Networks, geographies and travel: Travel between infrastructure and social structure

ABHANDLUNG
zur Erlangung des Titels

DOKTOR DER WISSENSCHAFTEN

der

ETH ZÜRICH

vorgelegt von

ANDREAS FREI

Dipl.-Ing. ETH Zürich

geboren am
24. Juni 1979
von
Zürich ZH, Zürich ZH

Angenommen auf Antrag von

Prof. Dr. Kay W. Axhausen, examiner
Prof. Dr. Juan A. Carrasco, co-examiner

2012
## Contents

1. **Introduction** 1  
   1.1. Rationale of travel between social structure and infrastructure 1  
   1.2. Delimitation and motivation 2  
   1.3. Objectives 4  
   1.4. Methodological approach 5  
   1.5. Thesis organisation 6  

2. **Social networks and transportation** 7  
   2.1. Social networks, space and travel 10  
      2.1.1. Social networks 10  
      2.1.2. Social network approaches in context of space 14  
      2.1.3. Travel, space and social network 16  
      2.1.4. Transportation planning and social networks 20  
   2.2. Social connected traveller framework 22  

3. **Review of previous research** 29  
   3.1. Social networks as an influencing factor on travel 29  
      3.1.1. Leisure travel as focus 29  
      3.1.2. General statistics of social networks and leisure travel 30  
      3.1.3. General statistics of social networks and tourism 32  
   3.2. Case studies: Travel and social networks 33  
      3.2.1. Local and non-local contacts 34  
      3.2.2. Mobility biographies and Social Networks, Ohnmacht et al. (2006) 35  
      3.2.3. The connected lives study, Carrasco et al. (2006) 37  
      3.2.4. Social networks and mobility, Larsen et al. (2006) 40  
      3.2.5. Personal network of usual places (NUPs), Flamm and Kaufmann (2006) 41  
      3.2.6. Contact and support beyond Netville, Hampton and Wellman (2001) 42  
      3.2.7. Communication and social support, Mok and Wellman (2007) 43  
      3.2.8. Social interaction travel diary, van den Berg et. al (2009) 43  
   3.3. Simulations of social networks and travel 44  
      3.3.1. BPDA 45  
      3.3.2. ERGM 46
6.2. Model of geographically embedded social networks ........................................... 142
  6.2.1. Basic assumptions ......................................................................................... 142
  6.2.2. Two dimensional BPDA .................................................................................. 142
  6.2.3. ERGM ........................................................................................................... 143
  6.2.4. Results and discussion .................................................................................. 145
6.3. Experiments .......................................................................................................... 146
  6.3.1. Joint activity participation organization model .............................................. 147
  6.3.2. Collective choice model .................................................................................. 152

7. Conclusion and closing views .................................................................................. 159
  7.1. Empirical evidences and hypothesis .................................................................... 159
    7.1.1. Sociodemographic influences ..................................................................... 161
    7.1.2. Mobility-biographical influences ................................................................ 162
    7.1.3. Mobility tool ownership influences ............................................................ 162
    7.1.4. Contacts in a shrunken world and the influence of ICT use ....................... 162
  7.2. Relevance and contributions .............................................................................. 163
    7.2.1. Transportation, policy and social networks ................................................ 164
    7.2.2. Social capital ............................................................................................... 165
  7.3. Research Perspectives and Outlook .................................................................... 166

A. Questionnaire ........................................................................................................ 187
# List of Figures

<table>
<thead>
<tr>
<th>Figure Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1. Spatial and social density</td>
<td>15</td>
</tr>
<tr>
<td>2.1.2. Road travel time: Scaled maps of Switzerland (both same scale)</td>
<td>17</td>
</tr>
<tr>
<td>2.1.3. Quality adjusted 2004 purchase prices for private cars 1906 - 2004</td>
<td>18</td>
</tr>
<tr>
<td>2.1.4. Real costs of US interstate and international telephone calls 1930 - 2000</td>
<td>19</td>
</tr>
<tr>
<td>2.2.1. The individual in a dynamic social context</td>
<td>25</td>
</tr>
<tr>
<td>2.2.2. The individual’s decision in the context of its social network</td>
<td>27</td>
</tr>
<tr>
<td>3.2.1. Cumulative distribution of social contacts by distance and mode in rural</td>
<td>36</td>
</tr>
<tr>
<td>Lincolnshire</td>
<td></td>
</tr>
<tr>
<td>4.4.1. Response rate and ex-ante assessment of response burden</td>
<td>75</td>
</tr>
<tr>
<td>4.4.2. Geo-coded home addresses of ego’s and alter’s home location</td>
<td>82</td>
</tr>
<tr>
<td>5.1.1. Distribution of the number of relationships and the share of core contacts</td>
<td>85</td>
</tr>
<tr>
<td>5.1.2. Frequency comparison plot of the degree distribution and the best-fit Poisson and negative binomial distribution</td>
<td>88</td>
</tr>
<tr>
<td>5.2.1. Share of relationships by country (n=3,792)</td>
<td>91</td>
</tr>
<tr>
<td>5.2.2. Distance distribution between respondents and their contacts (note the different scaling of the x-axes)</td>
<td>93</td>
</tr>
<tr>
<td>5.2.3. Contact and population shares by distance band around Zurich</td>
<td>94</td>
</tr>
<tr>
<td>5.2.4. Example social network geography</td>
<td>103</td>
</tr>
<tr>
<td>5.2.5. 95%-confidence ellipses of the social network geographies for a set of the respondents</td>
<td>104</td>
</tr>
<tr>
<td>5.2.6. Sum of contact distances against the size of network geographies</td>
<td>105</td>
</tr>
<tr>
<td>5.2.7. Distribution of the social network geometries measured as 95% confidence ellipses in km$^2$</td>
<td>107</td>
</tr>
<tr>
<td>5.2.8. Size against main/minor axis (R1/R2) ratio</td>
<td>112</td>
</tr>
<tr>
<td>5.2.9. Orientation of the social network geographies</td>
<td>114</td>
</tr>
<tr>
<td>Figure Reference</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>5.3.1.</td>
<td>Share of the contact mode by log distance in deciles and zero between the respondent’s and the contact’s homes (for friends, relatives, workmates only)</td>
</tr>
<tr>
<td>5.3.2.</td>
<td>Path diagram for the multilevel path analysis model</td>
</tr>
<tr>
<td>5.3.3.</td>
<td>Trellis plot of the frequency of contact by mode against distance</td>
</tr>
<tr>
<td>5.3.4.</td>
<td>Trellis plot of the residuals of the Poisson multivariate multilevel linear regression model</td>
</tr>
<tr>
<td>5.3.5.</td>
<td>Power transformation analysis</td>
</tr>
<tr>
<td>5.3.6.</td>
<td>$Q$–$Q$ plot of the optimal power transformation and the $log_{10}$ transformation</td>
</tr>
<tr>
<td>5.3.7.</td>
<td>Histogram of the residuals of the model presented in Table 5.11 for all contact modes and trellis histogram by contact mode</td>
</tr>
<tr>
<td>5.3.8.</td>
<td>Trellis plot of the residuals of the full model presented in Table 5.11 against distance from the model presented in Table 5.11. The left part shows the contact frequency for ties within 1 km, whereas the whole distance spectrum is shown in the right part.</td>
</tr>
<tr>
<td>5.3.9.</td>
<td>Predicted contact frequency per year plus the 95% confidence interval against distance from the model presented in Table 5.11. The left part of the Figure shows the contact frequency for ties within 1 km, whereas the whole distance spectrum is shown in the right part.</td>
</tr>
<tr>
<td>6.2.1.</td>
<td>Empirical data of the edge length distribution, exponential distribution and simulated distribution</td>
</tr>
<tr>
<td>6.3.1.</td>
<td>Number of accompanying travelers</td>
</tr>
<tr>
<td>6.3.2.</td>
<td>Survival function of the empirical duration of social contacts</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Share of (a) trips and (b) miles traveled by trip purpose</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Cumulative frequency distribution of the residential locations of the contacts of a sample of residents in three Swiss immigrant neighbourhoods</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Number and configuration of relationships in social networks variables</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Information and communications technology influences on contact frequency and distance</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Mobility biography influences on contact frequency and distance</td>
<td>56</td>
</tr>
<tr>
<td>4.4</td>
<td>Mobility tool ownership influences on contact frequency and distance</td>
<td>57</td>
</tr>
<tr>
<td>4.5</td>
<td>Sociodemographic attribute influences on physical distance and contact frequency in social networks</td>
<td>58</td>
</tr>
<tr>
<td>4.6</td>
<td>Expected cost comparison for different survey protocols</td>
<td>66</td>
</tr>
<tr>
<td>4.7</td>
<td>Recruiting success rate for a face-to-face interview</td>
<td>68</td>
</tr>
<tr>
<td>4.8</td>
<td>Key figures of the pre-test and pre-estimates</td>
<td>70</td>
</tr>
<tr>
<td>4.9</td>
<td>Costing structure of the pre-test</td>
<td>70</td>
</tr>
<tr>
<td>4.10</td>
<td>Response rate of the survey on social networks and mobility biography</td>
<td>74</td>
</tr>
<tr>
<td>4.11</td>
<td>Socio-demographic comparison between the characteristics of the Zürich respondents and the Zürich population</td>
<td>77</td>
</tr>
<tr>
<td>4.12</td>
<td>Comparison of different name-generators</td>
<td>79</td>
</tr>
<tr>
<td>4.13</td>
<td>Response information accuracy of alter’s home location</td>
<td>81</td>
</tr>
<tr>
<td>5.1</td>
<td>Sample size and mean number of relationships in ego-centric social networks from different studies</td>
<td>84</td>
</tr>
<tr>
<td>5.2</td>
<td>Number of relationships by socio-demographic characteristics</td>
<td>87</td>
</tr>
</tbody>
</table>
5.3. Parameter estimates for the negative binomial regression of the degree of ego-centric social networks ........................................ 89
5.4. Comparison of the power law model fits of the tie probability with distance by Butts (2000) of the data in Latane et al. (1995) and the Zurich data ................................................................. 96
5.5. Descriptive statistics of the great circle distances between home locations of contact and respondents [km] (with and without contacts within 25 km from the respondent) ........................................ 97
5.6. Multivariate multilevel analysis of the great circle distances between home locations of contact and respondents ................................................................. 100
5.7. Descriptive statistics of 95% confidence ellipses of the social network geographies [10^5 km^2] (with and without contacts within 25 km from the respondent) ................................................................. 108
5.8. Parameter estimates for the Tobit regression of the logarithm of the size of the 95% confidence ellipses and the associated Probit model of the Cragg approach Tobit model ........................................ 110
5.9. Characteristics of named contacts by type of relationship (excepting “other” and partners) ................................................................. 116
5.10. Standardized estimates of the multi-level path analysis model of the contact frequencies by mode ................................................................. 120
5.11. Parameter estimates for full multivariate linear log-model with interactions ................................................................. 130
5.12. ANOVA of all linear models ................................................................ 131
5.13. Model comparison ................................................................ 136
6.1. Resulting group organization by parameter proportion .................. 156
6.2. Descriptive statistics of the distance of each vertex to the chosen activity location in km ................................................................ 157
7.1. Summary of the empirical results .......................................................... 160
Abstract

Travel is the price we pay to be with others, at least it is a very substantial part of the generalised cost of meeting them. If we accept this proposition, and consider that the vast majority of travel serves activities with others, then one has to wonder, why the social content of activities and the constraints arising from coordination with others has received so little attention so far in transport research. This thesis complements the previous extensive but mostly qualitative research of social networks in a spatial setting with a novel approach to incorporate interactions among individuals within social networks into transportation planning. Bi- and multivariate analysis and application of social networks and travel are carried out, based on an extensive survey in Zurich of ego-centric social networks. The thesis combines in its methodological approach, analysis and modelling a perspective influenced by transport planning and social network analysis to give new insights in social aspects of travel and how social networks can be integrated in the framework of transportation planning. The social network analysis includes the

- social network degree,
- the tie distances,
- the social network geography,
- and the contact frequencies within social networks by different contact means (face-to-face, telephone, SMS and email).

The size and structure of social network geographies is measured by the confidence ellipse of the spatial distribution of the members in egocentric social networks. Young, well educated people with a low and middle income and an eventful
work- and education-biography tend to maintain bigger social network geographies. These parameters as well as a low ratio between the minor and the main axis and a strong east-west direction are the main determinants for big social network geographies.

The tie probability shows an inverse square relationship with distance. Geographical distance matters to how people interact today, as it did in the past (Mok and Wellman, 2007). The effort involved in the different modes of contact within social networks (face-to-face, phone, email and SMS) matches the strength of the distance decay in the frequency of those interactions. The patterns of the contact frequency by mode follow these differential distance decays:

- The frequency of face-to-face visits and their relative frequencies to the other modes (shares) fall quickest with distance. While email frequency is unaffected by distance, its share relative to the other modes rises fastest with it.

- The reverse is true for SMS messaging, which stays stable relative to the other modes, while only slowly falling with distance in terms of absolute frequency.

- The high starting frequency of phoning translates into growing shares, even if the absolute frequencies fall with distance.

- The interaction of the different contact modes and face-to-face meetings are complementary.

Based on the empirical distribution of the social ties, models are used to simulate spatially embedded networks to reproduce the tie length distribution in the empirical data. Then a methodology for incorporating social networks into models of leisure activity location choice is discussed. This is the foundation to build a collective location choice model to measure the influences of social networks on travel behaviour. The results suggest that using a joint location choice model based on a social network can explain to some extent the gap between observed leisure trip length and the distance of the nearest leisure activity location measured in micro simulations.
The complex interplay between space, society and transportation is demonstrated in this thesis by theoretical and empirical work. The empirical data collection and analysis results provide an important contribution to current research of transportation planning, space sociology and social capital research.

- sozialen Netzwerk Grösse,
- die physische Distanz der Beziehungen,
- das soziale Netzwerk Geographie
- und die Kontakthäufigkeit innerhalb sozialer Netze mittels verschiedener Kontaktmittel (Face-to-Face, Telefon, SMS und E-Mail).

Die Grösse und Struktur der sozialen Netzwerke Geographie wird durch das zwei-
Erklärung


Auf der Grundlage der empirischen Verteilung der sozialen Beziehungen, werden Modelle verwendet um gesamtheitliche, räumlich eingebettete Netzwerke zu simulieren, welche die Distanzverteilung der Beziehungen in den empirischen Daten reproduziert. Dann wird der methodische Einbezug sozialer Netzwerke in Standortwahlmodelle für Freizeitaktivitäten diskutiert. Diese Grundlage wird dann verwendet, um die kollektive Standortwahl zu modelieren, um die Einflüsse von sozialen Netzwerken auf das Reiseverhalten zu messen. Die Ergebnisse legen nahe, dass mit
einem kollektiven Standortwahlmodell, basierend auf räumlichen sozialen Netzwerken, die Lücke zwischen empirisch beobachtet Freizeitswegdistanzen und den aus Mikro-Simulationen resultierenden Distanzen weitgehend erklären lässt.

Preface and Acknowledgements

This dissertation is the result of a part of my work as a PhD student at the Institute for Transport Planning and Systems (IVT) at the Swiss Federal Institute of Technology (ETH) Zurich. External work was supervised by Prof. Dr. Juan A. Carrasco, Department of Civil Engineering, Faculty of Engineering, Universidad de Concepción, Chile.

Portions of the work described in this thesis is based on the work of the following publications:

**Papers in refereed journals**


**Papers in edited volumes**

Preface and Acknowledgements


Conference papers


Other papers


The reader is referred to these articles to get a streamlined overview at the following documented dissertation.

I would like to express my thanks to:

Supervision of the work  My special thanks go to my first supervisor and role model, Prof. Dr. Kay W. Axhausen, for investing plenty of time and effort to make my project a success. Prof. Dr. Juan A. Carrasco I would like to thank for the second opinion and helpful comments as well as his work to get recognition for the importance of social networks in the field of transportation.
Financial support of the work  The work reported here was possible because of the generous and gratefully acknowledged support of the Institut für Mobilitätsforschung, Berlin and the Swiss Bundesamt für Bildung und Wirtschaft for the project Changing Behaviour Towards a more Sustainable Transport as part of the COST Action 355, as well as the VW Stiftung support, Travel impacts of social networks and networking tools.

Exchange with colleagues and project partners  The work reported in this thesis was created by exchanging views with numerous colleagues. Prof. Dr. Vincent Kaufmann, I thank you for the exchange of ideas during the work for the project Changing Behaviour Towards a more Sustainable Transport. Prof. Dr. Hani S. Mahmassani, I would like to thank you for the exchange of ideas during my research stay at the Transportation Centre at the Northwestern University Evanston. My thanks also go to Mr. Matthias Kowald, Dr. Jeremy Hackney and Dr. Timo Ohnmacht for the friendly and pleasant work on joint projects on social networks and transportation at the ETH Zurich.

I would like to thank my parents, my brother and sister in law, for mental and full financial support from my very first day at the University.

Further my thank goes to the Open Source community, which enabled this research to be carried out using versatile, distributable, free software. Last but not least I thank my wife Ms. Charlotte Frei for reviewing the manuscript.
1. Introduction

1.1. Rationale of travel between social structure and infrastructure

Transport planning aims to describe, understand and model the choices people make during the execution of their daily lives, including the more or less frequent journeys outside their daily activity space (Ortúzar and Willumsen 2001). In aggregated models, each zone attracts trips from other zones based on a function of possibilities for activities in this zone and its attractiveness for another zone decreases as its distance increases. Disaggregated activity-based models have now become available for practical application. As part of these models, the underlying mechanisms of individual travel decisions has to be understood and modelled in much more detail. The dominant paradigm for transport planning work is the individual satisfying his or her needs while maximizing the utility derived from the activities undertaken. Activity based models take the distribution of land-use patterns, which define the type of activity available at a certain location. The activity distribution is then derived from the distribution of land-use patterns. The access to land-use patterns is then described by the the generalized costs on the available infrastructure. The individual makes decision within this given framework based on its personal constraints, mainly budget constraints and capability constraints (Hägerstrand 1970). The compromise of modelling the decision-makers independently has been reached mainly because of a lack of data dealing with decision-makers interactions and also because of computational and theoretical convenience in transportation and econometric models. The fact that people do not only travel within infrastructure but also within social structure has been ignored with a few
exceptions (see e.g. Kurani and Kitamura, 1996). Nearly all of these exceptions deal with the interaction of household members and their allocation of time and resources among each other: Who is using the car, or is the household car shared among members for a certain trip? Who is running errands and picking up the child from school? Outside of the household the interaction among individuals has been mostly ignored so far.

This thesis tries to incorporate interaction among individuals within social networks. The work presented here outlines a novel approach of field work, analysis and application of social networks and travel. It combines in its methodological approach, analysis and modelling a perspective influenced by transportation planning and social network analysis to give new insights in social aspects of travel and how social networks can be integrated in the framework of transportation planning.

1.2. Delimitation and motivation

Most problems in transportation arise from peak hour traffic on daily commutes. It is therefore not surprising that most work in transportation planning concentrated on these problems. However, there are also increasing problems which arise from leisure travel. For example, there are peaks in traffic during beginning and ends of holidays, or peaks on weekends for skiing trips. In fact, leisure travel dominates the travel market in terms of the miles travelled and journeys undertaken. In industrialized countries, leisure travel shares of 33% of daily travel and 40% of daily mileage are common, see (ARE/BfS 2007a, BTS 1995). The reason why leisure travel research is rather limited compared to research of commuting travel is partly also due to its complexity. To analyse and model trip making, transport researchers use measurable indicators such as socio-demographic attributes of the population, the spatial distribution of activity opportunities, and the generalized travel costs by mode and infrastructure. In most situations, data of this type is available from local administrations and can be used for modelling commuter traffic. However it has become clear that the location choice for executing leisure activities cannot be understood with just the attributes of the location, the available infrastructure
1.2. DELIMITATION AND MOTIVATION

to access the locations and the traveller’s characteristics without understanding how social networks of relatives and friends are distributed. As leisure travel is influenced by individual lifestyles, values and essentially is driven by social motivations, i.e., to visit friends or relatives or to join them in activities (Larsen et al. 2006b). The 1995 American Travel Survey classified 33% (BTS, 1995) of the reported leisure journeys as visiting friends and others, but a further 33% involved purposes which people rarely or never engage in alone, such as outdoor sports (e.g., golf, sailing) or family events (e.g., weddings, funerals), so the true share of journeys which are fully or partly motivated by the wish to see families and friends is even higher. Götz (2004) states that the highest traffic volume in leisure time is created by activities in context of social networks or activities that serve the needs of maintaining the social contacts. Not only are visiting relatives, children and friends social activities, but also disco, cinema, theater, concert, musical, opera, exhibitions visits and attractions such as amusement parks, zoos, wildlife parks, fairs, festival, carnivals are activities which are for the most part attended with members of the social network.

For all travel, household interactions are an important constraint, and for leisure travel the traveller’s social network becomes important. A big part of leisure activities are social activities. The derived demand of these activities has to be examined under the conditions of the social network involved in these activities. The use of information and communication technologies (ICT) is important to maintain and plan social activities (Axhausen et al. 2006a). Even with the available communication means today, this thesis assumes that maintaining a social network demands personal meetings and personal travel. This thesis examines the relationship of spatial and non-spatial attributes of social networks, social activities and its derived travel demand. To understand leisure travel, we must understand the allocation of resources among members of social networks. For leisure travel the allocation of resources includes mainly the allocation of time and space. The allocation of space as location choice can therefore only be modelled as a decision of of a group of people and not only of individuals independently from each other. This thesis focuses therefore on leisure travel and especially on the impact of the spatial attributes of social networks on social activities and vice versa.
1.3. Objectives

The objectives of this work are three-fold:

- Analysing and interpreting the spatial patterns of social networks,
- generating social networks based on spatial patterns,
- and evaluating the influence of the social networks’ spatial attributes on leisure travel decisions.

The first objective and goal of this thesis is to analyse the distribution of relatives and friends over space. If travel is generally about meeting others, then it is obviously important to know from where these others start their trip to the meeting point, what they know about the opportunities offered by a destination, which constraints limit their choices or availability and therefore the choices of the full group meeting. Therefore the patterns of these journeys cannot be understood without understanding how social networks of relatives and friends are distributed over space. In any study of these spatial, behavioural and social patterns the key questions from a transport policy and social point of view are: What is the form of the precisely measured distance distribution between the home locations of the respondents and the members of their ego-centric network? What are the patterns of the frequencies of the interactions by mode? Do the shares of the modes of interaction vary systematically with physical distance? What size and orientation do the geographies of the social network have?

The second objective arises from the spatial allocation of relatives and friends: Can we use this knowledge to construct whole connected social networks? Data of such type is very rare since on the one hand the realization of appropriate surveys is often very costly and extensive, and on the other hand respondents are required to disclose private data [Axhausen 2008]. A common way to model the social relation between individuals is to use graph theory. Research on social networks started in the 1960s [Wasserman and Faust 1994] and has gained increasing interest from sociologists as well as physicists in the last years. Although there is a huge amount of studies about social networks, their use for transport planning is rather limited. This has pragmatic reasons, as empirical social networks are very hard to explore in unlocalised environments, and therefore they are often embedded in a closed
environment (e.g., people working at the same company, pupils of the same school or same class, cinema actors or collaborating scientists). A social network which is representative in terms of travel behaviour is not localised and therefore samples of such networks are very rare. In addition, the majority of social network studies provide no exact information about spatial dimensions of the social network. The second goal for the work presented in this thesis is to develop a methodology to generate data that describes the social relations required for the leisure travel modelling process.

The third research objective is to incorporate the generated social networks from the data collection and analysis about spatial social network into transport planning. First to develop a theoretical model on how the data and gained knowledge can be incorporated and second on a practical level, by using an example study. The example study should provide on one hand the feasibility of incorporating social networks into transport planning, and on the other hand the resulting errors of neglecting the influences of social networks.

