COLLECTING SOCIAL NETWORK DATA TO STUDY SOCIAL ACTIVITY-TRAVEL BEHAVIOUR: AN EGOCENTRED APPROACH

Juan Antonio Carrasco, Bernie Hogan, Barry Wellman, and Eric J. Miller
University of Toronto, Toronto, Canada

INTRODUCTION

Within the study of travel demand under the activity-based approach (Axhausen and Gärling, 1992; EIRASS, 2004), there has been an increasing interest in the effect of social interactions in activity-travel behaviour (Bhat and Lawton, 2000), and in particular on how social networks influence activity-travel decisions. Complementary, there is a growing interest in the study of social activities, recognising their importance in the overall travel patterns, and their behavioural difference with more studied purposes, such as work and shopping (e.g. Bhat and Gossen, 2004). Despite this interest, an empirical understanding of these two aspects has been limited so far due to the lack of proper data collection efforts that explicitly link travel behaviour and social interactions. Moreover, although data collection and modelling techniques have gone very far understanding individual activity-travel decision making processes in time and space (e.g. Doherty and Miller, 2000), we still know very little about the linkages between social and spatial interactions.

Axhausen (2004) explicitly illustrates about the limits currently faced to incorporate the “social dimension” in travel behaviour:
“Transport planning and even more so transport modelling 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 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.” (p.3)

These empirical and data collection limitations not only involve the need of more detailed and large scale spatial representation of meeting patterns (as Axhausen argues), but also the necessity of knowing the specific characteristics of those interactions. In fact, although “with whom activities are performed” constitute now a common standard question in the transportation data collection literature (see the discussion by Arentze and Timmermans, 1997, and an example in Doherty et al., 2004), they do not necessarily capture the necessary information to understand the importance of the social dimension in activity-travel patterns1. In fact, asking only with whom the activity was performed does not necessarily reflect the relevance of those individuals in the overall activity-travel pattern, since the interaction with a specific person is collected only if the respondent performed an activity with that specific person in the time frame the instrument covers. This issue is particularly relevant since in general the set of activities collected is very limited, considering that variety seeking is present in long time spans, especially in the case of social activities (Schlich et al., 2004). This fact makes very difficult to understand and study aspects such as time and space fixity and recurrence in activity-travel patterns. Furthermore, asking only with whom the activities were performed does not inform about the overall set of people with whom individuals interact – their social network – showing a very partial picture of the individuals’ overall interaction patterns.

In addition, collecting only with whom the activity was performed with, has also a conceptual consequence from a travel behaviour perspective, since the social dimension in that case becomes an attribute of the social activity (at the same level as destination or time of the day, for example), rather than the cause of the social activity. That approach of collecting and analysing information about interactions as attributes may hide behavioural 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. I may not visit an old lady once a week, traveling one hour only because she is a nice lady, but because she is my grandmother; and I travel one hour not necessarily because trade-offs of distances and costs, but because she lives there. Although somehow silly, that example illustrates how the social dimension is cause of behaviour and not an attribute.

This paper presents a data collection effort designed to address some of these challenges, looking for a better understanding of the links between social travel behaviour and social interactions. The instruments designed in that work explicitly collect the individuals’ social

1 Goulias et al. (2004) develop an interesting approach to gain insight about the importance of interactions in activities, asking “for whom” the activity was done. However, their approach mainly concentrates in the dichotomy of self serving versus altruistic behaviour, capturing a different dimension than the one of interest in this work.
networks using an egocentred scheme, which – as it will be explained in more detail later – is constituted by the social structure of specific individuals (egos), and by the interplay between the individuals social activities and their social networks. More explicitly, the objective of the data collection is providing a data set that could help studying 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 use (ICT), such as telephone and Internet. Regarding the generation and spatial distribution of social activities, the major interest of the social network approach resides in the explanation that can provide to understand the individuals’ social patterns, testing the hypothesis of 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, especially in times when new technologies are increasingly embedded in ordinary life (Haythornthwaite and Wellman, 2002).

Also, a background motivation of this work is studying the general usefulness of social network approaches to understand travel behaviour, especially considering remarks such as those by Axhausen (2002), who argues 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 behaviour over long time horizons, anyway…” (p. 3). However, the study of social networks and travel behaviour is still in early stages, and their research concentrates only in the economics literature of interactions, and the study of social influence (e.g. Dugundji and Gulyás, 2003; Dugundji and Walker, 2005; Páez and Scott, 2005). The motivation of studying social networks is also based in the interest on not only outcomes, but behavioural processes (Lee-Gosselin, 2002), echoing the long discussed need of incorporating complementary explanations to the dominant microeconomic paradigm (Pas, 1990; Gärling, 1998). From a data collection perspective, the focus in social networks is related with the recent interest in exploring broader techniques such as qualitative data collection (Cliffton and Handy, 2001) that can help to have a better understanding of the underlying behavioural processes in activity and travel. Finally, this data collection effort and subsequent analysis has as a final objective of informing and enriching with the social network perspective the behavioural components of operational activity travel demand models (e.g. TASHA, Miller and Roorda, 2003; Miller et al., 2005) and integrated transportation and land-use models (e.g. ILUTE, Salvini and Miller, 2005).

The rest of the paper is divided as follows. The second section reviews the key concepts that served as background in the data collection effort; the third section presents the data collection, describing both the survey and interview instruments, focusing in most relevant parts for the interests of this paper; section four discusses the key elements of the data collection; and finally section five presents some conclusions and suggestions.

**Key Concepts**

This section presents the key background concepts that motivate the data collection effort. First, some of the central concepts of the social networks approach are briefly reviewed;
followed by the key concepts of the interplay between social networks and activity-travel behaviour that motivated the design of the instruments. Finally, the main characteristics, issues, and challenges in social network data collection are discussed, setting the context of the particular design chosen for this study.

