

Egos' horizons and behind it

Snowball sampling of personal leisure networks

Working Paper**Author(s):**

Kowald, Matthias; Axhausen, Kay W. 

Publication date:

2010-06

Permanent link:

<https://doi.org/10.3929/ethz-a-006104703>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Working Paper 626

Words = 8.600
Tables = 4
Figures = 6

Egos' horizons and behind it: Snowball sampling of personal leisure networks

M Kowald (IVT, ETH Zurich)

KW Axhausen (IVT, ETH Zurich)

June 2010

Egos' horizons and behind it: Snowball sampling on personal leisure networks

Matthias Kowald
Wolfgang-Pauli-Str. 15
Institute for Transport Planning and
Systems
ETH Zurich
CH – 8093 Zürich
Telefon: +41 44 633 30 89
Telefax: +41 44 633 10 57
kowald@ivt.baug.ethz.ch

Kay W. Axhausen
Wolfgang-Pauli-Str. 15
Institute for Transport Planning and
Systems
ETH Zurich
CH – 8093 Zürich
Telephone: +41 44 633 39 43
Telefax: +41 44 633 10 57
axhausen@ivt.baug.ethz.ch

June 2010

1 Introduction

Explaining leisure travel is challenging the field of transport planning in both, quantitative and qualitative senses. First, leisure travel is the fastest growing segment of travel in terms of time spent on trips and in terms of miles travelled at least for Western Europe and North America. The share of travel related to leisure activities is e.g. in Switzerland in 2005 with 52% dominant and has been growing for recent years (see BFS and ARE 2007). Also in the US (43%) and Germany (35%) leisure travel has the highest share of all travel segments (for the US see U.S. Department of Transportation, 2004; for Germany see Deutsches Institut für Wirtschaftsforschung, 2003). In the UK it is not dominant but with a share of 26% still very important (see DfT, 2006). Also there are hints from diary studies that leisure trips are on average longer than trips for other purposes (Schlich *et al.*, 2004; Silvis *et al.*, 2006). Secondly, leisure is challenging because its structure is different from other forms of travel. It can in comparisons to other purposes of travel, e.g. commuting, be described as less driven by constant needs and more influenced by changing conditions like weather, traffic and social contacts. Therefore it is performed more spontaneously (see Schlich *et al.*, 2004).

As transport planning describes itself as a scientific field trying to understand and model out-of-home movements of populations (see Ortuzar and Willumsen, 2006) it is crucial to find explanations for all types of travel. To do so, various procedures are used to approach and explain leisure travel. One approach amongst these is using the methods of social network analysis (SNA) (Axhausen, 2005; 2008). In recent years there were a few initial studies of leisure travel investigating the influence of social contacts. Whilst they were starting points and produced new empirical insights the use of SNA-techniques is still in an early stage in transport planning. Nevertheless some of these former studies can be characterized as cornerstones and have directly influenced the work presented in this chapter:

Jonas Larsen, John Urry and Kay Axhausen examined the link between social networks, locations and travel in a study called 'Mobilities, Networks, Geographies' (Larsen *et al.*, 2006). By analysing data from 24 respondents they found the structure of leisure travel being different than presented in many other studies. Understanding individuals as embedded in social networks makes it necessary to accept cooperations between them as enabling and constraining possibilities for individual actions. From this perspective

meeting others for leisure purpose is not, as it is described in many older studies, marginal or superfluous but necessary and important for social life. In terms of geographical distance they found most respondents' networks implying not only local but also nationwide and even international ties. Because of it's few respondents this study is limited in terms inferential generalizations. There were two other survey projects aiming for more general results. Both collected information on personal networks and additional information on geographical distances and communication modes and - frequencies between around 320 respondents and their contacts.

The first survey was conducted in the East York area of Toronto city. It focused on the evolution of joint activities and their spatial distribution. Survey methodology and empirical results are published in the dissertation of Juan Carrasco (2006). The second study was conducted at the Institute for Transport Planning and Systems (IVT) of ETH Zurich and collected data on the link between individual mobility biographies and the geographical spread of emotionally important contacts in Zurich city. Details on methodology and empirical results can be found in Frei and Axhausen (2007) and Axhausen and Frei (2007).

Using respondents to collect information about their relationships and social contacts by using an egocentric methodology is the common characteristic of these three studies. Whilst the project of Larsen *et al.* is clearly limited, both other projects used random samples from predefined geographical areas. Both used the name generator technique, implying one or several questions to focus egos on that part of their social network being of interest for the study. In addition, questions of the name generator provide stimuli to support egos in recalling their contacts (Marin 2004; Marin and Hampton, 2007). The project of Carrasco *et al.* (2006) enriched this data by not exclusively asking egos for their relationship to each alter but additionally using a 25% sub-sample of their survey population to request data on the assumed relationships between alters in personal interviews (see Hogan *et al.*, 2007). Nevertheless, both studies are limited to predefined geographical areas in which they found their egos. In addition, the egocentric methodology uses egos' horizons in terms of social contacts to collect data, but it does not allow to look behind this horizons. From this perspective questions about the global structure of connected egocentric networks cannot be addressed. Getting an idea about egocentric leisure networks and calculate statistical figures of connected personal structures is the aim of a new survey project resented in this chapter. Also this project is the first study in transport planning that uses a name generator, which asks for leisure contacts first rather than second.

To evaluate findings of the former studies and to implement them in agent based travel demand simulations it is necessary to look behind isolated egos' horizons to see how they are connected to each other in large populations.

2. Survey methodology and –protocol: A brief introduction

Collecting data on connected egocentric networks can be done by taking a snowball sample which belongs to the family of ascending sampling strategies. It uses a set of initial respondents, called ego-seeds, as starting points. With the help of a name generator all seeds are asked to report certain contacts from their network. The methodology aims to continue the data collection with these alters by asking them to participate. Those who accept this request become egos themselves and report their alters. This procedure is repeated for a number of predetermined iterations (Vogt, 2005; Gabler, 1992; Goodman, 1961). The snowball methodology is usually used to collect data on populations with certain characteristics. Particularly in cases where members of these populations are discriminated or hide their characteristics or habits for other reasons, the snowball design has advantages. It is efficient in finding hard-to-reach target population as it potentially only needs few seeds to find further members (see Atkinson and Flint, 2001;). For this reason link tracing methodologies are used in the field of sociology, anthropology and medical studies (see Salentin, 1999; Schweizer, 1998; Jones, 2003; for an overview on link tracing sampling strategies see Heckathorn, 1997, 2002; Gabler, 1992). Snowball sampling can also be used to survey connected egocentric structures and investigate indirect links between people. The method is feasible to focus on the distribution of social contacts and estimate the structure of a given population (Scott, 2007; Frank, 1979).

The present survey strategy is influenced by the study of Silvis *et al.* (2006). Using a three-day interaction diary this project aimed to collect information on the link between social interactions and travel behaviour. Using three seeds to start the snowball chains, respondents were asked to hand out postcards to alters they were meeting face-to-face. The postcards asked for volunteer participation in the survey project being possible by contacting the researcher team. In total information from 24 respondents were collected. The research team reported various problems resulting from both, survey instrument and -strategy. Beside a high item non-response and a low response rate

they mentioned selection bias resulting from egos that felt uncomfortable to hand out postcards to certain alters.

In general, snowball sampling includes several sources for bias. Selection bias, as in the postcard example of the Silvis *et al.*, (2007) study, results from the possibility for egos to mention certain contacts arbitrarily in the name generator. This bias is included in all studies using a name generator but is even more urgent in snowball sampling. In cases where selections are systematic, not only certain alters are missing but these people also have no chance to become egos and report their contacts as well. Another source for bias results from overrepresenting people with many social contacts. People with many contacts have a higher probability to be reached by the snowball chain as a result of their various communication channels. This leads on the other side to an underrepresentation of people with lower than average social contacts and absolute non-observance of isolated people on the iteration levels. Also there is a risk of being captured in a homogeneous cluster in terms of similarity between egos and alters in certain characteristics, which is a well documented phenomenon from various network studies (see McPherson *et al.*, 2001). These different kinds of bias result in samples that do not fit in the criterion of randomness (for an overview on bias see Coleman, 1958; Erickson, 1979).

A stratified random sample from the population of Canton Zurich is used to recruit ego-seeds and start the snowball chain. Being stratified in terms of sex, age and living location, whether urban or rural, this sample enlarges the chance of recruiting representative ego-seeds in comparison to a simple random sample. The ego-seeds are asked to fill out the questionnaire with the help of an interviewer who can, on the one hand, help to avoid misunderstandings resulting from the survey instrument and, on the other hand, ask the questions of the name generator several times to survey the network of interest as completely as possible (see Wolf, 2004). Only if the ego-seeds do not want the help of an interviewer they are asked to answer the questions on their own. For respondents on the following iteration levels the help of an interviewer is not provided as they can be out of budgetary reach for the survey team.