1.4. Methodological approach

The methodological approach for this thesis is relatively straightforward, as the structure of the thesis is given by the objectives in the previous section. As it is acknowledged since the mid-1970s, that the understanding of travel is under-socialized (Jones et al., 1983), not much is known about how social contacts influence the travel behaviour, and up to now only a few attempts have been made to collect data to quantify these influences (Larsen et al., 2006b; Carrasco, 2006; Hogan et al., 2007; Flamm and Kaufmann, 2006; Latane et al., 1995). Therefore the basis of this thesis is to collect new data to satisfy the objectives in Section 1.3. The appropriate fieldwork methodology to collect such data has to be developed and carried out with drawing experience on the extensive sociological experience in the capturing of ego-centric networks (See for example Marsden, 1990, 2005). By means of empirical data analysis the effects of social networks on travel behaviour are described, quantified and modelled. Then simulation techniques are used to extrapolate empirical social network data to build a model of social interaction in
context of travel and vice versa.

1.5. Thesis organisation

The structure of the thesis is as follows:

Chapter 2 takes a look at the theoretical background and performs a literature review from a sociological perspective and a transportation perspective. The aspects from both fields are then incorporated into the framework of the thesis.

Chapter 3 takes a closer look at case studies done before this thesis and gives an overview on to what extend this thesis can be compared to previous research.

Chapter 4 describes the fieldwork and the survey undertaken, which builds the basis for the analysis and modelling.

Chapter 5 analyses the collected data descriptively and draws first conclusions. Based on the descriptive analysis hypothesis and the structure of the more in depth data analysis and modelling are drawn. The in-depth multivariate analysis will model the degree, spatial and contact-frequency properties of ego-centric social networks.

Chapter 6 determines one practical implications of the collected data, by using some of the model results obtained in the previous chapters to extrapolate and simulate social networks and incorporate them in location choice models.

Chapter 7 draws conclusions and contains a summary of the contributions of the work.
2. Social networks and transportation

This chapter looks at the overlap of sociological research and transportation research, more explicitly travel behaviour research. Based on the views on human behaviour from a more sociological point of view and a travel behaviour research point of view the framework for this thesis is derived. The thesis’s key focus is travel behaviour research and therefore the sociological point of view is less strongly elaborated.

In travel behaviour research, traditional factors of interests are activity duration, frequency, location, travel mode, ownership of mobility tools (such as, car, bicycle, discount cards, periodic travel pass), socio-demographics, etc. The travel demand is derived from the activities undertaken. These behavioural demand models generate the supply, and network models assign traffic to the infrastructure. These models are in turn used to evaluate and optimize changes to the transport system undertaken by the owners of its various components, e.g. reductions or expansions of road capacity through interventions. Based on the characteristics of the system and their perceptions, travellers make their decisions. For example, as new information becomes available, travellers adjust their perception and adapt their travelling behaviour. Different factors of the system as well as decisions en-route have to be considered. People also have to decide where they want to live, where their workplace is, and whether they will own a car, a transit ticket, or both. They have to decide how often and where their everyday and less frequent journeys take them, mode of transport they choose (where multi-modal trip alternatives may be available), when to start trips and what route they want to take. The paradigm used to quantify these models is that the individual satisfies his or her needs
while maximizing the utility derived from the activities undertaken. This concept was further developed towards joint activity scheduling and rescheduling. Time and task allocation at the household level has been incorporated in models (e.g. Kostyniuk and Kitamura 1983; Fujii et al. 1999; Zhang et al. 2002). However, joint activities involving household members are only a part of joint activities. More generally they include members of social network, which may be household members. Leisure activity choices often involve negotiating with the other activity participants about where, when or what to do in a leisure activity. Who should host a party or organize a film night? At which cinema should we meet to watch which film? Each individual is part of a social network and individual behaviour is not independent from others’ behaviour. Especially spatial choices can not be understood without an understanding of the spatial dimensions of a social network. Therefore transportation planner started to realize and research how individuals organize their social space. This adds a new dimension of transport behaviour modelling research. Axhausen (2008) states:

“Instead of relying exclusively on the generalised costs of travel and the hedonic utility of a location as modulated by the sociodemographics of an individual and perhaps his or her values, attitudes and lifestyle, one can add as explanatory factors both the social network geography of the person and his or her biography and network-based decision making.”

On the other hand sociologists also started to develop interest in the influences space has on social capital. Social capital is founded by the utility a social connection has. This can be trust, mutual help or information flow. Bourdieu (1983) defined social capital as “the sum of the current and potential resources associated with the property of a durable network of more or less institutionalized relationship of mutual acquaintance or acknowledgement.” That is, social capital is a resource which is based on the membership of a network of people. New communitarian-based approaches to social capital research measure social capital in the (voluntary) participation in social activities in clubs and associations. They

---

1The generalised cost of a trip is the weighted sum of resources consumed, i.e. time and money as perceived
come to the conclusion that the membership numbers in recent years had shrunk considerably (for the USA [Putnam 2000]). In combination with shrinking political interest, growing divorce rates and increased crime it is concluded that modern societies disintegrate based on an institutionalized context. These research approaches have in common that social networks are considered as a source of social capital. In this context some sociologists see the spatial expansion as one of the reasons of the disintegration of modern societies. Cass et al. (2005) notes:

“We need to know more about the spatial and temporal properties of people’s social networks and about how these vary; only this will provide a point of reference against which to judge whether social-spatial exclusion, or “access” - by which we mean the ability to negotiate space and time so as to accomplish practices and maintain relations that people take to be necessary for normal social participation - is indeed improving or declining.”

The different viewpoint from which sociologists and transportation planners look at social networks comes together as this thesis’ hypothesis that the frequency and mode of interaction in social network is crucial for understanding leisure travel, and these basic characteristics of social networks are also crucial to the amount and maintenance of social capital. The social capital embedded in these networks finds expression in travel patterns and the associated communication via mail, e-mail, phone calls, SMS messages, electronic chats, and video or picture websites. The impact of these travel patterns in terms of energy consumption and emissions (which is expected to increase further with the continuing globalisation of production and labour forces) is important for transport policy and travel behaviour analysis, as it is for sociologists to understand the mechanisms through which travellers build and maintain social capital and social networks. Otherwise, transport policy is bound to optimise its instruments blindly, looking only at its own metrics, overlooking the impact on the social capital structure of society.

What follows is a discussion of the sociological approaches to social networks with a focus on spatial attributes. Then spatial expansion from a transportation planning view is examined, as well as social network. At the end of the chapter the framework for the reminder of the thesis is derived from the theoretical discussion
in this chapter.

2.1. Social networks, space and travel

Following the three main strands of social science network analysis research in the context of space and travel are discussed;

- The classical social network analysis,
- the small-world theory,
- and various work on social capital.

Below are all the approaches briefly presented and followed up to identify their relation to space and travel.

2.1.1. Social networks

Social network analysis focuses the attention in general to the “linking” of social actors, organizations or groups and the interrelation of nation states or international relations (Freeman, 2004). A more micro sociological view on social networks defines them as the specific weave of everyday social relations (Keupp and Röhrle, 1987). Simmel (1890) presented an initial approach to the study of social networks as early as 1890. His work on relational ties between individuals and groups can be seen as a forerunner of network analysis. Interestingly he already connected social networks with geography, describing how social relationships through expansion of markets can be both individualized and globalised.

The following is an overview of the different current approaches to social network research. After reviewing the literature to the broad topic of social networks it is quickly apparent that very different approaches to the social network concept are used. Therefore this chapter tries to characterize and delineate these different approaches in the context of transportation. In general it is focused on the the social network analysis approach and its properties, the so-called small-world theory and studies examining social capital.
2.1. SOCIAL NETWORKS, SPACE AND TRAVEL

At the end of this chapter, a social network framework in terms of a social connected traveller is outlined where the transport behaviour is socially integrated and is the object of analysis. This builds the theoretical starting point of this work in terms of concrete topic of social networks. Then several terms and expressions are explained and defined the way they are used in this thesis.

Social network analysis and properties

Social network analysis is based on the mathematical description of the role position and structures of such roles. In general a social network (graph) is a mathematical expression referring to a set of nodes (vertices) and links (edges) representing well-defined relationships. (see e.g. [Freeman 2004; Wasserman and Faust 1994]). The nodes represent the individual elements of the network. Depending on research interest, this can for example be individual persons, organizations or countries. The analysis is based on the sociometry, working with a relationship matrix by directed and undirected relations (unilateral or bilateral contact). Using this information many properties of social networks, for example the density, the ratio of actual connections to the number of possible connections, can be calculated. The type of connection between the nodes defines the relationship. A distinction of different relationships can be made e.g. on the strength of the contact: weak- and strong-ties (Granovetter 1973). The definitions for a relationship and its distinction must be established by the research it is used for. After collecting a predefined network - relationships between multiple people within a defined population of a whole population - analysis at various levels can be done. Typically this includes:

- Dyad: the dyad is the smallest unit of network analysis. In this case the relationship between two elements are examined.
- Triad: The Triad addressed the study of the interrelations between three elements.
- Ego-centric network: The investigation of an ego-centric network is made completely disintegrated from the whole network. It maintains the relations of a specific individual subject (the ego) and its links to others (alters) with
which a relationship is maintained.

- Global network: The description of total networks. It is based on the calculation of relations of specific measurements, which provide information about density, multiplexity and other measurements that describe the network cohesion.

It is at this point noted that the quantitatively oriented network analysis in the empirical part of this thesis works with ego-centric networks. Based on these analysis global networks are mathematically constructed, whose relations are the further used for analysis. A special approach to social network analysis is the small-world theory. In the next section the small-world-theory is described as it takes a spatial reference which connects social network analysis with attributes of transportation.

**Small world theory**

The second perspective on social networks is small-world theory (small-world phenomenon). This approach is a special case and very prominent within social network analysis. It is based on global inter-personal networks across the entire world. The small-world theory is based on the familiar idea of the so-called "six degrees of separation" by [Milgram] (1967). In the field of social psychology he represented the idea that everyone in the world is from anyone in average only six intermediate stations away. [Milgram] (1967) designed an experiment, in which he lets a packet travel from the the American west coast to the east coast. The package was sent out to different random persons with the basic information of the target recipient. If the holder of the package did not know the target recipient personally, he was told to send the package further to a friend or relative who was more likely to know the target recipient. The package travelled only along a chain of acquaintances. With the help of this experiment [Milgram] (1967) showed the basic assumption that the modern world is highly connected. Milgram’s studies were barely followed up until the late 1990s, where in the mathematical field of probability theoretical models were provided for small-world theory. The physicist [Watts] (1999) rediscovered the small-world theory against the background of increasing global interconnectedness in the wake of globalisation. His work provides
a mathematical model, in which he assumes that every person maintains both, short and long distance relationships. In the model itself, both types of contacts are determined by a random algorithm. This model illustrates that no matter what random acquaintance model was generated, almost always the small-world theory holds. The small-world theory provides an example of how social network analysis broaches the issue of spatially expansive networks, which in return did only come available with modern forms of transportation. It also clarifies how social networks today have an impact on transportation as they shape our personal world and the needs of transportation means which are connected to our personal world. Since places can also shape social networks, reference is made to another approach used in network analysis - social capital.

Social capital

Various social services are together summarized under the concept of social capital. Social capital is based on the benefits of social relationships which can be social support, as the embeddedness can be understood in mutual help and trust. First approaches to social capital research can be found in Bourdieu (1983), who integrates the social embedding of actors in social networks with social capital. He defines social capital as the totality of actual and potential resources associated with the property of a durable network of more or less institutionalized relationships. Social networks offer support and information flow. Newer communitarian-based approaches to social capital research measure social capital based on participation in social activities in clubs and associations (Putnam, 2000). In combination with decreasing interest in politics, increasing divorce rates and increased crime, Putnam concludes that in modern societies a social disintegration is happening. Participation in associations, clubs and churches is understood as an essential prerequisite for the functioning of democratic systems. The assistance and information sources based on social interaction are described as social capital. The social engagement of citizens is thus constitutive of the social capital within a community. The decline of this social capital by a decreasing willingness to participate provides as an explanation for the problems of modern societies, as well an instruction for the solution of these problems (Putnam, 2000). The social capital research ap-
CHAPTER 2. SOCIAL NETWORKS AND TRANSPORTATION

proaches have in common that social networks are considered as a source of social capital. In that context, social networks have a strong impact on transportation and society and vice versa, as more spatially expansive social networks and transportation systems, which allow faster forms of travelling, constitute different social networks and societies, that reflect the wishes of activity and location choices. If in today’s society friends, acquaintances and family members are increasingly spatially dispersed, it follows that the maintenance of personal social networks is only possible with the help of transportation means such as cars, airplanes and trains, as well as with different means of communication. The geography of a social network can have a decisive influence not only on the travel behaviour of people, but also on their social capital. For this reason, the before mentioned approaches on social networks are further investigated in the context of transportation.

2.1.2. Social network approaches in context of space

The topic addressed in this thesis demands the analysis of these different social network approaches under the inclusion of space and mobility. Below the approaches are presented from the viewpoint of the spatial arrangement of the networks by social density, and discussed with respect to the quality of relationships. To structure different perspectives of the approaches below, the typology Figure 2.1.1 is used.

In order to assign the approaches according to a typology, the terminology has to be clarified. Social density, spatial density but also relationship and social contact and other network statistics can give different indications about the whole network or parts of the network like an edge or link. Some definitions will help the following discussion.

- Here the the degree is the most heavily used network statistic, which describes the number of links entering or exiting a node.

- Further there are structural centrality measurements, which basically measure the relative importance of each individual in a social network. The clustering coefficient measures the proportion of instances in which three nodes
are connected in a triangle relative to the number of instances in which they are connected by fewer than three edges.

- The network “density”, describes the number of realized links relative to the number of possible links \( ((N \times (N - 1))/2) \) in the graph. Mathematically this is simple described as \( D_{\text{social}} = \frac{2C_a}{N(N - 1)} \), where \( C_a \) is the actual number of links in the network and \( N \) is the number of nodes. With many other network related measurements the network density is described e.g. in Wasserman and Faust (1994); Hanneman and Riddle (2005); Monge and Contractor (2003). It describes the connectedness of the network. The social density is an indicator on the degree of integration of a given population and is thus an indication of the network quality.

- For capturing the spatial density of a network, the links have to have a geographical attribute. Subsequently, the term network geography is used
which describes the spatial dispersion or the spatial compactness of the network. Spatial dispersion refers to the expansive geographically distributed population in a network, and the term spatial compactness is used to describe network populations, which are geographically close to each other. Regarding the network geography, it is important to note that the structural feature of space in classical representations of network typologies are rare. The network edges in Figure 2.1.1 indicate not only the general connection between network members, but also represents the respective geographical distance that exists between the various stakeholders. In this work distance is referring to physical distance as a link attribute and describes of how far apart two edges in physical space are “as the crow flies”, which measures the distance along the surface.

- The degree-correlation describes the dependence of the degree of one node with the degree of its alters.

- In addition to network structural measurements of the network and the link attribute distance, the quality of the connection can be classified as strong or weak  \(^\text{Granovetter 1973}\). The boundary between the two categories, whether it is a intense, close friendship, or rather a loose acquaintance is defined in the research context. The quality of the relationship is not shown in Figure 2.1.1

2.1.3. Travel, space and social network

Increasing travel speed has brought villages, cities, countries and continents closer together, especially since the beginning of industrialization. Economies of scale and economies of scope benefited then the drastically increase in accessibility based on better infrastructure and technological advances. These changes in spatial accessibilities also increased the social potential of modern societies, as the spatial patterns of social networks change. With the changes in spatial social network also the social capital changes, which is one of the links between sociology and transportation planning and the overlap of social network research in both fields. Unease about these changes in social network membership has been a topic in the
literature since at least the 1970s (for references see [Campbell, 1990, Campbell and Lee, 1992, Day, 2006]). Changes in accessibility have permitted travellers and residents to adjust the geography of their daily lives and their social networks (see also [Hampton and Wellman, 2001]). While [Friedman, 2006] speaks of a “flat world”, it is more appropriate to think of it as a shrunken world (see [Dicken, 1998]).

One of the strongest constraints of this shift is that the generalized costs of travel limit the social contacts, and also the travel patterns. Changes in generalized costs of travel are therefore fundamental to the changes of social contacts as well as to travel patterns. As public investment in motorways and roads and private investment in vehicles have, for example, scaled Switzerland down from 1950 to 1990 for the road user by a factor of two and a factor of 1.5-1.8 for public transport. These costs did not fall further from the 1990 onwards, and they even increased in the last years through fuel costs and congestion ([Axhausen et al., 2006b] see Figure 2.1.2).

Figure 2.1.2.: Road travel time: Scaled maps of Switzerland (both same scale)

Maps: [Axhausen et al., 2008]; highway investment and motorization have halved the size of the country in terms of travel time; Scale: 1 Stunde=1h

Costs of car ownership and car operation have fallen from 1950 to 1990 even faster
(see Frei, 2005b; Raff and Trajtenberg, 1996) by a factor of 3.9, but from 1990 onwards only by 10% as illustrated in Figure 2.1.3.

Figure 2.1.3.: Quality adjusted 2004 purchase prices for private cars 1906 - 2004

Data: Frei (2005b) for Switzerland 1950-2004; Raff and Trajtenberg (1996) for the USA, 1906 - 1940. The Swiss time series is very likely underestimating the drop in the quality adjusted prices since 1985, as appropriate time series data on vehicle electronics were not available for analysis.

The collapse of long-distance telecommunication costs over the period of 1950 to 1990 has been similarly dramatic and scaled down the costs by a factor of 3.1, but from 1990 until 2001 it was even more, by a factor of 5.8 (see FCC (2001) and illustrated in Figure 2.1.4).

And if we look at the voice-over-Internet-protocol (VOIP) telephony, the marginal monetary costs are zero. So if we want to understand the effects of these dramatic
changes in communication costs and take into account, that the generalized costs of travel did not change much since the 90ties, the interactions of the different modes are crucial. Note also how the mobile phone has reduced the effort required to find and access a phone. Since World War II, the joint effect of price and travel time reductions and frequency increases in air travel have been the same order of magnitude as telecommunication changes. The regionalisation process of air services, bringing lower access and egress costs and times, and the adoption of low-cost airline business models is not yet complete, especially in Europe or Asia. It has not yet started for intercontinental flights. This promises further reductions in the generalised costs of travel.
The best-documented popular reaction to this seminal reduction in generalised costs of travel is the suburbanisation of the post-World War II period, which coincided with a long-term increase in real incomes and with the motorisation of the adult population (see Tschopp et al. 2003 for Switzerland or Bruegman 2005 for the USA).

Putnam (2000) named commuting and television as reasons for the collapse of community he diagnosed for the USA. (See Offer 2006 for the temptations of passive entertainment and its parallel fast adoption via radio and television ownership; actually these were adopted faster than household labour-saving devices). He argues that local social capital has been lost, and he equates this with an overall loss of social capital. He measured social capital mostly through membership in formal organisations as mentioned in Chapter 2.1.1. According to Putnam, further reductions in generalised travel and communication costs should lead to further losses of social capital and associated increases in negative social externalities. This seems an unlikely outcome, as we will argue in the framework presented in this chapter, increased real income should provide people with more opportunities to build and maintain social capital.

2.1.4. Transportation planning and social networks

As already mentioned, most studies within transportation planning in context of social network are devoted to study particularly the effect of household attributes on joint activity travel. Therefore the first section in this chapter summarizes the household interaction studies.

Household interactions

Jones et al. (1983), Kostyniuk and Kitamura (1983) and more recently Lee et al. (2009) studied the effects of the presence of children on the activity behaviour of adults. They both found a strong relationship, where couples with children perform most joint activities at home, whereas couples without children are more likely to perform joint out-of-home activities. The employment status of both
parts of couples influences the starting point of joint activities; couples where both are employed tend to choose a start of the location out of home and vice versa. The same researchers also found that the availability of a car influences positively more individual time-use patterns of couples. But the influence of the presence of household members on activities is not only empirically measurable on the frequency and location choice, but also on the perceived satisfaction of activities (gain in utility) (see Fujii et al. (1999) for the ratings of time spent involved in joint leisure activities). People chose, if possible, to allocate time to joint rather than independent activities. Long-term decisions are also influenced by intra household interactions. Freedman and Kern (1997) researched the implications of two-worker household status on location choices, and concluded that wives’ commute burdens influence home and workplace location decisions. In a similar study of time use via detailed in-person interviews with 30 dual-career households in the UK, Green (1997) found similar results with longer commuting trips for men. Based on these findings and more, time and task allocation at the household level was incorporated in activity based demand models models (e.g. Zhang et al. 2002 or Ettema et al. 2004). However, as stated previously, joint activities do not only involve household members, but also include other members of social network outside the household.

A review of transportation planning studies follows, which include social network members outside the household.

**Social network**

The influence of social networks on travel behaviour has been recently attracted more and more attention within the transportation behaviour research field. Surveys and data collection efforts have been made outside the work presented in this thesis (see e.g. Carrasco and Miller 2006, Silvis et al. 2006, van den Berg et al. 2009 and Kowald et al. 2009b). A number of new models and simulation have shown the influence of social networks on travel behaviour, e.g., Hackney and Axhausen (2006), Páez and Scott (2007). These studies explored mostly the cross-sectional relationships between characteristics of social networks and travel. van den Berg et al. (2009) used a social interaction diary to study the factors influencing the planning of social activities. They found that social activities scheduled
later in the day are less likely to be routine. In contrast, social activities of longer
duration and taking place on the weekend are more likely to be routine or pre-
planned. Harvey and Taylor (2000) studied the influence of the work location on
joint activities with time use data. They argue that people who work at home
spend more time alone and therefore show a tendency to travel more to fulfil their
needs for social interaction. Carrasco and Miller (2006) describe the joint activity
participation with egocentric social structure effects (degree of a person), the use
of communication technology and socio-demographic variables. They found that
people with a high egocentric social network degree are more likely to perform joint
activities. Also, the availability of communication technologies such as telephone
and internet reduces the cost of coordination and influences the participation in
joint activities. Further, the information dissemination within social network could
change people’s attitudes and perceptions leading to changes in travel behaviour
(Molin et al., 2008). To date, not much is known about the influence of social
network on travel behaviour, and studies have not been integrated: Either they fo-
cused on the descriptive analytical work, or they performed numerical experiments
and simulations. The complexity and lack of data have hindered the incorpora-
tion of social network concepts into full transportation demand models beyond
speculation about their effects.

2.2. Social connected traveller framework

The network research group around Wellman assumes that communities are now
increasingly geographically dispersed, loosely attached and are slightly overlapping
(Wellman et al., 2001). Directly linked to this is the decoupling of the relationship
of social and spatial proximity. According to Wellman et al. (2001) this results
in different spatially compact social networks and individualized forms as well
as nomadic networks. The compact social networks are characterized by a high
degree of mutual knowing and acknowledging, while the networked individualism
by stakeholders indicates that self-determination establishes independent spatial
contacts, which reduces the degree of mutual familiarity and recognition among
individuals. As already discussed (see also Axhausen, 2006, 2007), the spatial
structure of the social network is sensitive to changes in the generalised cost of travel and communication. People today have a vastly increased spatial reach, which allows them to maintain existing relationships even after one party moves or to build new relationships with persons living distant from them. They can, and have to, make choices. Given general human risk aversion, we expect that people will try to maintain existing social capital/relationships, which in time will lead to a physical dispersion of social network geographies.

The small-world theory does not include a specific geography, but it is an approach that draws on the entire globe and assumes that the social network has members beyond national borders and cultures, and its embrace can be ubiquitous (Buchanan and Aldana-Gonzales, 2003). Thus, this approach addresses network connections in an extended space without taking predefined borders to examine closed groups. The small-world theory is the representation of the third part in Figure 2.1.1. The small-world theory is characterized by a high degree of integration over a few edges and is also characterized by a weak relationship between the quality of network participants.

