Insights into a spatially embedded social network from a large-scale snowball sample

Johannes Illenberger Matthias Kowald Kay W. Axhausen Kai Nagel

November 15, 2010

Abstract

Much research has been conducted to obtain insights into the basic laws governing human travel behaviour. While the traditional travel survey has been for a long time the main source of travel data, recent approaches to use GPS data, mobile phone data or the circulation of bank notes as a proxy for human travel behaviour are promising. The present study proposes a further source of such proxy-data: the social network. We collect data using an innovative snowball sampling technique to obtain details on the structure of a leisure-contacts network. We analyse the network with respect to its topology, the individuals' characteristics, and its spatial structure. We further show that a multiplication of the functions describing the spatial distribution of leisure contacts and the frequency of physical contacts results in a trip distribution that is consistent with data from the Swiss travel survey.

1 Introduction

Understanding human travel behaviour is a major research field in the community of transport science. The emerging research on complex networks reveals that the understanding of human mobility is not only required for urban planning and traffic forecasting but also to gain insights on the dynamics of the spreading of diseases, information, or social values and norms. Several tools to monitor human travel behaviour have been proposed, reaching from the traditional travel survey, such as the German MiD [1] or the Swiss micro-census [2], and travel diaries [3], to automated processing of GPS data [4], mobile phone data [5] or tracking of dollar bills [6]. The latter two approaches are promising as they usually provide much greater sample sizes compared to the traditional travel survey with relatively less effort. However, they do not directly monitor the travel patterns, but use other data as a proxy of travel behaviour. The circulation of bank notes describes rather a diffusion process than repeated movement patterns. Such data may by quite useful for the modelling of, say, virus spreading, yet the inference of individual travel behaviour is limited since the link between the circulating bill and the travelling individual is missing. For instance, it is unclear if the displacement of bank notes is congruent with the movement of individuals or if the displacement is caused by the transfer of bank notes between institutions. GPS data allows to obtain individual trajectories, yet characteristics of travellers and their motives remain unrevealed. While the traditional travel survey usually provides information about the individuals' motives and sociodemographic attributes, all of these monitoring tools have in common that they lack any information about the social connectivity of individuals. Table 1 provides an overview of the strengths and weaknesses of selected monitoring tools.

Both communication and face-to-face encounters are crucial for the maintenance of social networks [8]. In consequence, there is a reciprocal interaction between travel and communication on the one hand, and social networks on the other hand. Social networks cause travel and communication, yet travel and communication enable the spatial spread of the social

Table 1: Overview of selected monitoring tools (adopted from [7]).

aspect	paper & pencil [1, 2]	mobile phone data [5]	person-based GPS [4]	dollar bills [6]	present study
delimitation of journeys	easy	impossible	difficult	no	only leisure
coverage of journeys	with bias (depending on respondent)	incomplete	complete	no	with bias (depending on respondent)
duration of journeys	rounded (usually 5 mins)	no	exact	no	possible (with data from travel diary)
identification of locations	precise	zones	exact	no	precise (depending on respondent)
trip purpose	yes	no	imputed	no	yes (leisure)
social content	in part	no	no	no	rich

network.

With the increasing availability of large data sets, such as co-authorship networks [9], movie actor networks [10], or email networks [11], social network analysis has made great advances in understanding the dynamics of complex networks. Yet, the link to travel behaviour is missing as those networks provide no information about their spatial structure. In sociology, it has been already pointed out by Latané [12] that distance is a significant explanatory variable for the structure of social networks, but only recently research in sociology has focused on this aspect [13, 14, 15, 16].

To shed more light on the link between network structures and travel behaviour, some initial studies use the methods of social network analysis. By focusing on actors and their relations simultaneously, these methods prove productive, producing new empirical insights and results (for examples see [8, 17, 18, 19, 20, 21]). A common characteristic of these studies is their limitation to first degree relations. This means that the obtained network data is limited to isolated star-like network structures.

The present study widens the scope of network studies in transport planning by collecting data on iteratively connected personal leisure networks in Switzerland with a so-called "snowball sampling" [22] approach. This approach allows to make statements about the network structure beyond first degree relations, such as transitivity and average path length between individuals. We choose the context of leisure contacts, as this allows for the inference of leisure related travel patterns from the spatial distribution of leisure contacts and information about the frequencies of face-to-face meetings. We further show that the resulting travel patterns are consistent with other empirical studies.

The remainder of this article is organised as follows. Section 2 describes the survey instrument, the construction of the leisure network from the raw survey data and presents a technique to correct the inherent "degree bias" of the snowball approach. Section 3 presents an analysis of the network with respect to its topology, spatial structure and sociodemographic attributes of the individuals. Furthermore, travel patterns are inferred from the spatial structure of the network. The article is closed with a discussion in Sec. 4 and a summary in Sec. 5.

2 Data collection and parameter estimation

2.1 Survey instrument

Collecting data on contacts is possible with a personal network approach. Here respondents, called egos, are asked to report their social contacts, the alters. Usually, two different kinds of survey instruments are employed:

- Diaries aim to record all contacts an ego meets in a given time frame.
- Name generators use questions and stimuli to focus egos on their contacts.

Although, there are name generators aiming for the entire social network of a person [23] they are usually designed to collect data on a specific part of egos' networks. Being interested in leisure contacts makes a name generator questionnaire the appropriate design for the present study.

The questionnaire aims to collect data on egos' characteristics, ego-alter relations, alters' characteristics and alter-alter relations. The range of topics and the level of confidentiallity implied in the questions result in a high amount of response burden. Respondents are offered a 20 CHF incentive to compensate their efforts. The questionnaire is divided into four sections: (i) an introductory questionnaire, (ii) the name generator, (iii) the name interpreter and (iv) a sociogram. Details are given in the following sections.

A subset of respondents participate in a 8-days travel diary. The diary records daily activities together with the information with whom these activities are conducted, how frequently they are conducted and who initiated the activity. The travel diary is still in an early state and just little data has been collected. For that reason, its analysis is not included in this article. However, it is expected that the diary enriches the current data set.

2.1.1 The introductory questionnaire

A first set of questions asks for egos' socio-demographics. In addition, it asks for respondents' mobility biography by collecting postal addresses of former home locations, work and education places.

2.1.2 The name generator

The second part of the survey instrument is the name generator. It employs several stimuli in two questions:

1. "Please list the people with whom you make plans to spend free time (Examples: sports, club or organised activities, cultural events, cooking together or going out to eat, taking holidays or excursions together)."
2. "If there are other people with whom you discuss important problems, please list them here."

Both kinds of contacts can be considered as crucial in terms of leisure travel as the ego meets those persons frequently. Because both categories are overlapping and respondents are, to reduce response burden, not asked to mention alters who fit into both categories twice, multi-relational analyses are not possible. The questionnaire limits the number of contacts that can be named to 40. However, respondents are allowed to additionally name (note them on the back) further contacts if they feel so. Since this opportunity is only rarely used, the current analysis neglects the additional contacts.

2.1.3 The name interpreter

The third part of the questionnaire is a name interpreter asking egos to report characteristics of each alter mentioned in the name generator. Basically, the questions ask for

- alters' socio-demographics
- the relationship between ego and alter
- contact modes used and
- the contact frequency.

2.1.4 The sociogram

Finally, the last part of the questionnaire, a sociogram, asks egos to report groups of alters that make plans to spend free time together. This part of the survey instrument is influenced by the work of Carrasco et al. [18] in Toronto (see also [24]). Activity groups can be reported by mentioning the context of the activity, for example "hiking group", and identifying all alters from the name generator that join in this activity. Egos are allowed to mention up to 20 groups.