The questionnaire includes four parts: First, questions on egos' characteristics; secondly the name generator, followed thirdly by questions on alters' characteristics. The fourth part is a sociogram, asking for alter-alter relations. The name generator has two questions and several concrete stimuli for each question. The first question asks for leisure contacts, the second for emotionally important contacts. The questionnaire provides

space for 40 names. Respondents are not asked to mention names twice, even not if they also fit in both categories. For this reason multi-relational analysis is not possible. To collect information on the relationship between the alters as well, the egos are asked to mention groups of alters making plans to spend free time together. To allow easy participation for all egos a standardized form to report groups and group members was developed. The questionnaire allows for 20 groups whereby each group can have as many members as names were mentioned in the name generator.

To keep bias from the snowball methodology low several arrangements are employed. To decrease the amount of selection bias it is necessary to make respondents feel comfortable. To survey the network of interest as complete as possible trust is needed. To increase the transparency of the project a multi contact strategy between respondents and the research team is used. All persons get an announcement letter and recruitment call. If they participate and mention alters they can decide if they want to stay anonymous in the further recruitment process or if they want their alters to know who mentioned them. In case the egos want to be identified they are asked to sign a prepared greeting card which is usually sent to the alters by the research team together with the announcement letter. All egos receive a paper questionnaire with an accompanying letter. Afterwards a thank-you-letter or, if needed, a reminder letter is sent. All letters provide information about the project and offer help in case of questions. Also an Internet page was built, providing detailed information about the projects' background and about each researcher involved. To increase response rate and avoid non-response behaviour all respondents receive a 20 CHF incentive with the questionnaire or the equivalent in the currency of their country. Only the ego-seeds do not know about the incentive but get it once the interview is completed. The whole survey strategy follows closely the tailored design method describing successful strategies to increase response rates and decrease non-response as well as item-non-response behaviour (Dillman, 2000). To handle the languages problem the instrument is designed in German and English. To react to the possibility of being captured in a cluster of respondents with similar characteristics the entire snowball sample is divided into three sub-samples. These sub-samples, each started by 20 ego-seeds, are in the field consecutively. In cases of an overrepresentation of certain characteristics in one sub-sample the following sub-sample can use specific ego-seeds to react (for more information see Kowald *et al.*, 2009). Unfortunately there is no way to solve the problem of overrepresenting people with many contacts in the survey process. Estimating the amount of bias resulting from the chain method and weighting the sample to decrease its influ-

ence is the aim of a team working at the Institute for Sea- and Land-Transport at TU Berlin (for details see Illenberger *et al.*, 2008; Illenberger *et al.*, 2009).

In summary, the present survey uses the snowball approach to focus on leisure contacts and look on network connections behind egos' horizons. Respondents and their contacts are not limited to certain characteristics. In terms of the snowball methodology, the survey is not limited to a specific location or to certain kinds of communication modes. The only selection, whether a person will be part of the sample or not, is related to the kind of relationship: Alters have to be mentioned as leisure or emotionally important contacts by egos. To our knowledge the snowball approach is used for the first time in a study of this size. Information on around 10.000 people is expected, whereby the criteria to continue the chain are open in terms of not being limited to institutional settings. In transport planning this is the first study using SNA-methods that explicitly focuses on leisure contacts.

3. Response rate and data quality

Currently the survey is in the field. Two of three sub-samples are complete. Starting with 40 ego-seeds information on 568 alters were collected. These alters are the iteration 1 population. 28.3% of them were recruited as well and became iteration 1 egos, mentioning another 1462 alters, the iteration 2 population. Here 252 alters, 29.5%, became egos and reported 2305 alter of which 201 were re-identifies because they were already part of the sample. On this level the snowball chain was stopped. The overall response rate of 26.9% is satisfying given the nature of the sampling methodology and the necessity to ask for sensitive information like names and postal addresses of friends and family members. In addition the questionnaire implies a high amount of response burden that also decreases the response rate (for the link between response burden and rate see Axhausen and Weis, 2010). Table 1 gives an overview on the overall response rate. It clearly reflects the influence of the multi contact strategy. The response rate is much higher when potential respondents received a previous letter and a recruitment call, labelled here as 'participation with recruitment'. Contrarily did persons, for which no phone number was available, just receive the questionnaire with an accompanying letter. In their cases, 'participation without recruitment', the response rate is much lower.

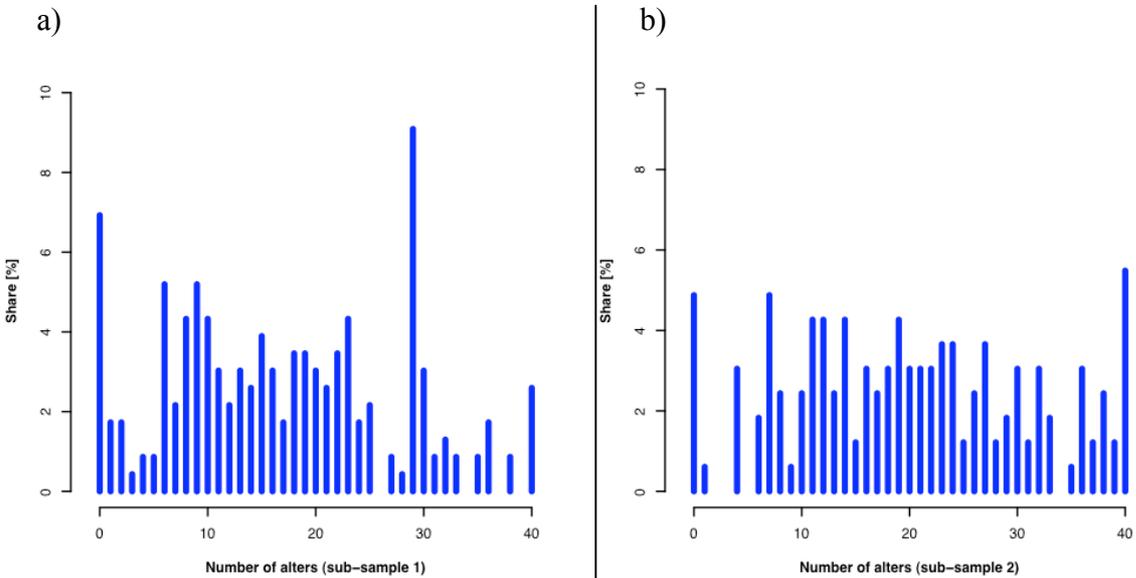
Table 1 Response rate

| | Ego-Seeds | | Iteration 1 | | Iteration 2 | | Whole sample | |
|-----------------------------------|-----------|-------|-------------|-------|-------------|-------|--------------|-------|
| | Abs. | [%] | Abs. | [%] | Abs. | [%] | Abs. | [%] |
| Sample size | 275 | | 568 | | 1462 | | 2305 | |
| Reidentified | - | | - | | 201 | | 201 | |
| Valid addresses | 247 | 100.0 | 377 | 100.0 | 855 | 100.0 | 1479 | 100.0 |
| Participation with recruitment | 40 | 16.2 | 94 | 24.9 | 229 | 26.8 | 363 | 24.5 |
| Participation without recruitment | - | | 13 | 3.4 | 23 | 2.7 | 36 | 2.4 |

A difference between the sub-samples can be observed in the number of names provided by the egos. In the first sub-sample, egos mentioned 16.9 names on average whilst those from the second sub-sample mentioned 20.1 names. The difference is due to a redesign of the name generator. In the first sample the questions were separated from each other. Respondents first were asked to write down their leisure contacts for which space for 29 names was provided. Afterwards they were asked to report emotionally important contacts; 11 names were allowed at maximum. This resulted in an artefact: Some respondents reported 29 leisure contacts and no emotionally important people. Figure 1a clearly shows the peek at 29 names. It can be assumed that this design caused fatigue- and satisficing effects. As name generators often imply several questions, fatigue effects result in lower numbers of names in case of questions that follow on former name generator issues. Satisficing effects result from respondents feeling they have already reported 'enough' names for the study (for an overview on these and further effects from name generators see Pustejovsky and Spillane, 2009). In its new design the name generator avoids these effects by asking both questions concertedly and providing space for 40 names as it can be seen in figure 1b¹. Considering both sub-samples each ego reported 18.2 names on average.

¹ The name generator form can be found in the appendix.

