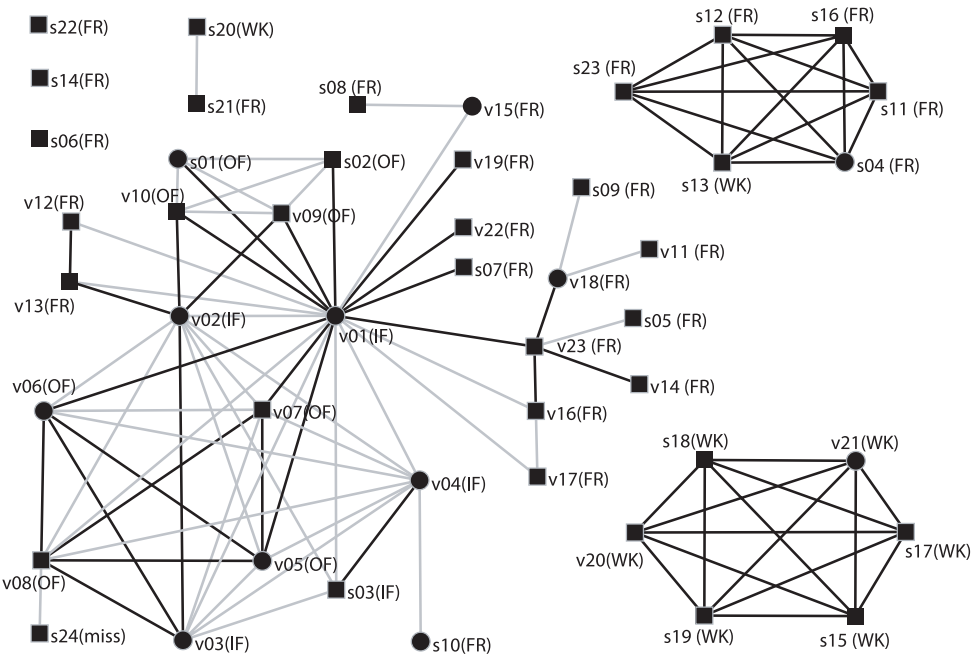
Specifically, respondents were prompted about social episodes with six selected social network members from the fifteen whose names had been elicited in the name-interpreter section (discussed above). The six network members were restricted to those with whom the respondents have socialized in the Greater Toronto area. These six alters were semirandomly chosen using a scheme that was biased towards the ego's most frequent social episodes with his or her network members and those with emotionally closer alters. The scheme kept the same balance of objectives as in the name-interpreter section: sampling the overall network, emphasizing the emotionally closer alters, and maintaining the consistency of alters sampled in each section. For each of the six alters respondents were asked about specific aspects of their last social episode with him or her: what the activity was about; when it had occurred (time of the day, day of the week, duration); where it was (detailed spatial location, feasible to be geocoded); which other network members were involved; and which transportation mode was used (see appendix). Other qualitative questions about the activity planning process were included, such as who triggered the episode, what media were used, and how far in advance the episode was planned. Respondents were also asked about the frequency of such an episode occurring and the fixity of such episodes in time and space. As in the name-interpretator section, the interview setting allowed respondents to add qualitative information about the behavioral context of each of the social episodes and patterns. The preliminary perception of the interviewers is that six social episodes seems to be an adequate number to provide a good overview of the individual's social patterns, which is in part reaffirmed by results in Europe that show eight locations capturing 80% of the overall leisure activities (Schlich et al, 2004).