2.2 Snowball sampling

2.2.1 Sampling technique

To draw a picture of connected personal leisure networks the survey population is sampled by employing a chain methodology called snowball sampling. It belongs to the family of ascending sampling strategies and uses an initial set of first respondents, called "seeds" to ask them about their social contacts. Instead of only collecting information on the alters, snowball sampling aims to also recruit them and again ask them to report their social network. This process is repeated for a number of predetermined iterations (for a more detailed discussions on this methodology see [22, 25, 26]). The methodology has the advantage that it requires only few seeds to find other members of a given population with similar characteristics [27]. Therefore snowball sampling or similar kinds of link tracing methodologies are often used to collect information on hidden or hard-to-reach populations like drug users [28, 29], persons with sexually transmitted diseases [30], or other special populations like migrants [31]. However, snowball sampling can also be used to survey information on more general networks and investigate their global structures in terms of connectedness [32, 33]. Schweizer et al. [34] uses snowball sampling to get an impression of social support in a multi-ethnic community and Jones [35] investigates regional economic exchange relations between villages in Ecuador.

2.2.2 Snowball induced biases

Several issues have to be considered because snowball sampling is well known for several sources of bias [36, 26]. First of all, snowball samples do not fit the criterion of randomness because the probability for becoming part of the sample is influenced by the egos as egos mention their alters selectively, whether unintentionally or intentionally (for a study with such problems see [19]). This selectivity limits the number of possible paths the chain can take to be continued. To survey the network of interest as completely as possible and reduce "selection bias" the survey employs several arrangements to establish trust between respondents and the research team, such as a multi-contact strategy, a greeting postcard which egos are asked to send to their alters and a web page providing detailed information about the survey and each researcher involved (for details see [37]). Apart from these arrangements selection bias is a problem and nothing can be done if a person decides to hide certain alters.

The second source of bias results from the snowball chain as it provides a higher chance for persons with many social contacts to be included in the sample. We address this source of bias, also called "degree bias", in Sec. 2.4.

Third, bias can result from similarities between egos and their alters. These characteristic similarities, also addressed as status homophily, are well documented in network studies [38, 39]. On the one hand, the present study aims to observe the spread of homophily in connected personal networks as persons with similar characteristics have a higher probability of establishing a relationship then dissimilar pairs of persons [40]. On the other hand, there is the danger of being captured in homogeneous clusters. If, for instance, a male seed reports male contacts exclusively and the reported contacts report only male contacts again, the sample would ultimately not be representative for any general target population. The study employs arrangements to react to "homophily bias" and conserve the sample's heterogeneity. The seeds are recruited with the help of a stratified random sample in terms of sex, age and home location, whether urban or rural, of Canton Zurich's population. Switzerland is chosen as the Swiss data protection law allows for snowball sampling and many additional data on the population are available to the study, from, for example the Swiss micro-census [2], a nationally representative sample conducted every fifth year. Canton Zurich is chosen because it includes the largest Swiss City as well as smaller towns and rural areas.

The advantage of starting the snowball with a random sample is that seeds effect the populations' heterogeneity which helps to counteract homophily bias. In addition, seeds were asked to fill out the questionnaire with the help of an interviewer to increase data quality and to ensure a as complete as possible coverage of the network of interest.

2.2.3 Application

A part of the survey is still in the field. Snowball chains are started with 40 ego-seeds, however, two seeds rejected to report their contacts and are excluded in further analysis. For 20 seeds the snowball is expanded up to the second iteration. The remaining seeds are expanded up to the fourth iteration. At the time of the writing of this article the snowball is expanding the third iteration (date of data: September 2010).

So far, a response rate of 27 % is achieved (calculated conforming to the guidelines of the AAPOR [41], see also discussion in Sec. 2.3) This is satisfying considering the amount of response burden of this study. Filling out the questionnaire requires, depending on egos' network size, between one and four hours. It is also satisfying considering that the survey asks for very confidential information such as names and postal addresses of friends and family members (for arrangements to increase the response rate see [37]). Assessing the instrument's response burden a tool from commercial survey research estimated a lower response rate [42].

To our knowledge, this is the first time a snowball sample approach is used to sample a survey population of this size (targeted are 800 egos reporting about 12'000 alters) with so few restrictions on the persons included. Of course, language or national frontiers are likely to affect the spread of the snowball. However, the survey instrument makes no limitations regarding institutional settings (such as workplace, school or clubs), personal characteristics, or communication modes.

2.3 Constructing the snowball graph

For the following analysis the raw survey data is transformed into a graph data structure. Vertices represent egos and alters and edges represent the reported leisure contact. Even if, strictly speaking, the survey data represents directed edges (from ego to alter) the graph is assumed to be undirected.[1]

The raw survey data comes in the form of an edge list, i.e., a listing of all reported leisure contacts. All vertices are assigned an id and are checked for equality. This means that, if multiple egos report the same alter, the identity of an alter is first tried to be verified on a name and address basis. However, the residential locations of approximately 25 % of alters are missing since the reporting egos did not disclose their addresses. In such situations, further attributes such as age, civil status and citizenship are used to identify an alter. In critical cases the respondents are contacted for clarification. Although, much effort is invested into the validation of alters, the number of unique alters remains error-prone. The data of the sociogram (Sec. 2.1.4) is for the present ignored. An inclusion of the data is discussed in Sec. 3.1.3.

In the remainder of this article the following notations will be used. The index of an iteration is denoted with i, where the 0-th iteration represents the initial random draw of the seed vertices. Vertices that represent an ego are called ego-vertices. Vertices representing an alter are called alter-vertices. Ego-vertices are those who participated in the survey, i.e., vertices that have filled out a questionnaire. For the statistical analysis, this distinction is crucial since the true degree of an alter is unknown. In most cases the observed degree of an alter is one since it

[1] Consider the situation where the alter participated the survey in the preceding iteration but did not report the back-link to the ego. It is now unknown if the missing back-link is intentionally in the sense of "this in not my friend", unintended in the sense of "i know that our friendship has already been reported, so i do not report it again" or just forgotten.

has been reported only by one ego. Moreover, in the case of an ego, vertex related attributes (such as age or income) are reported by the ego itself, whereas alter attributes are reported by the ego and thus represents a potential source of uncertainty. The reliability of the information of an alter reported by an ego is tested by comparing such information with the details the alter reports herself once she participated in the survey in the subsequent iteration. The comparison shows that the information matches in more than 90 % of all cases, which is consistent with the findings of other studies [43].

Quantities that are calculated based on different iterations are denoted with the iteration index in parentheses in the superscript. For instance, the number of ego-vertices sampled in iteration i is denoted with $n^{(i)}$, the number of vertices that have been sampled up to and including iteration i is denoted by $n^{(\leq i)}$. Symbols without an iteration index refer to the complete sample obtained in iteration 3, i.e., n corresponds to $n^{(\leq 3)}$.

Based on the vertices' ids the edge list can be merged into one graph. The resulting graph consists of 7311 vertices and 7716 edges, where 406 vertices represent egos. An overview of the graph size per iteration is given in Tab. 2. The response rate $\alpha^{(i)}$ is defined as the number of egos $n^{(i)}$ over the total number alters in the previous iteration $i-1$. Note that this definition is different from the one that is commonly used in sociology [41] in that the sociologist would rather use the total number of all enquired vertices in the denominator. This leads to higher response rates since due to missing contact information not all alters can be enquired. However, for the estimation technique that will be discussed in Sec. 2.4 it is not relevant for what reason a vertex has not been sampled and the first definition will be used.