Figure 1 Number of alters reported in sub-sample 1 (a) and sub-sample 2 (b)



To check whether data are representative for the target population they have to be compared to characteristics of this population. These comparisons are especially important in case of snowball samples because of the mentioned biases. They can easily be done for the ego-seeds as they are all living in Canton Zurich. The comparisons are more challenging for the alters because they can be located elsewhere. As nearly 88% of all alters are Swiss their characteristics are compared to the Swiss Microcensus, representative for the entire Swiss population as well as the cantons (ARE/BfS, 2007). This approximation allows to judge the sample quality in terms of both, ego-seeds and alters which is done in table 2.

The ego-seeds do not represent the Canton Zurich population properly. Females are over-, singles underrepresented. This is due to the small sample size of only 40 respondents. Focussing on respondents' age the distribution fits well. The alters characteristics fit much better to the Swiss population. However, males are underrepresented whilst the sample includes a higher share of married and a lower share of singles. Focussing on citizenship Swiss and Germans are overrepresented whilst French and Italians are underrepresented. The sample is representative in terms of age, which is divided in four categories whereby the mean age of each category is compared.

Table 2 The characteristics of egos and alters compared to the Swiss population

| Character | Snowball: Ego-seeds (n = 40) | Microcensus Canton Zurich (n = 5702) | Snowball: All alters (n = 7120) | Microcensus Switzerland (n = 33390) |
|---------------------------|------------------------------------|--|---------------------------------------|---|
| Sex [%] | | | | |
| Male | 30.77 | 48.80 | 42.78 | 48.70 |
| Female | 69.23 | 51.20 | 57.32 | 51.30 |
| Civil Status [%] | | | | |
| Single | 20.00 | 32.70 | 20.32 | 29.90 |
| Married | 55.00 | 50.30 | 65.02 | 54.50 |
| Divorced | 10.00 | 9.30 | 8.00 | 7.60 |
| Widowed | 15.00 | 6.40 | 5.56 | 6.60 |
| Married living seperately | 0.00 | 1.40 | 1.10 | 1.40 |
| 1. Citizenship [%] | | | | |
| - Swiss | 90.00 | 78.10 | 88.70 | 80.00 |
| - German | 0.00 | 3.50 | 4.46 | 2.30 |
| - French | 0.00 | 0.20 | 0.27 | 1.30 |
| - Italian | 0.03 | 4.60 | 1.54 | 4.60 |
| Age [Ø years/category] | | | | |
| 0 – 20 | - | 13.32 | 13.54 | 13.31 |
| 21 – 40 | 31.57 | 31.51 | 33.36 | 31.41 |
| 41 – 60 | 50.05 | 49.84 | 49.48 | 49.96 |
| 61 – 80 | 72.17 | 69.28 | 68.50 | 69.37 |
| 81 + | 81.00 | 84.46 | 84.05 | 84.61 |

Source: Microcensus data taken from ARE/BfS, (2007).

Collecting data on connected egocentric networks allows to check for another attribute of data quality: the validity of information egos provide about their alters. As some alters become egos themselves it is possible to compare information an ego provides about them to information they report about themselves. The test can be done for 228 ego-alter relations. Asking for gender egos' answers match in 98% of all cases. Also in terms of age, categorized as alters' self-declared age plus/minus 5 years the fit is with

97% very good. In 95% of all cases egos reported the correct civil status and in 93% the correct first citizenship of their alters. The lowest share of matching characteristics can be found in terms of education with 64%. These findings are consistent with those reported in literature. Usually egos have precise information about socio-demographics of their contacts. Reliability decreases when asking for softer issues like political sentiments (see Marsden, 1990; Wolf, 2004).

These comparisons show that the survey instrument as collecting valid data on both, egos and alters. Problems resulting from the design of the name generator were solved. The snowball methodology is feasible and working well. It results in representative data with a high level of reliability. The overall response rate of 26.9% is good given the nature and response burden of the survey.

4. Network topologies

The reported personal networks of leisure and emotionally important contacts have 18.2 alters on average. During the live course this number is fluctuating. Respondents under 40 years report a network size of 16.4 contacts on average. People between 40 and 60 years have the largest networks with 19.9 alters. This number is decreasing to 17.5 for respondents between 60 and 80 and to 11.0 for those older than 80 years. Whilst this first result shows people being connected to each other it does not say too much about the topology of the networks: Who is in contact with whom, where do these people live and how do they stay in contact.

In terms of network topology it first has to be mentioned that 304 of 395 egos complemented the information from the name generator by reporting alter-alter relationships in the sociogram. 29 of the remaining egos, 7.3%, reported one or zero social contacts and were not able to report those relations. The other 62 egos with more than one alter did not report any alter-alter relations. Most of these networks are of small size: 29 include 10 or less alters, another 11 networks contain 11 to 20 alters. Whilst asking for alter-alter relations is the last part of the questionnaire and fatigue effects cannot be excluded completely it is assumed that these 62 egos, 15.7% of all egos, have sparse networks. In all, these 91 egos without filled out sociograms are excluded from the analysis of the social networks' topologies to avoid bias in their direction.

Investigations on network topologies use concepts from graph theory. Table 3 provides an overview on different descriptive indices that are briefly introduced here.² As the egos and their relation to each alter are always present in personal networks, due to the mode of data collection, they are removed from the networks. This procedure is done commonly as otherwise all figures would be biased by them (Scott, 2007; also see Wasserman and Faust, 2007). In all the present 304 leisure networks contain 20.5 alters being organized in 4.0 cliques with each clique having 4.4 members on average. By definition a clique is a network part in which each member is adjacent and directly connected to each other member. As it is not possible to see which member of a clique opens the contact or sends information to other members, the network structures are considered as undirected and, of course, alters can be members in several cliques. If this is the case, one or several alters connects two cliques into a component. A component is a sub-graph in which each actor can reach each other actor directly or through a chain of indirect links. There is no connection to actors not belonging to that component. Overlaps between cliques happen frequently as the average number of cliques is higher than the average number of components per graph, which is 2.5. Each network contains around 43.3 relations between the alters on average, with a high standard deviation and a maximum of 399 relations. In addition an average network implies 6.4 isolated people that do not spend their leisure time with any other alter from the graph. In such cases the isolates are not removed from the graph each of them is a component existing of only one member. The density of a graph is an index, which is calculated by dividing all possible connections between the actors of the graph through the connections actually present. The average density of 0.2 means that 20% of all possible connections between the alters are realized. Of course the density is higher, 0.4, when isolates are excluded. The degree is calculated for each actor in a network graph. It informs about the number of actors being adjacent with the person of focus. Whilst degree is a good measure to find the most central actor in a graph, the index of degree centralization is more appropriate to compare egocentric network structures. In its standardized form, it describes the extent to which a network is organized around its most central actor. An index of 1.0 suggests all people are exclusively connected to one central actor, like in a star graph, whilst an index of 0.0 suggest all actors being equally connected to each other, like in a circle graph. As a result the degree centralization can be interpreted as a proxy for the activity level between alters. In a star graph the central actor is the only active element. The lower the index, the more equal the alters are in showing efforts to perform joint activities (see Carrasco and Habib, 2009). The present

² All indices are calculated with the sna-package in R (R development core team, 2009; Butts, 2008)

index of 0.2 on average shows most networks as including nearly equally connected actors and decentralized structures. Only few exceptions are strongly centralized with a maximal index of 0.8. Also the concept of betweenness centralization belongs to the global graph indices. Like nodal degree betweenness it can be calculated for each actor. It informs about the number of times the actor of focus lies on the shortest path between two other people, their geodesic. Whilst actors' degrees are calculated using exclusively adjacent people, betweenness uses nonadjacent pairs of actors being connected indirectly. Betweenness centralization is complementing degree centralization as people lying on the geodesic between others might have control over or might be able to influence their interaction. The more often an actor lies on geodesics the more potentially powerful is the position. In its standardized form the global index of betweenness centralization informs about the extent to which all nonadjacent actors of a graph have to pass one 'gate keeping' alter. If one alter connects all others the index equals 1.0. If the nodal betweenness scores are distributed equally the index equals 0.0 (for details on the calculation of degree- and betweenness centralization see Freeman, 1979). An average leisure network includes no powerful gatekeepers. Even though the maximum value is 0.40, the distribution does not suggest that connecting actors having much influence on the relations of others are commonly present. Table 3 also shows the heterogeneity of personal leisure networks. All characteristics and particularly the indices for density, degree- and betweenness centralization have a wide range and high standard deviations.