3.4.4 *An illustration*

Figures 4(a), 4(b), and 4(c) show an example of an individual social network from the data collected in the interview; figures 4(a) and 4(c) were constructed using the software visualization tool GUESS (Adar, 2005), and figure 4(b) was constructed using ArcGIS (ESRI Inc., Redlands, CA). We will use the pseudonym Francis to identify the ego and 'v' or 's' followed by numbers to identify very-close or somewhat-close alters, respectively. Figure 4(a) shows the network diagram of the forty-seven people elicited in the name generator (ego not included); dark links represent very-close ties

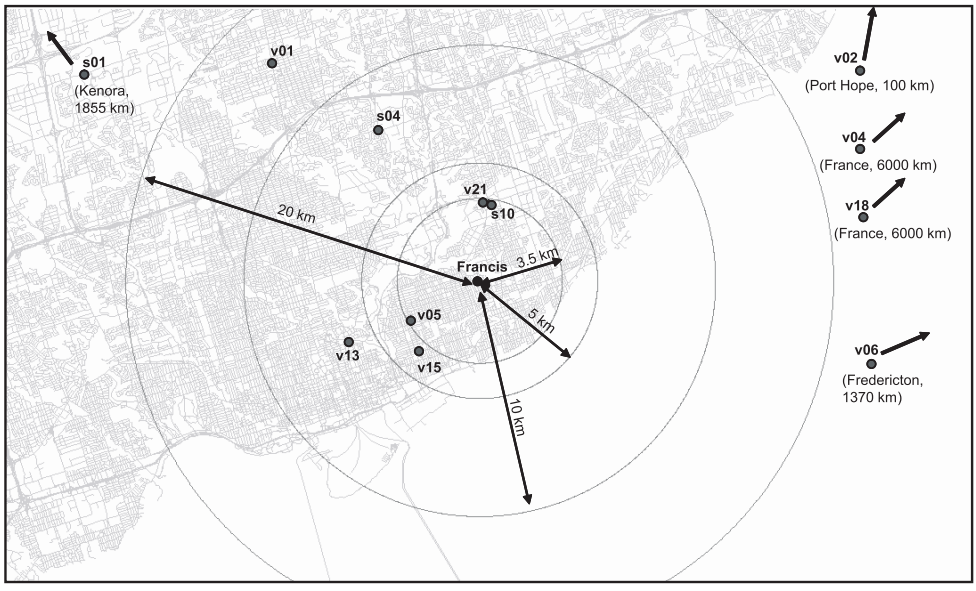
between alters, and light links represent somewhat-close ties, as reported by the respondent in the third step of the name-generator section. The role of each alter with respect to the ego is shown in parentheses. From this diagram we can see that Francis's network is constituted by a large component (a subnetwork with all nodes connected to each other, with mainly family and friends), two smaller cliques (subnetworks with all nodes connected, each of them mainly constituted by workmates), and one dyad and a few isolated alters. Excluding the isolated alters and dyads, the three main subnetworks immediately reflect independent potential sets of social activities, which need to be analyzed to obtain a good picture of Francis's overall social patterns.

Circled nodes in figure 4(a) show the alters sampled in the name-interpreter section; we can see how the sampling approach has the capability of capturing a good overview of Francis's social patterns. First, very important people in the network are elicited, not only in terms of closeness and frequency of interaction with respect to the ego, but also with respect to their degree (number of ties with other alters)—such as in the case of v01 and v06—which potentially reflects their importance in Francis's overall social interaction patterns. Second, the name-interpreter sampling approach is capable also of sampling people not necessarily in the core of the largest subnetwork (eg v18 and v15), as well as people from the other two subnetworks (s04 and v21). This sampling