Table 2: Graph size per iteration. *By the time of the writing of this article iteration 3 has not been fully expanded.

iteration (i)	egos ($n^{(i)}$)	alters ($n^{(i)}_{\text{alter}}$)	edges ($m^{(i)}$)	response rate ($\alpha^{(i)}$)
0	38	568	-	1.0
1	103	1649	119	0.18
2	238	4586	303	0.14
3*	27	470	32	0.006
total	406	7273	454	0.06

2.4 Estimation

As mentioned in Sec. 2.2.2, there are several sources of bias in a snowball sample. To properly estimate topological characteristics of the network an approach to correct the degree bias is presented. Other approaches already exist [44, 45]. However, they all refer to different implementations of snowball sampling and thus apply different estimation techniques. For instance, in Respondent-Driven Sampling [28] the number of alters an ego tries to recruit is fixed, whereas in our implementation it is proportional to its degree (for simplicity it is assumed that the response rate is constant over all degrees). This section will summarise the ideas of our estimation method. For details on the characteristics of the estimator the interested reader is referred to [46].

The progress of a snowball sampling is heavily determined by the topology of the underlying network. Well connected vertices are covered relatively fast by the sampling algorithm whereas it takes a couple of iterations until the less connected vertices are reached. As a consequence, vertices with high degree are overrepresented in the early iterations. Even though this effect is undesired for network parameter estimation, it interestingly is of advantage for immunisation strategies: Randomly select a person to immunise, but also immunise her friends since it is likely that one reaches persons with higher connectivity [47].

In terms of estimation theory snowball sampling can be regarded as sampling with unequal inclusion probabilities. The inclusion probability π_v of a vertex v cannot be calculated directly; however, it can be estimated by the following considerations.

First, expand the notation of π_v to account for the iteration index. Thus denote by $\pi_v^{(\leq i)}$ the probability that vertex v is included in a snowball sample that has been run up to and including the i-th iteration.

This probability can be expressed as the probability that one of the vertex's neighbours has been sampled in or before the previous iteration $i-1$. More formally, the probability that vertex v is not sampled in iteration $\leq i$ is the joint probability that none of its neighbours w has been sampled in or before the previous iteration:

$$\pi_v^{(\leq i)} = 1 - \prod_{w=1}^{k} \left(1 - \pi_w^{(<i)}\right), \tag{1}$$

where k denotes the degree of vertex v. The probability $\pi_w^{(<i)}$ is, however, just as unknown as $\pi_v^{(\leq i)}$. An assumption that may be arguable at this point but shows to be sufficient is to ignore the details of the snowball sampling process and assume that all neighbours are equally and independently sampled. Thus, the inclusion probability of a neighbour can be approximated by

$$\pi_w^{(<i)} \approx \frac{n^{(<i)}}{N} \tag{2}$$

where N is the total number of vertices. Replacing $\pi_v^{(\leq i)}$ with the estimator $\hat{\pi}_v^{(\leq i)}$ Eq. (1) becomes

$$\hat{\pi}_v^{(\leq i)} = 1 - \prod_{w=1}^{k} \left(1 - \frac{n^{(<i)}}{N}\right). \tag{3}$$

Since Eq. (3) holds equally for all vertices of a particular degree it can be rewritten as

$$\hat{\pi}_k^{(\leq i)} = 1 - \left(1 - \frac{n^{(<i)}}{N}\right)^k. \tag{4}$$

Obviously, this estimator is only applicable for $i > 0$. In the 0-th iteration samples are drawn randomly, i.e., $\pi^{(0)} = n^{(0)}/N$.

To account for the response rate, one simply multiplies $\hat{\pi}_k$ with α. The inclusion probability of an edge (uv) is defined as

$$\hat{\pi}_{uv} = \hat{\pi}_u + \hat{\pi}_v - \hat{\pi}_u \hat{\pi}_v, \tag{5}$$

i.e., the probability that either u or v is sampled.

This estimator requires knowledge of the size N of the network which is strictly speaking unknown. However, the snowball shows to predominantly remain in the German-speaking part of Switzerland. This means that the language boundary indeed restricts the expansion of the snowball. It is thus plausible to set N to the size of the German-speaking Swiss population, approximately 5.2 million inhabitants. Moreover, since N is in the denominator of Eq. 2, its influence on $\hat{\pi}_v$ decreases with increasing N. As long as N is at least in the order of 10^6, variations of N are less significant.

Given the estimator for the inclusion probability we can obtain further statistical quantities. An estimator for a population total is

$$\hat{t}_y = \sum_v \frac{y_v}{\hat{\pi}_v}, \tag{6}$$

where y is the quantity of interest and $\sum_v$ denotes the sum over all ego-vertices. The weighted sample mean is an estimator for the population mean

$$\hat{\bar{y}} = \frac{1}{\sum_v 1/\hat{\pi}_v} \sum_v \frac{y_v}{\hat{\pi}_v}. \tag{7}$$

An other estimator is the so-called Horwitz-Thompson estimator [48] $\hat{\bar{y}} = \hat{t}_y/N$ which, however, is often inferior to the weighted sample mean, Eq. 7 [49]. The difference is that Eq. 7 replaces N with the estimator

$$\hat{N} = \frac{1}{\sum_v 1/\hat{\pi}_v}. \tag{8}$$

An estimator for the covariance is

$$\hat{S}_{yz} = \frac{1}{\hat{N}-1}\hat{t}_{yz} - \frac{1}{\hat{N}\left(\hat{N}-1\right)}\hat{t}_y \hat{t}_z, \tag{9}$$

where z denotes the second variable of interest and

$$\hat{t}_{yz} = \sum_v \frac{y_v z_v}{\hat{\pi}_v}. \tag{10}$$

This equation holds equally if one replaces $\hat{\pi}_v$ with the inclusion probability of an edge $\hat{\pi}_{uv}$, e.g., to obtain an estimator for the degree-degree correlation (Eq. 13). The estimator for variance S_{yy} and S_{zz} is analogous [49].

In the following section Eq. 7 and 9 will be used to obtain estimators for the mean degree, degree distribution, degree-degree correlation and the mean clustering coefficient.

3 Analysis

3.1 Topological network properties

3.1.1 Degree

According to Sec. 2.4, an unbiased estimator for the mean degree is

$$\hat{\bar{k}} = \frac{1}{\sum_v 1/\hat{\pi}_v} \sum_v \frac{k_v}{\hat{\pi}_v}, \tag{11}$$

where k_v is the degree of vertex v. The sum goes only over all ego-vertices, since the true degree of an alter is unknown. The estimated degree distribution is obtained with

$$\hat{p}(k) = \frac{1}{\hat{N}} \sum_{v_k} \frac{1}{\hat{\pi}_v}, \tag{12}$$

where $\sum_{v_k}$ denotes the sum over all ego-vertices with degree k.

Calculating the mean degree, without correction, for each iteration reveals the snowball bias. The mean degree for the initial random draw is $\bar{k}^{(0)} = 15$, for the sample after the first iteration is is $\bar{k}^{(\leq 1)} = 17.6$, $\bar{k}^{(\leq 2)} = 20.1$ and finally $\bar{k}^{(\leq 3)} = 20.1$. Using the estimator in Eq. 11 one obtains an estimated mean degree of $\hat{\bar{k}}^{(\leq 3)} = 13.3$, slightly less than the mean degree $\bar{k}^{(0)}$ of the initial draw.

Figure 1 shows the observed and estimated degree distribution. Both distributions are right-skewed and it is clearly visible that the estimator shifts probability mass from the high degrees to the low degrees. Different from other studies [10, 11, 50], the tail of the estimated distribution rather follows an exponential than a power law decay. We note that the survey questionnaire limits the number of contacts that can be reported to 40. (Egos can have higher degree than 40 if they named 40 contacts and then are additionally named by another ego.) Assuming that among the respondent with $k = 40$ there are actually respondents with $k > 40$, this would shift probability mass to degrees above 40.