A rather significant shortfall in the classic social network analysis is that the effort needed to maintain social contacts has hardly been made the subject of research. Unfortunately, sociology has so far had no reason to characterise and measure the geographies of the social networks. Its focus has been on the structure of the social networks and their impact on the social processes in need of explanation (see Wasserman and Faust, 1994). The social capital literature is notoriously inconsistent with regard to definitions. See Portes and Landolt (2000), Lin (1999), Burt (2000) or Sobel (2002) for reviews; see Jackson (2005) for a review of other economic approaches to social capital and social network formation. Putnam’s focus on local voluntary organisations stands next to Bordieu’s on sympathy and solidarity produced by common social and cultural markers. Burt (2000) and Granovetter (1973) take a structural view and provide a social capital definition which is based on specific social action, i.e. brokerage of information flows, which is, unfortunately, very narrow. Völker et al. (2007), Glaeser et al. (2002) or earlier Grecco (1987, 1996a) have adopted an alternative view which guides this thesis. Social capital, in this view, is first and foremost the property of the individuals
and the persons with whom they interact. \cite{Axhausen2007}, following this lead, defines social capital

“...as the stock of joint abilities, shared histories, understandings and commitments enabling the skilled performance of joint activity, even at a distance, of a pair or larger number of persons. In contrast to discussions on human capital or team work, the range of activity is not limited here to gainful employment, but explicitly includes activity which is purely social, enjoyable and hedonic. The return on this social capital is the above-average enjoyment, monetary gain and speed of the joint performance in comparison to conducting the activity with a randomly selected person, even a randomly selected person trained for the specific activity. In this sense, it is a true capital, i.e. a stock of past achievement and work, stored and put to use in future activity.”

Positive social phenomena, such as higher levels of general trust, are products of individual social capital. The increased productivity of societies with high social capital are also influencing the political processes as described by \cite{PutnamEtAl1994} for the common benefit. Nevertheless, see also \cite{Venkatesh2009} for also negative effects of rich social capital within societies of small monetary capital. If we understand social capital as a property social network connections, its construction and maintenance becomes the choice of the persons involved. \cite{Axhausen2007} also argues that this dispersion has negative social externalities locally, such as the local anomie documented by \cite{Putnam2000}. One of the task of this thesis is to observe and document the current social network geographies.

The presented conceptual instrument serves as a foundation for the theoretical framework of this thesis. Figure 2.2.1 describes the individual in a dynamic social context, as the networked traveller.
The strong and ongoing reduction in the costs of travel and the even steeper reduction in telecommunication has allowed the individual to decouple itself in many aspects of daily life from his immediate local environment. These individualised networks of relationships justify the description of today’s travellers as networked travellers, even more so as ICT plays a strong role to build, maintain or restructure their social network. They own mobility or networking tools including cars and bicycles, public transit passes, land line and mobile phones, computers and internet access facilities. The network travellers, as a rule, do not form communities with the neighbours any more. They share the same public and semi-public spaces around their front door, but it is not clear if their networks of friends, acquaintances or work colleagues are coupled to their immediate surrounding.

Still, the neighbourhood view of the social world still is very dominant in many policy discussions, especially in transport and urban planning. But this might be
no more true and should be debated with the new findings of empirical studies.

In the context of a transportation behaviour research framework, the paradigm of the individual maximizing its own utility has to be questioned as well, as decisions are no more only subject to one person, but to a group of networked people. The traditional framework

$$\text{Max} (U) = AX$$

$$\text{s.t. } B_i X \leq R_i$$

where $A$ is the utility of a certain alternative for person $X$, $R$ are the resources available to this person and $B$ the consumption of resources, has to be extended to

$$\text{Max} (U) = A \sum_{p \in N} X_p$$

$$\text{s.t. } B_i \sum_{p \in N} X_p \leq R_i$$

where the utility maximisation is the sum over the networked persons involved for a certain alternative, and the resources of persons can be shared among the network members. This means for the transportation behaviour research, which aims to understand the derived demand of activities, that a spatial decision of a person can not be understand without knowing the location of the network members involved in this decision. For modelling these decisions, we have to ask what we need to know. Figure 2.2.2 describes the decision process of a networked traveller. Number one in Figure 2.2.2 shows the observed trip, which includes among many decisions the decision to be modelled about the location choice. To be able to understand this decision, we need to know with whom the activities are undertaken (Denoted as 2 in Figure 2.2.2). With whom, describes the other participants in this activity. These might or might not be part of the ego’s network. To determine the influence the participants of an activity have on the decision process of Ego, we have to know the relationship between the households, denoted as 3 in Figure 2.2.2. In general the social network describes these relationships and is a mathematical expression referring to a set of nodes, representing people, and links representing well-defined relationships between the people. The term “global” network refers to sum of all nodes and edges in the system. The links are all undirected. These links can have attributes, for example the number of physical or other social contacts.
(meetings) within a defined time period. The person in focus is referred to as an “ego” and those he is connected to via links, “alters”. The ego together with all the alters he is directly connected with including the connections between those alters constitutes an “egocentric network”. The egocentric network without the links between the alters is called a “deintegrated egocentric network”. The geodesic distance is defined in graph theory as the smallest number of edges that must be crossed to reach one node from another node: thus directly connected nodes have geodesic distance 1. In this work distance is referring to physical distance as a link attribute and describes of how far apart two edges in physical space are “as the crow flies”, which measures the distance along the surface. Network statistics can give different indications about the whole network or parts of the network like an edge or link.

Figure 2.2.2.: The individual’s decision in the context of its social network

The next chapter will discuss case studies related to this thesis. Then the empirical part will draw on survey undertaken to empirically analyse the spatial properties of social networks and their implications on the decision process of transportation research behaviour.
3. Review of previous research

3.1. Social networks as an influencing factor on travel

3.1.1. Leisure travel as focus

Focusing on socially motivated travel, this thesis is about the relationship between social activities and the derived mobility needed to perform these activities. Beside physical travel to meet friends face-to-face, this also includes the use of ICT as an integral part of maintaining social capital (Axhausen et al., 2006a). The maintenance of social contacts requires personal travel, and activities, which serve the maintenance of social network are responsible for the highest share in leisure travel (Götz, 2004). And although leisure travel is a rather overlooked and under-explained research field within transportation research, it is responsible for the highest share by frequency and distance of personal travel in developed countries (e.g. ARE/BfS [2007a] for Switzerland; or BTS (1995) for the US). In a non-negligible way the purpose to meet other people or at least to be co-present with others is responsible for a large share of today’s mobility. On the other hand leisure travel and the possibility of today’s mobility is an important contribution to social integration. Empirical indicators presented next, outline and underline the importance of leisure travel and the importance of understanding the spatial attributes of social networks to plan, estimate and understand leisure travel.
3.1.2. General statistics of social networks and leisure travel

In Switzerland leisure traffic accounts for the largest share of traffic volumes, not least because of social motivated activities. The Swiss transport Microcensus 2005 showed that 45 percent of the distance travelled in Switzerland, which is 95 billion passenger kilometres, is due to recreational activities (ARE/BfS 2007a). But similar numbers are also outside of Switzerland observed, as shown in Table 3.1. Conservative scenarios projected by the Federal Office for Spatial Development (ARE) expect a growing share of leisure travel and extrapolate the share of leisure travel to account for 50 percent of the total transport volume by 2030 (ARE 2006).

Members of the Swiss resident population travel everyday in average 6.16 km for recreational purposes (ARE/BfS 2007b). A comparison between the results of the Swiss transport Microcensus 2005 and the first Microcensus from 1984 show an increase of passenger kilometres by 15 percent in leisure travel (ARE/BfS 2007b). Compared to 1970, the leisure travel in Switzerland has even grown by 60 percent (Grotrian 2007). The increasing travelled distances in leisure travel are over-proportional to other travel purposes, and can therefore not only be explained by
the decreasing costs for mobility (see Figure 2.1.2 and 2.1.3). The additional cost decrease in telecommunication (see Figure 2.1.4) has decreased the maintenance cost for social networks further, and it can therefore be argued, that the over-proportional demand growth for leisure activity related travel is an expression of individual mobility needs of people who may have changed their mobility behaviour by altering the spatial arrangement of their social network. While the amount of time commuting in Switzerland has remained relatively constant over the past 25 years, travel time for leisure activities have increased steadily (Simma 2002). Since the time budget for commuting is roughly constant, while it has increased for leisure travel, it can also be assumed that the monetary budget for leisure as well as the general time budget allocated towards leisure has increased. Leisure traffic reflects a phenomenon of modernity. (In the Swiss travel survey Microcensus 2005 the leisure activity “visits” as main activity purpose was the most common). In 2000, 18.3 percent of all leisure travel mentioned “visits” as the main reason, and 21.5 percent in 2005 (cf. ARE/BfS, 2001, 2007a). Other studies clearly show that leisure and activities with relatives and friends are closely linked (see e.g. Fastenmeier et al., 2003; Schlich et al., 2004). In a 12-week travel diary survey in Thurgau Switzerland Schlich et al. (2004) found that meeting friends and relatives accounts for 14 percent as the main leisure activities. Note that in the Thurgau travel diary 18 types of activities were given, where the “real” number of social motivated activities is likely higher than the reported ones. In Germany the numbers are similar. The German travel survey “Mobility in Germany 2002” (MiD 2002) reports that 42 percent of the total traffic volume in Germany is caused by leisure activities. Compared to 1982, this also means an increase of 15 percent (Götz, 2004). And also in Germany the travel purpose leisure travel is dominated by visiting friends and relatives, which is reported to account for 31 percent of the leisure traffic (Follmer et al., 2004). It is important to remember that other activities such as restaurant visits, sports event visits or participation, or visiting cultural events, such as concerts, museums but also festivals etc. are usually also undertaken with other people, which biases the statistics of the social purpose leisure travel towards smaller numbers. Numbers in the US are similar. The 1995 American Travel Survey classified 33 percent of the reported journeys as visiting friends and

\footnote{Own analysis based on the data available at www.transtats.bts.gov}
relatives, but a further 33 percent involved purposes which people rarely or never engage in alone, such as outdoor sports (e.g., golf, sailing) or family events (e.g., weddings, funerals), so that the true share of journeys which are fully or partly motivated by the wish to see family and friends is even higher. The numbers are 29 percent for visiting friends and relatives and 34 percent for other pleasure/leisure activities in the 2001 US National Household Survey. As further evidence of the growth in leisure traffic because of spatially disperse social networks, the following numbers look at the composition of households in Switzerland. The number of single person households has increased drastically over the past 50 years. In 1960 only 14.2 percent of all households were single households. This number increased by 1980 to 32.4 percent and was 2005 reported to be 36 percent [BIS, 2007]. And again, the forecasts assume further increase in single households in the future, in this case it was predicted that single households are expected to increase to 41 percent by 2030 [BIS, 2008]. In terms of the spatial attributes of social networks, the increasing number of single households indicate that the social life is no more as localized as it once was. The social life is played outside of the home (apart from visits that are received at home), thus it contributes to the increase of leisure traffic.

3.1.3. General statistics of social networks and tourism

Another part of leisure traffic is tourism. The 2004 official UK tourism statistics reported for example that 37 percent of all non-regular travel was motivated by visits to friends and relatives. A share of 59 percent was reported for South Africa [Rule et al., 2003]. In Switzerland the tourism survey reported 47 percent of short private trips (less than three nights or more) of the Swiss population end with friends [BIS, 2005]. Other statistics show that in 2001 11 percent of the persons who crossed the Swiss border by land were motivated by visiting a known person [ARE, 2003], which was the dominant mentioned motivation criterion. The World Tourism Organisation defines tourism as travel beyond the range of daily life, generally involving an overnight stay away from home.

Own calculations based on Table 1 in [BIS, 2003], correcting for the commuting included in the table. 

http://www.staruk.org.uk/default.asp?ID=731&parentid=469

The World Tourism Organisation defines tourism as travel beyond the range of daily life, generally involving an overnight stay away from home.
Tourism Organization (WTO) reports 154 million arrivals worldwide in 2001 as visiting friends and relatives, health, religion and others. These are twice as many arrivals as late as 1990 (WTO 2004). In addition, there are in Switzerland reliable statistics on travel patterns, which show that meetings with friends and relatives represent a very common motif for the leisure market. The “Switzerland Travel Market 2002”, for example, shows that at 22 percent of the long distance journeys of the Swiss population stay with friends at the destination (Bieger and Laesser 2008). Larsen et al. (2006b) illustrate in their study that friends and relatives living spatially far away are visited less often, but visitors from further away stay for a longer period of time. In terms of local close relationships, visits offer the possibility of a (short) trip with a meeting of friends. This clearly indicates a shift in the everyday mobility to travel longer distances for maintaining the social network.

In summary it can be formulated at this point that given the importance of social network in leisure and tourist travel and their importance in volumes of traffic per se, it is inevitably to study the spatial attributes of social networks to understand this major segment of travel, but also to understand the importance of the role mobility has for the social integration of modern societies. If the geography of the social network changes, changes in the leisure mobility behaviour is inevitably, as leisure and tourism trips are conditional on participation in the daily social life and thus directly related to the spatial arrangement of social relations. In general, the spatial arrangement of social relations is a major influencing factor on the distance and frequency of leisure travel, and thus increasing spatial dispersed social network arrangement is a strong source for the increase in leisure travel.

3.2. Case studies: Travel and social networks

So far, macro-indicators were presented for leisure travel in Switzerland, underlining the importance of leisure travel for social integration. In the remaining section the few scientific projects that have examined the effects of social networks on the forms of mobility already will be presented. The selection of examples is based on different aspects of the influence of social networks on various forms of mobility.
Thus, all elements of interest can be taken into account. Moreover, the focus is on selecting case studies with related data collection or at least analysis, since such studies are surprisingly rare, especially considering that in sociology the influence of space on society is a topic which many researchers have studied for many years.

Given the effort involved in precise geocoding before the arrival of cheap georeferenced place-name or address data bases in the 1990s, it is understandable that sociologists have avoided this task. This was reinforced by a preference for samples which had a clear and limited spatial focus, such as firms, classes of school children or a neighbourhood of limited spatial extent (see Day, 2006 for a review of studies of "neighbouring"). There was no effort to account for the geography and shape of the distribution of the contacts.

3.2.1. Local and non-local contacts

The table from Wimmer (2004) is typical of studies in sociology to incorporate spatial attributes (Table 3.2). While a wide range of analyses are possible with such tables, it is very difficult to compare them between studies. The definitions of the spatial categories vary between them, and more importantly, they have different spatial scales. Just consider the category “canton” in the table below, which refers to two spatially large cantons (Bern and Zurich) and one tiny one (Basel-Stadt). See also Lauterbach and Pillemer (1996), Grossetti (2007), Henning (2007) or Litwak and Kulis (1987) for further examples, or McPherson et al. (2001) for further references. Nevertheless, note the mixture of local and non-local contacts of Wimmer’s respondents. Outside sociology, Rothenberg et al. (2005) recently employed detailed geocoding to analyse the distance between contacts at risk of HIV. Again we can note the mixture of local contacts (distance for the 25% quartiles) and non-local contacts (90% percentile) for this special population.

Work on contact frequencies is generally hampered by the categorical coding of distances. See for example Onnen et al. (1987); Lawton et al. (1994) or Hank (2005), who had to work with indicator variables such as “Within 1-hour drive” or “Less than 25 km distance”. Latane et al. (1993) provided estimates based on respondent estimates of distances. Clark and Unwin (1981) come closest to more...
3.2. CASE STUDIES: TRAVEL AND SOCIAL NETWORKS

Table 3.2.: Cumulative frequency distribution of the residential locations of the contacts of a sample of residents in three Swiss immigrant neighbourhoods

<table>
<thead>
<tr>
<th>Residence of the contacts in</th>
<th>Same part of neighbourhood</th>
<th>Same neighbourhood</th>
<th>Same district</th>
<th>Same city</th>
<th>Inside the same canton</th>
<th>Switzerland</th>
<th>World</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swiss respondents</td>
<td>30%</td>
<td>38%</td>
<td>45%</td>
<td>77%</td>
<td>86%</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td>Italian respondents</td>
<td>20%</td>
<td>33%</td>
<td>42%</td>
<td>81%</td>
<td>90%</td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td>Turkish respondents</td>
<td>24%</td>
<td>40%</td>
<td>45%</td>
<td>76%</td>
<td>88%</td>
<td>97%</td>
<td>100%</td>
</tr>
<tr>
<td>All respondents</td>
<td>24%</td>
<td>37%</td>
<td>44%</td>
<td>78%</td>
<td>88%</td>
<td>96%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: Adapted from Wimmer 2004

precise measured studies of distance, as their mid-1970s survey in rural Lincolnshire requested information about all interactions by mode (face-to-face, phone and mail) over a five-day diary period, but they did not identify the contacts involved individually (see Figure 3.2.1).

A more in depth and detail of surveys and studies to incorporate distance in social network and contact frequency studies are given following.

3.2.2. Mobility biographies and Social Networks, Ohnmacht et al. (2006)

As a preliminary study to the project, which this thesis is based on, Ohnmacht and Axhausen (2005) conducted a mostly qualitative study to explore the relationship between social networks and travel. For this purpose they conducted 30 in depth oral interviews. During the interviews they collected information about the geography of social networks, work, holiday destinations, and their means of transport throughout their lives Ohnmacht (2004). They measured the geography of social networks based on the interviewees’ social contacts home locations. These home
Figure 3.2.1.: Cumulative distribution of social contacts by distance and mode in rural Lincolnshire

Source: adapted from Clark and Unwin (1981) Figure 2
3.2. CASE STUDIES: TRAVEL AND SOCIAL NETWORKS

locations were geocoded and the resulting coordinates are then used to calculate a two-dimensional confidence interval, the 95%-confidence ellipse. Qualitative data is then used to differentiate between six identified cases: Interviewee(s) with a diasporic family, interviewee(s) with a recent change workplace because of unemployment, interviewee(s) with a city residence for lifestyle reasons, interviewee(s) evicted by a war, interviewee(s) with a mobile profession and localized interviewee(s). The extend of the network geography of these different is then discussed with anecdotes biographical facts from their life’s in a quantitative way. Based on this analysis the following conclusions are drawn:

- Mobility biographies are increasingly the subject of conception of life of people;
- Besides the own life conception, there are also external requirements effecting the social geographies (global recruitment, career specialization etc.). The appropriation of space is thus determined by a set of endogenous and exogenous factors.
- The advanced modernisation of societies involves a shift away from local socializing reference points toward more expansive spatial areas of common interest.
- Modern societies require attributes such as high education, mobility and flexibility from people in the “new” job world.
- Reduced or impeded access to mobility tools is a dimensions of social exclusion.

3.2.3. The connected lives study, Carrasco et al. (2006)

The Connected Lives project examined mobility in the context of social network (Wellman et al. 2005). Carrasco and Miller (2006) and Carrasco (2006) examine in their empirical work travel behaviour as a function of network geographies in particular, and in general depending on the ego-centric social network. In their work, they extend the activity-based transportation research framework to include a social dimension, as the social understanding of travel behaviour is underdeveloped. In their analysis, they examine the number and locations of meetings as
a function of socio-demographic characteristics of the spatial composition of the social network as well as the frequency of contact by means of internet and telephone communications. They use a small sample ($n = 78$) of Toronto residents to study the distance between ego and alters and their contact frequency. Their basic hypothesis is (Carrasco and Miller, 2006):

“individuals social network characteristics are relevant for their propensity to perform social activities and that these effects can be appropriately measured and used to understand the underlying decision making processes.”

Distance

The empirical findings of Carrasco (2006) show how a higher income is associated with greater distances travelled to visit social contacts. These results are consistent with U.S. studies that demonstrate with empirical evidence that people with higher income travel longer for socially motivated activities (Tardiff, 1977). This suggests that their social contacts are arranged on a more extensive spatial scale and/or that they have sufficient funds to bridge the distance, while poorer people, such as migrants with highly dispersed social networks, are not able to maintain their transnational social network face-to-face with personal travel because they lack the financial budget to do so. In addition, Carrasco shows that people who are living for a longer period of time at the same place, maintain geographically more compact networks in comparison with people who live less long at the same place. Persons living together with a partner, have a tendency to have spatially more compact social networks. People who work from home are as well characterized by spatially more compact social networks. In his study Carrasco shows, that parents are characterized by close social relationships within the community. On the other hand, groups with a extended social geography are characterized by internet access and an immigrant background.

Carrasco (2006) in his dissertation pursued to aim similar objectives to those of this thesis, but focussed primarily on the trip-generation effects of the structure of the social networks (For the underlying survey, see Carrasco, 2006; Hogan et al., 2007).
3.2. CASE STUDIES: TRAVEL AND SOCIAL NETWORKS

He reports for a neighbourhood in Toronto a left-skewed distance distribution with a mode of 1km and a mean of 1,017 km (median 741 km) for the reported distances between the homes of respondents and their contacts. The mode is consistent with what is known from the previous literature. The meaningful median and mean are only possible because of the geocoding, although these values have to be taken with some caution, as Carrasco coded international distances by the distance between the respondents and the capital of the contact’s country of residence.

Contact frequency

Based on the collected data in Toronto from the Connected Lives project [Wellman et al. (2005) and Carrasco and Miller (2006)] model the contact frequency with multivariate statistical methods. On the household and individual characteristics level, income shows a strong influence on contact frequency. The contact frequency increases with higher income levels. It turns out that people with lower incomes meet more often in their own home than people from higher income groups. The latter group is meeting their social network more frequently in bars and restaurants. Furthermore, the presence of children affects the frequency of contact negatively, especially in the case of meetings in bars or restaurants. Women meet less frequently with social contacts - both at home, as well as in restaurants, cafés and bars. The number of years spent at the time of the survey at the current home-location has a positive effect on meeting at home. According to these results, the number of meetings increases with close contacts in their own home with the number of years of residence in one place. [Carrasco and Miller (2006)] infer from these results, that a low residential and urban mobility could be a proxy for a more stable and settled social network, and thus more intimate people with whom to socialize. As a central finding [Wellman et al. (2005)] has to be noted that there are strong differences in network geography (distance between ego and alter) and contact frequencies, which can be attributed to socio-demographic and socio-economic influences.
3.2.4. Social networks and mobility, Larsen et al. (2006)

In the spring of 2004 Larsen et al. (2006a) analysed 24 male people from different professions and industry groups including architects, security personnel, managers and staff of fitness clubs in the age range from 24 to 32. In order to obtain the widest possible distribution of socio-demographic and socio-economic characteristics within this very specific sample, they used quota sampling. Geographically the sample was focused on North West England (Manchester, Liverpool and Lancaster). Behind the selection of these professional groups are two basic theses. Larsen et al. (2006b) states the first argument concerning the professional group of architects and managers in health clubs that those tend to have an educationally diverse biography. They can be very mobile and be well equipped with networking and mobility tools. In addition, members of this group are expected to travel long distances to visit their social network. The second thesis concerns the safety and fitness club staff. Larsen et al. (2006b) assume that this group is characterised by a more local life with a geographically local and dense network, as their education requires less specialization and is therefore often conducted within the same city they also work now. The theoretical starting point of Larsen et al. (2006b) is that personal travel is often made out of a strong “compulsion to proximity” to be physically close to important social contacts and from the need to fulfil cultural commitments to key people. This pressure for proximity requires geographically distant social relations for especially big travel expenses. Spatially extended relationships are according to Larsen et al. (2006b) relationships that are not in the immediate environment, be it for example, in the same neighbourhood or in the same village/city. Against the background of the compulsion for proximity for meeting any such contacts longer distances have to be overcome. Larsen et al. (2006b) define close social relationships as contacts with especially important fellow human beings such as the nuclear family, close relatives and friends and working colleagues. Such contacts have a cohesive force for social actors and develop normally continuously over a long period of time. According to Granovetter (1973) close social relationships lasting, deep bondings with frequent interaction and a strong emotional component. Weak relationships on the other hand can be according to Granovetter (1973) described as loose everyday relationships with a
small “compulsion to proximity”, which would for example be true for co-workers. Similar to the previously presented Connected Lives project by Wellman et al. (2005), respondents were asked to report location, type and frequency of meetings (face-to-face and virtual meetings) with friends, acquaintances and relatives. Larsen et al. (2006b) asked the study participants for a maximum of ten close and ten weak social contacts. According to the results of Larsen et al. (2006b) the professional group of architects and managers of health clubs reflected as hypothesised that they maintain in fact spatially more expansive social networks in comparison to the group of security personnel and health club staff. The first group is characterized by higher incomes and longer education in comparison to the second. In addition, Larsen et al. (2006b) show that spatial distances play a role in the case of close relationships especially, since they have more frequent mutual visits and frequent communication. It is shown that spatially distant relationship visits are of longer duration than spatially close ones, and visits are often combined with holiday activities (Larsen et al., 2007). (Larsen et al., 2006b) found empirical evidence that there are differences in the network geographies, which can be explained as a function of education degrees. This empirical finding is, according to the authors, also reflected in income, as education and income are highly correlated.