Table 3 Egocentric network attributes, only egos with filled out sociogram (n = 304)

| | Mean | Std.- Deviation | Minimum | Maximum | 25- Percentile | Median | 75- Percentile |
|-------------------------------|-------|--------------------|---------|---------|-------------------|--------|-------------------|
| Number of alters | 20.48 | 9.80 | 4.00 | 40.00 | 12.00 | 19.00 | 29.00 |
| Number of relations | 43.32 | 59.71 | 1.00 | 399.00 | 9.00 | 22.50 | 50.00 |
| Cliques | 3.98 | 2.63 | 1.00 | 20.00 | 2.00 | 3.00 | 5.00 |
| Number of clique members | 4.42 | 2.53 | 2.00 | 28.00 | 2.89 | 3.78 | 5.33 |
| Isolates | 6.35 | 5.66 | 0.00 | 30.00 | 2.00 | 5.00 | 9.00 |
| Components (with isolates) | 8.88 | 5.95 | 1.00 | 31.00 | 5.00 | 8.00 | 12.00 |
| Components (without isolates) | 2.53 | 1.46 | 1.00 | 9.00 | 1.00 | 2.00 | 3.00 |
| Density (with isolates) | 0.20 | 0.18 | 0.01 | 1.00 | 0.08 | 0.15 | 0.26 |
| Density (without isolates) | 0.44 | 0.27 | 0.09 | 1.00 | 0.24 | 0.38 | 0.56 |
| Degree centralization | 0.22 | 0.14 | 0.00 | 0.80 | 0.13 | 0.17 | 0.28 |
| Betweenness centralization | 0.03 | 0.06 | 0.00 | 0.40 | 0.00 | 0.00 | 0.03 |

Comparisons between personal networks are difficult. The name generators have to be the same or at least similar to each other to result in comparable types of networks. Whilst the study of Carrasco *et al.*, (2008) asked for alters the respondents feel ‘close’ or ‘somewhat close’ to and comparisons are therefore limited, the instrument of Frei and Axhausen (2007) is similar to that used here, only the questions’ positions were exchanged. First they asked for respondents’ ‘core contacts’, emotionally important alters, which egos use to discuss important problems or which they can ask for help. Such alters are also referred to as ‘close’ or ‘strong’ contacts. In addition they asked for leisure contacts, which were expected to be more ‘weak’. On average, networks reported in this study were with 12.4 alters smaller than those in the present study. As the survey instrument did not ask for alter-alter relations, further information on the network structure itself are not available. However, the alters’ characteristics can be compared. The share of relatives in the Zurich City-study was 31.0% and is higher than in the present study with 27.3%. In addition the first study reports a share of 48.0% weak ties compared to 67.3% in the present project. It has to be mentioned that in the later survey emotional closeness was not, as in the study of Carrasco *et al.* (2008) and Frei and Axhausen (2007), directly measured in the name generator but with the help of proxy-

questions asking if alters provide help in difficult situations or are used to discuss important problems. Considering the Zurich City-study highlighted emotionally important contacts in first place, these findings are meaningful. An average network resulting from this name generator formulation is smaller than a network in which leisure contacts are highlighted, by including more relatives and substantially more strong ties. In literature on personal networks the relation between network size and the share of strong contacts is well documented. The higher the degree of a person, the less emotional on average the relations are (Roberts *et al.*, 2009). In terms of relationships' duration, both projects are similar to each other. Axhausen and Frei (2007) found 20.6 years on average whilst it is 22.7 years with a standard deviation of 16.0 in the present study.

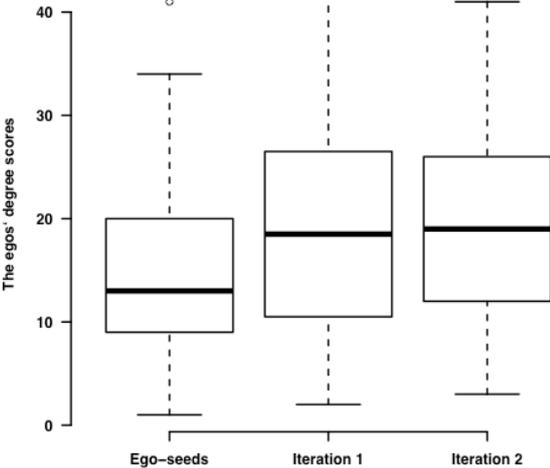
Before investigating relations between egos and alters more deeply, the structure of the snowball graph, in which all agents and their relations are included, is introduced. The number of included agents and relations differs from the sum of the personal networks figures because it has to be approached differently. Whilst on personal networks' level all alters are considered the snowball graph has to be checked for re-identified people, alters that have been mentioned by two egos, which is very time consuming work. Once the survey is completed these differences will fade away. Re-identifications are a source for transitivity, a term describing the probability of two people to establish a relationship, in cases where both are connected to a third person³. A high rate of transitivity results from agents being embedded in a single group and knowing each other. A low rate indicates people being connected less densely. It results from few local groups and a number of gatekeepers that bridge between them. The concept of transitivity is related to the small world-phenomenon, originally describing the effect of two people meeting by chance and detecting that they have an acquaintance in common (Milgram, 1967). The second source for triadic closure in the present survey is the sociogram, the alter-alter relations. Whilst the snowball graph includes more clique structures than re-identifications the later are of high importance for the project because of their quality. Some re-identification may bridge between components of which each was started by an ego-seed. These bridges connect large number of people indirectly and therefore they are an important element of cohesion in society. To our knowledge the present project is the first investigating global network figures for leisure and emotionally important contacts. The study started with 40 components. 2 ego-seeds were isolates with no contacts at all. They are excluded from the analysis. The graph cur-

³ The concept of transitivity follows the definition of Barrat and Weigt (2000). For more details also see Newman (2003).

rently includes two bridges between two different components. In all, this makes 2774 people being connected by 16.454 relations and structured in 35 components. Person re-identifications occurred in 216 cases. As the egos are included in this graph as well as all cliques from the sociogram, the level of triadic closure of 61% is high. In comparison, the average transitivity of the population of the United States is estimated to be around 52% (Christakis and Fowler, 2009). It is expected that component overlaps will occur in significant numbers in the future, as there are still many names and addresses that have to be checked for re-identifications. It is also planned to start one more sub-sample.

The degree scores of the egos are used to detect the bias of the snowball methodology. As mentioned above, people with many social contacts are expected to have a better chance to become part of the sample. Clearly this is verified in the present study. Figure 2 shows the actors' degree-distribution differentiated by iteration levels. Focussing on the median shows them as increasing with each iteration step. The scores seem to follow a logarithmic curve.

Figure 2 Number of alters reported in sub-sample 1 (a) and sub-sample 2 (b)



In personal network studies the phenomenon of homophily is well documented. It is also referred to as homogeneity bias or network autocorrelation and describes a ten-

dency of similarity between people that are, in terms of network structure, closely related to each other, that are e.g. adjacent. The phenomenon is subdivided in status-, resulting from similar socio-demographic characteristics, and value-homophily, describing similar kinds of behaviours. Homophily is influencing the social world of network agents: Information spread by people with substantially different socio-demographic attributes or ways of behaviours have to pass more intermediate alters on average, than information spread by people with similar attributes to reach a certain receiver (for a general overview see McPherson *et al.*, 2001). Two sources of homophily in personal networks can be distinguished. First, people can choose others with similar attributes as contacts. These choices are limited by structural opportunities, which are the second source of homophily. An often mentioned issue of these opportunities is geographical distance as people are not distributed randomly in space but are clustered in geographical areas, e.g. in urban districts, depending on socio-demographics like income or number of children. As network distances are related to homophily and similar people are geographically clustered, similarity is also an important factor influencing the probability of agents to establish and maintain new contacts. In a study on friendship networks of around 30.000 students in the US Kossinets and Watts (2009) describe the probability of tie formation as much higher than average when people are similar under equal structural constraints. People are similar because they share either attributes or friends. A discussion on whether similar behaviour results from selecting similar others or is a consequence of influence from others in the network can be found in Steglich and Snijders (2010).

High degrees of status-homophily are also observed in the present study. In all characteristics that are compared, the attribute similarity is significant when employing a chi-square test of independence. The highest value of homophily can be observed in age (70%), followed by gender (69%). It is lower in terms of civil status (62%) and education (60%), attributes that usually change during the live course. Homophily becomes even stronger when excluding kin-contacts as these, except for spouses, are not chosen. In course of life the similarity between egos and alters seems to be robust. An exception is age, where similarity is decreasing constantly. In addition all egos of age-class 80 and older report lower degrees of similarity for all characteristics than younger people except for gender. Whilst gender similarity is nearly equally high in all age-classes it

seems quite intuitive that other forms of homophily decreases with increasing age as contacts die or become widowed.