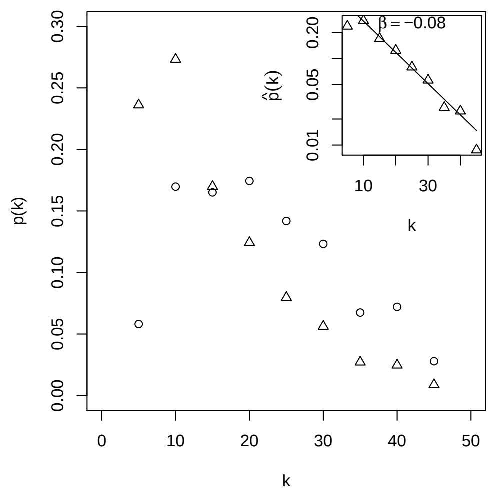


Figure 1: Observed ($p(k)$, circles) and estimated ($\hat{p}(k)$, triangles) degree distribution. Results are aggregated in bins of width 5. Inset: Log-plot of the estimated degree distribution. The solid line indicates a fit of the function $p(k) \sim \exp(\beta k)$, resulting in $\beta = -0.08$.

3.1.2 Degree-degree correlation

A further interesting property is the degree-degree correlation which can be expressed as the Pearson correlation coefficient of the degrees of the vertices on either ends of an edge [51]. Social networks are known to exhibit a positive degree-degree correlations indicating that vertices of same degree tend to be connected [52].

The nature of the snowball sampling technique also biases the degree-degree correlation, so that the observed degree-degree correlation is close to zero ($r_k = 0.08$). Using the estimator for the (co-)variance (Eq. 9) results in

$$\hat{r}_k = \frac{\hat{S}_{k_u k_v}}{\sqrt{\hat{S}_{k_u k_u}}\sqrt{\hat{S}_{k_v k_v}}} \tag{13}$$

as an estimator for the degree-degree correlation, where k_u and k_v denote the degrees on either ends of an edge. Analogous to the calculation of the estimated mean degree, only edges between ego-vertices are considered. Applied to the survey data, Eq. 13 results in $\hat{r}_k = 0.26$. This means that the present sampled network is assortative with respect to degree. Degree-degree correlations of similar magnitude are observed in networks of movie actors ($r_k = 0.208$ [10]) or company directors ($r_k = 0.276$ [53]).

3.1.3 Transitivity

Social networks are often an example for complex networks with high transitivity, i.e., a lot of triangular configurations [9, 10, 53]. There are two methods to measure transitivity in a network and one should be precise about which method is used. Transitivity can be quantified with the network clustering coefficient

$$C = \frac{3 \cdot n\,(\text{triangles})}{n\,(\text{connected triples})} \tag{14}$$

or by the mean clustering coefficient over all (ego-)vertices

$$\bar{C}_v = \frac{1}{n}\sum_v \frac{2m_v}{k\,(k-1)}, \tag{15}$$

where m_v denotes the edges connecting neighbours of vertex v and n, in this specific case, the number ego-vertices. The latter definition tends to weight the values of low-degree vertices more heavily [52]. Both definitions exhibit no significant transitivity: C=0.018 and $\bar{C}_v$=0.05.

Considering the estimation technique, it proves to be better applicable to $\bar{C}_v$ than to C: Since C_v is a vertex-local property, the mean value $\hat{\bar{C}}_v$ can be estimated according to the weighted sample mean:

$$\hat{\bar{C}}_v = \frac{1}{\sum_v 1/\hat{\pi}_v}\sum_v \frac{C_v}{\hat{\pi}_v}. \tag{16}$$

Analogous to Eq. 11, the sum goes only over all ego-vertices. Yet, even after the correction by the estimation technique, the clustering $\hat{\bar{C}}_v$ remains at 0.08.

A drawback of the snowball approach is that alter-alter relations remain undetected unless one alter participates in the survey in following iterations. Consequently, it can be assumed that the number of edges m_v connecting alters (Eq. 15) is vastly underestimated. This effect is particularly pronounced in the final iteration where no alters are surveyed.

An attempt to shed more light on alter-alter relations is made in capturing information about cliques. A clique is defined as a fully connected set of vertices and is obtained from the sociogram data. Respondents are asked to define activity-groups (for instance "hiking group" or "soccer club") and assign their alters to those groups. Connecting all alters within an activity-group with each other results in a clique. One may argue that alters within an activity-group are not necessarily connected to each other. Especially for large groups, the probability of being connected is likely to decrease. However, half of all reported cliques contain less than 4 persons. This is consistent with the findings of Dunbar [54] that groups of core contacts are rarely larger then four persons. Given the clique information, a meaningful approximation of m_v is obtained, and the estimated mean clustering coefficient increases to $\hat{\bar{C}}_v = 0.22$. The network clustering C increases considerably to 0.55.

Figure 2 shows the size of cliques as well as the average number of cliques in relation to the degree of the reporting ego. Both quantities increase nearly linearly with the degree. This means that the few high degree respondents contribute a lot of cliques containing many persons, and consequently contribute many triangles increasing network clustering C. Assuming that it is less likely that all alters in a large activiy-

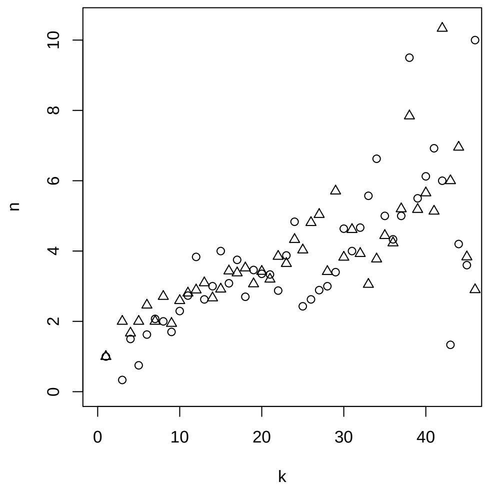


Figure 2: Average number of cliques (circles) and size (triangles) of cliques over the degree of the reporting respondent.

Table 3: Seed-vertex connection matrix. Row/column names are the ids of the seed-vertices. The entries represent the path length between both seed-vertices. Listed are only those seed-vertices that are reachable by at least one other seed-vertex.

id	31	207	754	89	799	241	845	329	297	559	63	709
31	0	-	17	-	8	-	-	9	13	-	-	4
207	-	0	-	-	-	-	5	-	-	-	-	-
754	17	-	0	-	17	-	-	8	4	-	-	13
89	-	-	-	0	-	4	-	-	-	-	-	-
799	8	-	17	-	0	-	-	9	13	-	-	4
241	-	-	-	4	-	0	-	-	-	-	-	-
845	-	5	-	-	-	-	0	-	-	-	-	-
329	9	-	8	-	9	-	-	0	4	-	-	5
297	13	-	4	-	13	-	-	4	0	-	-	9
559	-	-	-	-	-	-	-	-	-	0	6	-
63	-	-	-	-	-	-	-	-	-	6	0	-
709	4	-	13	-	4	-	-	5	9	-	-	0

group are connected, it is arguable that $\bar{C}_v$ provides a more meaningful indicator of tranisitivity since $\bar{C}_v$ assigns the same weight to the few high-degree vertices as to all other vertices.