3.2.5. Personal network of usual places (NUPs), Flamm and Kaufmann (2006)

Flamm and Kaufmann (2006) describe often visited places by person as a “personal network of usual places” (NUPs). Looking at ego-centric networks from a socio-material perspective, the nodes of NUPs represent the residence, meeting places with the family, friends, partners, classmates, and often visited places like restaurants, or the fitness centre for the bi-weekly training. In addition, the networks are represented by transfer nodes such as the short shopping trips or the stop at the station to switch from rail tram. Within the theoretical concept of Flamm and Kaufmann (2006) the network nodes are connected with each other; and the connecting arc represent paths that are characterized by the choice of mobility tool, as well as the the travel duration and travel frequency. In addition, the network nodes
represent the centres of social practices. [Flamm and Kaufmann (2006)] examined the social geographies using the "6-week Mobidrive travel diary" (see [König et al., 2000] for a description of the data set). Their results show that social structural differentiation (which was calculated based on [Levy et al., 1997]) have no significant distinctions in the geographical extension of the activity-space represented by the NUPs. However, the network geographies of frequently visited places by people with access to private cars are spatially statistically significant more extensive. While self-employed persons and full time employed persons have spatially more extensive NUPs, housewives or students have spatially more compact organized NUPs. (Flamm and Kaufmann, 2006) conclude that due to the small differences of network geographies by socio-economic and socio-demographic characteristics, the differences of network geographies should instead be studied more in depth in connection to lifestyle. The central finding of the study by Flamm and Kaufmann (2006) is that in addition to variables of the situation in life (students, housewives, employment status), there are no differences in the network geographies by socio-demographic and socio-economic factors. However, their study show significant influences of mobility tool ownership: According to their result, people with private car have more extensive spatial NUPs.

3.2.6. Contact and support beyond Netville, Hampton and Wellman (2001)

Hampton and Wellman (2001) raise the question of how the internet affects social capital within networks. Interaction without physical presence between persons is widespread available at least since letter postal services. The traditional postal mail service has been extended by telegraph, telephone, and later e-mail; and more recently through mobile phones, Skype, etc. To maintain relationships across long distances, these communication tools became self-evident. Hampton and Wellman (2001) research the effects of the internet use on social capital by an empirical study of residents living in the suburbs of Toronto, before and after they had access to internet and thus before and after they had access to communication via e-mail. As a result, the contact and social exchange with distant relatives increased,
3.2. CASE STUDIES: TRAVEL AND SOCIAL NETWORKS

while the meetings with local contacts stagnated at the same level. Specifically Hampton and Wellman (2001) researched the interaction between the contact and physical meetings. Their empirical results show that the e-mail contact was used for additional interactions, without increasing the number of face-to-face meetings or telephone calls, respectively, to substitute them. Hampton and Wellman (2001) show how new means of communication help to maintain extensive social network geographies without substituting or complementing face-to-face contacts.

3.2.7. Communication and social support, Mok and Wellman (2007)

Mok and Wellman (2007) examine a small sample of 29 ego-centric social networks from 1979. They show that long spatial distances undermine comprehensive assistance, especially when face-to-face interactions can not be substituted by letters or phone calls (Mok and Wellman, 2007). Their study shows again, that the frequency of contact and spatial proximity are strongly dependent on each other and in a strong relationship with mutual assistance services. Long distance interactions can undermine comprehensive assistance, as these can not be replaced by more impersonal forms of interaction (Mok and Wellman, 2007). From the work of Putnam (2000) and Mok and Wellman (2007) one can conclude that social capital expressed as information and emotional support are affected by the increased mobilization of socially relevant persons. The underlying thesis for this is that social relations can experience drastic changes with geographic changes.

3.2.8. Social interaction travel diary, van den Berg et. al (2009)

van den Berg et al. (2010) used a social interaction diary to study the factors influencing the planning of social activities. They collected data from 116 respondents about their social network, as well as a two day travel diary. The distance between ego’s and alter’s home location was reported by the respondents in categories. They also asked about the frequency of contact by face-to-face, telephone, SMS
and email in categories. They found that social activities scheduled later in the day are less likely to be routine. In contrast, social activities of longer duration and taking place in the weekend are more likely to be routine or preplanned. They also found that beside certain socio demographic factors, mainly car ownership and work status as well as the spatial distance between ego’s and alter’s home location influence the degree of a social network (van den Berg et al. 2009). Their analysis of the frequency of contact by different means indicates a complementary effect of ICT use with face-to-face contact frequency (van den Berg et al. 2009).

3.3. Simulations of social networks and travel

Research on models of spatially embedded networks is even more sparse than literature on the analysis of the spatial dimension of social contacts. Recent coupled models include work from Arentze and Timmermans (2008), Hackney and Axhausen (2006), and Marchal and Nagel (2005). Arentze and Timmermans (2008) present a fully developed concept for social interactions and activity patterns based on the ego-centric (personal) network, including abstractions of homophily (McPherson et al. 2001), social need, and satisfaction. Their utility functions maximize the value of ego networks within the total discretionary time budget of the agent. Tests were limited to four social groups of five agents each, so there are no summary statistics of the social networks, and it is noted that the complex results of even the small population tested are difficult to summarize and understand.

Marchal and Nagel (2005) have modelled the spread of information about secondary location choice (shopping) along the affiliation network of co-located co-workers.

Hackney and Marchal (2009) propose a model which is based on the idea that there is a certain probability that people become friends if they remain at the same place in an overlapping time interval. To extract the trajectories of peoples’ mobility patterns, they use a microscopic traffic simulation, which simulates daily traffic in the greater area of Zurich, Switzerland. Hackney and Marchal find that
distance distribution of social contacts is heavily right-skewed, but they provide no detailed analysis on the shape of the distribution.

Besides the model of Hackney and Marchal, there are several social network models, which are based on probability, but which are not especially used to model geographically embedded social networks. However, Toivonen et al. (2009) classifies these social network models in two main categories: those were the link updates depends on local network structures (network evolution models, NEMs) and those were the update of the links is dependent on the attributes of the vertices or nodes (nodal attribute models, NAMs). A third family of model, which can be interpreted as NEM or NAM, is the exponential random graph model (ERMG). The difference between NEMs and ERMGs are that the ERMGs do not include a link update step based on probabilities. The ERGMs link existence is defined by the probability distribution. Markov Chain Monte Carlo (MCMC) sampling methods, which are used in ERMGs, can also be used to model the evolution of social network, as proposed by Snijders et al. (2010).

All of these models generate non-bipartite, acquaintance networks, which are undirected.

Since the spatial dimension of a social network is defined as an attribute of each link and not as a network structure, a recent NAM (Boguna et al. 2004 (BPDA) and an ERGM Wong et al. 2006 (ERGM)) are studied following in more depth.

3.3.1. BPDA

Boguna et al. (2004) propose a class of models based on the concept of social distance, e.g. based on the idea, that friendship links are established whenever individuals feel close in some sense to each other. This leads to a notion of social distance, which rule the establishment of friendship in a way that individuals with short distances have a large probability of being related, and vice versa for long distances. Their model has 3 free parameters $N$, $\alpha$ and $\beta$. The mechanism can be described as:

- Distribute $N$ nodes with uniform probability in a (one-dimensional) social space (a segment of length $h_{max}$);
Link nodes with probability \( p = \frac{1}{1+(d/\beta)} \), where \( d \) represents the distance in the social space and \( h_{\text{max}} \) can be absorbed within \( \beta \).

If the model is treated many-dimensionally, similarity along one of the social dimensions is sufficient for the nodes to be seen as similar.

### 3.3.2. ERGM

Wong et al. (2006) propose a spatial random graph model. The model uses four free parameters, \( N, H, p \) and \( p_b \). The model’s mechanism can be described as: Distribute \( N \) nodes according to a homogeneous Poisson point process in a (two-dimensional) social space of unit area. Create a link between each node pair separated by distance \( d \) with probability \( p + p_b \) if \( d < H \), and with probability \( p - p_{\text{delta}} \) if \( d > H \) (where \( p_{\text{delta}}(p, p_b, H) \) is such that the total fraction \( p \) of all possible links is generated). They model the probability of an edge formation as a simple step function that only differentiates between edges within and beyond a so-called neighbourhood radius. As a simulation scenario they use randomly scattered vertices. Wong et al. study the effect of the neighbourhood radius on small-world properties and community structures, but they make no statement about the edge length distribution.

### 3.3.3. Integration of social networks in a large scale travel behaviour microsimulation, Hackney (2009)

Hackney (2009) studied in his dissertation the effects of coupled social influences and travel behaviour within a well established multi agent based transport simulation tool (MATSim). He uses “only” simulation techniques without data about social interactions and activity-travel behaviour. The social networks are created based on the short-term traffic flows, which determine the exact time and location of each agent within a day. Based on the overlapping time and space, each agent pair is assigned a probability to form a relationship. These social networks then guide the interactions of agents in space, as the agents receive different utility rewards for face-to-face meeting, compared to activities performed alone. The social
network mechanism incorporated in provide measurable emergent coordination of activity travel in time and space, which is not possible with a centrally-governing mechanism.

The results from incorporating social networks in MATSim are analysed towards changes in activity timing and secondary (non-work and home) location choices.

**Activity timing effects**

While Hackney could not show any effects from the incorporation of social networks on the aggregate travel statistics such as shifts in departure time, the incorporation of social networks causes peak spreading of travel volumes in the evening. This effect is due to more activities toward the evening of leisure activities. Leisure activities are given a higher priority for those agents, with a higher number of friends.

**Secondary location choice effects**

In his experiments, Hackney does not detect any significant differences in traffic volumes with the incorporation of social networks in the utility function of activity participation. However, the activity chain and trip length for secondary activities are significantly longer when the effects of social networks are considered. The increase in distance is dependent on the model parameters chosen in the utility formulation. An other interesting insight from the incorporation of social networks in the models is the effect on location choice based on the information exchange between agents. By allowing of increasing information to spread, Hackney discovered variation in the number of agents per location. This means that locations are more popular because they are more likely to be mentioned socially.

**3.3.4. Short conclusions simulation of social networks and travel**

In general one may conclude, that simulation models which incorporate social networks into travel demand models are rare. This can be partly attributed to the
lack of empirical data to use and calibrate such simulation models. Most of the models use either hypothetical scenarios, or use existing models to simulate the process of building social networks. New empirical data gives new opportunities to build spatial embedded social networks based on real data, and then use them to feed the travel demand models.

3.4. Conclusions from the review of previous research and case studies

Social sciences and mobility research discussed theoretical expectations of the consequences of spatially expanding social networks for a while now. In contrast, only a few systematic studies were carried out approaching this field of study empirically. This partial review has also demonstrated that previous work has been handicapped by the absence of precise distance estimates. Knowledge could not accumulate or be compared, as the distances between the respondents and their contacts were coded categorically in non-comparable ways and measured in erroneous ways. Only the most recent studies are overcoming this limitation through geocoding, which is now a nearly trivial expense for the researcher. The few studies with empirical findings and more precise measured social network geographies were presented in this chapter. These studies show effects of social network on different forms of mobility. The overview of macro figures about leisure travel stated that the spatial arrangement of social contacts is a possible factor influencing the growth of leisure traffic, and therefore has to be taken into account. With regard to the structure of the social network, the empirical findings indicate, that differences in socio-demographic and socio-economic influences contact frequency and network size (Wellman et al., 2005; Carrasco, 2006; Larsen et al., 2006b). In other studies, socio-economic differentiation has no effect (Flamm and Kaufmann, 2006), whereas mainly vehicle ownership has a positive impact on the geography of the social network (Flamm and Kaufmann, 2006; van den Berg et al., 2009). Further results show that physical meetings can not be replaced by ICT contacts, in contrast, ICT contacts are rather a complement to face-to-face contacts (Hampton and Wellman, 2001). However, it is clear that effects of biographical
3.4. CONCLUSIONS FROM THE REVIEW OF PREVIOUS RESEARCH AND CASE STUDIES

dimensions on social networks has so far not received much attention. Also the precisely measured distances between members of social network are still not much explored, as well as the frequency of contacts and the interaction between physical meetings with ICT contacts. Thus, the connection of mobility biographical influences on the network geography, which in turn can affect the leisure travel behaviour, remains mostly unexplored. Especially if the effects of social network go beyond the trip generation effects. It remains particularly unclear how existing social networks and mobility behaviour interact, and how, for example, the spatial distribution of network nodes influences the decision making process within leisure travel. It is important to note that there are in Switzerland no systematic studies on the relationship between mobility and social networks. In addition, no studies on contact frequency have been found, which measures the distance and contact frequency continuously and precisely. The empirical findings presented here indicate that there are systematic differences in the geographical arrangement of social networks, which can in turn be assigned to other social differentials. In the next sections the main focus will be to address these research gaps and implement them into a survey, whose results will then be analysed. The aim of the analysis should be to deliver new insights into the differences in network size and network geography for Switzerland, as well as how different communication means influence each other at different distances. Because differences in the conquest of space will just be socially critical, it can be assumed that these differences in the appropriation of space can be explained. This is attempted in the following sections by representing the socio-spatial worlds as network geographies. Using the reviewed simulation techniques, data about spatial social networks can be incorporated in practical travel demand models. The influence of social network and their spatial attributes can shed new light on the underlying behaviour in leisure travel. These new insights into behaviour processes can then be used to model crucial decisions in the travel demand process with greater accuracy and influences of social network geographies can be quantified and made useful for policy.
4. Fieldwork

4.1. Preface

The fieldwork and data collection and its subsequent analyses are based on the work done for the project “Changing Behaviour Towards a more Sustainable Transport” (Frei 2005a, 2007b, Frei et al. 2009, Frei and Axhausen 2010). As part of this project it was promised to conduct a representative survey of 300 respondents in the city of Zurich. The survey conducting process from this project is described in the following sections, and the collected data is then used as the basis for the empirical data analysis in this thesis.

4.2. Research gaps, hypothesis and questionnaire content

This Section gives first an overview of the different research gaps and topics found in the literature and case study review. Based on the findings of the review and the gaps in the literature broader hypothesis are formed. These Hypothesis are used to derive the questionnaire content and the design, which is then discussed in the following sections.

Topics and research gaps

Leisure travel is heavily based on activities that are performed together with friends and relatives. Highlighted is the traffic that is generated for the purpose of visit,
but also activities which are performed together are of interest, because their destination can only be understood with the underlying spatial distribution of the social network (see Schlich et al., 2004; Axhausen, 2002; Götz, 2004). Further, do spatially distant social contacts also represent the target for tourist activities (cf. ARE, 2003; WTO, 2004; Larsen et al., 2006a)? In the context of the sociological literature it can be observed that spatially expansive networks have higher organizational and financial maintenance levels, and in addition is the exchange of favours and emotional assistance suffering from increasing distances. This observation is supported further from empirical studies which show, that the frequency of contact and proximity are in a strongly associated with mutual assistance (Mok and Wellman, 2007). So far, there has been no earlier systematic study in Switzerland about social network and their spatial representation. Further there are also no earlier studies which examine the influence of social network on leisure travel in Switzerland. So, both of these topics are novelties and will be studied in depth for the first time with this thesis.

For designing the content of the questionnaire, the different topics discussed in the literature review will be categorized together with the literature. With this list, the relevant findings in terms of variables is then summarized to build the base for the questionnaire together with additional questions to address further research gaps. The different identified topics are not exclusive and overlap often, but nevertheless, they make it easier to grasp the many different influences from the different fields discussed here.

### 4.2.1. Number and configuration of relationships in social networks

Structural information about the egocentric social network are new to travel behaviour research and build the main focus here. For our purposes, merely transferring the approaches and objectives which sociologists working on social network structures have developed would not suffice (see Wasserman and Faust, 1994 for an overview and Freeman, 2004 for a history of social network analysis). Survey methods for obtaining lists of social contacts (name generators) from respondents
tend to focus on relatives and primary contacts only, whereas secondary contacts could also be of interest in this case.

The number and configuration of relationships in social networks has implications on leisure travel, tourism and social capital as discussed in the literature review. In the case of leisure travel, this is mainly because leisure activities are strongly related to friends, relatives and other social contacts with whom the activities are jointly performed (compare Schlich, 2004; Axhausen, 2002; Götz, 2004; ARE/BfS, 2007a). So far there are no research studies available, which examined the growth of leisure travel frequency and distance (as shown in the time series analysis of the micro-census Switzerland, 2000, 2005, 2010, 2015) together with the distance and the number of relationships and the number of visits. As a time series analysis is not possible to build with one survey, nevertheless the analysis should be able to link the variables number of relationships, frequency of contact and distance among network members with leisure travel and distance traveled for meeting social contacts.

Spatially more distant relationships are the goal and motivation for tourism destinations (compare ARE, 2003; WTO, 2004; Bieger and Laesser, 2008; Larsen et al., 2006b). There have been only aggregated statistics available, which show the influence of social relationships on tourism. Even tough not a main goal of this thesis, the spatial distribution of relationships over long distances could shed light into the destination choice question in tourism. For that purpose the detailed home location of social contacts over longer distances is interesting.

Maintaining spatially extensive social networks demands a more complex timely and monetary organization. The exchange of emotional and physical assistance suffers from increasing distance and therefore contact frequency and physical closeness are in a close relationship with mutual assistance (Mok and Wellman, 2007). There have been only a few empirical studies to address the the question of physical distance and emotional closeness. Especially there are no studies in Switzerland about the physical distances within social networks and the influence of income to enable the maintenance of these relationships. Further the emotional support within social contacts and distance is not the main interest in this thesis, but the study should be able to give detailed information about differences of distance
Table 4.1.: Number and configuration of relationships in social networks variables

<table>
<thead>
<tr>
<th>Question content</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of relationships</td>
<td>The number of relationships which are important to leisure activities.</td>
</tr>
<tr>
<td>Distance between home locations</td>
<td>Crow-fly distance</td>
</tr>
<tr>
<td>Frequency of visits</td>
<td>Contact frequency, e.g. twice a week, every Sunday, etc. normalized by year</td>
</tr>
<tr>
<td>Detailed location on the surveyed persons home location (Ego)</td>
<td>Geo-coded street-address</td>
</tr>
<tr>
<td>Detailed location on the surveyed persons contacts home location (Alters)</td>
<td>Geo-coded location, as exact as possible, e.g. street-address, corner, close by landmark etc.</td>
</tr>
</tbody>
</table>

between emotional more important contacts and vice versa, as this is important to understand housing location choices and where to settle down.

Table 4.1 summarizes the questionnaire content from the topic number and configuration of relationships in social networks.

4.2.2. Information and communications technology (ICT)

The information about the structure of social networks as discussed in Section 4.2.1 should be complemented by questions about the frequency of other forms of interaction, such as phone calls, e-mail, SMS, and letter writing. Information about the costs of these interchanges would be useful as well. In most contexts it will not be possible for budgetary reasons to conduct long-duration diaries covering both travel and communication, nor is it likely that we will receive access to the telephone bills and e-mail archives of large numbers of respondents. Answers to these questions will depend on respondents generalizing over longer time periods.

Impacts of information and communication technology on distance among network members, their contact frequency, but also on the social capital have been discussed in the literature. Carrasco (2006) showed that persons with access to internet have spatially more extended social networks. Hogan et al. (2007) showed that
Table 4.2.: Information and communications technology influences on contact frequency and distance

<table>
<thead>
<tr>
<th>Question content</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-face contact</td>
<td>Contact frequency, e.g. twice a week, every Sunday, etc. normalized by year</td>
</tr>
<tr>
<td>Phone contact frequency</td>
<td>&quot;</td>
</tr>
<tr>
<td>Email contact frequency</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

Face-to-face contacts are more frequent, if the persons’ home locations involved in the interaction are no further away than one hour driving distance. But so far detailed studies about the influences of different contact means on each other with distance have not been undertaken. Hampton and Wellman (2001) determine that E-mail contacts are complementary to Telephone contacts, but do not affect Face-to-face visits, but Hogan et al. (2007) shows that with increasing frequency of ICT interactions, the number of Face-to-face visits increases as well.

Table 4.2 summarizes the questionnaire content from the topic Information and communications technology influences on contact frequency and distance.

4.2.3. Mobility biography

The third new element beside the social network structure and the use of technology to maintain the contacts is the mobility biography of the respondent (see Beige and Axhausen, 2006). This retrospective reconstruction of the major stages of the respondent’s life, with an emphasis on aspects relevant to travel, is in itself nothing unusual. Life course and biography research has employed such instruments for some time. The range of questions is obviously open, but the few previous empirical studies hint at the possible mobility biographical influences on social network geography, contact frequencies or social capital. Carrasco (2006) finds that settled persons have in average shorter distant social relationships and have more frequent contact with their social contacts. Viry et al. (2009) state that the further a person commutes, as further their social relationships home locations are from each other. Putnam (2000) goes even further and states that longer distant commutes lead to social anomie. These studies look only at the duration at the
4.2.4. Mobility tools

Since the maintenance of social network requires Face-to-face visits, mobility tools, such as car availability or public transit passes reduce the resistance to overcome
4.2. RESEARCH GAPS, HYPOTHESIS AND QUESTIONNAIRE CONTENT

Table 4.4.: Mobility tool ownership influences on contact frequency and distance

<table>
<thead>
<tr>
<th>Question content</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver license ownership</td>
<td>-</td>
</tr>
<tr>
<td>Car availability</td>
<td>Car availability such as always, often, never etc.</td>
</tr>
<tr>
<td>Transit pass ownership</td>
<td>Transit pass ownership, such as the GA travel-card</td>
</tr>
<tr>
<td></td>
<td>Half-tax, Regional pass etc.</td>
</tr>
</tbody>
</table>

physical distances. [Flamm and Kaufmann (2006)] shows that car owners maintain spatially more extensive social networks. But so far there have been no studies to also include transit pass ownership and other mobility tools to measure their impacts on the distance among network members. There have also been no studies about the frequency of contacts and how they are affected by mobility tool ownership.

Table 4.4 summarizes the questionnaire content from the topic mobility tool ownership influences on contact frequency and distance.

4.2.5. Sociodemographic attributes

The sociodemographic attributes of the interviewee showed influences in the different empirical studies available on the physical distance between members of social networks and also on the contact frequency and the mean of contacts. In the studies of [Carrasco (2006); Larsen et al. (2006b); Tardiff (1977)] income has a positive influence on physical distance between network members, which are travelled to maintain the relationship. Persons living in a relationship, parents and home workers have spatially more compact networks [Carrasco (2006)]. People with lower incomes meet more at home, whereas people with higher income tend to meet more outside of the home [Carrasco (2006)]. Flamm and Kaufmann (2006) show, that persons which are self-employed and full-time employees tend to have spatially more extensive networks compared to students and housewives. Sociodemographic factors on contact mean and frequency are in the study of Carrasco (2006) the income, which is positively correlated with contact frequency (all means), and persons with children living in the same household and women tend to meet less often. But in the study of Boneva et al. (2001) women tend to write more
Table 4.5.: Sociodemographic attribute influences on physical distance and contact frequency in social networks

<table>
<thead>
<tr>
<th>Question content</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic profile</td>
<td>Gender, age, marital status, income, education, etc.</td>
</tr>
<tr>
<td>Migration</td>
<td>Migration background, such as double citizenship, or foreign citizenship</td>
</tr>
<tr>
<td>Work status</td>
<td>Employment status, such as full time or self-employed</td>
</tr>
</tbody>
</table>

frequent E-mails then men. Further, [Wellman et al. (2005)] show that relationships, which emerge from common work or education meet more frequent than relationships with a different origin. Also the migration background is important. [Grieco (1996b)] shows that people with migration background have local contacts, but comparable more non-local contacts as people with no migration background (see also [Carrasco and Miller (2006)]). Besides the few cited studies, which studied the relationship between sociodemographic variables and distance among network members and contact frequencies, there have been no studies so far conducted in Switzerland.