**The Social Networks Approach**

The social networks approach used in this work is more than metaphorical, and draws from a long tradition in Sociology; for a further revision of early work and origins of the paradigm and techniques, see Wellman (1988), and for an in depth revision about the approach see for example Wellman and Berkowitz (1988), Knoke and Kuklinski (1988), Scott (1991), Wasserman and Faust (1994), and Carrington et al. (2005). Drawing from Tindall and Wellman (2001), this subsection briefly describes the key ideas of this paradigm. Those authors 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 network, that is, a set of actors (nodes) and a set of relationships connecting pairs of these actors” (p. 1-2).

Thus, two key components define this paradigm: **actors**, who represent different entities, such as groups, organizations, nations, as well as persons; and **relationships**, 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, behaviours, and cognitions”. According to Tindall and Wellman, this focus is based in the belief that “knowledge about the structure of social relationships enriches explanations based only on knowledge about the attributes of actors”. In other words, social network analysis conceives the overall behaviour as more than the sum of individual behaviours, and contrasts with “explanations that treat individuals as independent units of analysis”, as those traditionally used in travel behaviour research.

In addition, the social network approach is inherently **multilevel**, since takes into account the combined effect of two levels of analysis: the level of the overall social structure, and the level of each pair relationship between network members. From a data collection and analysis perspective, this multilevel approach implies the focus in two kinds of characteristics:

- **Structural characteristics** or attributes of the social structure, which reflect the overall pattern of interactions among individuals, such as network size and density, and
- **Composition characteristics** or attributes of each individual who belong to the network and their relationship with other members, such as gender and interaction patterns

Thus, behaviour is explained not only through personal attributes, but by using social structure and composition attributes that incorporate the interaction among the different social network members. Also, this multilevel vision remarks the fact that the whole is more than the sum of its parts, that is, social phenomena cannot be understood by solely studying
individual characteristics (such as socio-economic attributes), but also by analysing the social structure characteristics that emerge from the interaction of those individuals. In that regard, values, attitudes, and norms emerge from where individuals are situated in the social structure (Tindall and Wellman, 2001), as well as their overall interaction patterns with others network members.

Social Networks and Activity-Travel Behaviour

This subsection describes the main hypothesis underlying in the data collection effort, and their relationship with the motivations previously exposed in the introductory section. This hypothesis and those motivations, as well as the previous social networks definitions, serve as a background to define the key conceptual definitions that determine the specific aspects gathered in the data collection effort.

Main hypothesis. As it was discussed before, the main motivations to collect this data is to study the effect of the interplay between social networks and activity-travel behaviour, exploring the importance of this interplay in aspects such as the generation and spatial distribution of social activities, and the overall relationships between different communication media, such as the Internet, telephone, face-to-face, and social interaction. Given those motivations, 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 the multilevel individuals’ social network characteristics (both composition and structure attributes). This hypothesis implies a specific approach to conceptualise the generation and spatial distribution of social activities, and the communication media patterns among individuals.

From the point of view of the generation of activities, the previous hypothesis gives a central role not only to the individual’s socioeconomic and lifestyle attributes (e.g. Lu and Pas, 1999), but also to the people with whom the individual performs social activities, and who constitutes the individual’s social network. Then, ties or links between the individual and other people in the network represent a flow of potential activity and travel generated by the interaction between those “nodes” of people and the individual of interest. Furthermore, the multilevel perspective of social networks implies that sources of explanation of activity and travel are not only the characteristics of each individual, but also the structural characteristics of the network they are embedded. Similarly, in the case of the spatial distribution, the above hypothesis implies studying the activity destinations from another, complementary perspective. In this perspective, the observed people’s activity patterns are understood as a direct consequence of not only people’s preferences and restrictions, but also of the location of their social network. In other words, people’s spatial distribution of activities is assumed as partly determined by the spatial location of the individuals who belong to the network. Finally, regarding communication patterns, the previous emergence hypothesis conceives interpersonal relationships (expressed in the individuals’ social network) as the focal element from where different communication interaction media emerge, including not only telephone and Internet, but also face-to-face and specific socialising episodes. This focus of communication subjected to the social network potentially sets a more intuitive and consistent
Collecting Social Network Data to Study Social Activity-Travel Behaviour: An Egocentred Approach

Conceptual definitions. A key step in the design of the data collection effort was identifying the key conceptual definitions, derived from the previous hypotheses, that determine the operationalisation of the information gathered in the study instruments. Four key sets of definitions were developed here: social networks, social interaction episodes, generation of social episodes, and spatial distribution of social episodes.

Social networks are composed by two levels of characteristics: structural characteristics, defined by the overall network attributes (e.g. size or density), and composition variables, defined by the characteristics of network member, and of each link or tie that joins them. A key characteristic of those links is the strength of the tie, which defines the level of “proximity” or closeness between two given network members. Other tie classifications are according the frequency of interaction between two network members, and the roles of each of these individuals.

A social interaction episode is defined in this work as an activity or a set of activities performed by two or more individuals primarily for recreational or support purposes, and that can be performed face-to-face or virtually (telephone or the Internet in the latter case). A social episode can be recurrent or non-recurrent according to the regularity it is performed. In addition, social episodes have a duration, start time, and location, which can also be recurrent (i.e. individuals perform regularly that activity at that time, duration, and location) or non-recurrent (i.e. attributes vary from occasion to occasion). Social episode’s location can also be concurrent or non-concurrent, depending on whether all members are in the same place or not. A classification scheme chosen in this data collection divides social episodes into two broad categories according to the location they are performed: at a specific actor’s home (e.g. hosting / visiting), and at public places (e.g. pubs or restaurants).