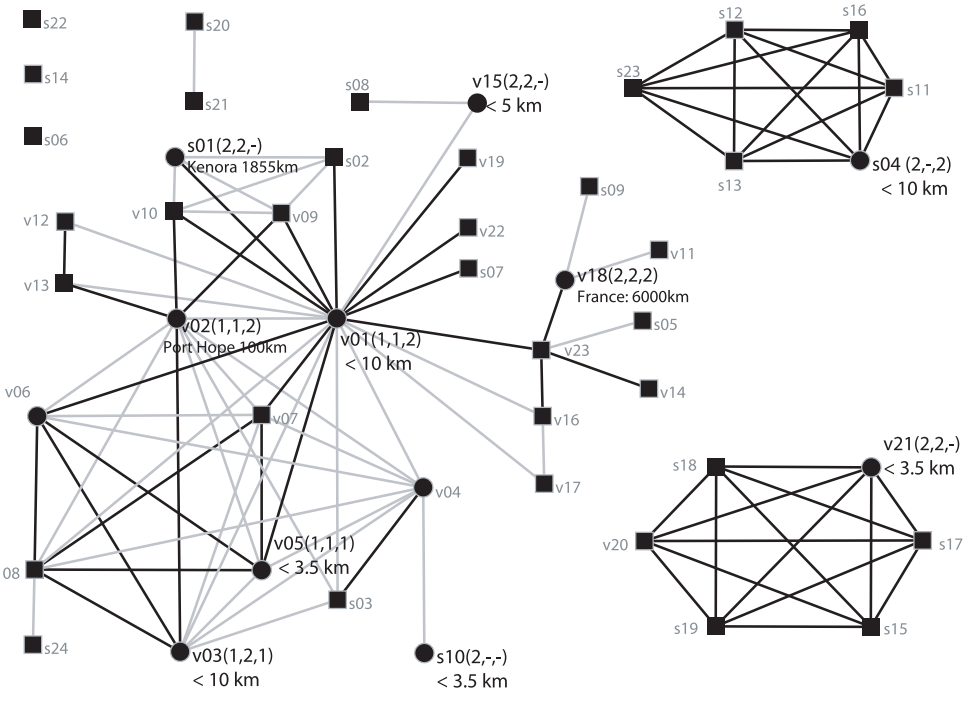


(a)

Figure 4. (a) Example of an egocentric network (ego not shown, the role of each alter with respect to the ego is shown in parentheses—IF: immediate family, OF: other family, FR: friends, WK: workmate, NG: neighbor); (b) example of the spatial location of an egocentric network (the ego, Francis, and alter locations in Toronto and outside Toronto are shown); (c) example of an egocentric network showing frequency of interaction with sampled alters. Letters 'v' or 's' followed by numbers identify very-close and somewhat-close alters, respectively. Dark lines represent very-close ties and light links represent somewhat-close ties. Circle nodes represent sampled alters in the name-interpreter section. The frequency of face-to-face social interaction, telephone, and e-mail between the ego and selected alters, respectively, is shown in the three-dimensional vector. The frequency categories in each medium are 1: more than once a month, 2: between once a month and less than once a year, 3: once a year or less, and -: never.



(b)



(c)

Figure 4 (continued).

capability potentially captures very different sets of social interaction patterns with respect to those alters in the largest subnetworks.

Finally, we explore some of the name-interpreter questions. Figure 4(b) shows the geocoded locations of Francis's alters within Toronto (study area), and the locations of those outside Toronto. Four buffer rings at 3.5, 5, 10, and 20 km with respect to Francis give a sense of the spatial dispersion of the seven alters living in Toronto. Three of the other five alters live at greater distances within Canada—Kenora (Ontario), Port Hope (Ontario), and Fredericton (New Brunswick)—whereas the other two alters live in France. An immediate interesting finding is the varied spatial dispersion of Francis's alters at the city, national, and worldwide scale. This information about the spatial distances between Francis and his or her alters, complemented with the structure and frequency of interaction with his or her social network give a rich overview of Francis's social interaction patterns. To illustrate this aspect, figure 4(c) presents Francis's social network, focusing on the alters sampled in the name-interpreter section, and combining the spatial information of his or her alters with his or her interaction patterns with each of them. A vector of three ordinal numbers shows his or her frequency of interaction with each alter, representing social face-to-face, telephone, and email interaction, respectively. The frequency categories in each media are 1: more than once a month, 2: between once a month and less than once a year, 3: once a year or less, -: never. For convenience, figure 4(c) also includes the approximate distance between Francis and each alter. The figure shows how the data allow us to gain insights about the interplay between distance, frequency, and media in social interactions and social activity-travel behavior. For example, we can see in this particular network that:

- Central alters (such as v01, v02, and v05) have a higher frequency of contact compared with Peripheral alters (such as s10);
- both sampled alters in the subnetworks (s04 and v21) have, in general, lower frequency of contact than those in the larger subnetwork;
- an alter's distance with respect to Francis plays a role (eg the frequencies of face-to-face social interaction with both French contacts are among the lowest), but are mediated by their location in Francis's social network (eg the frequency of social face-to-face interaction is very different between v05, s10, and v21, although the three alters live less than 3.5 km from him or her).