Table 4.5 summarizes the questionnaire content for the sociodemographic attributes on physical distance and contact frequency in social networks.

4.3. Survey design

The additional question items identified above are mostly additions to the person descriptions through the mobility biography and the social-network questions. Name generators and mobility biographies are a substantial challenge as the respondents’ willingness to cooperate is expected to be lower the more intrusive the questions are. The issue of the response burden is another one and has been subject of the literature and will be discussed below in the following sections.
4.3. SURVEY DESIGN

4.3.1. Mobility biographical questions

While it is possible to build on the existing literature, the first transport experiences show the difficulties with social-network and mobility biography survey elements. To capture the mobility biography, Lanzendorf (2003); Beige and Axhausen (2006); Ohnmacht and Axhausen (2005) used a year by question grid in a face-to-face interview, which was filled out by or with support from the interviewer. Beige and Axhausen (2006) employed a grid in a postal survey, following examples in the literature (see Kluge and Kelle, 2001). The quality of the responses was good, but the response rate was low. Larsen et al. (2006b) replaced the grid with independent lists, which seemed to be easier for the respondents, many of whom regarded grids as awkward. Schiffmann (2005) also turned to lists, which he found were well filled in. For this questionnaire a list style was used as well following the previous survey designs. These single lists were collected continuously as retrospective data along major life anchor-points or adjustments over the entire life. Firstly, it concerned important events of the personal and family history, such as the move from the parental home, marriage, divorce, births, deaths, retirements, etc. These events were used as the anchor-points for the memory of the respondent, to facilitate the reconstruction of the life course and to increase the accuracy of the remembered information Brückner (1990). Second, the existing residences, including workplaces including income were indicated. In addition, possession or in possession of mobility tools in the form of the availability of cars and possession of various public transport season tickets were asked. In addition the respondents were asked about their memberships in clubs and the nature of the association.

4.3.2. Name-generator

Problems with name generators are well known: conveying the type of social contact which is of interest to the study to the respondents; the trade-off between the respondents’ burden and the number of contacts retrieved; and the level of detail of the contact descriptions (for a comprehensive review, see Marsden 1990, 2005). The type of contact of interest will vary with the study: some focus on relatives, some on physical neighbours, some on fellow students or employees, and some on
friends. But the studies have the personal network approach in common, which focuses on specific individuals (egos) and their social contacts (alters). The number of contacts a person has is estimated to range into the thousands (see e.g. Pool and Kochen 1979). To define the network boundaries, usually a set of name generating questions is used. Different name generating approaches can be distinguished (Molin et al. 2008), each of them capturing a certain portion of the entire network. For instance, the interaction approach asks for a record of all alters with whom the respondent (ego) interacts in a certain period (including casual and unknown contacts). Another approach is the role relation approach: a record of people with whom the individual has a certain role relationship, such as immediate family, relatives, neighbours or friends. Thirdly, the affective approach asks respondents to record the people with whom they have a close personal relationship, or who are especially important to them. Finally, the exchange approach concentrates on people with whom the individual has social exchange, such as help, social activities or talks about worries. In general, once all alters are elicited, additional questions are used to gather more information on the characteristics each alter and the ego-alter relationship. These additional questions are called name interpreters. Carrasco et al. (2008); Wellman et al. (2005) for example used an affective approach. Respondents were asked to name the persons who live outside their household, with whom they felt very close or somewhat close. Very close people consist of those persons with whom the respondent discusses important matters or regularly keeps in touch with, or are there for them if they need help. Somewhat close people were described as those persons who are more than just casual acquaintances, but not considered to be very close.

None of the approaches above really focuses on those contacts which have an impact on the respondents’ travel. It is apparent that comprehensive, or at least very long, listings can only be retrieved from written records of some kind, including personal sources such as e-mail archives, telephone bills, address books, telephone numbers stored on mobile phones, Christmas mailing lists, and publications lists; and institutional sources, such as yearbooks, membership lists, internal telephone directories, and minutes of meetings. A comprehensive fusion of all relevant administrative sources and diary data has not yet been attempted due to budget and
response burden constraints. Still, the core of the personal networks should be retrievable if Aiello and Dunbar (1993) and Zhou et al. (2005) are correct. They suggest that the core should have five to fifteen members, of which only three to five are the central reference group. The problem, as observed above, is that this central group might not be relevant to some of the travel the analyst is interested in. These journeys are often to contacts which are strong enough to host the respondent, but are not necessarily to members of the core group. A survey strategy which tries to amass a large number of contacts is likely to backfire, as the attendant response burden increases the likelihood of systematic survey nonresponse. Starting from the rich experience of sociology with name generators, a name-generator appropriate to a stimulus was chosen. With a stimulus a certain kind of activity is given, e.g. discussing important matters, for which the interviewee names alteri. With this approach the possible lack of clarity in the understanding of specific social contexts such as friendship can be avoided. Name-generators and name-interpreters together are the network-instrument. The best known and used instruments are Burt’s (Burt, 1984), Fischer’s (Fischer, 1982) and Wellman’s instrument (Wellman, 1979). Burt’s instrument was used for the first time 1985 in the General Social Survey (GSS). A random sample of 1,534 persons was interviewed with the following name-generator: “From time to time, most people discuss important personal matters with other people. Looking back the last six month – that would be back to last August – who are the people with whom you discuss an important personal matter?” (Burt, 1984). Fischer’s instrument was designed for Northern California Community Study (NCCS). 1,050 persons were asked with Fischer’s instrument which uses ten questions appropriated to stimulus 3. Coates and Wellman obtained

3 Fischer’s name generator asks for the first names of those who:
   1. respondents would ask to look after their homes when they go out of town:
   2. had helped with tasks around the house in the previous three months:
   3. they talk with about how they do their jobs (asked only of employed respondents):
   4. they do various social activities with - sharing a meal, visiting, going out socially, etc.;
   5. they get together with to talk about hobbies;
   6. they date seriously or consider fiancé(e) (asked only of unmarried respondents);
   7. they talk with about personal matters of concern;
   8. they rely on for advice about important decisions; and
ego-centric social networks 1968 in East York, Toronto. As Burt they used only one name-generator: “I’d like to ask you a few questions about the people outside your home that you feel closest to; these could be friends, neighbours or relatives” (Wellman 1979). Only the first six of the named alteri were recorded and described in detail. For the research goal here of measuring the size and structure of social network geographies influencing travel behaviour, destination choice or residential choice, it is not so important to survey alters from all different social context spheres but to obtain ego’s most important alteri. Therefore we used an adapted and appropriate set of name-generator as stimuli. The respondents were handed two lists with two different name-generators. The first name generator is adapted from Burt’s and Fischer’s instrument where we asked for persons with whom the respondents “discuss important problems, with whom you stay in regular contact or which you can ask for help”. These questions cover the ten stimuli from Fischer in an abstract way as Burt’s generator does and should cover the “very close” or “most important” core contacts. The second name generator asked for persons with whom the respondents spend and plan leisure time. This generator targets less strong ties against the background that leisure travel makes up the largest share of long distance travel. The name-interpreter asked then for all of the named contacts how they got to know each other, how long the relationship exists, the frequency of contacts by different modes (face-to-face, telephone, email and SMS – short message service via mobile phone), where they met the last time and the contact’s place of residence. The list of modes was left incomplete by design, as we assumed that letter writing has become marginal and that computer-supported chat would follow the patterns of other asynchronous modes (email, SMS). There was also concern about the total response burden. The origin of the acquaintance was categorized as family, subdivided in first degree, relatives or partner, from/of the work, education or partner or others. The frequency of contacts should be specified as accurately as possible e.g. every week, 2 times per year. The contact’s place of residence should be indicated with as much detail as possible (post code,

9. they would or would ask for a sizable loan.

10. Is there anyone who is important to you who doesn’t show up on this list? (Fischer 1982)

4With whom they make plans to spend free time
municipality, street and house number).

The individual questionnaire included general questions about age, gender, nationality, education or employment (nature, scope, location) and for possession of a mobility tools.

The whole questionnaire is available in Appendix A.

4.3.3. Survey protocol design

Given the difficulties in winning the respondents’ cooperation for even simple surveys, the range of questions listed above, which add to the usual, quite numerous questions, raises the problem of respondent load. The rather personal nature of the questions about social networks does not simplify the issue. The face-to-face interview is the best method of establishing the trustworthiness of the interviewer and, by extension, the study. It would be possible to include an incentive to the protocol; however, this might become large enough to raise worries about the motivation of the participants.

At present, it is not clear which survey mode (written, oral by phone, or face-to-face) with which support (motivation call, written explanatory notes and examples, presence of an interviewer) is most appropriate for which theme. A telephone interview could be a robust enough approach for the expanded diary, as the interviewer can explain and motivate the new items. With careful recruitment and initial instruction, a paper or web-based self-completion questionnaire might also be possible. See for example the experience with demanding six-week and twelve-week diaries by Axhausen et al. (2002); Schlich and Axhausen (2003) and Löchl et al. (2005).

Pre-test

The survey is a modular written questionnaire instrument, which is composed of three parts. The individual questionnaire includes general questions about age, gender, nationality, education or employment (nature, scope, location) and the possession of a mobility tools. The main two parts of the questionnaire consisted of
questions on mobility biography and social networks. The questionnaire is enclosed with a cover letter, which explained the survey purpose, questions and protocol. The time required for filling in the questionnaire depends heavily on the occurrence of events in the life of the respondent and the number of social contacts reported based on the stimuli. The duration to complete the questionnaire for different survey protocols is ranging between one hour to two hours.

Due to resource limits and a clear boundary, the study area was limited to the city of Zurich. The respondents were randomly selected, as the response rates are expected to be low, which makes a quota sampling difficult, as variations in the response rate by sociodemographic characteristics are possible. In a random sample one assumes that all units have the same chance of being included in the selection, which should equal the sample to the population. The pre-test was carried out from 1 to 21 December 2005. During the pre-test, three different survey methodologies (self-completion; face-to-face, mixed face-to-face and self-completion) were tested to identify a survey format that would minimise missing values (due to fatigued interviewees) and reduced recall problems for retrospective survey items.

The information in the following tables are based on a survey of 300 complete responses. The ratios which the calculation of the costs are based on are assumptions that must be examined in the pre-test and adjusted as necessary. Following are the expected costs listed for the different protocols, which are independent of the survey instrument:

- Hourly rate for Research assistant: CHF 28.-/h
- Mailing postage: CHF 0.70/letter (same amount for the return postage)
- Printing, Packaging and Shipping: CHF 1.60/questionnaire
- Printing and mailing of “non-response” - postal cards: CHF 1.50/postal card
- Purchase of addresses (up by 10% due to old stock): CHF 0.17/address; 0.23/address including telephone number

The cost estimates from the pre-test are based on the following tentative results. As the size of the pre-test is very small, these cost estimates have to be taken with caution.
4.3. SURVEY DESIGN

**Paper and pencil questionnaire**  For the purely paper and pencil questionnaire the response-rate is estimated to be 15%. The proportion of “non-response” - cards is therefore estimated at 85%.

**Paper and pencil questionnaire - phone call**  For the paper and pencil questionnaire in combination with a motivation call by phone and an incentive, the response-rate is expected to be 25%. The proportion of “non-response” - cards is therefore estimated at 75%.

**Paper and pencil survey - telephonic interview**  The combination of a paper and pencil questionnaire and a telephone interview survey is estimated for the following protocol sequence: The written part of the survey includes the socio-demographic characteristics of respondents and the recruitment for the interview as well as the date, including a reminder call. The telephonic interview includes the mobility biographical questions and the social network questions. For the paper and pencil part in combination with a telephonic reminder a response rate of 45% is expected. The utilization of telephonic interviews is estimated at 80%. The proportion of “non-response” - cards is therefore estimated at 55%.

**Recruiting - "Face to Face" interview - paper and pencil survey**  Following are the assumptions for the combination of a written questionnaires and a face-to-face interview. The interviewees are recruited by phone, and the recruitment rate is expected to be 10%. The written part of the survey includes the socio-demographic characteristics of respondents and their mobility biography. The face-to-face interview refers to an introductory conversation, and then fill out the social network questionnaire together with the interviewer, whereas the interviewee has time afterwards to complete missing items, such as addresses. The interview duration is expected to be 1.5 hours on average. The response rate of the interviews is estimated at 90%, representing a share of non-response cards of 10%.

A summary and overview of the estimated cost of the various survey instruments is in Table 4.6.
Table 4.6.: Expected cost comparison for different survey protocols

<table>
<thead>
<tr>
<th>Costs in [CHF/300 responses]</th>
<th>Paper and Pencil</th>
<th>Motivation Call</th>
<th>Telephone Interview</th>
<th>Face-to-Face</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print and postage</td>
<td>6,000.00</td>
<td>3,600.00</td>
<td>1,600.00</td>
<td>668.00</td>
</tr>
<tr>
<td>Call the interviewee</td>
<td>5,040.00</td>
<td>3,500.00</td>
<td>12,600.00</td>
<td>14,000.00</td>
</tr>
<tr>
<td>Face-to-Face interview</td>
<td>2,240.00</td>
<td>1,680.00</td>
<td>2,240.00</td>
<td>560.00</td>
</tr>
<tr>
<td>Response control and reminders</td>
<td>2,240.00</td>
<td>1,680.00</td>
<td>2,240.00</td>
<td>560.00</td>
</tr>
<tr>
<td>Telephone interview</td>
<td></td>
<td></td>
<td>13,860.00</td>
<td></td>
</tr>
<tr>
<td>Non-response-cards</td>
<td>2,550.00</td>
<td>1,350.00</td>
<td>688.05</td>
<td>50.10</td>
</tr>
<tr>
<td>Incentive</td>
<td>6,000.00</td>
<td>12,000.00</td>
<td>3,000.00</td>
<td></td>
</tr>
<tr>
<td>Data input</td>
<td>2,100.00</td>
<td>840.00</td>
<td>2,100.00</td>
<td></td>
</tr>
<tr>
<td>Address costs</td>
<td>340.00</td>
<td>276.00</td>
<td>191.82</td>
<td>76.82</td>
</tr>
<tr>
<td>Total costs</td>
<td>13,230.00</td>
<td>20,046.00</td>
<td>34,919.87</td>
<td>33,054.92</td>
</tr>
</tbody>
</table>

The total expected costs are between 13,230 CHF and 33,055 CHF based on 300 completed questionnaires. These adds up to costs between 44.10 CHF and 110.18 CHF per completed questionnaire.

Because of the cost estimate for the pre-test, two survey instruments were selected in order to verify the cost estimates and to compare the cost to the quality of the survey answers to decide on the survey protocol for the main study.

Already carried out travel surveys including mobility biographical questions (see Beige and Axhausen [2006]) have shown that due to the complexity of the survey and the intrusive personal questions required of the respondents, a purely paper and pencil survey, is not the best survey protocol, despite the low expected costs.

The computer aided telephone interview survey protocol is not considered for the pre-test because of the missing flexibility of tools such as photo albums, address books or the like (see Ohnmacht [2004]) that support the memory performance compared to a face-to-face interview.

For the pre-test two versions of the survey protocol were combined: Motivational phone call with a paper and pencil survey and a face-to-face interview combined with a motivational recruitment call. The expected findings to be extracted from the pre-test regarding the costs, are divided into two parts:
4.3. SURVEY DESIGN

1. Does the pre-test verify the expected cost and survey measures and if necessary how much they have to be corrected, and

2. it is to examine whether any additional costs of the face-to-face protocol are justified based of the quality difference in the obtained data.

Pre-test results As described in the previous paragraphs, the pre-test selects two different survey instruments: A paper and pencil survey combined with a motivation call and a recruitment call combined with a face-to-face interview.

Pre-test: Recruitment - written questionnaire - face-to-face interview The pre-test survey protocol including a face-to-face interview is divided into three main sub-steps:

1. Recruitment: The recruitment takes place by telephone. The recruitment phone call is simultaneously utilized to arrange the interview date and time.

2. The questionnaire will be sent in advance to give the participants the chance to prepare for the interview and filling in the more simple questions. This also gives the interviewee the chance to prepare possible clarification questions during the interview.

3. The personal standardized interview is conducted with the interviewees and the interviewees are given the choice to return the complete survey to the interviewer or to rework or complete the written questionnaire and then send in the complete survey at a later date.

The recruitment success rate in Table 4.7 is broken down by telephone accessibility and shows the shares reached by phone, the survey response rate and the total of addresses used. In total 399 addresses and telephone numbers were randomly drawn from the Zurich population. 302 persons (75.7%) were successfully contacted within five days between Monday and Friday, with attempts made distributed during 9 am and 6 pm. Of these 302, 35 persons (11.6%) were willing to participate in the survey, which equals a total response rate of 8.8%. All the numbers are presented in Table 4.7.
Table 4.7.: Recruiting success rate for a face-to-face interview

<table>
<thead>
<tr>
<th>Able to contact by phone</th>
<th>Survey participation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number [-]</td>
</tr>
<tr>
<td>No</td>
<td>97</td>
</tr>
<tr>
<td>Yes</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

Pre-test: Paper and pencil questionnaire - motivation call The pre-test of the written questionnaire combined with a motivation call is divided into the following steps:

1. The questionnaire will be sent by postal services to a randomly drawn person.
2. Three days after the surveyed person is expected to have received the questionnaire, the surveyed person is attempted to be contacted by phone to motivate the participation in the survey. A maximum of 5 attempts at different times of day and on various days during the week are made. This intends to increase the willingness to complete the questionnaire. In addition, the respondents should be given the opportunity to ask clarification questions.

The response rate to the postal survey is broken down by successful motivation call attempts. For the written survey 150 people were contacted. Of these, 81 people were successfully contacted by telephone within the maximal five call attempts. 9 persons out of the 81 persons reached with the motivation call completed the survey (11.1%). Out of the 69 people not reached with the phone, only 3 people returned the questionnaire (4.3%). This sums up to a total response rate of 8.0%.

Revised cost calculation The cost calculation is revised based on the pre-test survey result. The cost compared to the quality of the responses builds the base to chose the survey protocol for the main study. In addition, the breakdown of the cost may be useful for future surveys.

The types of costs are classified in survey instrument-dependent and-independent items. This results in the following outline:
4.3. SURVEY DESIGN

Independent survey instrument costs:
- Hourly rate for Research assistant: CHF 28.-/h
- Mailing postage: CHF 0.70/Brief (same amount for the return postage)
- Printing, Packaging and Shipping: CHF 1.60/questionnaire
- Printing and mailing of “non-response” - postal cards: CHF 1.50/postal card
- Purchase of addresses (up by 10% due to old stock): CHF 0.17/address; 0.23/address including telephone number

Survey instrument dependent measures:
- Response rate [%]
- Entrance control and reminders [CHF/questionnaire]
- Data Entry [h/questionnaire]
- Reached by telephone [%]
- Cost per completed call [h/completed call]
- Recruitment rate [%]
- Cost per interview [h/interview]

Table 4.8 shows the difference between the expected costs and the resulting costs from the pre-test.

The change in the cost parameters on the overall cost is summarized in Table 4.9.

The total estimated costs are based on a desired number of 300 fully completed questionnaires.

Data quality Data quality in relation to a survey refers to the quality, i.e. importance, relevance and accuracy of the information contained in the answers of the respondents.

The quality of data from a survey can be measured in terms of properly completed questions to the whole number of questions. An important factor is the high temporal effort required of the respondents by this particular survey as well as the
### CHAPTER 4. FIELDWORK

Table 4.8.: Key figures of the pre-test and pre-estimates

<table>
<thead>
<tr>
<th>Key figures</th>
<th>Face-to-face</th>
<th>Paper and pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Pre-test</td>
</tr>
<tr>
<td>Response rate [%]</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>Response control and reminders [CHF/questionnaire]</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Data Entry [h/questionnaire]</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Reached by telephone [%]</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Cost per completed call [h/completed call]</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td>Recruitment rate [%]</td>
<td>10</td>
<td>8.77</td>
</tr>
<tr>
<td>Cost per interview [h/interview]</td>
<td>1.5</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Table 4.9.: Costing structure of the pre-test

<table>
<thead>
<tr>
<th>Costing unit</th>
<th>Paper and pencil</th>
<th>Face-to-face</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print and postage</td>
<td>4,500.00</td>
<td>947.37</td>
</tr>
<tr>
<td>Response control and reminders</td>
<td>1,215.00</td>
<td>255.79</td>
</tr>
<tr>
<td>Non-response cards</td>
<td>1,800.00</td>
<td>23.68</td>
</tr>
<tr>
<td>Data entry</td>
<td>2,025.00</td>
<td>2,025.00</td>
</tr>
<tr>
<td>Address costs</td>
<td>345.00</td>
<td>968.42</td>
</tr>
<tr>
<td>Call the interviewee</td>
<td>4,556.25</td>
<td>16,200.00</td>
</tr>
<tr>
<td>Interviews</td>
<td>12,150.00</td>
<td>20,935.01</td>
</tr>
<tr>
<td>Incentives of 5 CHF</td>
<td>1,500.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14,441.25</td>
<td>32,570.26</td>
</tr>
<tr>
<td>Cost per interview</td>
<td>48.14</td>
<td>108.57</td>
</tr>
</tbody>
</table>

70
intrusive private nature of the questions. It can therefore be assumed that the willingness to complete the questionnaire accurately decreases with time. Therefore, the quality of the responses due to the number of social contacts named and the accuracy of the information given with each of the named contacts can be used as a proxy for measuring the data quality. Following the data quality of the two survey protocols used in the pre-test is measured according to the number of named contacts and the alter’s home address accuracy.

In the paper and pencil survey were in average 23 contacts named. In the face-to-face interview in average 22. The median named contacts in the paper and pencil survey is 12 and in the personal interview 16. This is due to fact that in the written survey, a single person used all the possible fields (40) to name contacts. The spread of responses is lower than in the personal interview with the written survey.

The quality of filling in the questionnaires between the two methods seems not to be large, although there are advantages of personal interviews. The personal interview also offers advantages in terms of the quality of responses. Thus, certain questions, such as for example the question about the frequency of meetings can be open and general in a way that the interviewee can specify it in own words, while it must be specified as individual cohorts in the written survey, so the respondent has to tick the right categorie, in order to obtain comparable answers.

The accuracy of the named alter’s home location is measured by the rate of complete street addresses compared to less accurate information such as street intersection, corner, landmark, etc. For the paper and pencil protocol 64% of the alter’s home location were given as street address, whereas for the face-to-face interview 79% of the alter’s home locations were completed. The difference of 15% is mostly due to the motivation in the interview to look up the addresses in an address book or directory. Although the accuracy of the addresses can not be verified, the interview builds confidence and trust on both sides of the interview process.

**Conclusions from the pre-tests** Concluding on the pre-test, the following observations can be made to decide on the survey protocol for the main study:
Advantages and disadvantages of the paper and pencil protocol compared to the face-to-face interview protocol. The solely written survey compared to a personal interview was expected to have an especially strong cost advantage. The pre-test results corrected this perception due to the low response rate leaving the cost comparison in the pre-test close at CHF 104 per questionnaire for the paper and pencil version compared to CHF 129 per questionnaire for the face-to-face interview protocol. Other benefits for the solely written protocols are that no interference can take place by the interviewer, and the interviewee has a free choice of response time, which is especially noticeable in hard to reach people.

The disadvantages of the written survey are the very low response, which is noticeable in the cost comparison. Moreover, such a low response rate biases in the sample are expected to affect the representativeness of the survey. Another disadvantage of the written survey compared to a personal interview is the restriction to relatively simple questions to ensure comparability. In addition, written surveys have a relatively slow response, and the lack of control options with regard to the seriousness, response duration, sequence, information retrieval, question understanding and the influence of third parties in completing the survey.

Disadvantages of the personal interview protocol are the relative high costs. Other possible problems are the interviewer bias, and the problem of contacting hard to reach respondents.

The advantages of personal interviews are the high representativeness and the high completion rate (no terminations). It is possible to respond flexible to respondents, misunderstandings can be clarified and the interviewer can control the response behaviour to a certain degree. Moreover, the interview reduces the influence by third parties, and once the interview date is set, quick responses can be expected.