The generation or decision to perform of social episodes can be explained by the individuals’ propensity and opportunity to engage in a social activity (Hägerstrand, 1970; Chapin, 1974). Propensity depends on the individuals’ intrinsic characteristics and needs, which include not only socioeconomics and lifecycle, but also their social network characteristics, and their interaction patterns with other people (e.g. frequency and type of media used). Opportunity of engaging in social activities is less a focus in this data collection effort, and includes time and space opportunities and constraints, mainly related by scheduling processes, although can be also defined in the medium term by the spatial distribution of the individuals’ social networks.

Finally, the spatial distribution of social episodes can be understood using the concept of social anchor points, which describe the main places where the individual “moves around” when he/she interacts with other members of his/her social network. Several anchor points can be defined a priori, such as social network individuals’ home, work and school locations; public places locations (e.g. pubs or restaurants), and other places of people related with those who interact. These anchor points generate the key pivotal points of the social activity space (Horton and Reynolds, 1971; Schönfelder and Axhausen, 2002).
Social Network Data Collection

A more practical, but not less important set of concepts are those related with the specific techniques and issues of collecting social network data. There is a long tradition in these kind of efforts; for a review in overall data collection methods see Marsden (1990; 2005). Overall, social network data collection is a complex process; some of its challenges are:
- Network boundaries are difficult to define (Marsden, 1990)
- People does not easily recall their network members (Brewer, 2000), and need appropriate “prompts” to elicit them,
- Networks are in general very large and need an appropriate sampling strategy to make the data collection feasible,
- Different social network members may have different importance depending on the phenomenon studied. For example, for the purposes of social activity-travel behaviour, not all workmates may be relevant; however, a more complete list of workmates may be important if another kind of phenomenon is studied, such as interorganizational relationships.

This subsection performs a brief review of the social network data collection techniques relevant for this study, especially concentrating in the chosen method and the challenges enumerated above. From a theoretical perspective, any data collection needs to appropriately represent the multilevel nature of social networks, that is, it needs to capture both the level defined by the individual features of population members, and the structural characteristics associated to the overall social network; and also the level defined by the individual-individual interaction, that incorporates the specific characteristics of each relationship.

Whole versus egocentric networks. Most of social network data collection can be divided into “whole” and “egocentric” networks. Whole network studies examines actors “that are regarded for analytical purposes as bounded social collectives” (Marsden 2005: 8), that is, their boundary is predefined; in other words, actors in whole network studies are named in a closed (usually pre-defined) “list”, known a priori. However, the possibility of studying whole networks in urban settings, like those of main in activity-travel behaviour research, is unpractical since they often involve very large populations where the network members are not known in advance; this makes egocentric data collection the only feasible method. Egocentric network studies concentrate in one specific actor or ego and those who relate with him, called alters; that is, from the participant’s perspective (the ego), constitutes a “network of me” or a network of actors (the alters) with whom the participant has some relationship. From the multilevel perspective discussed before, egocentric data collections imply gathering data at the following two levels: i) an ego-network level, gathering the ego characteristics, and the overall structural features of the ego’s network; and ii) an ego-alter level, gathering alter characteristics, and alter-ego relationship features.

Boundaries. Another important challenge in both egocentric and whole network studies is specifying the boundaries, or which actors to include within the network. For egocentric networks, the problem can be seen as twofold: selecting “appropriate” egos, and selecting “appropriate” members in each egocentric network. While an adequate ego sampling offers no more challenges than having a good representation of the context studied (urban setting in this
Collecting Social Network Data to Study Social Activity-Travel Behaviour: An Egocentred Approach

case), selecting an appropriate boundary for the egocentric network requires a more complex process. In fact, as it was discussed before, the issues resides not only from the fact that networks are in general very large, but also from the other problems enunciated before, that is, the need to help respondents to recall the most complete information as possible, and the need of sampling the adequate members for the phenomenon studied. In egocentric methods, the instrument that sets network boundaries is the name generator, which consists in free recall questions that define the members elicited from the network (Burt, 1984; Marsden 2005).

Name generators. Although some attempts to capture “total networks” have been made (e.g. McCarty et al., 1997), name generating questions elicit “only a fraction of a respondents’ contact” (Marsden, 2005: 12). The key decision then is choosing the appropriate specific question(s) that will elicit the network members relevant for specific phenomenon of interest, also 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 (e.g. the American General Social Survey, e.g. Marsden, 1987) or can be unlimited (as in the data collection described in this paper). There is an extensive literature that compares different name generators (e.g. Bernard et al., 1987; 1990; Marsden, 1990; 2005; Burt, 1997; Bailey and Marsden, 1999; Brewer, 2000; Straits, 2000; Feld and Carter, 2002), discussing aspects such as their influence in network size, the number of “core” and extended network members that each elicit, the importance of the instrument’s context, the relevance of the order and wording of questions, and the forgetting phenomena.

In addition, name generators can be used to measure the tie strength between ego and alters and between alter-alter pairs, as is done in the Connected Lives Study. Tie strength is a very relevant measure regarding the alter’s importance in the overall social network. As before, which indicator is chosen will impact on the reliability of tie strength measures, being emotional closeness the most usual and accepted (Marsden and Campbell, 1984). An important aspect in egocentric networks is the fact that tie strength is only from the ego perspective; as a consequence, in the case of ego-alter pairs, tie strength reciprocity is not known, and in the case of alter-alter pairs, tie strength is gathered according to the ego’s perception.

Finally, although name generators are by far the more extensively used techniques to gather network data, other two ways of eliciting network characteristics exist, which have been constructed to measure very specific purposes: position generators (e.g. Nan et al., 2001) and resource generators (Van der Gaag and Snijders, 2004). Position generators elicit people known by the participant who have specific social positions (e.g. occupations, race, gender), whereas resource generators elicit people known by the participant who have specific resources that can be useful for the participant (e.g. ability to repair a car, knowledge of a specific subject, certain income level).