3.1.4 Components

Starting with 38 seed vertices, the snowball graph of the 0-th iteration consists of 38 isolated components. A component denotes a graph where every vertex can be reached by every other vertex. Already within the first iteration the snowball detects bridge-persons that connect components. Three seed-to-seed paths, i.e., a path connecting a pair of seed vertices, can be identified, each with a length of four edges (see bold edges in Fig. 3). A fourth path is an indirect path, composed of two of the above paths and thus connecting two seeds through the original component of a third seed.

Within the second iteration the snowball graph merges to 30 isolated components. A total of 18 seed-to-seed paths can be identified, where ten paths are again indirect paths composed of eight direct paths. There are two paths with a length of 17 edges, which, however, overlap for 13 edges (see bold path in the centre Fig. 3). Both paths connect each five seed vertices, where four seed vertices are covered in the overlapping part. Within the third iteration no further paths are found. The average path length is 8.4 edges. It may be too early for statements about small world properties. Assuming that there exists a path form each seed-vertex to each other seed-vertex only 18 of 703 (≈ 2.5 %) possible paths are detected.

Figure 3 visualises the hitherto sampled network. A giant component is identified with 3404 vertices. It is a composition of components that originally emerged from six seed-vertices. Additionally, three further components each containing two seed vertices are identified with a size between 400 and 700 vertices. All remaining components are still the isolated egocentric networks emerging from the seed-vertex.

3.2 Spatial properties

Given the residential locations of more than 75 % of all reported egos and alters the edge length distribution is calculated (Fig. 4(a)). The distribution appears to break up into a short range domain up to approximately 20 km and a long range domain including transcontinental contacts up to 16.000 km

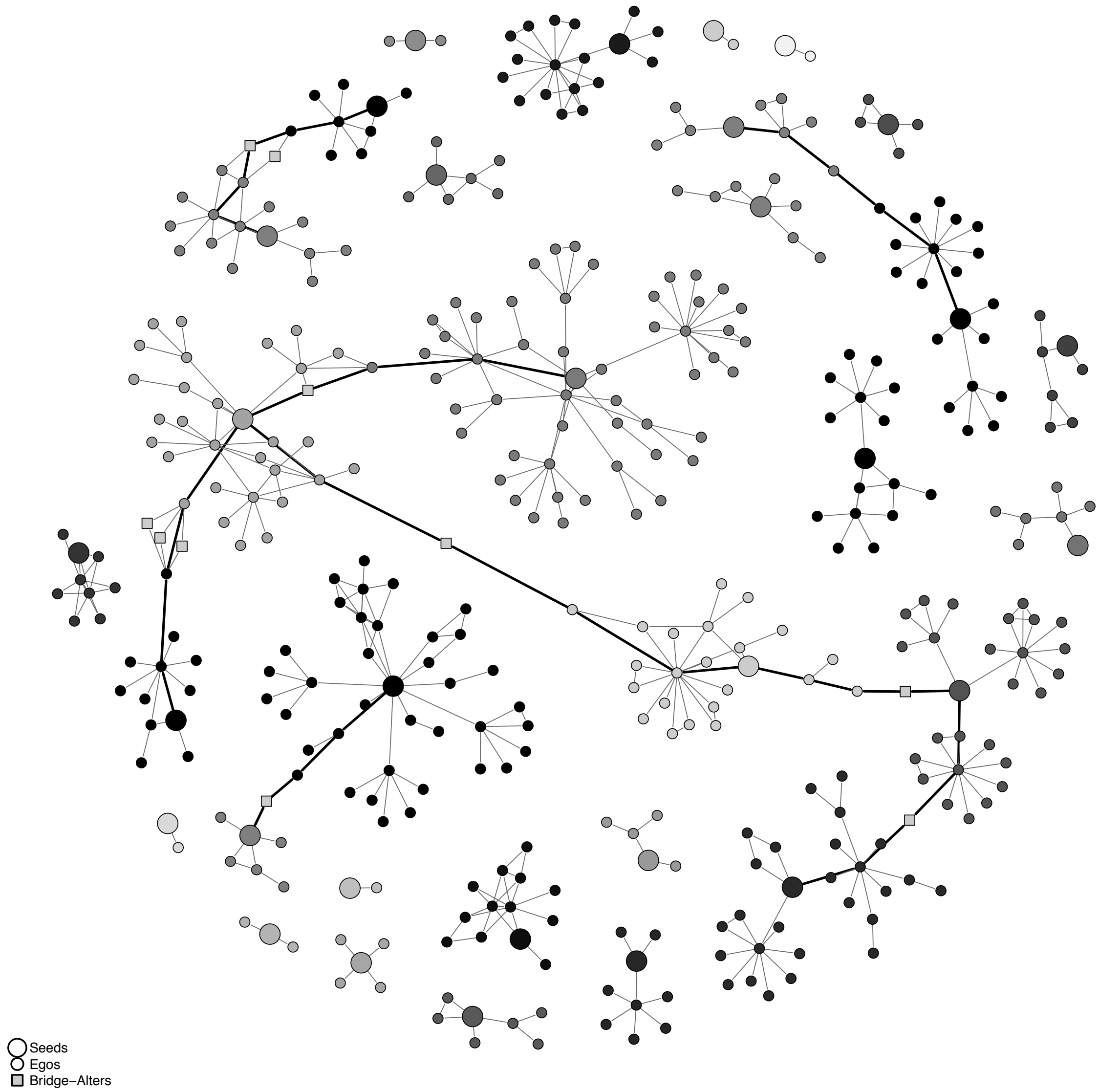


Figure 3: Sampled leisure network. Drawn are only ego-vertices, and alter-vertices if they act as a bridge vertex (gray squares) connecting two components. Paths connecting seed-vertices are highlighted with bold edges.

distance. Both domains follow a power law distribution $p_m(d) \sim d^{\beta_{1/2}}$ with $\beta_1 \approx -0.5$ for the short range domain and $\beta_2 \approx -2.1$ for the long range domain. Half of all connected individuals are located within a distance of 11 km to each other.

We assume that the observed edge length distribution is a multiplication of an individual's probability $p_{accept}(d)$ to accept a contact at distance d and the number of opportunities $M(d)$ at distance d, so that

$$p_m(d) = p_{accept}(d) \cdot M(d) \ . \tag{17}$$

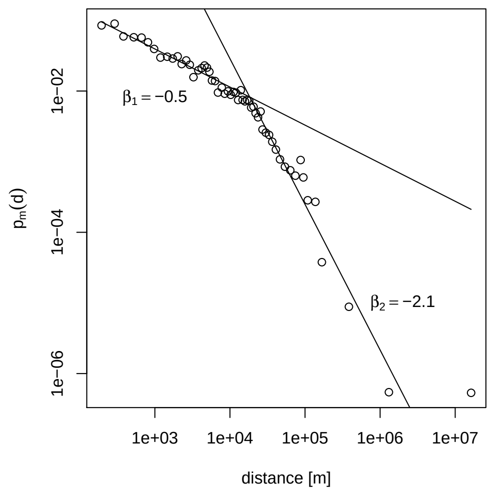
Using land use data to obtain $p_m(d)$ it is possible to extract $p_{accept}(d)$ from the survey data (Fig. 4(b)). For this, it is necessary to re-weight every occurrence of an edge connected to ego v by $1/M_v(d)$. Here, every $M_v(d)$ is individually computed for every ego as the sum of opportunities at distance d. Areas outside Switzerland contribute zero opportunities.

The function $p_{accept}(d) \sim d^{\alpha}$ with $\alpha \approx -1.6$ fits well to the resulting distribution. This may be an indicator that the change of the exponent in the edge length distribution $p_m(d)$ is induced by boundary effects. In fact, the initial seeds of the snowball are drawn within Canton Zurich, i.e., samples are concentrated within the metropolitan area of Zurich, while the southern border of Germany is approximately 20 km north of Zurich city.