Table 4 Share of similarity between egos and alters [%]

| Type of contact | Age | Gender | Civil status | Education |
|--|------|--------|--------------|-----------|
| All contacts (n = 7120) | 0.70 | 0.69 | 0.62 | 0.60 |
| Non-kins (n = 4907) | 0.78 | 0.74 | 0.64 | 0.63 |
| Non-kins, ego <= 39 years (n = 722) | 0.85 | 0.71 | 0.68 | 0.66 |
| Non-kins, ego = 40-59 years (n = 2546) | 0.82 | 0.74 | 0.65 | 0.63 |
| Non-kins, ego = 60-79 years (n = 1530) | 0.71 | 0.77 | 0.62 | 0.63 |
| Non-kins, ego >= 80 years (n = 83) | 0.45 | 0.74 | 0.39 | 0.54 |

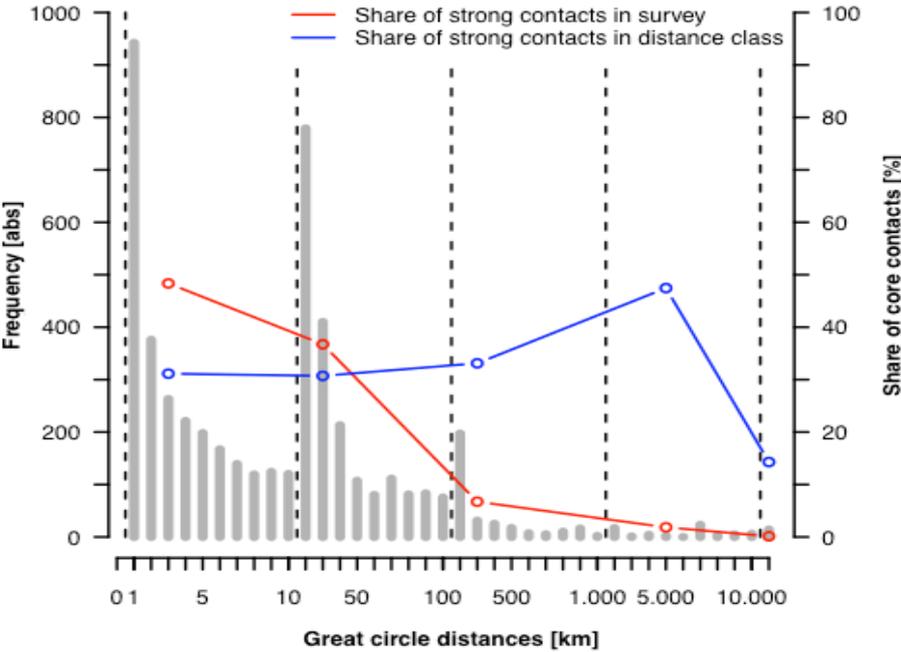
In summary, homophily seems to be an important factor when focussing on leisure and emotionally important networks, too. The findings match to those reported in network literature (see e.g. McPherson *et al.*, 2001). As homophily is related to opportunities for getting known to each other and the theory of homophily suggested people clustered locally it strongly influences the probability of establishing local relationships. Beside this factor there is the counterpart of few people bridging between local communities as shown in the results of the snowball network.

5. Network geographies

In total the egos reported 5007 postal addresses: addresses of all egos' and of 70.3% of alters' home locations are known. Using concentric circles to investigate the share of egos' contacts by distance between egos' and alters' homes shows 18.8% of all alters living within a one km radius. A ten km radius includes more than half of all contacts, 53.3%. More than two thirds of all contacts live in 25 km distance and 92.0% in a 100 km radius. Following a definition from in the Zurich-City study (Frei and Axhausen, 2007) considering the geographical position and size of Switzerland around 74.8% are local contacts, living in 25 or less km distance. Another 17.3% are of regional and national character (26 – 100 km). 6.1% of all alters are intereuropean contacts (101 – 1000 km) and 1.9% can be labeled as intercontinental relations with more than 1001 km distance

between egos and alters. Figure 3 provides a graphical overview on the share of alters living in certain distance classes. The distances are given in a log-scale. The shares of strong relations in terms of emotional closeness, persons that egos use to discuss important problems or ask for help, in relation to all strong contacts from the entire survey are plotted in red and in relation to all strong contacts in the certain distance classes in green. Both lines are plotted at the average distances of each distance class. The distribution of alters in each distance class follows approximately an exponential decay function. The same shape can be seen when plotting the entire graph without dividing it in certain distance classes. Because people are not equally distributed in space, e.g. mountain regions are usually sparsely populated and oceans are not populated at all, the distribution is not simply parametric. In figure 4a small peak can be seen at 7.000 km distance, which represents the distance between Switzerland and the east-coast of the United States. In comparison to Frei and Axhausen's investigation (2007) the present study collected more local contacts and fewer more regional and national relations. It surveyed less intereuropean and international contacts. Also the distribution of emotionally important contacts looks different. In the present study the share decreases rapidly when compared to all strong contacts from the study. This was expected as the overall number of alters is decreasing the further the distance between egos and alters. More surprising is the finding of an increasing share of strong contacts in comparison to all contacts in the distance classes. Starting with a share of 30% at 3 km distance it begins to increase slowly until a distance of 300 km. Here it begins to increase fast until it reaches a peak of nearly 50% at 5000 km. This seems to be a threshold as the share of strong contacts decreases again at greater distances. In the Zurich city study the share of emotionally important contacts per distance class was highest at 1 km. From here it decreased slowly with the exception of a local maximum at 1000 km. Summarized, there are differences in the spatial distribution of contacts between networks resulting from name generators that highlight emotionally important- or leisure contacts in first place.

Figure 4 Share of alters and strong contacts in distance classes (the x axis is given in a log-scale)



The present analysis confirms the general finding presented in many transport planning studies that use the methods of social network analysis. Although they partly used other name generators those studies found respondents mixing local, regional and international ties in their personal networks (see Larsen *et al.*, 2006; Carrasco, 2006; Frei and Axhausen, 2007). This is also the case when focussing on leisure networks. Whilst most alters live in ego’s nearest environment there are also contacts at great distances. This was expected as spatial opportunities for meetings between people are necessary to establish relationships. In connection with high degrees of homophily the spatial distribution of contacts confirms the suggested influence of space and similarity for the case of leisure contacts. The trend of an increasing share of strong contacts up to a threshold at the quite large distance of 5000 km shows that if people maintain long distance relations these contacts are often of certain importance for them.

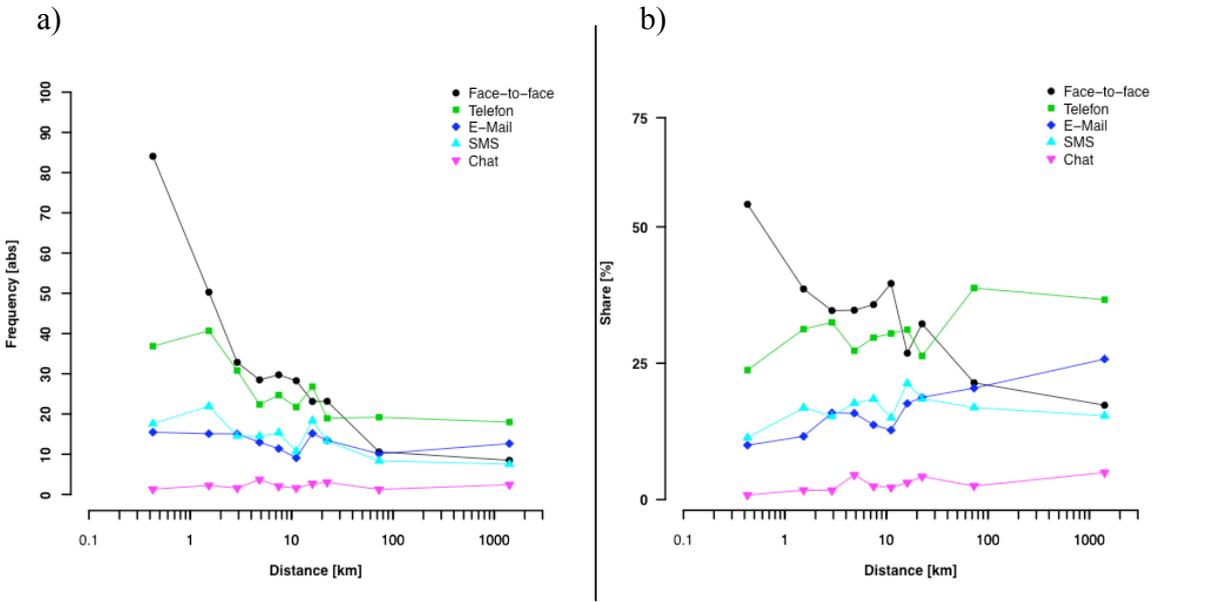
In literature on the link between personal networks and spatial distances two trends are hypnotised. First, distances limit the opportunities for getting known to each other. In addition it takes more efforts to maintain long distance relationships as man-made and natural barriers inhibit information flows and limit personal action spaces (Stutz, 1973). On the other side travel patterns have changed over the last decades. On average leisure time budgets have increased (Schlich *et al.*, 2004) and travelling got substantially cheaper (Larsen *et al.*, 2006). In addition there are various kinds of electronic information and communication technologies (ICT), which allow long distance contacts without the necessity of physical meetings. Whilst there were discussions about the influence of ICT on travel patterns, particularly if they can replace face-to-face meetings or are performed complementary to them there are hints for both. ICT can help to maintain contacts at greater distances but face-to-face meetings are still of high importance even at those distances (for this discussion see e.g. Mokhtarian and Salomon, 2002; Axhausen and Frei, 2007; Mok and Wellman, 2007; Larson *et al.*, 2008). They are perceived as more intense as they force interacting agents to share the same environment whilst ICT does not. Also face-to-face meetings often take longer than other forms of communication (Tilahun and Levinson, 2009).