Name interpreters. After eliciting network members, a second set of questions is usually performed to obtain more information about the social network; specifically about the characteristics of each alter (e.g. socioeconomics, relationship with the ego), and the ego-alter relationship (e.g. frequency and characteristics of interaction). From a practical point of view, a key challenge of name interpreters is gathering an adequate amount of information in a non-
tedious, relatively short, and reliable way; issues especially critical when the number of alters is not defined beforehand, as in this study. In order to avoid this issue, a sampling strategy is usually performed (Marsden, 2005); however, to the authors’ knowledge, no firm guidelines can be found from the literature about this issue.

**Instrument.** A final important aspect in social network data collection is the instrument to be used. Both paper and pencil questionnaires (e.g. Burt, 1997) and computer-based (Bernard et al., 1990) techniques have been used in the past, with little knowledge about the effect of that election and the effect on the results of name generators and interpreters (Marsden, 2005). As it will be discussed after, this data collection adopted a more “analog” approach, that is, paper and pencil, and personal interview instruments (Hogan et al., 2005a).

**THE CONNECTED LIVES STUDY**

This section describes the overall study and specific instruments used to capture the link between social networks and social activity-travel behaviour. After a brief presentation of the overall study, the survey and interview instruments are described in detail.

**The Study**

The data was gathered in the East York area of Toronto, Canada, as part of the “Connected Lives Study”, a broader study composed by surveys, interviews, and observations about people’s communication patterns. The first author conducted the study in conjunction with sociologists (among them, the second and third author) and social workers of the NetLab, part of the Centre of Urban and Community Studies at the University of Toronto, and lead by Barry Wellman. The multidisciplinary setting allowed for a rich crossed-fertilisation in data collection techniques, and a broad set of information gathered. For an overview of the overall study see Hogan et al (2005b). The study was composed by two stages: surveys to a random sample of 350 people from the East York area in Toronto, and interviews and observations to a subsample of 87 people. The East York area is located in the East side of downtown Toronto, and is fairly representative of the overall inner city characteristics regarding sociodemographics and general transportation level of service (see more details in Hogan et al., 2005b).

**Survey**

The survey was designed between November 2003 and June 2004, and was administered between July 2004 and March 2005 to 350 people who acceded from a total random sample of 700 people. The survey covered a variety of aspects that include: individual and household characteristics, individual and household computer and Internet use; work patterns outside and inside the home; division of household chores; interaction and communication patterns within the household (including those with children); social network composition; relationships between the individual social network and aspects such as planning patterns,
sources of new information, use of devices to maintain contact with network members, and social support; and measures of extroversion and openness. For the purposes of this paper, social network composition is the most relevant part of the survey, and it will be discussed more in detail now.

As it was discussed before, a key design aspect in the data collection is the name generator section, that is, the way members of the individual network are prompted to the respondent, especially considering that social networks tend to be very large, and eliciting methods are capable to gather only a portion of them. The option made in the study was concentrating in the individual’s affective network or a network of people the respondent defines as close, approach that seems to be useful for understanding communication and sociable activity-travel patterns. Concretely, the respondents were asked to name the people who live outside the household who they feel very close and somewhat close. Very close people was defined as “people with who discuss important matters with, or regularly keep in touch with, or there for you if you need help”. Somewhat people were defined as “more than just casual acquaintances, but not very close”. This “closeness” approach defines two key aspects. First, closeness becomes the measure of strength of the tie: strong ties corresponding to very close people, and weak ties corresponding to somewhat close. Second, closeness also defines the “boundary” of the social network and the communication and sociable activity-travel patterns captured in the data; concretely, the social network is defined only by people who are more than acquaintances.

Table 1 shows the specific items prompted in the survey. Specifically, respondents were asked how many very close and somewhat close people living outside their households they have in their networks in each of the following roles: immediate family, other relatives, neighbours, work or student mates, people that know only online, people from organisations, other friends, and others not included in the previous. As an aid, respondents were provided with a working sheet that helped them to write the names of the people in each category, sheet that was kept by the respondents. After prompting the number of network members in each role, further questions included how many very close and somewhat close network members in each gender, with different ethnic heritages, living outside Canada, and living in Canada but at more than one hour’s travel (this last as a proxy for distance). Finally, a set of questions about frequency and media of interaction were done. Concretely, respondents were asked how many very close and somewhat close people they typically: i) call by cell phone, ii) call by regular phone, iii) send an email, iv) send an instant message, v) talk with face to face, vi) meet at restaurants or bars, and vii) visit or host them as a visitor. Each of the previous questions differentiated between two time horizons: at least once a week, and between once a week and once a month.

The approach adopted in the survey consisting in capturing the number of individuals of the respondent’s social network and those with whom respondents communicate and have activity-travel patterns in their networks, could be defined as a “meso” approach. This “meso” approach in the data collection implies capturing structural features, such as size of the network, approximate density, and aggregated composition by aspects such as role and gender. This approach contrasts with a “micro” approach that also captures composition variables or “individual” characteristics, such as each alter’s gender, spatial location, and the
interaction patterns between each alter and the respondent; this kind of “micro” data was gathered with the interview instrument. Although that valuable “micro” information is lost with the “meso” approach, the survey provides a quick and feasible way of prompting structural features in a paper and pencil instrument, without using the complex and expensive method used in the following interview section.