The previous observations are of course specific for Francis's network, and cannot be generalized; future work analyzing all respondents' networks will provide more general insights. In addition, the multilevel structure of egocentric networks (ego–network and ego–alter) gives the opportunity to analyze how the different aspects explored in this data influence travel behavior using quantitative models such as multilevel models (eg Duijn et al, 1999).

4 Discussion and conclusions

A data collection effort designed to incorporate the social dimension into travel behavior has been presented: the background hypothesis of this effort is that overall individuals' communication and social activity patterns emerge and can be inferred in part from their social networks. With that hypothesis in mind, the data collection instruments are designed to incorporate the key characteristics of the interview respondent's social networks—remarkably its multilevel structure, which considers the ego–network and ego–alter levels, and the interactions between egos and their alters. The instrument explicitly collect data about the interplay between social networks, the characteristics of social episodes, and their generation and spatial distribution. The study involved a multi-instrument strategy, consisting of a paper-and-pencil survey,

and an interview, balancing aggregated and less in-depth data from a large sample with more disaggregated and more in-depth data from a subsample.

The study involved a number of design options to address the intrinsic challenges of social network data collection. The first key design option corresponded to the name-generator questions, which defined the way in which alters are elicited from the respondent. The option of focusing on emotional proximity sets the boundary of each respondent's elicited social network and the consequent activity-travel behavior patterns captured. This boundary choice builds on previously successful network data collection methods, and seems to be adequate for the social episodes focus. In addition, the use of emotional closeness and role relationships provides a useful way of capturing adequate network sizes for the purposes of the study, addressing issues such as respondent's forgetting and fatigue.

The second key design option involved the interview's sampling scheme, which combined multiple objectives, such as (i) gathering consistent samples during the interview, (ii) capturing a relevant subset of the respondent's network, (iii) capturing a relevant set of their activity-travel episodes, and (iv) lowering the risk of a heavy burden on respondents.

Besides those design options, the interview involved another two aspects worth remarking on. First, the sociogram used to build the respondents' social networks helped to increase their comprehension of the questions, and—we believe—also increased their motivation, especially because of the step-by-step procedure. In fact, the sociogram created emotional impact (individuals could 'see' their network), and empowerment (they could 'build' their networks)—elements that were very useful in prompting the subsequent name-interpreter and social-episode questions. This approach to sociograms was an invention of this study.

No less important, the fact that the sociogram was built in an interview setting (allowing the interviewer's help), and was constituted by simple interfaces (face-to-face and simple elements, such as Post-it™ papers), lowered the technical burden on the respondents, helping them to concentrate on their networks and behavior. Second, the use of semistructured questions in the interview setting encouraged the gathering both of systematic quantitative information about the respondents' network and behavior and of qualitative data about the context within which the studied phenomena occurred. This design supports the rich insights that mixed qualitative and quantitative techniques can potentially provide to travel behavior research.