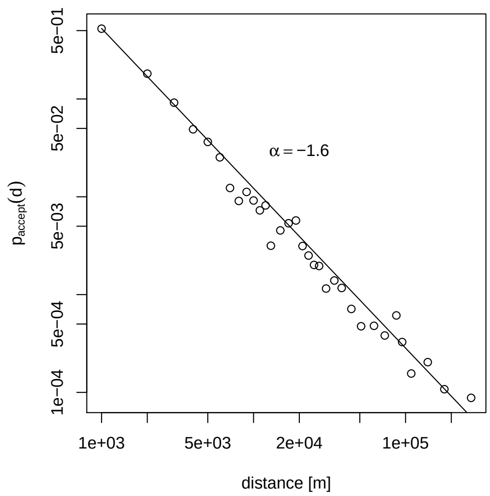
The spatial distribution of social contacts defines possible origin-destination relations of leisure related travel but makes no statement about the actual number of trips made. Therefore, it is assumed that reported physical contacts, i.e., face-to-face meetings, are located at either one actor's residential location.[2] Then, given the frequency distribution $f(d)$ of physical contacts the distribution of trips $p_{trip}(d)$ is obtained by

$$p_{trip}(d) = f(d)\, p_m(d) \ , \tag{18}$$

i.e., a multiplication of the functions describing the frequency distribution and the spatial distribution of leisure contacts. Figure 5(a) shows the trip distance distribution $p_{trip}(d)$. Similar to the edge length distribution (Fig. 4(a)), $p_{trip}(d)$ also exhibits a short



(a)



(b)

Figure 4: (a) Edge length distribution $p_m(d)$, (b) acceptance probability distribution $p_{accept}(d)$. Samples are aggregated into distance bins each containing 100 samples.

[2]Once the data from the 8-days travel diary (Sec. 2.1) is available the precise locations of face-to-face meetings are known.

range and long range domain. Both domains follow again a power law $p_{trip}(d) \sim d^{\gamma_{1/2}}$, however, with smaller exponents $\gamma_1 \approx -1.1$ and $\gamma_2 \approx -3.5$. The qualitative similarities between $p_m(d)$ and $p_{trip}(d)$ indicate that the frequencies of visits given a contact follow the same basic scaling law, i.e., $f(d) \sim d^{\eta}$, as the probability of a contact given an opportunity. Moreover, there should exist a relation of the exponents, so that $\gamma = \beta + \eta$. Figure 5(b) shows the frequency distribution $f(d)$ with respect to the face-to-face contact mode. The distribution does not (and should not) exhibit the two distance domains as $p_{trip}(d)$ and $p_m(d)$, but, apart from some outliers at very long distances, follows roughly $f(d) \sim d^{\eta}$ with $\eta \approx -0.4$.

We further conclude that leisure contacts do not only occur more frequently with short distances, they are also activated more frequently at short distances. Contacts that are met at least once a week have an average length of less than 10 km, whereas contacts that are met just once per year are more than 100 km distant.

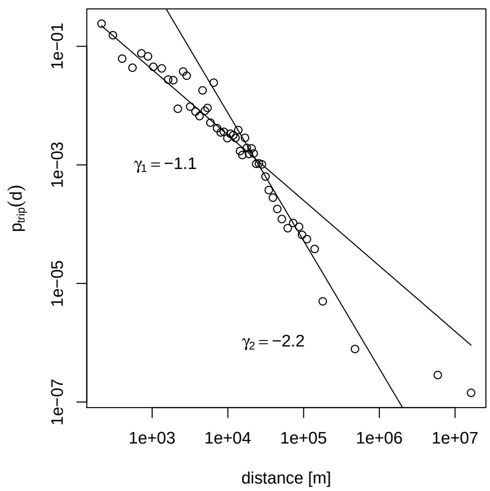
3.3 Homophily

Analogous to spatial distance, decreasing "social distance" between two actors increases the probability of being connected, where "social distance" denotes a measure of how much two individuals differ in their socio-demographic attributes. In social network analysis this phenomenon is known as homophily [38].

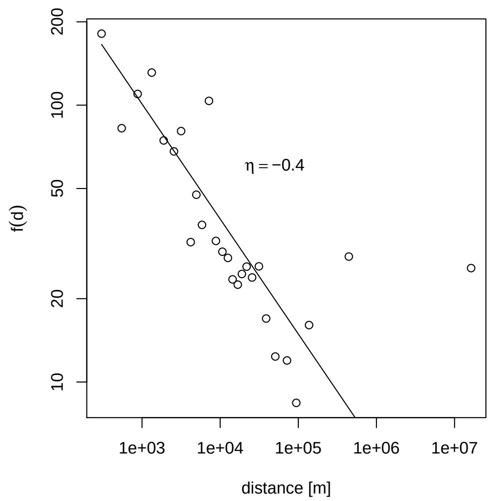
The attribute which induces the strongest degree of homophily is age. It can be quantified with the Pearson correlation coefficient of the age values at either end of an edge (uv):

$$r_a = \frac{S_{a_u a_v}}{\sqrt{S_{a_u a_u}}\sqrt{S_{a_v a_v}}}, \tag{19}$$

where a_u and a_v denote the age [years], $S_{a_u a_v}$ denotes the covariance, $S_{a_u a_u}$ and $S_{a_v a_v}$ the variance respectively. A correlation coefficient of $r_a = 0.55$ indicates a strong correlation. Interesting details on how homophily with respect to age changes during the course of life can be revealed if one looks at the alters' age distribution. For respondents of age below 30 years the distribution is rather narrow (Fig.



(a)



(b)

Figure 5: (a) Trip distribution $p_{trip}(d)$, (b) frequency distribution of physical contacts $f(d)$. Samples are aggregated into distance bins each containing 100 samples.

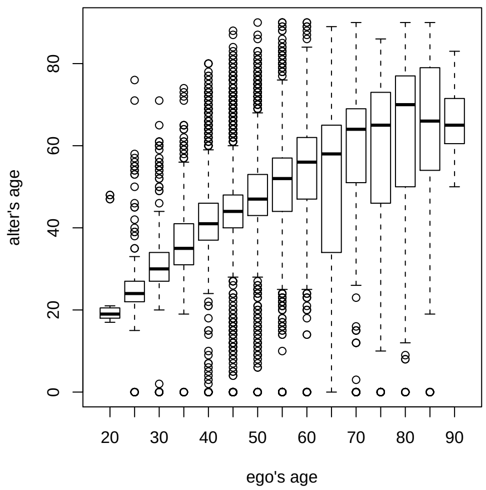


Figure 6: Box plot (conforming to [55]) of the distribution of alters' age. Lower and upper box bounds are first and third quartile respectively.

6). This means that alters are of nearly same age as the ego. With increasing age the distribution becomes broader which is rather intuitive since the absolute age difference becomes less important as one becomes older. Also remarkable is that at an age of approximately 40 years more outliers at the lower and upper end of the scale occur. The outliers at the lower end represent the young generation that grows up, the outliers at the upper bound indicate that the parental generation becomes more relevant considering leisure contacts.

Apart from age, homophily with respect to gender plays a significant role. Interestingly, the degree of homophily of this attribute differs between female and male. The probability that a contact of a female ego has the same gender is 0.72. For male respondents the probability that the alter has the same gender decreases to 0.64. As a consequence, the sample is biased such that 58% of the vertices are female, whereas in the Swiss micro-census is only 51%. It is also observed that female respondents do have more leisure contacts ($\bar{k}_{\mathrm{female}} = 20.8$) compared to male respondents ($\bar{k}_{\mathrm{male}} = 19.1$), i.e., women name more contacts than men.