Summarizing it can be said that the added costs of a personal interviewing protocol can be compensated by its benefits. Especially the very low response rates in the paper and pencil protocol show that the intrusive questions about social contacts makes it necessary to build a trusting relationship between the interviewer and the interviewed person to collect reliable data. Moreover, given the low response rate of the solely written protocol, it is very difficult to obtain a representative sample. An increase in the response rate in return to the slightly higher costs are
therefore very desirable. The mixed method was the most effective approach with an acceptable cost per response. In this pre-test experiment we found that the additional cost of the face-to-face interview was offset by the better response rate in comparison to a paper-instrument plus motivation-call survey.

4.4. Main study

The main study adopted the mixed face-to-face protocol described in the pre-test. For the survey, 4,200 Zurich residents with available addresses and telephone numbers were chosen randomly. Following an announcement letter, the subjects were contacted on different days of the week and times of day, and then recruited during the telephone interview, including arranging appointments for the face-to-face interviews. With the reminder notice for the interview, respondents received the written form allowing them to raise questions during the upcoming interview. The written part consists of a person form and a form with mobility biographical questions about relocations, former and current job locations, usage of mobility tools, important life events and memberships in groups that meet periodically (Beige and Axhausen, 2006, for detailed information about mobility biographies). The one to two hours long face-to-face interview covered the social contact questionnaire, but was also used to detect and address respondent difficulties and to establish rapport with the respondent.

4.4.1. Response rate in the main study

The main study goal was to collect 300 complete surveys. From a population of about 100,000 purchasable addresses representing the population of persons over 18 years old living in the city of Zurich, a gross sample of 4,200 persons was drawn. The net sample size contained 2,714 people (64.4%), after deducting the households not reachable by telephone within five attempts and false addresses. The recruitment was based on a pure probability sampling. Finale, of these 332 persons were recruited, of whom 307 (11.3%) were interviewed and completed the
Table 4.10.: Response rate of the survey on social networks and mobility biography

<table>
<thead>
<tr>
<th>Phase</th>
<th>Pre-test</th>
<th>Main study</th>
<th>Share [%]</th>
<th>Reached by phone [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>150</td>
<td>4,200</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Wrong addresses</td>
<td>0</td>
<td>56</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Not reached by phone</td>
<td>36</td>
<td>1,486</td>
<td>34</td>
<td>-</td>
</tr>
<tr>
<td>within five attempts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reached by phone</td>
<td>114</td>
<td>2,714</td>
<td>64</td>
<td>100</td>
</tr>
<tr>
<td>Recruited</td>
<td>14</td>
<td>318</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Interviewed</td>
<td>13</td>
<td>310</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Complete survey</td>
<td>13</td>
<td>307</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

questionnaire. The interviewees received no incentive. See Table 4.10 for more details.

Due to the high response burden, the response rate is acceptable and within expectations (see Axhausen and Weis, 2009). Figure 4.4.1 shows a clear, nearly linear relationship between the response burden and the response rate in a recent set of surveys at the Institute for Transport Planning and Systems at ETH in Zurich. The ex-ante estimated response burden of the written-only format of the pre-test can be found in lower right corner of Figure 4.4.1. The comparable survey in Toronto mentioned above was able to reach a response rate of about 50 % with an incentive of 50 CAN$ (Carrasco et al., 2007), but drawing on known responders from the first wave of the overall study.

4.4.2. Representativeness of the data collection

Table 4.11 shows the socio demographic characteristics of the respondents in comparison with the Zurich population, as observed in the Swiss Microcensus Travel 2005, its representative national travel diary survey (ARE/BfS, 2007b) and the Swiss Census 2000 (BfS, 2000). The income information is not directly comparable because the Microcensus and the Census measures the household income while this study is person-based. The comparison shows that the Zurich sample is a little bit older, slightly better educated and has a higher share of public transport.
Figure 4.4.1.: Response rate and ex-ante assessment of response burden

Source: Axhausen (2007), Figure 1.
season tickets. The mean age for the respondents with a university degree is with 43.61 years below the population mean and for those with an obligatory schooling higher with 56.77 years respectively 59.66 years for those with a vocational training. Respondents with no public transport season tickets are overrepresented in the group of people with vocational training and a university degree. Based on the sample size, the deviation in the sample from the population is within the expected range. For e.g. the share of male people in the sample, the standard deviation is calculated as $\sigma = \sqrt{1 - \frac{n}{N}} \sqrt{\frac{p_m(1-p_m)}{n-1}} = \sqrt{1 - \frac{307}{347,517}} \sqrt{\frac{0.436(1-0.436)}{307}} = 0.0285$. For a confidence interval of 95% the $t - value$ is 1.96 and therefore the confidence interval from the sample is 43.6% $\pm$ 5.54, which is within the population mean. Overall a re-weighting of the data seems not necessary given the relatively small deviations.

It is also likely that the sample from a total of around 300 individuals provides enough variance for the subsequent multivariate analyses. Because the goal of the analysis is primarily to verify causal hypotheses, the extrapolation of the values of individual indicators on Zurich’s whole population is of secondary importance.

### 4.4.3. Quality measurement and data preparation

The main part of the survey are the questions about the social network. The answer quality should therefore be measured against the received information in terms of relevance and accuracy. The number of properly complete given information about the egocentric social network and it’s accuracy can be used to measure the data quality.

**Answer validation: Number of relationships**

As it is not possible to compare the surveyed egocentric social network data to official data, it requires special care, as it is a key indicator of survey quality. Therefore the number of the contacts obtained are analysed in detail and compared with the

---

$^5$Where $N = \text{Population } (347,517 \text{ in Zurich for the year 2005, see }), n = \text{samplesize } (307)$  $p_m = \text{share of males in the sample } (43.6\%)$
### 4.4. MAIN STUDY

Table 4.11.: Socio-demographic comparison between the characteristics of the Zürich respondents and the Zürich population

<table>
<thead>
<tr>
<th>Variable</th>
<th>Survey Mean</th>
<th>Population Mean</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>50.76</td>
<td>46.76</td>
<td>8.50%</td>
</tr>
<tr>
<td>Males</td>
<td>43.60%</td>
<td>47.90%</td>
<td>-4.30%</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>5.20%</td>
<td>12.50%</td>
<td>7.30%</td>
</tr>
<tr>
<td>Obligatory schooling</td>
<td>8.00%</td>
<td>19.20%</td>
<td>-11.20%</td>
</tr>
<tr>
<td>Vocational training</td>
<td>31.80%</td>
<td>31.30%</td>
<td>-0.50%</td>
</tr>
<tr>
<td>Highschool diploma</td>
<td>8.30%</td>
<td>9.20%</td>
<td>-0.90%</td>
</tr>
<tr>
<td>Further technical training</td>
<td>20.80%</td>
<td>10.70%</td>
<td>10.10%</td>
</tr>
<tr>
<td>University degree</td>
<td>26.00%</td>
<td>17.10%</td>
<td>8.90%</td>
</tr>
<tr>
<td><strong>Car available</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>44.60%</td>
<td>42.80%</td>
<td>1.80%</td>
</tr>
<tr>
<td>Frequently and rarely</td>
<td>17.00%</td>
<td>18.40%</td>
<td>1.40%</td>
</tr>
<tr>
<td><strong>Public transport season tickets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% discount card (Halbtax)</td>
<td>49.50%</td>
<td>37.90%</td>
<td>11.60%</td>
</tr>
<tr>
<td>National season (GA)</td>
<td>24.60%</td>
<td>14.20%</td>
<td>10.40%</td>
</tr>
<tr>
<td>Regional season</td>
<td>13.80%</td>
<td>18.70%</td>
<td>-4.90%</td>
</tr>
<tr>
<td><strong>Personal income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>12.80%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0-1999</td>
<td>13.80%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2000-5999</td>
<td>46.40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6000</td>
<td>27.00%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
results of the literature. To get a clear picture when surveying social network size, it is necessary to review the methodology of information gathering from respondents. The survey focuses on ego-centric (personal) networks, meaning that the respondent (ego) are the core of the social network and specifically capture his or her social contacts (alters). To survey ego-centric networks, name-generators and name-interpreters are used. The name generator specifies the type of relationship that the survey wants the respondent, the ego, to list. Often, researchers set an arbitrary maximum number of contacts to be listed (see for example [Diaz-Bone, 1997]). The name-generator defines the ego-centric network and is the basis for further analysis. The name-interpreters then pose further questions to detail the description of the contact, alter, e.g. socio-demographic data or characteristics of the relationship.

A first point of discussion concerns answer validation. Respondents were able to name a total of 17 relationships with the first name-generator and 32 additional non-overlapping relationships with the second, producing a possible total of 49 alters. In fact, the lists could have been extended if necessary. The range of named relationships is 1 to 49. The maximum number of reported relationships was reached once and the mean was 12.35 relationships. Compared to the possible number of 49 relationships, the exhaustion rate (mean relationship divided by the possible number of 49 relationship) is 25.2%, indicating that respondents had sufficient possibilities to cite relationships.

A comparison of the instrument used with the best established instruments to survey social networks is given in Table 4.12. A single name-generator causes small ego-centric social networks. The sizes of the networks surveyed with the Fisher-instrument are remarkably larger than those with the two other instruments. The instrument in this study (Zurich-instrument) uses name-generator questions with four prompts and is in between the Fischer and the two single prompt name-generators. The other instruments report higher shares of relatives and proportionally higher shares of strong ties. The share of relatives seems to be an indicator for the social network size as the share of weak ties decreases with them, which is visible in the comparison with the other surveys, except that the instrument used seems to be biased against reporting family relationships, which is
### 4.4. MAIN STUDY

Table 4.12.: Comparison of different name-generators

<table>
<thead>
<tr>
<th>Variable</th>
<th>East York</th>
<th>NCCS</th>
<th>GSS</th>
<th>Zurich</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrument</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name-generator</td>
<td>I’d like to ask you a few questions about the people outside your home that you feel closest to.</td>
<td>11 prompts</td>
<td>Looking back over the last six months, who are the people with whom you discuss an important personal matter?</td>
<td>4 prompts</td>
</tr>
<tr>
<td>Generator limitation</td>
<td>6</td>
<td>No limitation</td>
<td>5</td>
<td>No limitation (17/32)</td>
</tr>
<tr>
<td><strong>Ego-centric network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size ($\hat{O}$)</td>
<td>4.70</td>
<td>18.48</td>
<td>3.01</td>
<td>12.35</td>
</tr>
<tr>
<td>Share of relatives ($\hat{O}$)</td>
<td>0.50</td>
<td>0.44</td>
<td>0.61</td>
<td>0.31</td>
</tr>
<tr>
<td>Share of weaker ties ($\hat{O}$)</td>
<td>0.18</td>
<td>0.32</td>
<td>0.23</td>
<td>0.48</td>
</tr>
<tr>
<td>Duration of the relation in years ($\hat{O}$)</td>
<td>&gt;10 for 57% of the ties</td>
<td>16</td>
<td>(6 years and more, was the biggest class)</td>
<td>20.6</td>
</tr>
<tr>
<td>Contact frequency per year ($\hat{O}$)</td>
<td>150.4</td>
<td>-</td>
<td>194.6</td>
<td>59.0</td>
</tr>
</tbody>
</table>

Source: Adapted from Diaz-Bone (1997) p. 78 and own data

probably caused by the generic formulation “persons with whom the respondents spend leisure time” for weaker ties. The duration of the relationships cannot be compared across the instruments because of different measurement approaches. It seems that if this question is asked in classes, the classes should have a wider range than just 10 years, especially if we expect a high share of relatives. The mean of the contact frequency per year is strongly dependent on the share of relationships which meet daily. The small share of weaker ties in the small networks causes the big differences between the two single name generators and the Zurich-instrument.
Address accuracy and geocoding

Accuracy The accuracy of the named alter’s home location is measured by the rate of complete street addresses compared to less accurate information such as street intersection, corner, landmark, etc. Even tough the interviewees were asked to give information of their named alter’s as accurate as possible, not all alter’s home locations can be identified down to the street-address. This has several reasons, but mainly because of lack of memory and privacy concern. The first reason was tried to counter by encourage the respondents to prepare for the interview by having personal sources such as e-mail archives, telephone bills, address books, telephone numbers stored on mobile phones and Christmas mailing lists at hand during the interview. Further the respondents were encouraged to complete missing information after the interview if possible, and then sending in the completed survey. The privacy concern was addressed during recruiting, in the introduction letter sent with the questionnaire and the interview itself. It was ensured, that all the given information will only be used for anonymous statistical analysis, and that the address data would not be kept electronically nor physically. Nevertheless, about 12% of the interviewees refused to reveal the exact street-addresses, but gave the most exact geocodable description of the alter’s home location, so that they felt still comfortable giving the information away. As acceptable entries were used in the following order:

- the next crossing street on the street of the alter’s home location (incl. zip code, city, country),
- identifiable landmarks (incl. zip code, city, country),
- neighbourhood (incl. zip code, city, country),
- village (incl. zip code, city, country) and
- city (incl. zip code, city, country).

In Table 4.13 are the corresponding numbers of received information for alter’s home location listed and split up by accuracy. The information accuracy seems reasonable although slightly worse than in the pre-test.
4.4. MAIN STUDY

Table 4.13.: Response information accuracy of alter’s home location

<table>
<thead>
<tr>
<th>Accuracy measurement</th>
<th>Number of addresses [-]</th>
<th>Share [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postal address</td>
<td>2601</td>
<td>68.6</td>
</tr>
<tr>
<td>Next crossing street</td>
<td>315</td>
<td>8.3</td>
</tr>
<tr>
<td>Identifiable landmark</td>
<td>610</td>
<td>16.1</td>
</tr>
<tr>
<td>Neighbourhood</td>
<td>118</td>
<td>3.1</td>
</tr>
<tr>
<td>Village</td>
<td>106</td>
<td>2.8</td>
</tr>
<tr>
<td>City</td>
<td>42</td>
<td>1.1</td>
</tr>
<tr>
<td>Total</td>
<td>3,792</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Geocoding The effort involved in precise geocoding was before the arrival of basically free web-based services, such as Google, MSN and Yahoo expensive and very time consuming. Georeferenced place-name or address data bases are expensive and huge, and do not allow for mistakes in the postal addresses. Further were those systems not intelligent enough to detect landmarks and descriptions, like the algorithm used by Google and others. So it is understandable that researchers have avoided the geo-referencing so far and used self reported distances.

Google released Google Maps (http://maps.google.com) in 2005 together with Google Earth (http://earth.google.com/), a standalone desktop application that offers anyone with internet access to geographic information.

Google Maps API (http://www.google.com/apis/maps/documentation/) lets developers embed Google Maps in their own Web pages or access their service through http requests. The Google Maps API was used to geocode intersection, postal addresses, identifiable landmarks, neighbourhoods, villages and cities. Google Maps uses an algorithm to allow the matching of locational text to locations on the street network. Google Maps allows users to geocode address lists using Google’s matching algorithm and a current TeleAtlas-based street network. However, Google Maps does not permit a high of level of customization, so have e.g. umlauts and other special characters to be replaced.

A total of 3,972 locations were geocoded. Overall, 3,614 (91%) were successfully automatically geocoded within eight runs of re-correcting addresses, which were not recognized. The remaining 9% had to be corrected to next higher level of available information, such as e.g. zip code or city. Figure 4.4.2 shows that contacts are
Figure 4.4.2.: Geo-coded home addresses of ego’s and alter’s home location

...to a large proportion closely located to the survey area Zurich. Other are spread numerous across Switzerland. In addition, the contacts are distributed worldwide with a higher density in Central Europe and in the wider European area, as well as in North America (an analysis differentiating by country is made in the following chapters).
5. Empirical Data Analysis

The previous chapter presented the possible influences on different endogenous attributes a social network, and the developed data collection methodology as well the field work executed. The purpose of this chapter is to analyse the collected data and to make evidence-based statements about the previous chapters discussed theories and developed hypotheses.

In the following section, the central empirical findings on mobility and social networks are presented. First, an analysis is made for the size of the ego-centered network (degree). Following that, the analytical focus on the spatial properties and arrangement of social relations. Then the influence of the mobility biographical variables on social network attributes is measured, followed by the analysis of contact frequencies in comparison with distance between members of social networks. Also the contact frequencies of different contact means is analysed and how they interact with each other. In general the analyses of the different social network attributes are structured that first a descriptive bivariate analysis is applied followed by a multivariate analysis using different statistical models.

5.1. Number of social relationships

This section is with permission of the co-authors an updated and extended version of parts of Frei (2007a); Frei et al. (2009).

A first point of discussion is the number of relationships named in the survey. As discussed in Section 4.4.3 the number of social relationship, also called degree, is highly dependent on the used name generator, and therefore absolute numbers do not have much of a meaning, unless compared within the name generator used.
Table 5.1.: Sample size and mean number of relationships in ego-centric social networks from different studies

<table>
<thead>
<tr>
<th></th>
<th>NCCS(1)</th>
<th>Zuma(2)</th>
<th>JV(3)</th>
<th>MCS(4)</th>
<th>Haifa(5)</th>
<th>Toronto(6)</th>
<th>Zurich(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>1,050</td>
<td>239</td>
<td>98</td>
<td>99</td>
<td>262</td>
<td>87</td>
<td>307</td>
</tr>
<tr>
<td>Mean degree</td>
<td>18.5</td>
<td>7.8</td>
<td>21.8</td>
<td>10.1</td>
<td>14.0</td>
<td>11.7</td>
<td>12.4</td>
</tr>
</tbody>
</table>

(1) North California Community Study (Fischer, 1982)
(2) Zuma-Study (Pfenning and Pfenning, 1987)
(3) Jacksonville Study (Bernard et al., 1990)
(4) Mexico City Study (Bernard et al., 1990)
(5) Haifa Study (Fischer and Shavit, 1995)
(6) Toronto Study (Hogan et al., 2007; Carrasco et al., 2008)
(7) Zurich Study (Frei, 2007b)

But as a first analysis, it has to be made sure, that the used survey instrument does not bias the number of social relationships in the collected data. A first point of discussion concerns answer validation. Respondents were able to name a total of 17 relationships with the first name-generator and 32 additional non-overlapping relationships with the second, producing a possible total of 49 alters. In fact, the lists could have been extended if necessary. The range of named relationships is 1 to 49. The maximum number of reported relationships was reached once and the mean was 12.35 relationships. Compared to the possible number of 49 relationships, the exhaustion rate (mean relationship divided by the possible number of 49 relationship) is 25.2%, indicating that respondents had sufficient possibilities to cite relationships. The number of relationships is within the expected region of other studies as shown in Table 5.1.

5.1.1. Bivariate degree analysis

Figure 5.1.2a shows the distribution of the number of relationships. The distribution is left skewed and has a variance of 73.0, compared to the mean of 12.35. The share of important relationships is 52% and drops, as expected, with an increasing
5.1. NUMBER OF SOCIAL RELATIONSHIPS

Figure 5.1.1.: Distribution of the number of relationships and the share of core contacts

(a) Degree density distribution  (b) Share of core contacts against degree

number of reported relationships (see Figure 5.1.2b). As shown in Figure 5.1.2b, the share of strong ties drops with the numbers of relationships reported, which is as well visible in the other surveys (see 4.12), except that the survey instrument used for the data collection seems to be biased against reporting family relationships, which is probably caused by the second name generator, to ask especially for relationships, with which leisure time is spent for weak ties.

The number of social relationships reported and sociodemographic attributes of the respondents is compared in Table 5.2. Age seems to make a rather large difference, with the younger people cultivating more relationships than the older ones. The share of the important relationships seems to grow with the age but has a small drop in the oldest age class. Gender seems to have no influence on the numbers of relationships. The education level indicates that a higher level of education increases the number of relationships, but there is no clear trend visible as the differences are rather small. The share of important relationships is slightly higher for persons with higher levels of education. There seems to be no income dependence. This is surprising at first sight, since education and income are often highly correlated. In the case of this dataset they are significantly positive correlated ($\chi^2$-test significant at 0.05-level at Cramer’s $V = 0.15$, also significant at 0.05-level). Reason for the different effects of education and income may be the
relatively low contingency between the two status variables on the social network degree. The education level is as opposed to income has a positive significant effect on the degree of the network, but the strength relatively low.

5.1.2. Multivariate degree analysis

To capture the wide range in the number of social relationships in further analysis it necessary to determine the probability distribution which represents the data set. Figure 5.1.2 shows that the data follows a left skewed bell shaped curve. A possibility to represent this shape is the discrete Poisson distribution. The variance for the number of relationships is with 73.0 much bigger than the mean of 12.3. This is an indicator for over-dispersion. As the Poisson distribution is described only by one parameter, it is symmetric. To deal with the skew of the number of relationships a negative binomial distribution can be used instead of a Poisson distribution. To obtain a distribution that fits the six persons reporting very high numbers (above 30 relationships) were removed as potential outliers. Figure 5.1.2 compares the number of relationships reported, and best-fit negative binomial distribution and the Poisson distributions. Figure 5.1.2 compares the number of relationships reported, and best-fit negative binomial distribution and the Poisson distribution.

Sociodemographic, travel related, biographical and survey-specific dummy variables are employed to explain the number of social contacts using a negative binomial regression. After removing variables which correlate highly with each other (limit = 0.5; e.g. working status and place of work), variables with a significance level lower than 0.05 were removed stepwise. The parameter estimates are reported in Table 5.3.

In the interpretation of significant effects below influence does not mean direct causality. The question of cause and effect can not be answered based on a multivariate analysis or the data available. The effects are considered to be a reciprocal. This is especially true with regard to transit availability and use. Because the availability and use of transport means has an influence on maintaining a social network and at the same time it's use is caused by participating in activities, where here is
### 5.1. NUMBER OF SOCIAL RELATIONSHIPS

Table 5.2.: Number of relationships by socio-demographic characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>Mean</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Important</td>
<td>All</td>
<td>Important</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up to 30</td>
<td>6</td>
<td>12.5</td>
<td>7.1</td>
</tr>
<tr>
<td>30 to 40</td>
<td>5</td>
<td>14</td>
<td>6.8</td>
</tr>
<tr>
<td>40 to 60</td>
<td>5</td>
<td>10</td>
<td>6.7</td>
</tr>
<tr>
<td>60 and older</td>
<td>5</td>
<td>9</td>
<td>5.6</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>11</td>
<td>6.4</td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>11</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>5</td>
<td>11</td>
<td>4.9</td>
</tr>
<tr>
<td>Obligatory schooling</td>
<td>5</td>
<td>8</td>
<td>5.5</td>
</tr>
<tr>
<td>Vocational training</td>
<td>5</td>
<td>11</td>
<td>6.1</td>
</tr>
<tr>
<td>High-school diploma</td>
<td>5</td>
<td>12</td>
<td>6.5</td>
</tr>
<tr>
<td>Further technical training</td>
<td>5</td>
<td>8.5</td>
<td>6.1</td>
</tr>
<tr>
<td>University degree</td>
<td>5</td>
<td>13</td>
<td>7.2</td>
</tr>
<tr>
<td><strong>Income [CHF/month]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 1,999</td>
<td>5</td>
<td>10</td>
<td>6.0</td>
</tr>
<tr>
<td>2,000 to 5,999</td>
<td>5</td>
<td>12.5</td>
<td>5.9</td>
</tr>
<tr>
<td>6,000 +</td>
<td>5</td>
<td>11</td>
<td>6.6</td>
</tr>
<tr>
<td>All</td>
<td>5</td>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>
Figure 5.1.2.: Frequency comparison plot of the degree distribution and the best-fit Poisson and negative binomial distribution

Best-fit parameters of the negative binomial distribution: 4, Probability: 0.26, $\chi^2$-test: 13.53 (Interval width: 2; 14 degrees of freedom) (Estimated with ExpertFit - Version 7.00 - Averill M. Law & Associates [2006]).
5.1. NUMBER OF SOCIAL RELATIONSHIPS

Table 5.3.: Parameter estimates for the negative binomial regression of the degree of ego-centric social networks

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St.dev.</th>
<th>Beta</th>
<th>St. err.</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.092</td>
<td>10.169</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age [years]</td>
<td>53.283</td>
<td>19.163</td>
<td>-0.040</td>
<td>-3.113</td>
<td>0.002</td>
</tr>
<tr>
<td>Age²/1,000 [years²/1,000]</td>
<td>3.208</td>
<td>2.081</td>
<td>0.352</td>
<td>2.806</td>
<td>0.005</td>
</tr>
<tr>
<td>Annual monthly public transit ticket [y/n]</td>
<td>0.853</td>
<td>0.893</td>
<td>0.242</td>
<td>2.036</td>
<td>0.042</td>
</tr>
<tr>
<td>Number of relocations [-]</td>
<td>5.963</td>
<td>3.116</td>
<td>0.038</td>
<td>3.023</td>
<td>0.003</td>
</tr>
<tr>
<td>University degree [y/n]</td>
<td>0.247</td>
<td>0.430</td>
<td>0.178</td>
<td>1.921</td>
<td>0.055</td>
</tr>
<tr>
<td>Part time employed [y/n]</td>
<td>0.170</td>
<td>0.382</td>
<td>-0.256</td>
<td>-2.321</td>
<td>0.020</td>
</tr>
<tr>
<td>Retiree [y/n]</td>
<td>0.327</td>
<td>0.469</td>
<td>-0.302</td>
<td>-2.003</td>
<td>0.045</td>
</tr>
<tr>
<td>Children &lt; 18 living in household [y/n]</td>
<td>0.250</td>
<td>0.434</td>
<td>0.177</td>
<td>2.307</td>
<td>0.021</td>
</tr>
</tbody>
</table>

N 300

Adjusted $R^2$ 0.13

assumed that social relationships are a main driver for participating in activities.