The importance of face-to-face meetings is also highlighted by the present dataset. Figure 5 provides an overview on different modes used by respondents to maintain contact to their alters. Each mode is plotted at the mean of distance deciles whereby the first class was fixed to people living within a one km radius around egos' home locations. The x-axis is given in logarithmic scale. Figure 4a shows the data in absolute communication frequencies. It clearly shows decreasing communication frequencies in dependence from growing spatial distances between people. Whilst this trend is strong in the beginning, mainly because face-to-face and phone contacts become less often, it seems to stabilize at distances over 70 km. In general the contact modes follow different trends. Clearly physical meeting decrease strongest. Also the phone loses much influence but does by far not decrease as much as face-to-face. E-mail and SMS seem to be influenced by distance similarly. Both lose little with growing distances. Chat, defined as all kinds of Internet-communication with exception of E-Mail, is not affected at all. In a pre-test the questionnaire also included the category of postal contacts. This category was chosen by only very few respondents and the reported frequencies were very low too. Because of that it was removed after the test.

Figure 4b shows the shares of modes for the deciles-classes. At a distance of around 20 km phone contacts get more important than face-to-face meetings. Also E-mails start

having higher shares than contacts per SMS. At around 100km E-mail become more important than physical meetings too. At great distances over 1000km face-to-face has only little more influence than SMS. Most of the share of physical meetings is absorbed by phone- and E-mail contacts. Again chat-contacts are the steadiest category. Overall this mode is of little influence but it seems to increase slowly the larger the distance between two people is.

Figure 5 Absolute frequencies (a) and shares (b) of communication modes (the x axes is given in a log-scale)

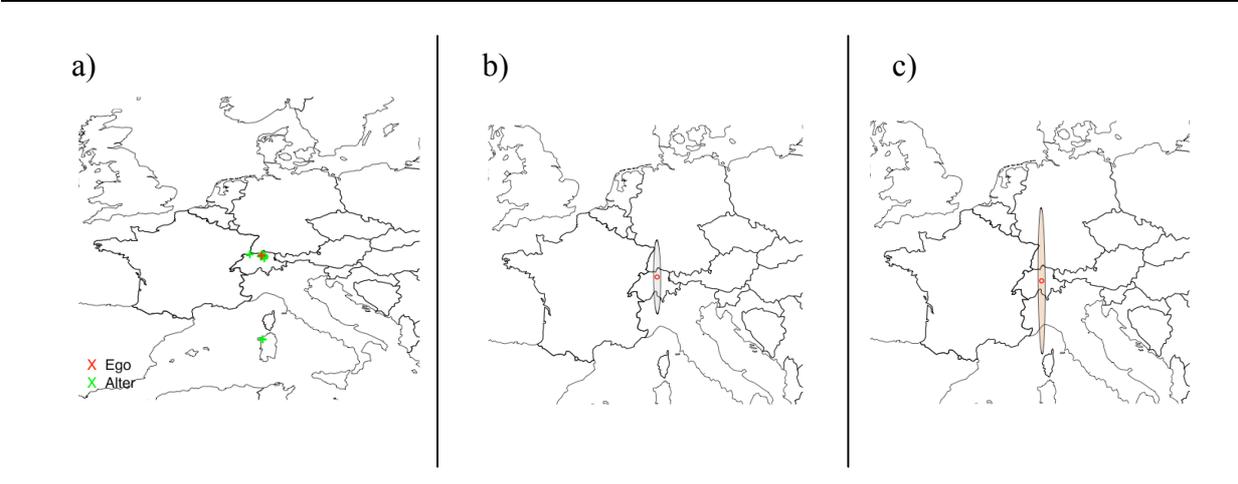


When focussing on personal networks to explain travel patterns it is crucial to consider distances as well as heterogeneities in alters' home locations. Whilst great circle distances and contact modes provide information on spatial distances between egos and alters, they do not allow investigations on spatial clustering or -diversities of social contacts. To see if leisure networks are clustered, in terms of many alters at a certain distances living in the same region, or spatially diverse, confidence ellipses are calculated. The approach is often used because it can be applied easily and causes only little computation costs. But there are also problems. First, the approach is parametric and assumes a bivariate normal distribution of contact distances, conditions which are hardly fulfilled in the present data as shown above. Secondly, it uses ego's home location as an initial point and the alters' home locations weighted by the annual contact frequencies to adjust the smallest elliptic area in which, with a predetermined likelihood, here 95%,

all of ego’s activities take place. Whilst the calculated areas can be interpreted as an ego’s interaction space, the symmetry of the ellipses often leads to covering empty areas, e.g. mountains and oceans, which results in overestimations of their sizes (for an overview on confidence ellipses see Schönfelder, 2006; Vikrant, 2005). Figure 6 shows an example for this from the present survey data. An ego from Canton Zurich reports several alters living in the nearer spatial environment. In addition one person lives in north-west Switzerland and one person on the west-coast of Sardinia (Figure 5a). Plotting the 95% confidence ellipse for physical meetings, shows it as reaching from south Germany to north Italy with only a small expansion in west-east direction (Figure 5b). This space can be defined as an egos activity space as it is based on face-to-face meetings. The ellipse for ego’s social space is larger. It is not only weighted by the frequencies of physical meetings but by the summed frequencies of all contact modes. In north-south direction it is stretched from mid Germany to the coast of Corsica, in west-east direction it is, in comparison to the activity space, less influenced by the alter living in north-west Switzerland.

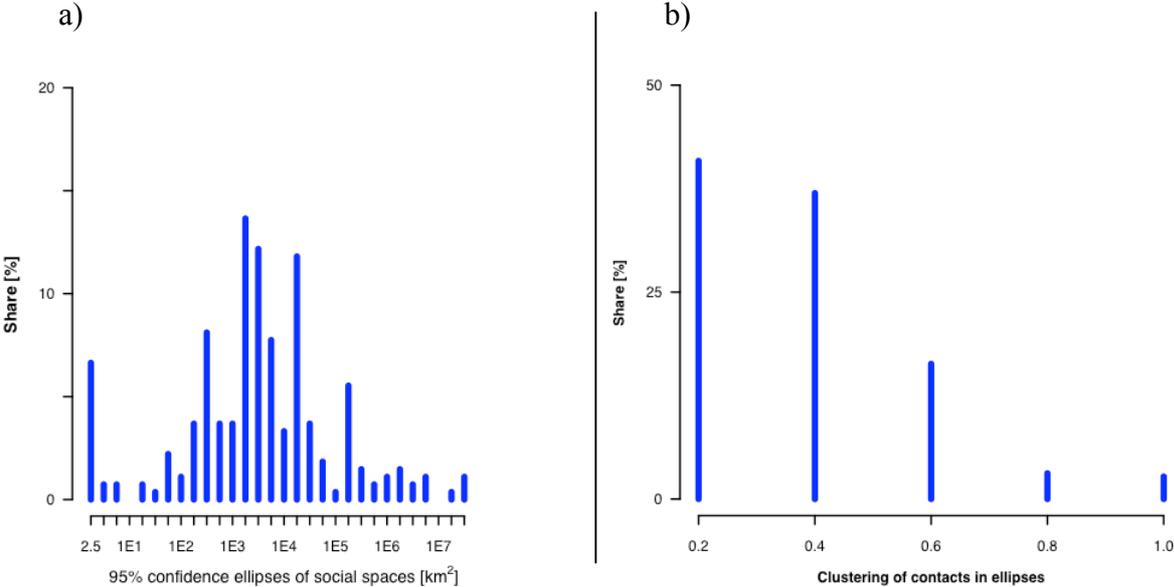
Confidence ellipses can be used to estimate the size of certain kinds of activity spaces, depending on the contact mode under investigation. Secondly, the main axis’ ratios can be used to show the spatial diversity of personal networks. The first main axis is defined as the most distant point from an ego’s home location on the contour of the elliptic polygon. The second main axis is found by using the right angle to the first main axis, again with its origin at ego’s home location. The ratio between the size of the two axis equals one if the ellipse has the form of a circle which means that the alters are equally distributed in space. The lower the ratio becomes, the more biased the circle is in direction of an ellipse and the more clustered the alters are in a certain region.