Table 1: Survey questions

<table>
<thead>
<tr>
<th># of very close and somewhat close people who are</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate family</td>
<td></td>
</tr>
<tr>
<td>Other relatives</td>
<td></td>
</tr>
<tr>
<td>Neighbours</td>
<td></td>
</tr>
<tr>
<td>Work or school mates</td>
<td></td>
</tr>
<tr>
<td>People you know only online</td>
<td></td>
</tr>
<tr>
<td>People from organisations</td>
<td></td>
</tr>
<tr>
<td>Friends not included above</td>
<td></td>
</tr>
<tr>
<td>Other people not included above</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td></td>
</tr>
<tr>
<td>Live in Canada and more than one hour’s travel away</td>
<td></td>
</tr>
<tr>
<td>Live outside Canada</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of very close and somewhat close people with whom the ego interacts typically i) at least once a week and ii) between once a week and once a month:</th>
</tr>
</thead>
<tbody>
<tr>
<td>By cell phone</td>
</tr>
<tr>
<td>By regular phone</td>
</tr>
<tr>
<td>By email</td>
</tr>
<tr>
<td>Using instant message</td>
</tr>
<tr>
<td>Meeting face to face</td>
</tr>
<tr>
<td>Meeting at a bar or restaurant</td>
</tr>
<tr>
<td>Visiting or hosting as a visitor</td>
</tr>
</tbody>
</table>

A key assumption in both the survey and interview stage is the record of “typical” communication and activity-travel patterns in the questions, rather than observed or stated patterns. This approach was adopted since the main interest was capturing the overall communication and social activity-travel patterns of the respondents, something that would have required very detailed and long time span instruments if actual observations of the realised patterns would have been recorded. In fact, previous experiences show the need of instruments that record more than one month of detailed records in order to properly capture social patterns (e.g. Schlich et al., 2004). Thus, in order to avoid more interviewee burden and due to cost restrictions, the approach taken consisted in asking the respondent’s perceived “typical” patterns. Although this approach potentially incorporates biases and possible differences with the actual respondent’s patterns, it also adds the opportunity of incorporating “perceived” or “usual” behaviour that can be contrasted with other more standard data that uses observed and stated patterns. More importantly, at the very least, recording typical patterns constitutes a reasonably adequate proxy of the actual respondent’s behaviour. Finally,
it is important to note that there are not a priori technical issues on including observed and stated patterns in conjunction with social network data collection; however, the length and potential burden for the interviewee prevented this study to do so.

A final remark is that these questions concentrate in individual behaviour, rather than household based. In fact, respondents were asked to prompt their social network members living outside the household, the same as in the case of their communication and activity-travel patterns. This option was made since other sections of the survey already contain a great detail of inter-household interaction patterns, and also as a way of better controlling the influence of individual characteristics in their social network and interaction patterns.

**Interview**

The interview was designed between October 2004 and January 2005, and was conducted between February 2005 and April 2005 to sub-sample of around a quarter of the people that answered the surveys (87 people). They consisted in an average two and a half hour session generally at the home of the individuals, and were conducted by graduate students, members of the NetLab team, and by the first author. The interview consisted in five parts: i) household information and relationships, ii) work at home, iii) name generator and interpreter, iv) social episodes, v) social support, and vi) health and cultural information seeking patterns. The parts of interest for this paper are the third (name generator and interpreter) and fourth (social episodes) sections, which took in average between 45 minutes to 1 hour. The following subsections describe the name generator, name interpreter, and social episode sections in detail.

**Name generator.** The goals of the name generator were three: create a participant-aided sociogram, maximise the size and richness of the egocentric network, and facilitate the record of network’s connectivity. A sociogram can be intuitively defined as a drawing that represents the respondent’s social network, which permits visualising the respondent’s social network, and which contains the name of the respondent’s alters, and the ties or links among them. In the case of the study, the sociogram is represented by a series of four concentric circles where the ego is at the centre, and the alters are situated around the ego, in any of the circles (see Figure 1). The importance of the sociogram is threefold. First, it is an instrument to collect social network data in an intuitive and easy way for respondents, lowering their burden – especially in the case of senior and less educated people – and facilitating the incorporation of the highest possible number of network members of interest. Second, the sociogram makes connectivity easier, helping to record ties among social network members in a more reliable and complete way. Finally, the sociogram serves as a cognitive aid to prompt further questions about communication and social activity-travel patterns, which are made in the name interpreter section. The sociogram was built in three steps, described in Table 2 below.

In the first step (shown in Figures 2a and 2b), respondents were asked to elicit their network members in a “name template”, composed by post-it papers where the respondents could name each alter. Network members were elicited using the definitions of very close and somewhat close people, discussed in the previous section; in all the sociogram building
process, these two definitions are differentiated by two different colours (blue for very close and purple for somewhat close in the figures). Contrasting with the survey, the list of names is elicited in a “free” recall order, that is, individuals’ role is not present in the name elicitation. Only after the respondent writes the entire list of people, the same set of roles used in the survey were presented, both as a “memory jogger” (in case the respondents have forgot someone important), and so that respondents could record each individual’s role. Alter’s roles can be multiple; e.g. a workmate can also be a friend, phenomenon called *multiplexity*, which is explicitly allowed and encouraged to be recorded. At this stage, two kinds of numbers are recorded for each name: a **rank** number, which represents the order in which the alter is elicited by the respondent (the smaller number, the higher the rank); and **role** number(s), which represent the code of the corresponding alter’s role(s) with respect to the respondent.

### Table 2: Steps to build the sociogram

1. **Names**
   1.1. Eliciting very close and somewhat close people
   - Close people
     - discuss important matters with, *or*
     - regularly keep in touch with, *or*
     - there for you if you need help
   - Somewhat close people
     - More than just casual acquaintances, but not ‘very close’
   1.2. Roles of each person, allowing for multiple roles (*multiplexity*)

2. **Locating names very close and somewhat close 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 (group and **foci** formation)

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

In the second step, respondents were asked to situate each post-it paper (representing each social network member) in one of four concentric circles, according to how “close” they felt about them; the closest they felt, the closer the ring from the centre, starting with the very close people, and followed by the somewhat close people (see Figure 3). Interestingly, this second measure of closeness was kept ambiguous, and does not necessarily coincide with the previously discussed definitions of closeness, adding a second and possibly complementary measure of emotional closeness. This step was entirely left to the respondent; the only
instruction given to the respondent was placing people that know each other nearby, in order to help the following step.