Considering the level of education, a similar effect is observed. Categorising respondents into "academics" (university or university of applied science) and "non-academics" reveals that the survey data contains surprisingly many academics. However, homophily within academics is less pronounced compared to non-academics. Regarding academics, alters' level of education is almost equally distributed over both categories. In contrast, for a non-academic respondent the probability that an alter belongs to the same category is 0.75. Consequently, the above average share of academics (45 % in survey data, 15 % in Swiss micro-census) can only be explained by the greater average degree ($\bar{k}_{\mathrm{academic}} \approx 23$ and $\bar{k}_{\mathrm{non\text{-}academic}} \approx 18$).

4 Discussion

The present study confirms a couple of findings from other studies but also reveals some new aspects of social networks and the link to travel behaviour. The average number of leisure contacts per individual is estimated to $\hat{\bar{k}} = 13.3$ and is comparable to, for instance, the study of Frei and Axhausen [20] ($\bar{k} = 12.4$) or Carrasco [18] ($\bar{k} = 12.1$). The tail of the degree distribution exhibits an exponential decay which is different from the often observed power law decay.

Unlike other studies, triangular configurations in the graph appear to be less frequent. Even when the links from the sociogram data are included, the mean clustering coefficient of, for instance, networks of company directors ($\bar{C}_v = 0.59$) [53] or physics co-authorship ($\bar{C}_v = 0.43$) [9] is two or three times as large as ours. However, one should recall that the present study is not embedded into any institutional setting, such as company directors or co-authors. As we observe in the sociogram data, alters are organised into several different communities: each ego reports at average 4.25 cliques. It is thus not surprising that leisure contacts show less transitivity.

The question if the present study supports the notation of "six degrees of separation" [56] is still open. The current average path length between two seed

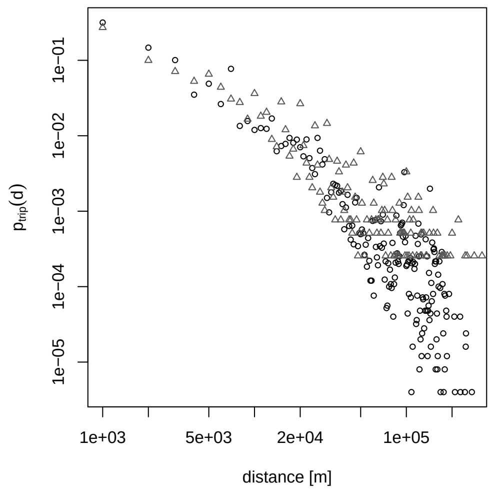


Figure 7: Comparison of $p_{trip}(d)$ between the present study (circles) with data from the Swiss micro-census (triangles). Samples are aggregated into 1 km bins.

vertices is 8.4. However, it is still possible that the snowball detects shorter paths in the next iterations. Non-responding vertices can, at first, disrupt the expansion of the snowball, but can still be reached from other components and act as bridge-vertices.

Given the spatial structure of the leisure network, we can show how to infer the leisure related trip distribution as a multiplication of the functions describing the spatial distribution of contacts and their frequency of physical meetings. A comparison with data from the Swiss micro-census reveals that the present approach is in fact able to infer reasonable leisure travel patterns. The micro-census includes a representative travel survey that also captures trip purposes. Figure 7 shows both, the trip distribution of the present study as well as the distribution of the micro-census with respect to trip purpose "visit". Considering that the sample size of the micro-census is smaller and thus scatters more, both distributions are roughly congruent.

Furthermore, the scaling laws of human travel identified in this study can be put into relation with the observations of Brockmann et al. [6] and González et al. [5]. Brockmann et al. propose that the circulation of bank notes can be used as a proxy for human travel. They show that the probability of a bill traversing a distance d (in a short time period) is well described with $p(d) \sim d^{\gamma}$ and estimate the exponent to $\gamma = -1.6$. The approach of González et al. in which they use trajectories obtained from mobile phone data shows that the probability of an individual to make a displacement of distance d is described by $p(d) = (d + d_0)^{\gamma} \exp(-d/\kappa)$ with $\gamma = -1.75$. Both studies show significant negatively greater exponents compared to the present study ($\gamma \approx -1.1$). Considering that the approaches of Brockmann et al. and González et al. do not capture trips, but random observations along a trajectory, i.e., all types of travel purposes, the different exponents indicate that leisure travel behaviour exhibits substantial differences. Quantitatively, the trip probability decays more slowly in distance in leisure travel compared to the average distances between two random observations along a trajectory.

The relations between the exponent in $p_{trip}(d)$ and the costs of travelling can be interpreted using a logit random utility model [57]. The logit model describes the probability of a trip between points u and v as $p_{uv} \sim \exp(U_{uv})$, where U_{uv} denotes the perceived utility of travelling from u to v. Equating this with the trip distribution $p_{trip}(d) \sim d^{\gamma}$, one obtains

$$U_{uv} = \gamma \ln d_{uv} + \text{const} \;, \tag{20}$$

i.e., that the (dis)utility scales logarithmically in the distance. The logarithmic form is quite plausible considering that, for instance, travelling longer distances usually involves a faster transportation mode or usage of a higher-ranked road network (see for instance [58]). Parameter γ is then the marginal (dis)utility. Although a strict comparison goes beyond the scope of this paper, it is interesting to note that in empirical estimations the marginal disutilities of travel for leisure are consistently smaller (in absolute terms) than the marginal disutilities of travel for all other travel segments (for instance Jong et al. [59] show that commuter and business travel have greater values of travel time savings compared to other travel purposes). Therefore, it is not implausible that the investigations by Brockmann et al. and González et al., averaging over all trip purposes, find slopes that are steeper than ours, for leisure travel only.

5 Summary

This article presents insights into the structure of a large-scale spatially embedded social network. The survey instrument accounts for both, revealing the topology of the network as well as collecting information about its spatial structure. While it seems practically impossible to obtain complete networks of regular leisure contacts, it is useful to go beyond the egocentric network by employing the snowball sampling approach. With the large sample size (currently 406 respondents naming more than 7700 contacts) the density of personal networks is so high that even paths connecting the initial seed vertices are found. Clearly, just having connecting components does not allow to estimate network-global parameters directly. The data, however, will provide evidence about the order of magnitude of the "degree-of-separation" distribution. This is not only useful for a much better estimate of statistical models for transport and communication modelling, but it particularly provides a sound basis for the spreading of diseases or rumours (see for instance chapter 5 of [60]).

Regarding the reciprocal interaction of social networks and travel, we focused in this article on one direction: from the social network to travel. We show that the trip distribution with respect to leisure travel is a multiplication of the functions describing the spatial distribution of leisure contacts and the frequency distribution of physical meetings. Our results are consistent with the Swiss micro-census and, moreover, the results provide further evidence that the value of travel time savings in leisure travel is substantially different from other travel segments. The latter aspect is crucial since leisure travel has become the dominating travel segment (at least in Switzerland [2], Germany [1], U.K. [61] and U.S. [62]) and detailed models are urgently needed.

The other direction, from travel to the social network, represents an aspect that is open for further research. Some initial work in this direction, by generating social networks with agent-based transport micro-simulations, has already been conducted [63, 64]

Once the snowball survey is completed it represents a large data set covering both, the topology of the social network and its spatial structure. The data from the travel diary will enrich the network data with details on the individuals' mobility patterns.

This work was funded by the VolkswagenStiftung within the project "Travel impacts of social networks and networking tools".