Figure 6 Egos’ contacts (a) and modelling them as activities- (b) and social spaces (c)



Sizes of the social spaces ellipses are plotted in Figure 7a. Some are of very small size, under 2.5 square-km. Most ellipses include more than 1.000 and less than 50.000 square-km. There are several peaks which is again influenced by alters living at certain distances like the US. Although there are some differences in size, the present ellipses are on average smaller, their distribution looks similar to those reported by Frei and Axhausen (2007). In terms of diversities in alters' home locations a high degree of clustering can be observed. Only few ellipses are like a circle, whilst most are clearly biased to ellipses with a long- and a short main axis.

Figure 7 The size of egos' social spaces (a) and the spatial clustering of alters (b)



Ellipses of social spaces that look like circles are with 1.147 square meters on average of small size. Their maximum size is 2.477 square meters. In comparison, extremely biased ellipses cover on average 902.000 square meters. It seems that networks with heterogeneous spatially distributed alters are those in which contacts live relatively near the ego. With growing distances the ellipses become biased, usually by those alters living in larger distances. The further away a certain contact is located the less likely is the probability of the egos to have other contacts living in similar distances but other directions.

6. Fazit

The field of transport planning tries to approach and explain leisure travel by using the methods of SNA, amongst others. Whilst there were some former studies, the present survey project aims to overcome their limitations in terms of small sample sizes, predefined geographical areas as well as limitations resulting from the survey methodology.

Collecting information on specific parts of personal networks can be done by employing the name generator technique which uses respondents' horizons in terms of social contacts. Whilst in transport planning name generators focussed on respondents' emotionally important contacts exclusively or at least in first place, the present project highlights leisure contacts and asks in addition for emotionally important contacts. Unfortunately it is not possible to distinguish between these kinds of contacts. Therefore a multi-relational analysis is impossible. To look behind egos' horizons and collect information on connected rather than isolated networks the egocentric approach is combined with the snowball methodology. The study aims to generate new empirical insights in personal leisure networks. In addition, it provides information on the global structure of connected personal networks, which is of importance to implement findings in agent based travel demand simulations. To our knowledge the snowball approach is used for the first time in a study of the planned size whereby the criterion to continue the sample chain is only limited to those kinds of contacts the name generator is asking for.

The egocentric network approach and snowball sampling are both known to include various sources of bias. Several arrangements are employed to reduce these biases and increase response rates. A bias in direction of egos with more contacts than average is detected and will be reduced by weighting data once the survey is completed. The overall response rate of 26.9% is satisfying considering the survey asking for very confidential information and implying a high amount of response burden. Comparing survey data against representative data of the Swiss population showed the fit as being good with only small problems of over- or underrepresentation. In addition, a test resulting from the snowball methodology showed the answers of the egos' being highly valid.

In all 395 egos in the database reported 18.2 alters on average. This number fluctuates in dependence from respondents' ages. It increases up to an age of 60 and decreases again for people in retiring age. In comparing egos' to alters' characteristics a high de-

gree of status-homophily was found which is important in terms of establishing relationships with others and information flows. It is highest for age, followed by gender, civil status and education. 7.3% of all networks are of very small size, including zero or one alter. Another 15.7% are very sparse including several alters that do not know each other but are only connected to egos. As the response burden of the survey instrument is high, it cannot be excluded that some of these egos simply refused to report alter-alter relations. However, 304 egos reported those relations. In all these personal networks of leisure and emotionally important persons have heterogeneous structures. On average they contain 4 cliques with each clique having around 4 members. Usually there are actors that participate in several cliques and connect them to larger components. In addition to alters knowing each other from leisure activities there are some isolates, alters that are exclusively connected to ego. On average 20% of all possible connections between the alters are realized. Focussing on those alters that are connected to each other shows them as having similar degrees of connectedness. Few networks have a more centralized structures, e.g. resulting from single alters that connect many cliques to one component but usually there are no central gatekeepers in powerful positions. In comparison to results from a former study using a similar name generator but asks for emotionally important contacts first rather than second shows an average 'leisure' network being bigger than an 'emotionally important contacts'-network by including less relatives and substantially less strong ties. These results are surprising and show name generators being very sensitive survey instruments.

Networks of leisure and emotionally important contacts are usually clustered in egos nearer spatial environment. Even though respondents mix local- with nationwide-, intereuropean- and even intercontinental ties, 75% of all contacts can be found in a 25km radius around egos' home locations. Considering the share of strong contacts at certain distance classes it was shown that contacts living in great distances are often important for egos in terms of discussing important problems and providing help in difficult emotional situations. The by far dominant communication mode to maintain relationships in the 25km-radius is face-to-face meetings. At thresholds of around 20km phone contacts become more important than physical meetings as well as E-mails at 100km distance. By calculating confidence ellipses it was shown that the contacts in a network are not equally distributed in space. Only in cases where all contacts live very near to ego there are some ellipses that look like circles. Usually they are highly biased as egos are spatially clustered in certain areas.

In terms of the global network a high rate of transitivity, 61%, is observed, describing the probability that a friend of a friend becomes also my friend. It results from exclusively focussing on leisure and emotionally important contacts that are likely to know each other as the average duration of contacts is with nearly 23 years quite long. In terms of community structure the snowball network is highly clustered in components of which each results from an ego-seed. So far only two bridges between those components were found. It is expected to find them in significant numbers as it is planned to use one more sub-sample with 20 ego-seeds. In addition, sub-sample two will be continued from iteration 2 to iteration 5, an undertaking that is in the field at the moment.

Whilst modelling network sizes and related questions is part of future work these findings suggest that there are three major sources influencing the probability whether a relationship is established between persons or not: First, persons have to be similar in terms of socio-demographics, secondly, they have to live in the same geographical area and thirdly they are more likely to become leisure contacts if they share a third person as a contact.

7 Literature

- ARE/BFS (2007) Mobilität in der Schweiz, Ergebnisse des Mikrozensus 2005 zum Verkehrsverhalten. *Federal Office for Spatial Development and Swiss Federal Statistical Office*, Bern and Neuenburg.
- Atkinson, R. and J. Flint (2001) Accessing hidden and hard-to-reach populations: Snowball research strategies, *Social Research Update*, **33**, University of Surrey, Guildford.
- Axhausen, K. W. (2005) Social networks and travel: Some hypotheses, in K.P. Donaghy, S. Poppelreuter and G. Rudinger (eds.) *Social Dimensions of Sustainable Transport: Transatlantic Perspectives*, 90 – 108, Ashgate, Aldershot.
- Axhausen, K.W. and C. Weis (2010) Predicting response rate: A natural experiment, *Survey practice*, 3 (2), <http://surveypractice.org/2010/04>.
- Axhausen, K.W. and A. Frei (2007) Contacts in a shrunken world, *Arbeitsberichte Verkehrs- und Raumplanung*, **440**, IVT, ETH Zurich, Zurich.
- Axhausen, K. W. (2008) Social networks, mobility biographies and travel: The survey challenges, *Environment and Planning B*, **35** (6) 981 - 996.
- Barrat, A. and M. weigt (2000) On the properties of small-world networks, *European Physical Journal B*, **13** (1) 547 – 560.
- Butts, C.T. (2008) Social network analysis with SNA, *Journal of Statistical Software*, **24** (6) 1 – 51.
- Carrasco, J.A. (2006) Social activity-travel behaviour: A personal network approach, *Dissertation*, University of Toronto.
- Carrasco, J.A., B. Hogan, B. Wellman and E.J. Miller (2008) Collecting social network data to study social activity-travel behaviour: an egocentric approach, *Environment and Planning B: Planning and Design*, **35** (6) 961 - 980.
- Carrasco, J.A. and K.N. Habib (2009) Understanding the social embeddedness of activity-travel participation: The case of frequency and duration of social activities, paper presented at the *12th International Conference on Travel Behaviour Research*, Jaipur, December 2009.
- Christakis, N.A. and J.H. Fowler (2009) *Connected. The surprising power of our social networks and how they shape our lives*, Little, Brown and Company, New-York.
- Coleman, J.S. (1958) Relational analysis: The study of social organizations with survey methods, *Human Organization*, **17** (4) 28 - 36.