The third and final step (Figure 4) consisted in recording the connectivity among the social network members. Since the relationship of each alter with the ego had been already recorded, this stage concentrates mainly in the relationship of alters among each other. First, respondents were asked to record – drawing a circle – the groups of people who they thought were all very close among them (left upper image in Figure 4). Second, respondents were asked to record very close and somewhat close ties among pairs of network members, called dyads, drawing a line between each name (right upper image in Figure 4). These two steps were repeated for the somewhat close people (left and right lower images in Figure 4). An important remark regarding the record of network connectivity is the fact that egocentred data collection tends to collect ties from the perspective of the respondent; in this concrete case, how close s/he feels with respect to her/his network members, and how close s/he thinks the network members are among each other. Although technically is possible to record ties in each direction –that is collecting the closeness from the alter’s perspective – the complexity and cost involved seems not necessary for the purposes and interests of this work.

The final result of the name generator section can be seen in the right lower image in Figure 4. The respondent has: i) elicited each social network name and their role, differentiating very close and somewhat close people, ii) located each alter in the sociogram according to a loosely defined emotional closeness, and iii) recorded the ties among all social network members, differentiated by strength. This sociogram will be the key cognitive aid for the questions about communication and activity-travel patterns of the name interpreter and social episode sections. In terms of the data collected, the method used in this section has provided a “step-by-step” procedure to gather the participants’ social network, base information to capture their consequent communication and activity patterns.

Finally, prior the next section, and in order to test the meaning of closeness for the respondents, questions about the reasons why respondents chose as very or somewhat close were asked to a sample of four alters (the highest rank number in each ring, as shown in the red arrows of Figure 5).

Name interpreter. The name interpreter section records the communication and social activity patterns between the participant and a selected number of network members; Table 3 shows the questions prompted in this question for each alter. Pre-tests of this section showed the necessity of sampling network members for whom to retrieve the information. The strategy used in this work contemplated a sample of fifteen alters, as a way of compromising both the need of a representative number of people, and keeping the length of the interview (and thus the tiredness of the participant) at a reasonable level. The fifteen alters were chosen the following steps: i) pick the three alters with highest rank number in the inner ring; ii) starting from the inner ring, pick the alter with the highest ring (if not alters already sampled remain), and then move outwards to the following ring; iii) repeat the previous step until fifteen alters have been sampled or there is no more alters in the network. Figure 6 shows an example of the sampling scheme adopted; orange arrows represent the sampled alters. Then, the sampling scheme elicits names from all the rings on the sociogram and from both very close and
somewhat close people, thus “covering” the overall network in the best possible way, but at the same time privileging the emotionally closer alters (i.e. those in the inner rings). From the communication and travel patterns point of view, this sampling scheme pursues balances two objectives: on the one hand, eliciting representative alters of the overall network, in order to have a vision of the overall ego’s patterns; and on the other hand, capturing a high proportion of the more “relevant” communication and activity patterns of the respondent, assuming that those are with the more emotionally closer people. In sum, as can be seen from Figure 6, within this sampling scheme the overall network is covered, the inner rings contain more sampled alters, and there is consistency between sampled alters in the previous section (about reasons of closeness in red arrows), and name interpreter questions (orange arrows).

Table 3: Name interpreter questions

<table>
<thead>
<tr>
<th>Alter’s characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, relationship, job, ethnic heritage, home location, most frequent place of interaction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Face-to-face</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (# per year, month, week, day)</td>
</tr>
<tr>
<td>On average, how long spend together (# minutes or hours)</td>
</tr>
<tr>
<td>Who go to see the other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Socialising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (# per year, month, week, day)</td>
</tr>
<tr>
<td>On average, how long spend together (# minutes or hours)</td>
</tr>
<tr>
<td>Who invites</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Telephone contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (# per year, month, week, day)</td>
</tr>
<tr>
<td>On average, how long conversations are (# minutes or hours)</td>
</tr>
<tr>
<td>Who calls who (Scale 1 – 5; 1 means “me”, 5 means “him/her”)</td>
</tr>
<tr>
<td>Landline or cell phone use (ego and alter)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Email contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (# per year, month, week, day)</td>
</tr>
<tr>
<td>On average, how long emails are (scale 1 – 5; 1 means “short”, 5 means “long”)</td>
</tr>
<tr>
<td>Who sends emails to who (Scale 1 – 5; 1 means “me”, 5 means “him/her”)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instant message contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (# per year, month, week, day)</td>
</tr>
<tr>
<td>On average, how long the conversations are (# minutes or hours)</td>
</tr>
<tr>
<td>Who starts the conversation (Scale 1 – 5; 1 means “me”, 5 means “him/her”)</td>
</tr>
</tbody>
</table>

After the sampling is done, the name interpreting questions about each alter and associated communication and activity-travel patterns were asked; table 3 shows the specific information gathered in this section. First, respondents were asked to record information about the
characteristics of each alter, including age, relationship, job, and ethnic heritage; also two spatial locations were recorded in this section: the alter’s home location and the most frequent place of interaction with the respondent. Second, respondents were asked a set of questions about their communication and interaction patterns with each alter, considering face to face, socialising, telephone, email, and instant messaging, and focusing in their *usual* frequency and duration, who usually triggers the interaction, and technology use (e.g. cell phone versus landline). Face to face and socialising were explicitly separated since a main interest is precisely differentiating between instrumental interactions (such as those existing in workplaces) and social interactions (visiting, hosting, going to pubs and restaurants); also capturing the circumstances when face to face interactions are different from socialising (e.g. how much the alter’s role influences this aspect). In the case of information and communication technologies (telephone, email, and instant message), the focus is on *usual* frequency and duration of interactions in the same way as in the survey section. The questions also focus in the “direction” of the interaction, that is, who starts or “triggers” the interaction, prompting whether the ego or the alter has more propensity to initiate the contact.