Still, some issues, challenges, and assumptions inherent in these kinds of data collection need to be explicitly considered. First, the people (alters) and patterns that are elicited are highly dependent on the name-generator questions, which can be sensitive to aspects that are difficult to manage, such as the respondent's interpretation of questions. Specifically, the use of social 'closeness', although one of the most adequate approaches available, is not free from these kinds of potential bias. Second, even with the efforts employed in this survey, the captured network size is always limited and represents a small portion of the respondent's total social network, which generally ranges upward of 200 persons. As a consequence, capturing most of the overall individual's communication and activity-travel patterns heavily depends on *how many* and *which* network members are prompted. This issue is also linked with a third aspect, which is the collection of usual rather than observed communication and activity-travel patterns. Although much research is needed to understand the biases involved, usual patterns at the very least give an overview of the individual overall behavior, since the related questions are rooted in objective experience, namely 'with whom' those activities were performed. A final issue worth mentioning is that the complexity of the design and relatively simple interface with the respondent implies

that most of the technical burden (and potential bias) is passed to the interviewer; and that the transcription stage can be complex (eg coding connectivity between alters) and costly (in terms of time and money).

Despite these challenges, the data collection effort presented in this paper constitutes a promising way of incorporating the 'social dimension' in travel behavior, linking in novel ways aspects that have been rarely studied together. Further research is needed to integrate further the rich social-network-based information with more explicit travel-based data—beyond frequency of interactions—such as those collected in travel diaries. However, an implementation of this idea needs a careful pretest, since it is not clear how much extra information is worth the extra respondents' burden, especially considering the need of very long-span surveys to capture social activities appropriately (Schlich et al, 2004). In this regard, any implementation must take into account the balance between large-scale, simple data collection and more focused studies. Overall, although it is still early days in terms of our knowledge about the importance of the social dimension in activity-travel behavior, the method provided here can serve as a base to gather relevant social network and frequency-of-interaction information in a reliable way.

In fact, some initial results using the survey section (briefly discussed here, and presented in detail elsewhere: Carrasco and Miller, 2006) show how aggregated measures of the individuals' social networks and interaction patterns can provide insights about travel behavior, especially regarding the propensity to perform face-to-face social activities. In addition, the illustration presented in the interview section shows the extra rich insights that the egocentric social network approach can provide in understanding the interplay between social and physical space, ICT interactions, and their interplay with social face-to-face interaction. Overall, the approach taken here—of incorporating social network data, theory, and method—is expected to provide a solid base with which to give new insights about social activity-travel behavior, and is potentially useful for future generations of travel-demand models.

Acknowledgements. The authors would like to thank Jeffrey Boase, Kristen Berg, Rochelle Côté, Dimitrina Dimitrova, Jennifer Kayahara, Tracy Kennedy, and Inna Romanovska, members of the NetLab group at the Centre for Urban and Community Studies, at the University of Toronto, who were also part of the research team who designed and collected the data of the Connected Lives Study. Thanks also to Phuoc Tran, and students who participated in the data collection stage of the survey. The authors would also like to acknowledge the financial support received from the Social Sciences and Humanities Research Council of Canada (SSHRCC), Major Collaborative Research Initiative (MCRI) and a Regular Research Grant.

References

Adar E, 2005, "The graph exploration system: GUESS", http://graphexploration.org/

Arentze T, Timmermans H, Hofman F, Kalfs N, 1997, "Data needs, data collection and data quality requirements of activity-based transport demand models", resource paper for the 5th International Conference on Travel Survey Methods, Grainau, Germany, http://gulliver.trb.org/publications/circulars/ec008/workshop_j.pdf

Axhausen K W, 2002, "A dynamic understanding of travel demand: a sketch", Arbeitsberichte Verkehrs- und Raumplanung, 119 [WP119], Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT), ETH Zurich, Zurich

Axhausen K W, 2005, "Social networks and travel: some hypotheses", in *Social Aspects of Sustainable Transport: Transatlantic Perspectives* Ed. K Donaghy (Ashgate, Aldershot, Hants) pp 98 – 110

Axhausen K W, Gärling T, 1992, "Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems" *Transport Review* **12** 323 – 341

Bhat C R, Gossen R, 2004, "A mixed multinomial logit model analysis of weekend recreational episode type choice" *Transportation Research B* **38** 767 – 787

Bhat C R, Lawton T K, 2000, "Passenger travel demand forecasting", in *Transportation in the New Millennium* CD-Rom, Transportation Research Board, Washington, DC