- Deutsches Institut für Wirtschaftsforschung (2003) Mobilität in Deutschland 2003, http://www.kontiv2002.de/pdf/mid2002_tabellenband_basis.pdf, accessed July 2009.
- DfT (2006) Transport Statistics Bulletin: National Travel Survey: 2006, *Department for Transport*, London.
- Dillman, D.A. (2000) *Mail and Internet Surveys. The Tailored Design Method*, Wiley & Sons, New York.
- Erickson, B.H. (1979) Some problems of Inference from chain data, *Sociological Methodology*, **10** (1) 276 – 302.
- Frank, O. (1979) Estimation of population totals by use of snowball samples, in: Holland, P. and S. Leinhardt, *Perspective on social networks*, Academic press, New York, 319 – 346.
- Frei, A. and K.W. Axhausen (2007) Size and structures of social network geographies, *Arbeitsberichte Verkehrs- und Raumplanung*, **439**, IVT, ETH Zurich, Zurich.
- Freeman, L.C. (1979) Centrality in social networks: Conceptual clarification, *Social Networks*, **1** (1) 215 – 239.
- Gabler, S. (1992) Schneeballverfahren und verwandte Stichprobendesigns, *ZUMA-Nachrichten*, **16** (31) 47 – 69.
- Goodman, L.A. (1961) Snowball sampling, *The Annals of Mathematical Statistics*, **32** (1) 148 – 170.
- Heckathorn, D. (1997) Respondent-driven sampling: A new approach to the study of hidden populations, *Social Problems*, **44** (2) 174 – 199.
- Heckathorn, D. (2002) Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations, *Social Problems*, **49** (1) 11 – 34.
- Hill, P.B. (1988) Unterschiedliche Operationalisierungen von egozentrierten Netzwerken und ihr Erklärungsbeitrag in Kausalmodellen, *ZUMA Nachrichten*, **22** (1) 45 – 57.
- Hogan, B., J. A. Carrasco and B. Wellman (2007) Visualizing personal networks: Working with participant-aided sociograms, *Field Methods*, **19** (2) 116 – 144.
- Illenberger, J., G. Flötteröd and K. Nagel (2008) An approach to correct biases included by snowball sampling, *VSP Working paper*, **08-16**, TU Berlin, Berlin.
- Illenberger, J., G. Flötteröd, M. Kowald and K. Nagel (2009) A model for spatially embedded social networks, paper presented at the *12th International Conference on Travel Behaviour Research*, December, Jaipur.

- Jones, E.C. (2003) Class-based social networks in regional economic systems, *Research in economic anthropology*, **22** (1) 3 – 23.
- Kossinets, G. and D.J. Watts (2009) Origins of homophily in an evolving social network, *American Journal of Sociology*, **115** (2) 405 – 450.
- Kowald, M., A. Frei and J.K. Hackney, J. Illenberger and K.W. Axhausen (2009) Using an ascending sampling strategy to survey connected egocentric networks: A field work report on phase one of the survey, *Arbeitsberichte Verkehrs- und Raumplanung*, 582, IVT, ETH Zurich, Zurich.
- Larsen, J., J. Urry and K.W. Axhausen (2006) *Mobilities, Networks, Geographies*, Ashgate, Aldershot.
- Larsen, J., J. Urry and K.W. Axhausen (2008) Coordinating face-to-face meetings in mobile network societies, *Information, Communication and Society*, **11** (5) 640 – 658.
- Marin, A. (2004) Are respondents more likely to list alters with certain characteristics? Implications for name generator data, *Social Networks*, **26** (1) 289 – 307.
- Marin, A. and K.N. Hampton (2007) Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators, *Field Methods*, **19** (2) 163 – 193.
- Marsden, P.V. (1990) Network Data and Measurement, *Annual Review of Sociology*, **16**, 435 – 463.
- McPherson, M., L. Smith-Lovin and J.M. Cook (2001) Birds of a feather: Homophily in social networks, *Annual Review of Sociology*, **27**, 415 – 444.
- Milgram, S. (1967) The small world problem, *Psychology today*, **1** (1) 61 – 67.
- Mok, D. and B. Wellman (2007) Did distance matter before the internet?, *Social Networks*, **29** (3) 430 – 461.
- Mokhtarian, P.L. and I. Salomon (2002) Emerging travel patterns: Do telecommunications make a difference?, in H.S. Mahmassani (eds.) *Perceptual motion: Travel behaviour research opportunities and application challenges*, 143 – 182, Elsevier, Oxford.
- Newman, M.E.J. (2003) The structure and function of complex networks, *Siam Review*, **45** (2) 167 – 256.
- Ortuzar, J.D. and L.G. Willumsen (2006) *Modelling Transport*, Wiley & Sons, New York.
- Pool, I. and M. Kochen (1978) Contacts and Influence, *Social Networks*, **1** (1) 5-51.
- Pustejovsky, J.E. and J.P. Spillane (2009) Question order effects in social network name generators, *Social networks*, **31** (1) 221 – 229.

- R development core team (2009) *R: A language and environment for statistical computing*, R foundation for statistical computing, Vienna.
- Roberts, S.G.B., R.I.M. Dunbar, T.V. Pollet and T. Kuppens (2009) Exploring variations in active network size: Constraints and ego characteristics, *Social Networks*, **31** (1) 138 – 146.
- Salentin, K. (1999) Die Stichprobenziehung bei Zuwandererbefragungen, *ZUMA-Nachrichten*, **23** (45) 115 – 135.
- Schlich, R., S. Schönfelder, S. Hanson and K.W. Axhausen (2004) Structures of leisure travel: Temporal and spatial variability, *Transportation Review*, **24** (2) 219 – 228.
- Schönfelder, S. (2006) *Modelling the rhythms of individual travel behaviour Dissertation*, ETH Zurich, Zurich.
- Schweizer, T., M. Schnegg and S. Berzborn (1998) Personal networks and social support in a multiethnic community of southern California, *Social Networks*, **20** (1) 1 – 21.
- Scott, J. (2007) *Social Network Analysis. A Handbook*, Sage, Thousand Oaks.
- Silvis, J., D. Niemeier and R. D'Souza (2006) Social networks and travel behaviour: Report from an integrated travel diary, paper presented at the *11th International Conference on Travel Behaviour Research*, August, Kyoto.
- Steglich, C. and T.A.B. Snijders (2010) Dynamic networks and behavior: Separation selection from influence, *Social Methodology*, forthcoming.
- Stutz, F.P. (1973) Distance and network effects on urban social travel fields, *Economic Geography*, **49** (2) 134 – 144.
- Tilahun, N. and D. Levinson (2010) Contacts and meetings: Location, duration and distance travelled, paper presented at the *Annual Meeting of the Transportation Research Board*, January, Washington.
- U.S. Department of Transportation (2004) *Summary of Travel Trends: 2001 National Household Travel Survey*, Federal Highway Administration, Washington.
- Vikrant, S.V., S. Schönfelder and K.W. Axhausen (2005) Continuous space representations of human activity spaces, *Arbeitsberichte Verkehrs- und Raumplanung*, **295**, IVT, ETH Zurich, Zurich.
- Vogt, W.P. (2005) *Dictionary of Statistics and Methodology. A Nontechnical Guide for the Social Science*, Sage, Thousand Oaks.
- Wasserman, S. and K. Faust (2007) *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.
- Wolf, M. (2004) Egozentrierte Netzwerke: Erhebungsverfahren und Datenqualität, *Kölner Zeitschrift für Soziologie, Sonderheft* **44**, 244-273.

Appendix A: The name generator

Part B: Social Contacts

In this section of the survey, we ask you for the names of people with whom you spend your free time. Different situations are prescribed to frame these interactions. Please add the names with whom you have contact in the manner described.

Please keep in mind the following:

- There is an example entry at the top of the list.
- If you cannot think of an answer to a question, household objects or photo albums could perhaps be useful to jog your memory.
- Please think to include family members in addition to other social contacts.
- Each entry is associated with a number (in the column before each name). The numbering will be used in Part C.
- Enter as many names as occur to you in the respective prescribed situation.
- If you need to write more names down than fit in the list, please continue the numbering (beginning with 41) on another sheet and attach it to the survey.
- Names of people who fit in more than one situation need only be mentioned once.
- If you cannot remember the proper name of a person, simply use a nickname.

| Please list the people with whom you make plans to spend free time. (Examples: errands, sports, club or organized activities, cultural events, cooking together or going out to eat, taking holidays or excursions together) | | | | | |
|---|----------------------------|-----|------|-----|------|
| Nr. | Name | Nr. | Name | Nr. | Name |
| <i>Bsp</i> | <i>Bsp. John Q. Public</i> | 13 | | 27 | |
| | | 14 | | 28 | |
| 1 | | 15 | | 29 | |
| 2 | | 16 | | 30 | |
| 3 | | 17 | | 31 | |
| 4 | | 18 | | 32 | |
| 5 | | 19 | | 33 | |
| 6 | | 20 | | 34 | |
| 7 | | 21 | | 35 | |
| 8 | | 22 | | 36 | |
| 9 | | 23 | | 37 | |
| 10 | | 24 | | 38 | |
| 11 | | 25 | | 39 | |
| 12 | | 26 | | 40 | |

If there are other people with whom you discuss important problems, please also list them.