Finally, although the name interpreter section is part of the interview, a paper and pencil “mini-survey” was used to record the information, leaving up to the respondent whether they answer the questions by their own, or whether the interviewer would help them. Yet the section was also recorded, strategy that proved to be very useful since in general respondents tended to add contextual extra information about the alters and interaction patterns, adding potentially very useful “stories” about their behavior. This method also provides an interesting mix between quantitative / structured data, and qualitative / contextual information.

*Social episodes.* The last section of interest for the interview consisted in recording social episodes the participants engaged with some of the alters previously elicited in the name generator and interpreter sections. As before, social episodes are defined as those involving visiting, hosting visitors, or meeting in restaurants, pubs or similar places; this definition defines clearly the boundaries of an activity type that otherwise becomes too fuzzy and ill-defined. The already complex design of the study did not leave room for an extensive collection of all the respondent’s social episodes, also considering that a complete account of social activities requires collecting data for long time horizons. Instead, the method used in this section involves a small sample of social episodes that potentially serve as a proxy to understand the overall respondent social patterns, without pretending being a detailed and extensive account of the individual’s social activities. With that purpose, six participants’ social episodes were recorded, using a sampling strategy consistent with the main background assumption of all this data collection effort, that is, communication and social activities emerge from the individuals’ social networks.

Specifically, participants were prompted of social episodes with selected six social network members from the previous fifteen elicited in the name interpreter section, with whom they socialised in the Greater Toronto Area (the boundaries chosen for the study). These six alters were chosen using the following sampling scheme: i) starting from the inner ring, pick the alter with whom the ego *most frequently* performs a social episode within the Greater Toronto Area, and move outwards to the following ring; ii) after covering the four rings, move to the
Collecting Social Network Data to Study Social Activity-Travel Behaviour: An Egocentred Approach

inner ring; ii) repeat steps ii) and iii) until six alters have been chosen. This sampling scheme tries to keep the same balance of objectives as in the name interpreter section, between sampling the overall network and emphasising the more emotionally close alters, also maintaining the consistency of sampled alters in each section (see green arrows in Figure 7). For each of the six alters, participants were asked the following characteristics of their last social episode with them: what the activity was about, when it was (time of the day, day of the week, duration), where it was (detailed spatial location), who other network members were involved, and which transportation mode was used (see Table 4). Also, other qualitative questions about activity planning were made, such as who triggered the episode, what media was used, and how far in advance was planned. Finally, the participants were asked about the episode in general, gathering whether the episode was frequent or not, and its recurrence and fixity in time and space (i.e. whether time and place were the same or whether there were changes or rotation in the spatial and time patterns).

As in the name interpreter section, the interview setting provided and opportunity for not only gather the more structured information mentioned before, but also the opportunity to understand in a more unstructured way the context of each of the episodes and social patterns. In addition, interestingly enough, the perception of interviewers is that six social episodes seem to be an adequate number to have a good overview of the individual’s social patterns; this is in part reaffirmed by results in Europe that show eight locations capturing 80% of the overall leisure activities (Schlich et al., 2004).

Table 4: Social episode questions

<table>
<thead>
<tr>
<th>About the specific social episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>What it was about</td>
</tr>
<tr>
<td>When it was (time of the day, day of the week, duration)</td>
</tr>
<tr>
<td>Where it was (detailed spatial location)</td>
</tr>
<tr>
<td>Who else was involved</td>
</tr>
<tr>
<td>Transportation mode</td>
</tr>
<tr>
<td>Planning: how was planned (routine, media), how far in advance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>About the episode in general</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td>Place fixity / recurrence</td>
</tr>
<tr>
<td>Time fixity / recurrence</td>
</tr>
</tbody>
</table>

**DISCUSSION AND CONCLUSIONS**

This paper has presented a data collection effort designed to incorporate the “social dimension” in travel behaviour, specifically using the social networks approach to study social activity-travel behaviour. The main hypothesis in the background of the overall data collection was that individuals’ communication and social activity patterns can be properly elicited from their social networks. Having this hypothesis in mind, the design of the data
Collecting Social Network Data to Study Social Activity-Travel Behaviour: An Egocentred Approach

Collection instruments incorporate the key characteristics of the social network paradigm, remarkably its multilevel structure, which considers both the structural and composition characteristics, remarking that communication and activity-travel activity behaviour does not depend only in individual actors, but also in their interactions. In addition, the design of the data collection incorporated explicit conceptualisations about the phenomenon of interest, especially with respect to the interplay between social networks, the characteristics of social episodes, and their generation and spatial distribution.

An interesting characteristic of the overall study is the use of two different instruments to measure the individuals’ social networks and patterns: a paper and pencil survey, and an interview. The survey gathered network data (especially structural characteristics) in a relatively simple and short way to a large sample of individuals, linking their main network attributes with aggregated “meso” estimators of their communication and activity-travel patterns. In contrast, the interview instrument elicited more “micro” information to a fourth of the same previously surveyed people, exploring more in depth and in detail the linkages between individual network characteristics (both structural and composition), and their communication and activity-travel patterns. These two instruments are complementary in two different aspects. First, they have some overlapped information (especially regarding structural characteristics and aggregated social patterns) which can serve to assess their reliability. Second and more importantly, the survey collected data that served as base information for the interview stage, allowing the study to balance aggregated and less depth information from a large sample, with more disaggregated and in-depth information from a sub-sample.