Burt R S, 1984, "Network items and the general social survey" *Social Networks* **6** 293 – 339

Carrasco J A, Miller E J, 2006, "Exploring the propensity to perform social activities: a social network approach" *Transportation* **33** 463 – 480
Carrington P, Scott J, Wasserman S, 2005 *Models and Methods in Social Network Analysis* (Cambridge University Press, Cambridge)
Chapin F S, 1974 *Human Activity Patterns in the City: Things People do in Time and in Space* (John Wiley, New York)
Clifton K, Handy S, 2003, "Qualitative methods in travel behaviour research", in *Transport Survey Quality and Innovation* Eds P Stopher, P Jones (Pergamon Press, Oxford) pp 283 – 302
Doherty S, Miller E J, 2000, "A computerized household activity scheduling survey" *Transportation* **27** 75 – 97
Doherty S, Nemeth E, Roorda M, Miller E J, 2004, "Design and assessment of the Toronto Area Computerized Household Activity Scheduling Survey" *Transportation Research Record* number 1894, 140 – 149
Dugundji E, Walker J, 2005, "Discrete choice with social and spatial network interdependencies: an empirical example using mixed GEV models with field and 'panel' effects" *Transportation Research Record* number 1921, 70 – 78
Duijn M A J, van Busschbach J, Snijders T A B, 1999, "Multilevel analysis of personal networks as dependent variables" *Social Networks* **21** 187 – 209
Gärling T, 1998, "Behavioural assumptions overlooked in travel choice modelling", in *Travel Behaviour Research: Updating the State of the Play* Eds J d D Ortúzar, D Hensher, S Jara-Díaz (Pergamon Press, Oxford) pp 1 – 18
Goulias K, Kim T-G, 2005, "On activity type classification and issues related to the *with whom* and *for whom* questions of an activity diary", in *Progress in Activity-based Analysis* Ed. H Timmermans (Elsevier, Amsterdam) pp 309 – 334
Granovetter M, 1973, "The strength of weak ties" *American Journal of Sociology* **78** 1360 – 1380
Hägerstrand T, 1970, "What about people in regional science?" *Papers of the Regional Science Association* **24** 7 – 21
Haythornthwaite C, Wellman B, 2002 *The Internet in Everyday Life* (Blackwell, Oxford)
Hogan B, Carrasco J A, Wellman B, 2007, "Visualizing personal networks: working with participant-aided sociograms" *Field Methods* **19** forthcoming
Horton F, Reynolds D, 1971, "Action space formation: a behavioral approach to predicting urban travel behavior" *Highway Research Record* number 322, 136 – 148
Kayahara J, Wellman B, 2005, "Finding culture online and offline", Heritage Canada, 5 Blackburn Avenue, Ottawa K1N 8A2
Lu X, Pas E, 1999, "Socio-demographics, activity participation and travel behavior" *Transportation Research A* **33** 1 – 18
McCarty C, Killworth P D, Bernard H R, Johnsen E C, Shelley G A, 2000, "Comparing two methods for estimating network size" *Human Organization* **60** 28 – 39
Marsden P V, 1987, Core discussions networks of Americans" *American Sociological Review* **52** 122 – 131
Marsden P V, 1990, "Networks data and measurement" *Annual Review of Sociology* **16** 435 – 463
Marsden P V, 2005, "Recent developments in network measurement", in *Models and Methods in Social Network Analysis* Eds P Carrington, J Scott, S Wasserman (Cambridge University Press, Cambridge) pp 8 – 30
Marsden P V, Campbell K, 1984, "Measuring tie strength" *Social Forces* **63** 482 – 501
Miller E J, Roorda M, 2003, "A prototype model of household activity/travel scheduling" *Transportation Research Record* number 1831, 114 – 121
Miller E J, Roorda M, Carrasco J A, 2005, "A tour-based model of travel mode choice" *Transportation* **32** 399 – 422
Páez A, Scott D, 2007, "Social influence on travel behavior: a simulation example of the decision to telecommute" *Environment and Planning A* **39** 647 – 665
Salvini P, Miller E J, 2005, "ILUTE: an operational prototype of a comprehensive microsimulation model of urban systems" *Networks and Spatial Economics* **5** 217 – 234
Schlich R, Schönfelder S, Hanson S, Axhausen K W, 2004, "Structures of leisure travel: temporal and spatial variability" *Transport Reviews* **24** 219 – 237
Scott J, 1991 *Social Network Analysis* (Sage, London)
Tindall D, Wellman B, 2001, "Canada as social structure: social network analysis and Canadian sociology" *Canadian Journal of Sociology* **26** 265 – 308
Wasserman S, Faust K, 1994 *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge)