The age of the respondents shows a U-shaped influence. Younger people maintain many contacts and then the number declines with increasing age, whereas every additional year causes a lower decrease of the number of social relationships. The ownership of public annual or monthly transport ticket has a positive influence on the number of social relationships. Maintaining a bigger social network seems to be influenced by the ownership of mobility tools, but only the annual or monthly subscription to public transport tickets was highly significant. The number of relocations influences the number of relationships. The positive influence of the number of relocations indicates that people keep their important friendships after relocation even at a certain distance. By building up a social network at the new location and keeping in touch with “old” friends, numbers of social contacts increase. A higher education, at least a university degree, leads to a larger number of social contacts. But there is no clear trend visible between the number of relationships and education. A rather high influence on the numbers of social relationships has the working status. In 27.3% of cases the original context of the acquaintance is work (41.0% friends, 25.9% family, 4.9% partner and 0.9% others)
and therefore is the second most frequent original context. The big influence of the work status is therefore not surprising. Part time employees and retired people have fewer social relationships than full-time employees and equivalents e.g. students. It is surprising, that children have a positive influence on the amount of social contacts, because it would be expected that the additional workload for parents would decrease the number of social relationship, but children open up the possibility to meet new possible contacts e.g. other parents with small children, parent-teacher conferences etc., which here outweighs the first effect.

5.2. Spatiality of social networks

This section is with permission of the co-authors an updated and extended version of parts of Ohnmacht et al. (2008); Frei et al. (2009, 2010).

In Figure 5.2.1 are 3,792 geocoded contacts according to their distribution to individual countries listed. Overall, 92% of the relationships are settling in Switzerland. This is surprising, as Zurich is an international city. The most common interconnections of respondents outside of Switzerland live in neighbouring Germany. A total of three percent of these contacts can be found there. This is followed by Great Britain, the United States of America and France, in each of which live between one half and one percent of the collected relationship’s home location. Less than half a percent of the responses locate in the following countries (arranged by descending absolute number of observation): Austria, Japan, Italy, Netherlands, Portugal, Canada, Spain, Russia, Romania, Poland, Peru, Norway, New Zealand, Croatia, Greece, Turkey; Czech Republic, Thailand, Pakistan, Monaco, Lichtenstein, Kenya, Iran, Denmark, Brazil and Australia.

5.2.1. Measuring and analysing the geography of social networks

The most obvious way to measure the spatial dispersion of the egos relationships is to measure the distance between their home locations. The addresses were
5.2. SPATIALITY OF SOCIAL NETWORKS

Figure 5.2.1: Share of relationships by country (n=3,792)
georeferenced and the distances between the home addresses were calculated as great circle distances using equidistant cylindrical projection to account for the shape of the earth.

**Distance properties**

The distribution of the great circle distances (Figure 5.2.2) between the respondents’ residence and their relationships has three elements. Nearly two-thirds of the alters live locally within 25 km. The bulk of the remaining distances are divided into regional and national relationships (within 26-100 km 13%) and international relationships in Europe (within 101-1,000 km, 15%). A noticeable share of inter-continental links makes up the rest of 3% (Figure 5.2.3d). The peak at about 10,000 km marks the intercontinental distance between Zurich and the USA. The respondents mix the local/regional contacts of daily life with a multitude of non-local and often long distance contacts.

Attachment to personally more important contacts Figure 5.2.3b shows that people are not only emotionally closer attached to personally more important contacts, but also spatially.

The distance has a strong impact (note the log scale! in Figure 5.2.3a) which is clearly visible in the histogram (Figure 5.2.3c), but the distribution does not follow a simple parametric distribution, which seems reasonable because people are not equally distributed over space because of places which are not habitable e.g. mountains or the sea and because people tend to cluster in cities.

Figure 5.2.3 shows a comparison between the shares of the population around the centroid of the respondents’ residences near the main station of Zurich and the share of relationships in the same distance bands. The population shares were calculated using the Swiss census hectare raster data of the Federal Bureau of Statistics (BfS). This comparison is limited to Switzerland, because of missing comparable data for the surrounding countries, but this is not so critical as the population in Southern Germany, the nearest neighbour, which is at least 30 km away, is rather sparse. The decay of the relationships shares in the close-up range is remarkable in contrast to the population shares. Still, the share of contacts is over-proportional as visible in the very high ratios. The population share would
Figure 5.2.2: Distance distribution between respondents and their contacts (note the different scaling of the x-axes)

(a) Over the whole distance spectrum
(b) Share of core contacts by distance
(c) Up until 100 km
(d) Over 100 km
be expected to grow with square of the distance at an equal distribution. But it decreases in the close-up range slightly with the distance to the inner city of Zurich. From the distance band of 10 km on, the bigger cities of Switzerland and their agglomerations are clearly visible with the peak of Winterthur and Rapperswil and then Basle and Bern. The last two peaks are also slightly visible at the share of relationships. The ratio of these shares is independent of the distance bands and shows first a steep and then steady decrease with the distance from home. While people still select more then proportionally from those close-by, the range is today well beyond the distance of a 30 minute walk.

Note the change in the width of the distance classes at 10 km; due to lack of data, the contribution of the population in southern Germany had to be omitted.

Figure 5.2.3.: Contact and population shares by distance band around Zurich

Note the change in the width of the distance classes at 10 km; due to lack of data, the contribution of the population in southern Germany had to be omitted.
5.2. SPATIALITY OF SOCIAL NETWORKS

Distance comparison

As mentioned in Section 3.2, many of the previous studies lack in measuring the distance precise and accurate between the egos’ and alters’ residences. Nevertheless, the distance distribution is one of the only social network measurement unit, which can be compared among different empirical studies. Butts (2000) summarizes the studies gathered by Latane et al. (1995) with simple power law models. These models fit the datasets better, compared to an initial model which also included an initial radius of constant tie probability. Table 5.4 summarizes power law models presented in Butts (2000) together with the power law model based on the Zurich data truncated to distances below 100 km. The distance effect is both strong and consistently present. All fits are power laws in distance, with different base probabilities of a social tie to account for the network degree (not shown in Table 5.4). As mentioned, the samples, which are compared to the Zurich data are the best available classic sociological studies, but not ideal for the task of fitting distance models, as all of these data sets use reported distances in classes or high aggregated distances.

The distance decay is entirely consistent among the data sets, and indicates that very different studies results in similar distance probability distributions. The exponent of roughly -1 shows an inverse square probability of a relationship probability with distance. The effect of heterogenous population distributions in Zurich region is not important here: the same inverse-square distance relationship results from the data if the incidence of a social ties is normalized for the population density (see Figure 5.2.3). This indicates, that egocentric social network evolution is not primarily dependent on the probability of meeting somebody, but on the cost of maintaining the social network, which is solely dependent on the distance and not on the population density as well.

5.2.2. Bivariate social relationship distance analysis

The great circle distances between the respondents’ residence and their relationship’s residence and sociodemographic attributes of the respondents is compared
Table 5.4.: Comparison of the power law model fits of the tie probability with distance by Butts (2000) of the data in Latane et al. (1995) and the Zurich data

<table>
<thead>
<tr>
<th>Data</th>
<th>Exponent of distance $\beta$</th>
<th>N distance bins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encounters/distance in Florida</td>
<td>-1.01</td>
<td>11</td>
</tr>
<tr>
<td>Encounters/distance in China</td>
<td>-1.05</td>
<td>10</td>
</tr>
<tr>
<td>Encounters/distance between academics in sociology</td>
<td>-0.93</td>
<td>8</td>
</tr>
<tr>
<td>Zurich data</td>
<td>-1.20</td>
<td>10</td>
</tr>
</tbody>
</table>

All models are of the form $y = \alpha x^\beta + \varepsilon$

Source: Butts (2000) and own calculations.

in Table 5.5. The socio-demographic differences are driven by age, education and their interactions plus the attachment to personally more important contacts.

Age makes a difference, with the younger people cultivating more distant relationships than the older ones. Gender seems to have also influence on the relationship distance, with men cultivating longer distant relationships. The education level indicates that a higher level of education increases the distance of relationships, and unlike with the social network degree, higher income people tend to have more long-distant relationships. The type of contact distance difference is driven by family with the longest distance mean and partner with the shortest distances. This is expected, as we maintain contact with important people also at longer distances, but on the other hand want to be close to the most important people.

5.2.3. Multivariate social relationship distance analysis

To analyse the distances between ego’s and alter’s home location, the hierarchical structure of the data has to be considered. The datasets have an unbalanced hierarchical structure, resulting from respondents having different number of social contacts. The characteristics of ego-alter relationships cannot be treated as independent observations as they also depend on individual characteristics of the respondent. Therefore, multilevel linear regression modelling techniques are used to account for this structure, to estimate unbiased coefficients(for detailed infor-
### Table 5.5.: Descriptive statistics of the great circle distances between home locations of contact and respondents [km] (with and without contacts within 25 km from the respondent)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median (w)</th>
<th>Mean (w)</th>
<th>St. err. of the mean (w)</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up to 30</td>
<td>13.3</td>
<td>92.8</td>
<td>436.2</td>
<td>707.0</td>
</tr>
<tr>
<td>30 to 40</td>
<td>9.5</td>
<td>166.5</td>
<td>480.8</td>
<td>1489.5</td>
</tr>
<tr>
<td>40 to 60</td>
<td>8.9</td>
<td>95.7</td>
<td>138.0</td>
<td>440.1</td>
</tr>
<tr>
<td>60 and older</td>
<td>8.4</td>
<td>121.9</td>
<td>219.5</td>
<td>773.9</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7.6</td>
<td>96.0</td>
<td>246.1</td>
<td>784.5</td>
</tr>
<tr>
<td>Male</td>
<td>12.4</td>
<td>124.3</td>
<td>318.7</td>
<td>893.2</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>6.6</td>
<td>92.8</td>
<td>574.2</td>
<td>2751.3</td>
</tr>
<tr>
<td>Obligatory school</td>
<td>4.3</td>
<td>75.8</td>
<td>210.6</td>
<td>826.3</td>
</tr>
<tr>
<td>Vocational training</td>
<td>8.4</td>
<td>86.5</td>
<td>130.4</td>
<td>523.0</td>
</tr>
<tr>
<td>High-school diploma</td>
<td>7.3</td>
<td>128.6</td>
<td>256.6</td>
<td>714.9</td>
</tr>
<tr>
<td>Further technical training</td>
<td>15.0</td>
<td>151.1</td>
<td>322.2</td>
<td>817.9</td>
</tr>
<tr>
<td>University degree</td>
<td>10.3</td>
<td>124.6</td>
<td>364.3</td>
<td>920.5</td>
</tr>
<tr>
<td><strong>Income [CHF/month]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>13.3</td>
<td>156.6</td>
<td>264.8</td>
<td>733.2</td>
</tr>
<tr>
<td>0 to 1,999</td>
<td>4.9</td>
<td>123.9</td>
<td>355.8</td>
<td>1341.8</td>
</tr>
<tr>
<td>2,000 to 5,999</td>
<td>8.5</td>
<td>108.7</td>
<td>233.3</td>
<td>679.8</td>
</tr>
<tr>
<td>6,000 +</td>
<td>11.7</td>
<td>96.4</td>
<td>317.1</td>
<td>952.0</td>
</tr>
<tr>
<td><strong>Type of contact</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>26.2</td>
<td>60.9</td>
<td>242.2</td>
<td>452.8</td>
</tr>
<tr>
<td>Friends</td>
<td>7.1</td>
<td>127.6</td>
<td>277.3</td>
<td>1030.1</td>
</tr>
<tr>
<td>Partner</td>
<td>2.2</td>
<td>112.1</td>
<td>150.6</td>
<td>910.6</td>
</tr>
<tr>
<td>Family</td>
<td>21.0</td>
<td>96.6</td>
<td>304.7</td>
<td>637.6</td>
</tr>
<tr>
<td>Work mates</td>
<td>8.7</td>
<td>123.8</td>
<td>272.1</td>
<td>877.1</td>
</tr>
<tr>
<td>All</td>
<td>9.0</td>
<td>108.7</td>
<td>276.4</td>
<td>833.2</td>
</tr>
</tbody>
</table>

97
CHAPTER 5. EMPIRICAL DATA ANALYSIS

Information on multilevel modelling, see [Bryk and Raudenbush, 1992; Goldstein, 2011; Snijders and Bosker, 1999]. The data for the distance analysis is structured in two levels: Level 1 includes information that depends on each ego-alter relationship, and Level 2 includes the egos socio-demographics as well as personal network characteristics, such as number of contacts and proportions of alters with common characteristics. Distances between egos and alters are skewed, resulting in outliers and extreme values. A residual maximum likelihood estimator (REML) is preferred over the least squares method (OLS) as OLS estimation is known for having disadvantages when dealing with skewed distributions including multiple outliers. Although both REML and Maximum Likelihood (ML) estimators fit parameters to the overall dataset, REML is preferred since is more appropriate when estimating mixed models including fixed as well as random effects [Snijders and Bosker, 1999]. A logarithmic transformation was used to address the skewed distance distribution in the dataset. In order to avoid infinite values and loose observations with zero km distance in the transformation, all zero km distances were set to the minimum distance in the data. This procedure helps to avoid adding a constant to each distance, which has a relative higher influence on small distances compared to large ones.

The models include interactions and non-linear effects to represent the characteristics of the data as accurately as possible. Table 5.6 provides an overview of the parameter estimates, standard errors and t-values. After removing variables which correlate highly with each other (limit = 0.5; e.g. education and income), variables with a significance level lower than 0.05 were removed stepwise.

Generally, it can be seen, that there is much more variance on Level 1 (Ego-Alter, Var[e] = 0.754) compared to the variance on Level 2 (Ego, Var[u] = 0.165), which means that the Level 1 variables can explain much more in the overall variance compared to the variables on Level 2. This also means, that in general most of the people mix their local and non-local contacts, whereas there are only very few cases of respondents with either just local or just non-local contacts.

Regarding level 2 effects, distance decrease with age with an U-shaped influence (see the quadratic non-linear effect). Counter intuitive is the effect of the number of relocations, as one would expect, that people with many relocations maintain
their previous contacts and thus maintain a more extensive network. The effect results from egos living with a partner in their households has a strong negative effect on distance. People living with their partner have smaller distances in their personal network, suggesting that those who care for others primarily maintain small distance relationships. Certainly, from an egos’ perspective this effect can be related with the trade-off between the benefits from emotional contacts and distance costs, since nearby alters can be reached without spending much time and money on travel costs. Those savings in time and money can be invested into emotionally important relations, such as partnership or parenthood. The availability of internet involves an increasing effect on network distances. In general, the data show that people with internet resources maintain longer distance relations. In addition, higher education levels in egos has an increasing effect on network distances. Considering that institutions for education are good places to meet others and establish relations, this result is not surprising.

Regarding level 1, immediate relatives live closer to egos than friends, while extended family members live further away. This suggests that people stay in contact to their family members independently from geographical distances, which could be expected since family members are in general a reliable source of help and support, and may offer kinds of support that other social contacts may not. This is accentuated by the strong negative influence of important contacts on distance, as well as the negative influence on distance for contacts known longer than 10 years. This result also suggests, that contacts over longer distance might be maintained for a certain period of time, but the likelihood of maintaining such costly and high effort relationships over a longer period of time is lower compared to close distance relationships. There are two significant interaction effects with contacts known for more than ten years and extended family as well as number of relocations. The first interaction effect suggests that extended family relationships are over-proportionally more maintained, when they are closer distant, and that respondents who move often maintain their long-time relationship also at a longer distance.
### Table 5.6: Multivariate multilevel analysis of the great circle distances between home locations of contact and respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>2.878</td>
<td>0.536</td>
<td>5.372</td>
</tr>
<tr>
<td><strong>Ego’s characteristics (Level 2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.138</td>
<td>0.064</td>
<td>-2.146</td>
</tr>
<tr>
<td>Age [years]</td>
<td>-0.023</td>
<td>0.011</td>
<td>-2.183</td>
</tr>
<tr>
<td>Age$^2$ [years$^2$]</td>
<td>0.001</td>
<td>0.000</td>
<td>2.004</td>
</tr>
<tr>
<td>Number of relocations [-]</td>
<td>-0.121</td>
<td>0.057</td>
<td>-2.132</td>
</tr>
<tr>
<td>Living with partner [y/n]</td>
<td>-0.312</td>
<td>0.074</td>
<td>-4.212</td>
</tr>
<tr>
<td>Internet access [y/n]</td>
<td>0.237</td>
<td>0.077</td>
<td>3.081</td>
</tr>
<tr>
<td>Further technical training [y/n]</td>
<td>0.177</td>
<td>0.121</td>
<td>1.458</td>
</tr>
<tr>
<td>University degree [y/n]</td>
<td>0.324</td>
<td>0.133</td>
<td>2.436</td>
</tr>
<tr>
<td><strong>Alter characteristics (Level 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate family [y/n]</td>
<td>-0.522</td>
<td>0.053</td>
<td>-9.85</td>
</tr>
<tr>
<td>Extended family [y/n]</td>
<td>2.321</td>
<td>0.932</td>
<td>2.491</td>
</tr>
<tr>
<td>Very close tie (emotionally) [y/n]</td>
<td>-0.256</td>
<td>0.052</td>
<td>-4.972</td>
</tr>
<tr>
<td>Known &gt; 10 years [y/n]</td>
<td>-0.973</td>
<td>0.465</td>
<td>-2.094</td>
</tr>
<tr>
<td>Known &gt; 10 years [y/n] *</td>
<td>-2.824</td>
<td>0.931</td>
<td>-3.033</td>
</tr>
<tr>
<td>Extended family [y/n]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Known &gt; 10 years [y/n] *</td>
<td>0.145</td>
<td>0.056</td>
<td>2.593</td>
</tr>
<tr>
<td>Number of relocations [-]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents = 265</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations = 3,156</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC = 8,547</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var[e] = 0.745, Var[u] = 0.165</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2. SPATIALITY OF SOCIAL NETWORKS

5.2.4. Social network geographies: Confidence ellipses

The analysis above focused on the distance between the respondents’ and their contacts’ residence, but the distance alone ignores the pattern of the contact, e.g. the agglomeration of contacts, which can not be measured just by distance and its distribution parameters. Biologists and more recently transport planners had to address the identical question of how to measure spatial distributions in their analysis of the daily activity spaces. They proposed parametric, semi-parametric and non-parametric approaches to measure the size of the activity spaces (see Schönfelder, 2006, for a review). The most popular, but also problematic approach is to calculate the size of the confidence ellipse, i.e. the two-dimensional generalization of the confidence interval. The Confidence ellipses are frequently used as an exploratory and hypothesis test-oriented method for the analysis of the relationship between two variables. They can be helpful to visualize and quantify the spatial distributions if the x- and y-coordinates are defined as locations of two (mutually dependent) variables. Confidence ellipses are analogous to the confidence interval in univariate distributions, where it defines the smallest possible (sub) region, in which a defined portion - 95% for a t-value of 1.96 - of the observations can be found. The calculation of the ellipses, however, is conditional on the assumption that the distribution follows a bivariate normal distribution. This is shown for at least approximately for the distribution of activity locations (Moore, 1970). But this does not necessarily hold for the spatial distribution of social relationships, which is not true as shown above. This leads to an overestimation of the calculated area. Also the symmetry of the confidence ellipse leads often to cases where half of the area covered by the ellipse is empty of locations and therefore too big. Nevertheless, the confidence ellipses are here used as a descriptive representation of the spatiality of ego-centric social networks and analysed are the relations between the confidence ellipses, which counts at least parts of the bias caused by the normal distribution assumption out.

The ellipsis is defined by the covariance-matrix $S$ all home locations of one ego’s alters, whereas the home locations can be weighted by the frequency of face-to-face contact. The covariance-matrix is defined as
CHAPTER 5. EMPIRICAL DATA ANALYSIS

\[ S = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy} \end{pmatrix} \]

where the single covariances are defined as

\[ s_{xx} = \frac{1}{n-2} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

\[ s_{yy} = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \bar{y})^2 \]

\[ s_{xy} = s_{yx} = \frac{1}{n-2} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \]

The area of the confidence ellipses is given by

\[ A = 6\pi|S|^2 \]

The slope of the ellipses is given by the linear correlation between the x and y values of the coordinates of each home location. Figure 5.2.4 and 5.2.5 show the graphical representation of the confidence ellipses.

The 95% confidence ellipse (compared with 95% confidence interval), which by definition encompasses 95 percent of the geocoded points per event, is chosen to represent the geography of social networks. To avoid bias, the alter’s residence locations are weighted by the annual frequency of face-to-face meetings. This has the purpose to represent the activity space in which the respondents physical social world is most likely to take place. This also avoids to put too much emphasis on geographically distant contacts, which are not visited often. This method is also used in this approach for the identification of mass numbers of the spatial distribution of mobility biographical measurements, such as residency, employment and education places for each respondent. The areas are calculated using the Behrmann-projection\(^1\), which is an equal-area cylindrical map projection. This is

\(^1\)http://mathworld.wolfram.com/BehrmannCylindricalEqual-AreaProjection.html
5.2. SPATIALITY OF SOCIAL NETWORKS

Figure 5.2.4.: Example social network geography

The respondent is female, 35 years old, full time employed and has moved 8 times in the last 22 years. The red circle tags the current home location and the black crosses are the home locations of the acquaintances. Important as the calculated geographies vary strongly and are therefore prone to distortions.

However, the easiest way to capture the geography would be to sum-up the distances of the egocentric social network ties. Unfortunately, this fails mainly due to the fact, that people tend to cluster, which is especially true in the case of social networks, where a strong homophily is expected. This homophily together with the strong distance decay observed in the previous paragraphs suggests a strong tendency of clustering within social network geographies. And this clustering matters especially to transport planning, as it makes a big difference for implied travelled distances e.g. if a social network consist only of two geographical agglomerations compared to if the members of the ego-centric social network are evenly spatially distributed. The frequency weighted sum of the contact distances correlates only weekly with the size of the 95% confidence ellipses as shown in Figure 5.2.6.

The confidence ellipse measures geographical patterns with just three parameters (length of the main axis, ratio between the two axes and angle of the main axis).
Figure 5.2.5.: 95%-confidence ellipses of the social network geographies for a set of the respondents
5.2. SPATIALITY OF SOCIAL NETWORKS

Figure 5.2.6.: Sum of contact distances against the size of network geographies

This makes it an easy to use instrument for the analysis of the spatial distribution of social relationships. For analysing the patterns of social geographies we use the following measures:

- The area measures the spatial expanse and incorporates the frequency of contacts as well as the spatial density as a confidence interval in two dimensions around the regression line of the residences. A disadvantage of the confidence ellipse as a measurement of size is that it covers the relative frequency and not the absolute frequency, which can lead to bigger areas for social networks which could be maintained with lesser effort (measured by great circle distance) as others under certain circumstances. Although the positive correlation of the number of relationships with the size leads to the assumption that this happens rather rarely.