However, as it has been described in this paper, collecting social network data has a number of option and challenges that need to be addressed. A first key design option described in this paper is the way alters are elicited from the respondent, which depends from the name generator questions. The option of focusing in emotional closeness sets the “boundaries” of each respondent’s social network elicited, and thus the activity-travel behaviour patterns captured in these instruments. As it was discussed in previous sections, this boundary seems an adequate compromise between on the one hand methods that have been proven useful in the past in social network data collection, and on the other hand, the specific focus of the study (social episodes). In addition, link in the name generator between emotional closeness and role relationships provides a useful way of capturing reasonable network sizes, addressing issues such as forgetting and context bias.

A second key design option taken in the data collection, specifically in the interview, was the sampling scheme for the specific alters’ attributes and communication and activity-travel patterns with the respondent. As it was described before, three samples were taken: a first, about the reasons of closeness, a second about alters’ attributes and frequency of interaction with the respondent by different media, and a third about specific social episodes between the respondent and some alters. The chosen sampling strategy combined multiple objectives: gathering “consistent” three samples, capturing a relevant subset of the participants’ network (i.e. a relevant sample in their social space), and capturing a relevant set of their activity-travel episodes. In addition, the number of sampled balanced the need of gathering the more possible information with the risk of a high respondent’s burden, also considering that there
are diminishing returns in the amount of the respondent’s communication and activity-travel patterns with respect to the number of alters sampled.

Besides from the previous overall design options, the interview involved other two aspects worth of remark. First, the sociogram used to build the individuals’ social network helped to diminish the burden on respondents and – we believe – also increased their motivation, especially because their step-by-step procedure. In fact, the sociogram created emotional impact (individuals could “see” their network), and empowerment (individuals could “build” their networks), both elements that were very useful to prompt the further questions in the name interpreter and social episode sections. Not less important, the fact that the sociogram was built in an interview setting (allowing the help from the interviewer), and was constituted by simple interfaces (within an interview, and using simple elements such as post-it papers) collaborated to lower the technical burden on the respondents, helping them to concentrate in their networks and behaviour. Second, the used of semi-structured questions in the interview setting encouraged on the one hand to gather systematic quantitative information about the respondents’ network and behaviour, but also encouraging the possibility of gathering more qualitative data about the context where the phenomena studied happen (“stories”). Thus, this design potentially supports the use of mixed qualitative and quantitative techniques.

However, some issues, challenges and assumptions – some inherent to these kind of data collection – need to be explicitly considered. First, the people and patterns prompted is highly dependent on the name generator questions, which can be sensible on aspects difficult to manage, especially the wording of those questions, and the interpretations that respondents may give. Although closeness seems to be one of the most adequate approaches, it is not free from these kinds of biases. Second, even with the efforts employed in this data collection, the network size captured in egocentric techniques is very limited, and represents a small portion of the total respondent’s social network. As a consequence, this approach is useful to capture only a part of the overall individual’s communication and activity-travel patterns; how much these patterns are collected will depend on how many and who network members are prompted. This issue also links with a third aspect, which is the collection of perceived “usual or typical” rather than actual observed communication and activity-travel patterns. While this approach may incorporate some bias, the already long instrument prevented to capture actual patterns, although there is no more deep technical reason to make it in the future. In any case, “usual or typical” patterns at the very least give an overview of the individual overall behaviour. Finally, the complexity of the design and relatively simple interface with the respondent implies that most of the technical burden (and potential bias) is trespassed to the interviewer; and that the transcription stage can be complex (e.g. coding connectivity between alters) and potentially costly (in terms of time and money).

Although these challenges, the data collection effort presented in this paper constitutes a useful way of incorporating the “social dimension” in travel behaviour, linking aspects that have been rarely studied together in a novel way. The approach taken here, regarding explicitly incorporating social network theory and method, provides a hopefully sound base of theory and method to provide more insights about social communication and activity-travel behaviour.
ACKNOWLEDGEMENTS

The authors would like to thank Jeffrey Boase, Kristen Berg, Rochelle Coté, Dinma Dimitrova, Jennifer Kayahara, and Tracy Kennedy, 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 Inna Romanovska and students who participated in the survey data collection stage. The authors would also like to acknowledge the financial support received from the Social Sciences and Humanities Research Council of Canada (SSHRC), Major Collaborative Research Initiative (MCRI).

REFERENCES


Collecting Social Network Data to Study Social Activity-Travel Behaviour: An Egocentred Approach


Pas, E (1990), "Is travel demand analysis and modelling in the doldrums?," in Developments in dynamic and activity-based approaches to Travel Analysis (P Jones Ed.). Aldreshot, UK: Gower.


Berkowitz, Eds.) Cambridge [Cambridgeshire]; New York: Cambridge University Press.

FIGURES

Figure 1: Blank sociogram
Figure 2a: Eliciting names in name template (Step 1 of sociogram building)
Figure 2b: Overview of the name template (Step 1 of sociogram building)
1) Cover of somewhat close number template, 2) post-it notes somewhat close people,
3) sheet to separate both templates, 4) post-it notes very close people, 5) cover very close
number template, 6) Alter’s name, 7) Alter’s role number, 8) Alter’s rank number

Figure 3: Alters are placed in sociogram
(Step 2 in sociogram building; role number is omitted for clarity)
Figure 4: Connectivity among alters (Step 3 of sociogram building)
Figure 5: Reasons of closeness (Sample 1)

Figure 6: Name interpreter questions (sample 2)
Sample 3: Social episode questions

Figure 7: Social episode questions (sample 3)