Wellman B, Berkowitz S D (Eds), 1988 *Social Structures: A Network Approach* (Cambridge University Press, Cambridge)
Wellman B, Hogan B, Berg K, Boase J, Carrasco J A, Côté R, Kayahara J, Kennedy T, Tran P, 2006, "Connected lives: the project", in *Networked Neighbourhoods: The Connected Community in Context* Ed. P Purcell (Springer, Berlin) pp 161 – 217

Appendix
Survey and interview questions

Survey section

Number of strong-tie and weak-tie network members who are

Immediate family
Other relatives
Neighbors
Work or school mates
People you know only online
People from organizations
Friends not included above
Men/women
Live in Canada at more than one hour's travel away
Live outside Canada

Number of strong-tie and weak-tie network members with whom the ego usually interacts (i) at least once a week and (ii) between once a week and once a month

By cell phone
By regular phone
By email
Using instant message
Meeting face-to-face
Meeting at a bar or restaurant
Visiting or hosting as a visitor

Interview section

Name generator and sociogram building

1 *Generating names*
1.1 Eliciting strong and weak-tie network members:
Very-close people (strong ties): discuss important matters with, *or* regularly keep in touch with, *or* there for you if you need help;
Somewhat-close people (weak ties): more than just casual acquaintances, but not 'very close'.
1.2 Roles of each person, allowing for multiple roles (multiplexity).

2 *Locating very-close and somewhat-close names in sociogram*
2.1 Locate very-close and somewhat-close people according to how 'close' they feel.
2.2 At the same time, locate people that know each other close to each other.

3 *Tie connectivity*
3.1 Draw ties among groups of people who all are very close to each other.
3.2 Draw ties among groups of people who all are at least somewhat close to each other.
3.3 Draw very-close ties between two people.
3.4 Draw somewhat-close ties between two people.

Name interpreter questions

1 *Alter's characteristics*
Age, relationship, job, ethnic heritage, home location, most frequent place of interaction?

2 *Face-to-face*
Frequency (per year, month, week, day)?
On average, how long spend together?
Who goes to see the other?

3 *Socializing*
Frequency (per year, month, week, day)?
On average, how long spend together?
Who invites?

4 *Telephone contact*
Frequency (per year, month, week, day)?
On average, how long are conversations?
Who calls (scale 1 – 5; 1 = 'me', 5 = 'him/her')?
Landline or cell phone use (ego and alter)?

5 *E-mail contact*
Frequency (per year, month, week, day)?
On average, how long are e-mails (scale 1 – 5; 1 = 'short', 5 = 'long')?
Who sends e-mails to whom (scale 1 – 5; 1 = 'me', 5 = 'him/her')?

6 *Instant message contact*
Frequency (per year, month, week, day)
On average, how long are the conversations?
Who starts the conversation (scale 1 – 5; 1 means 'me', 5 means 'him/her')?

Social episode questions

1 *About the specific social episode*
What is was about; when it was (time of the day, day of the week, duration); where it was (detailed spatial location); who else was involved; transportation mode; planning: how it was planned (routine, media), how far in advance?

2 *About the episode in general*
Frequency?
Place fixity/recurrence?
Time fixity/recurrence?