References

[1] Institut für angewandte Sozialwissenschft GmbH (infas), Deutsches Institut für Wirtschaftsforschung (DIW), *Ergebnistelegramm Mobilität in Deutschland* (2002)

[2] ARE/BFS, *Mobilität in der Schweiz, Ergebnissse des Mikrozensus 2005 zum Verkehrsverhalten* (2007)

[3] K.W. Axhausen, A. Zimmermann, S. Schönfelder, G. Rindsfüser, T. Haupt, Transportation **29**(2), 95 (2002)

[4] N. Schüssler, K.W. Axhausen, Transportation Research Record **2105**, 28 (2009)

[5] M. González, C. Hidalgo, A.L. Barabási, Nature **453**, 779 (2008)

[6] D. Brockmann, L. Hufnagel, T. Geisel, Nature **439**(26), 462 (2006)

[7] K.W. Axhausen, N. Schüssler, *Improving and replacing travel diaries using mobile tracing?*, presentation at Mobile Tartu (2010)

[8] J. Larsen, J. Urry, K.W. Axhausen, *Mobilities, Networks, Geographies* (Ashgate, Aldershot, 2006)

[9] M.E.J. Newman, Physical Review E **64**(016131) (2001)

[10] L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Proceedings of the National Academy of

Sciences of the United States of America **97**(21), 11149 (2000)

[11] H. Ebel, L.I. Mielsch, S. Bornholdt, Physical Review E **66**(3), 1 (2002)

[12] B. Latané, J.H. Liu, A. Nowak, M. Bonevento, L. Zheng, Personality and Social Psychology Bulletin **21**(8) (1995)

[13] C. Johnson, R.P. Gilles, Review of Economic Desgin **5**(3), 273 (2000)

[14] L.H. Wong, P. Pattison, G. Robins, arXiv:physics/0505128v2 (2005)

[15] D. Mok, B. Wellman, R. Basu, Social Networks **29**, 430 (2007)

[16] G. Daraganova, Ph.D. thesis, University of Melbourne, Department of Psychology (2008)

[17] J. Larsen, J. Urry, K.W. Axhausen, Information, Communication and Society **11**(5), 640 (2008)

[18] J.A. Carrasco, Ph.D. thesis, University of Toronto (2006)

[19] J. Silvis, D. Niemeier, R. D'Souza, *Social networks and travel behaviour: Report from an integrated travel diary*, in *Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR)* (Kyoto, Japan, 2006), see www.iatbr.org

[20] A. Frei, K.W. Axhausen, Working Paper 439, ETH Zürich, Institute for Transport Planning and Systems (2007)

[21] P. van den Berg, T. Arentze, H. Timmermans, *Social networks, ICT use and activity travel patterns: Data collection and first analyses*, in *Proceedings of the International Conference on Design & Decision Support Systems in Architecture and Urban Planning* (Eindhoven, 2008)

[22] L.A. Goodman, The Annals of Mathematical Statistics **32**(1), 148 (1961)

[23] C. McCarty, H.R. Bernard, P.D. Killworth, G.A. Shelley, E.C. Johnsen, Social Networks **19**(1), 303 (1997)

[24] B. Hogan, J.A. Carrasco, B. Wellman, Field Methods **19**(2), 116 (2007)

[25] W.P. Vogt, *Dictionary of Statistics and Methodology. A Nontechnical Guide for the Social Science* (Sage, Thousand Oaks, 2005)

[26] S. Gabler, ZUMA-Nachrichten **16**(31), 47 (1992)

[27] R. Atkinson, J. Flint, Social Research Update **33** (2001)

[28] D.D. Heckathorn, Social Problems **44**(2), 174 (1997)

[29] D.D. Heckathorn, Social Problems **49**(1), 11 (2002)

[30] C. Mathews, N. Coetzee, M. Zwarenstein, C. Lombard, S. Guttmacher, A. Oxman, G. Schmid, *Strategies for partner notification for sexually transmitted diseases*, Vol. 2 (Wiley, the Cochrane library, 2001)

[31] K. Salentin, ZUMA-Nachrichten **23**(45), 115 (1999)

[32] J. Scott, *Social Network Analysis: A Handbook* (Sage, Los Angeles, 2007)

[33] O. Frank, *Estimation of population totals by use of snowball samples* (Academic Press, New York, 1979), pp. 319–346

[34] T. Schweizer, M. Schnegg, S. Berzborn, Social Networks **20**(1), 1 (1998)

[35] E.C. Jones, Research in economic anthropology **22**(1), 3 (2003)

[36] B.H. Erickson, Sociological Methodology **10**(1), 276 (1979)

[37] M. Kowald, A. Frei, J. Hackney, J. Illenberger, K.W. Axhausen, Working Paper 582, ETH Zürich, Institute for Transport Planning and Systems (2009)

[38] M. McPherson, L. Smith-Lovin, J.M. Cook, Annual Review of Sociology **27**, 415 (2001)

[39] C. Steglich, T.A.B. Snijders, Sociological Methodology (2010)

[40] G. Kossinets, D.J. Watts, American Journal of Sociology **115**(2), 405 (2009)

[41] *The American Association for Public Opinion Research* (2009)

[42] K.W. Axhausen, C. Weiss, Survey Practice **3**(2) (2009)

[43] P.V. Mardsen, Annual Review of Sociology **16**, 435 (1990)

[44] O. Frank, T. Snijders, Journal of Official Statistics **10**(1), 53 (1994)

[45] D.D. Heckathorn, E. Volz, Journal of Official Statistics **24**(1), 79 (2008)

[46] J. Illenberger, G. Flötteröd, K. Nagel, VSP working paper, TU Berlin, Transport Systems Planning and Transport Telematics (forthcoming), see www.vsp.tu-berlin.de/publications

[47] M. Andre, K. Ijaz, J.D. Tillinghast, V.E. Krebs, L.A. Diem, B. Metchock, T. Crisp, P.D. McElroy, American Journal of Public Health **96**(11), 1 (2006)

[48] D.G. Horwitz, D.J. Thompson, Journal of the American Statistical Association **47**(260), 663 (1952)

[49] C.E. Särndal, B. Swensson, J. Wretman, *Model Assisted Survey Sampling* (Springer-Verlag, 1992)

[50] W. Aiello, F. Chung, L. Lu, *A Random Graph Model for Massive Graphs*, in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (Association of Computing Machinery, New York, 2000), pp. 171–180

[51] M.E.J. Newman, Physical Review Letters **89**(20) (2002)

[52] M.E.J. Newman, SIAM Review **45**(2), 167 (2003)

[53] G.F. Davis, M. Yoo, W.E. Baker, Strategic Organization **1**(3), 301 (2003)

[54] R. Dunbar, Behavioral and brain sciences **16**, 681 (1993)

[55] J.W. Turkey, *Exploratory data analysis* (Addison Wesley, 1977)

[56] S. Milgram, Psycology Today **2**, 60 (1967)

[57] M. Ben-Akiva, S.R. Lerman, *Discrete choice analysis* (The MIT Press, Cambridge, MA, 1985)

[58] FGSV, *Richtlinien für integrierte Netzgestaltung – RIN*, Köln (2008)

[59] G. de Jong, H. Gunn, Journal of Transport Economics and Policy **35**(2), 137 (2001)

[60] T. Smieszek, Ph.D. thesis, ETH Zurich, Switzerland (2010)

[61] Department for Transport, *Transport statistics bulletin: National travel survey: 2006*, London (2006)

[62] U.S. Department of Transportation, *Summary of travel trends: 2001 national household travel survey*, Federal Highway Administration, Washington (2001)

[63] F. Marchal, K. Nagel, Transportation Research Record **1935**, 141 (2005)

[64] J. Hackney, Ph.D. thesis, ETH Zurich, Switzerland (2009)