- The ratio of the axes measures how geographically directed the relationships are. In the case of egocentric social networks this can be interpreted as how spatially divers a social network is. E.g. a low ratio represents clusters of
social contacts in just two regions or along one axis.

- The angle of the main axis represents the geographical orientation of the ellipses. At a certain length of the main axis it could be interpreted as a cultural diversity of the social network. E.g. it could be assumed that in Zurich the German language - and the fact that the biggest minority in Zurich are immigrants from Germany (17.4% of all immigrants (Statistik Stadt Zürich, 2006)) will lead to south–north directions of the ellipses, whereas in the intercontinental group an east–west direction should prevail.

5.2.5. Bivariate analysis of social network geography

The distribution of the size of the 95% confidence ellipses seems to follow a log-normal distribution (Figure 5.2.7), if we ignore the third of the respondents who have a local set of contacts. The fit statistics for a log-normal distribution are good and do not reject this distribution at the 0.05-level (Chi-Square, Kolmogorov-Smirnov and Anderson-Darling Test) (Estimated with ExpertFit - Version 7.00 - Averill M. Law & Associates 2006). Other distributions, such as the Weibull, Gamma, log-Logistic and many more performed less well. The patterns of the socio-demographic differences follow generally the patterns of the distances to the contacts’ home location above (Table 5.7). An analysis of variance shows that none of these differences are statistically significant. Only if one takes the logs of the sizes of the geographies, then the age differences are significant.

The age of the average shows a U-shaped effect, where the middle ages have the largest network geographies. However, 50% of young people below 29 years old have at least 20,793 m² large network geographies, which is compared to other age groups the largest median value. If the age in years is cross tabulated with the log of the network size, there is a low Pearson’s R-value of -0.24, significant at a 0.05 level. Gender shows no influence on the size of the network geography. Income and education-related differences show nearly no effects. High and low income categories can have spatially more expansive network geographies, but the test statistics attest no significant correlation (Spearman R = -0.22, not significant).
5.2. SPATIALITY OF SOCIAL NETWORKS

Figure 5.2.7.: Distribution of the social network geometries measured as 95% confidence ellipses in km$^2$

Social Network geographies of less than 10 km$^2$ were coded as zero.
Table 5.7.: Descriptive statistics of 95% confidence ellipses of the social network geographies [10^5 km²] (with and without contacts within 25 km from the respondent)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>Mean</th>
<th>St. err. of the mean</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with w/o</td>
<td>with w/o</td>
<td>with w/o</td>
<td>with w/o</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up to 30</td>
<td>0.27</td>
<td>0.27</td>
<td>66.73</td>
<td>68.29</td>
</tr>
<tr>
<td>30 to 40</td>
<td>0.12</td>
<td>0.18</td>
<td>36.19</td>
<td>41.76</td>
</tr>
<tr>
<td>40 to 60</td>
<td>0.14</td>
<td>0.25</td>
<td>17.74</td>
<td>20.80</td>
</tr>
<tr>
<td>60 and older</td>
<td>0.16</td>
<td>0.29</td>
<td>17.75</td>
<td>20.68</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.15</td>
<td>0.21</td>
<td>20.54</td>
<td>23.57</td>
</tr>
<tr>
<td>Male</td>
<td>0.21</td>
<td>0.31</td>
<td>37.68</td>
<td>42.64</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>0.33</td>
<td>0.33</td>
<td>6.62</td>
<td>6.62</td>
</tr>
<tr>
<td>Obligatory schooling</td>
<td>0.04</td>
<td>0.14</td>
<td>1.71</td>
<td>2.14</td>
</tr>
<tr>
<td>Vocational training</td>
<td>0.06</td>
<td>0.15</td>
<td>5.74</td>
<td>7.03</td>
</tr>
<tr>
<td>High school diploma</td>
<td>0.16</td>
<td>0.19</td>
<td>51.37</td>
<td>55.65</td>
</tr>
<tr>
<td>Further technical training</td>
<td>0.21</td>
<td>0.34</td>
<td>39.51</td>
<td>39.39</td>
</tr>
<tr>
<td>University degree</td>
<td>0.38</td>
<td>0.40</td>
<td>55.04</td>
<td>59.34</td>
</tr>
<tr>
<td><strong>Income [CHF/month]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.A.</td>
<td>0.54</td>
<td>0.93</td>
<td>4.86</td>
<td>5.46</td>
</tr>
<tr>
<td>0 to 1,999</td>
<td>0.14</td>
<td>0.18</td>
<td>52.74</td>
<td>57.39</td>
</tr>
<tr>
<td>2,000 to 5,999</td>
<td>0.15</td>
<td>0.20</td>
<td>22.34</td>
<td>24.55</td>
</tr>
<tr>
<td>6,000 +</td>
<td>0.09</td>
<td>0.22</td>
<td>41.18</td>
<td>51.83</td>
</tr>
<tr>
<td>All</td>
<td>0.16</td>
<td>0.25</td>
<td>28.53</td>
<td>32.52</td>
</tr>
</tbody>
</table>
5.2. SPATIALITY OF SOCIAL NETWORKS

5.2.6. Multivariate analysis of social network geography

A model of the logarithm of the size of the 95% confidence ellipses as a dependent variable has to be considered as the values are all non-negative, with 33 zero value observations. Conventional regression-methods, as the ordinary least square method, are not adequate for such censored values (Greene and Zhang, 2003).

A model which is able to differentiate between limit-observations and non-limit-observations is the Tobit Model. The Tobit model assumes that the limit outcome is determined by the level of the non-limit outcome. To test this assumption, a different model which is also appropriate for the data can be compared to the Tobit model. This is Cragg’s Model for Censored Data (Cragg, 1971). It is used, when the assumption of the Tobit model, that the non-limit outcome is determined apart from the level of the non-limit outcome, is not true. Cragg’s Model is a combination of the Probit model (for y=0) and the truncated regression (for y>0). The zeroes in our data have their origin in two different problems, the first problem is, that only around 70% of the geocodes have street address accuracy, while the rest has only zip-level, neighbourhood, or village/city accuracy, which leads to just one geocode for several contacts and the second problem is, that the confidence ellipse needs at least three spatially distinct locations to be calculated. The origin of the zeroes in the data leads to the assumption, that the non-limit outcome is determined by the same level as the limit outcome, which is shown through the Probit and Tobit results (Table 5.8). The Tobit model was calculated after removing variables which correlate highly with each other (limit=0.5), and variables with a significance level lower then 0.05 were removed stepwise. The parameter estimates are reported in Table 5.8.

The analysis of the Tobit results shows that there are different factors which influence the social network geographies. The first group consists of socio-demographic variables. The model results indicate that young people with a high education and with a low- or a middle income tend to maintain a more spatially distributed social network. The influence of the age and the education is similar to their influence on the numbers of relationships. The influence of the income seems to be unexpected as a spatially more distributed social network is expensive to maintain. An interpretation of the negative influence could be that a higher income is often
Table 5.8.: Parameter estimates for the Tobit regression of the logarithm of the size of the 95% confidence ellipses and the associated Probit model of the Cragg approach Tobit model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>St. dev.</th>
<th>Beta</th>
<th>Sign.</th>
<th>Beta</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-</td>
<td>-</td>
<td>9.929</td>
<td>0.000</td>
<td>2.487</td>
<td>0.03</td>
</tr>
<tr>
<td>Age [years]</td>
<td>53.43</td>
<td>19.305</td>
<td>-0.296</td>
<td>0.000</td>
<td>-0.114</td>
<td>0.01</td>
</tr>
<tr>
<td>Age$^2$/1,000 [years$^2$/1,000]</td>
<td>3.226</td>
<td>2.099</td>
<td>2.946</td>
<td>0.000</td>
<td>1.091</td>
<td>0.01</td>
</tr>
<tr>
<td>Car ownership [y/n]</td>
<td>0.472</td>
<td>0.500</td>
<td>1.609</td>
<td>0.010</td>
<td>0.187</td>
<td>0.37</td>
</tr>
<tr>
<td>Number of relationships [-]</td>
<td>12.406</td>
<td>8.454</td>
<td>0.201</td>
<td>0.000</td>
<td>0.084</td>
<td>0.00</td>
</tr>
<tr>
<td>Education/workplace changes [-]</td>
<td>3.336</td>
<td>2.475</td>
<td>0.289</td>
<td>0.020</td>
<td>0.053</td>
<td>0.28</td>
</tr>
<tr>
<td>Further technical training [y/n]</td>
<td>0.213</td>
<td>0.410</td>
<td>2.485</td>
<td>0.000</td>
<td>0.581</td>
<td>0.04</td>
</tr>
<tr>
<td>University degree [y/n]</td>
<td>0.245</td>
<td>0.431</td>
<td>2.617</td>
<td>0.000</td>
<td>0.397</td>
<td>0.16</td>
</tr>
<tr>
<td>Income &gt;6,000 CHF/month [y/n]</td>
<td>0.262</td>
<td>0.441</td>
<td>-1.643</td>
<td>0.028</td>
<td>-0.279</td>
<td>0.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Tobit model</th>
<th>Probit model</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>286</td>
<td>241</td>
</tr>
<tr>
<td>Goodness-of-fit</td>
<td>Adjusted R$^2$ = 0.25</td>
<td>$\chi^2$(8 df) = 47.31</td>
</tr>
</tbody>
</table>
linked to a higher workload and more responsibility which leads to a higher value of time for these persons. As travel costs have decreased over the decades (See for example [Fröhlich and Axhausen 2002] for Switzerland or [Bruegman 2005] for the USA), the time costs seem to exceed the financial costs of travelling. The ownership of a car has a positive influence on the size of the social network geographies, even if the ownership of a car does not contribute to the maintenance of contacts over the distance (see above), it is an indicator of mobile people. The number of relationships has an influence as mentioned above, as it is correlated with the share of non-core contacts, as it is now possible to maintain spatially more widely distributed network of weaker ties (see Figure 5.2.2) with less frequent face-to-face contacts in combination with telecommunication contacts (see Section 5.3). The number of education or workplace moves is a biographical influence on the social network geographies. It seems that being less anchored in space and being professionally flexible have a positive influence on the size of the social network geographies, while surprisingly, the spatial distribution of the education and workplace changes, measured by their confidence ellipses, has no significant influence on them. Overall the model explains 25% of the variance of the social network geographies.

The parameters of the Probit model follow exactly the parameters of the Tobit model (see Table 5.8). The resulting predictions are 100% correct for the 1s (y > 0) and 22.5% correct for the 0s, which results in overall 89.2% correct values. As the parameters of the Probit model show, that the limit outcome is determined by the level of the non-limit outcome, the estimates of the truncated model for the non-limit observations are omitted.

The shown results of the Tobit and Probit models explain the size of the egocentric social network. However, the area of the ellipse is one out of three measurable parameters describing an ellipse. For example, a circle (which would be an ellipse with a ratio of one between the minor axis (R2) and the main axis (R1)) with the same area as a narrow ellipse (which is a ellipse with a low R2/R1 ratio) are not the same and therefore barely comparable in substantive terms. This could be one of many reasons for the relatively low explanatory power of the Tobit model. To analyse the influence of the ratio of the two axis on the area, the relation of the
size of the ellipses and the ratio between the minor axis (R2) and the main axis (R1) are illustrated in Figure 5.3. The x-axis represents the logarithm of the size of the confidence ellipses (km 2), the values of the R1/R2-ratios are their means of each deciles of the area and the ratios are subdivided in quintiles of the main axis (km). Three trends are visible: Regional contact geographies with a main axis shorter than 54 km have no small R2/R1 ratios and do not show a strong spatial orientation. For large geographies only small ratios are observed and the longer the main axis (R1) is, the smaller the maximum of the observed ratios are. This means that as social network geographies become larger, the more spatially concentrated in a certain direction the contacts are. It seems to be necessary for the maintenance of large social network geographies, to make it possible to visit a substantial part of the contacts with one longer trip, e.g. overseas journey, and some shorter trips from the new base location.

The third parameter of the confidence ellipse is the orientation of the main axis measured as the rotation deviation from the east-west direction in degrees. Figure 11 shows the mean of the rotation by deciles of the main axis length. As the ratio between the axes is important for analysing the rotation, the values are subdivided
by the terciles of the ratios. As mentioned above, the rotation represents the geographical orientation of the ellipse. The rotation is stronger, when the ratio between the main axes is small. Therefore the analysis concentrates on the blue and green points in Figure 5.2.9 Overall the deviations from the east-west axis are not as large as expected and have their maximum in these categories at 67 degrees. However, there are two remarkable breaks around 100 km and 1,000 km which divide the figure in three sections. The first section between 10 km and 100 km shows a drop of the rotation. The very local and regional geometries tend to have a higher ratio than the geometries at national level. The deviation from east-west on the regional level around Zurich (∼40km) is similar to the catchment area of the commuters to Zurich, which is concentrated along the Lake Zurich (Botte 2003). On the national level, around 100 km, the orientation follows a clear east-west direction which is determined by the larger Swiss cities, Basel and Bern in western- and St. Gallen in the eastern-direction. From 100 km till 1,000 km the rotation-degree increases till 42 degrees which indicates a dominant share of contacts in Germany at this level of the main axis length. As expected the rotation decreases for intercontinental social networks, which is caused mainly by the ties to the USA.

5.3. Social contacts in a shrunken world

This section is with permission of the co-authors an updated and extended version of parts of Frei and Axhausen (2010).

The respondents were not only asked to name their contacts, but also to describe them: especially, the original contexts of the acquaintance; and the frequency of communication by mode (face-to-face, phone, email, short text message (SMS)). Of interest here is whether the frequency of contacts can be explained, and in particular whether the distances between the respondents and their contacts play a role. A-priori, one would expect that distance will play a role if the effort involved with the interaction increases with distance. This should be true for:

- face-to-face interactions, which involve coordination efforts, travel time, travel costs, jet lag, accommodation costs, hospitality and gifts for the hosts visited;
Figure 5.2.9.: Orientation of the social network geographies

![Graph showing orientation of social network geographies.]

- phone calls, which involve coordination efforts, time zone differences and fees, which in Switzerland increase with distance beyond the local call area and even more so beyond the national boundaries. At least there is a shadow price if the person has a prepaid budget of minutes under the contract signed with the telephone service provider. At the time of the survey, VOIP’s market penetration was not yet high enough, and was not available yet for mobile phones to be a major factor;

- SMS messages, which involve flat fees, especially to recipients abroad, but no coordination costs. (The same shadow price argument holds here as well.)

- Email is essentially free at the point of use, either at home or at work, as flat rate schemes dominate and the contribution of email to any bandwidth charging scheme is negligible. There are no coordination costs.

Given the orders of magnitude of the cost differences between travel, calling and texting, one would expect that the strength of distance decay is different for these three modes. While this distance decay has been documented for travel since Lill’s analysis of railway travel between Vienna and Prague in 1889 and as well as for telephone calls, for example in Rietveld and Janssen (1990), it has not been shown
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

at the level of interactions between respondents and their contacts.

The rich literature on the interaction between travel and electronic communications has highlighted the potential for substitution and complementarity between the modes of contact, which necessitates a modelling approach that can account for these possible interactions.

5.3.1. Bivariate analysis of the number of contacts in social network

Table 5.9 shows the characteristics of relationships in comparison with the type of relationship. This includes the comparison of the type of relationships with the number of contacts by contact means. Partners (4.9% of all) and those classified as “other” (0.9%) are excluded from further analysis, as the former are special cases and the latter are rare. It has been shown that respondents overestimate rare events but assess high frequencies accurately [Schlich and Schönfelder, 2001].

The implied annual travel distances (the product of the great circle distance multiplied by the reported frequencies) does not confirm this point. These implied distances are comparable to the more careful estimates for annual leisure travel in the national travel surveys. Acknowledging that the national travel surveys often omit travel outside national boundaries and fail to capture journeys with mixed purposes, e.g., business travel mixed with a visit to relatives, the implied travelled distances come very close to the expected ones. In 2005 45% of the 13,000 annual kilometres were leisure travel in Switzerland, which is around 5,850 annual kilometres. 80% of leisure travel is declared visiting relatives and friends, which makes about 4,680 km of annual travelled distance per person socially motivated. If we consider, that leisure activity group sizes are around 2.3 in average (see Schlich and Axhausen, 2003), the 11,000 implied travelled kilometres from the face-to-face contact frequency and distance between home locations becomes 4,783 kilometres, which is surprisingly comparable to the 4,680 km from the Swiss national travel survey. If we add half of the estimated 6,000 km travelled outside the country (ARE/BIS, 2007a) (around 2,000 km for visiting and friends), we are even below the expected value. But we do not know who is visiting whom or to what
Table 5.9.: Characteristics of named contacts by type of relationship (excepting “other” and partners)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Statistic</th>
<th>Friends</th>
<th>Family</th>
<th>Workmates</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of relationship [years]</td>
<td>Mean</td>
<td>14</td>
<td>35</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>11</td>
<td>34</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Face-to-face meetings per year</td>
<td>Mean</td>
<td>47</td>
<td>49</td>
<td>54</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Telephone calls per year</td>
<td>Mean</td>
<td>27</td>
<td>48</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>12</td>
<td>24</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Email written per year</td>
<td>Mean</td>
<td>10</td>
<td>9</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SMS sent per year</td>
<td>Mean</td>
<td>12</td>
<td>16</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Core contact</td>
<td>Share</td>
<td>47%</td>
<td>57%</td>
<td>55%</td>
<td>52%</td>
</tr>
<tr>
<td>Great circle distance [km]</td>
<td>Mean</td>
<td>278</td>
<td>313</td>
<td>272</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>7</td>
<td>20</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Annual implied km travelled</td>
<td>Mean</td>
<td>7,816</td>
<td>16,712</td>
<td>7,675</td>
<td>10,669</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1,918</td>
<td>1,466</td>
<td>1,063</td>
<td>1,325</td>
</tr>
</tbody>
</table>

N = 3,302 for all variables except distance, where N = 3,168

extent people combine visits or see their contacts in groups during holidays at third locations. Accounting for all of this, the means are as expected. The much lower median highlights the importance of the rare, and probably overestimated frequencies of visits to contacts very far away.

Figure 5.3.1 shows, that the distance decay of contact frequency varies considerably by mode, which will lead to rather different conclusions when the modal shares are analysed. Figure 5.3.1 shows the shares of the modes as a function of the distance between the respondents and their contacts. The share of the face-to-face visits decreases quickly, as does the share of SMS, though rather more gently (statistically insignificant). Email gains share fastest with increasing distance because it stays constant in terms of frequency over the distance range. Phone calls gain share as
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

Figure 5.3.1.: Share of the contact mode by log distance in deciles and zero between the respondent’s and the contact’s homes (for friends, relatives, workmates only)

well because they overtake all other modes in terms of frequency.

5.3.2. Multivariate analysis of the number of contacts in social network

To analyse the frequency of contacts between ego’s and alter’s, the hierarchical structure of the data has to be considered. The datasets have an unbalanced hierarchical structure, resulting from respondents having different social network degree. The characteristics of ego-alter relationships cannot be treated as independent observations as they also depend on individual characteristics of the respondent. Therefore, multilevel linear regression modelling techniques are used to account for this structure, to estimate unbiased coefficients (for detailed information on multilevel modelling, see [Bryk and Raudenbusch 1992, Goldstein 2011])
Snijders and Bosker (1999). Further are the contact frequencies not independent of each other, so we have to allow the model to let the different contact means correlate with each other. Approaching these data structures with an OLS estimator and an ordinary linear regression model would be a model misspecification. The results would be inexact parameter estimations with both unreliable standard errors and hypothesis tests. The assumption that all of the group structure is represented by the explanatory variables is unfounded. To capture these possibilities, two different modelling approaches are used, a multilevel path analysis model and a 3-Level multivariate multilevel linear regression model.

The data for the frequency analysis is structured in two levels: Level 1 includes information that depends on each ego-alter relationship, and Level 2 includes the egos socio-demographics as well as personal aggregated network characteristics. For the 3-Level model, a third level denotes the different contact means.

**Path analysis model**

In the two-level path analysis model we allow for within- and between-respondents variation in contact frequency, which includes residuals at the respondent level and the ego-alter link level. The respondent residuals, often called “cluster effects”, represent unobserved respondent characteristics that affect the outcomes for relationship characteristics as they lead to correlation between outcomes for relationship characteristics of the same ego. (see Hox and Roberts, 2011, For an in-depth review of multilevel (structural equation) models). In the model used here, the respondents’ personal characteristics are used to explain the between-level patterns. The relationship characteristics are used to explain the variance at the Level 1. The path diagram of the model is illustrated in Figure [5.3.2]. The pathways (arrows) in the path-model represent the hypothesized effects. The variables on the right are endogenous variables, hypothesized to be influenced by the exogenous variables on the left. The gray arrows represent the influences at the between Level (Level 2 on Level 1), and the black arrows represent the influences within Level 1. At the within level, the endogenous variables, the four frequencies, are allowed to be regressed on each other.
The estimates were performed with MPlus 5 \cite{Muthen2007}. The dataset only includes the contacts with whom the respondents still are in contact, a self-selected set. The parameter estimates reported below have to be interpreted with this in mind, as we unable to correct for the self-selection bias, as we do not have information about the non-chosen contacts \cite{Heckman1979}. Table 5.10 presents the with-in and between-level maximum likelihood results of the analysis across all types of contacts. The overall model goodness of fit statistics are highly significant. The bulk of the explanatory power is due to the Level 2 (ego) differences, but for the face-to-face meetings where the within-level differences contribute substantially. Still, all of the equations contribute significantly (see $R^2$-values). At both levels the various modes of contact are strongly complementary, as one would expect given that the various channels are used to prepare, co-ordinate and later discuss the face-to-face meetings. As expected above the distance effect drops from emailing to face-to-face visits. The distance effect is actually insignificant for email in line with the hypothesis above.

On Level 1, the within level parameter estimates are of interest and the associated
### Table 5.10.: Standardized estimates of the multi-level path analysis model of the contact frequencies by mode

<table>
<thead>
<tr>
<th>From/to</th>
<th>Face-to-face</th>
<th>Telephone</th>
<th>E-mail</th>
<th>SMS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.30**</td>
<td>2.13**</td>
<td>0.74*</td>
<td>1.10**</td>
</tr>
<tr>
<td>Phone</td>
<td>0.38**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td>0.15**</td>
<td>0.30**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.20**</td>
<td>0.37**</td>
<td>0.20**</td>
<td></td>
</tr>
<tr>
<td>Duration of contact</td>
<td>-0.00</td>
<td>0.00</td>
<td>-0.01**</td>
<td>-0.00*</td>
</tr>
<tr>
<td>Work mate</td>
<td>-0.14*</td>
<td>-0.10*</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Relative</td>
<td>-0.07</td>
<td>0.16**</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>ln(distance)</td>
<td>-0.11**</td>
<td>-0.04**</td>
<td>-0.00</td>
<td>-0.03**</td>
</tr>
<tr>
<td>R²</td>
<td>0.13**</td>
<td>0.03**</td>
<td>0.01**</td>
<td>0.02**</td>
</tr>
<tr>
<td>N</td>
<td>3,364</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male ego</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.30*</td>
<td>-0.00</td>
</tr>
<tr>
<td>Number of relocations</td>
<td>0.09*</td>
<td>0.06*</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>College degree</td>
<td>0.14</td>
<td>0.24</td>
<td>-0.04</td>
<td>0.21</td>
</tr>
<tr>
<td>Highschool diploma</td>
<td>0.42</td>
<td>0.56</td>
<td>1.24*</td>
<td>0.26</td>
</tr>
<tr>
<td>Primary school only</td>
<td>-0.18</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>&lt; 1999 sFr/month</td>
<td>0.25</td>
<td>-0.11</td>
<td>-0.46*</td>
<td>0.08</td>
</tr>
<tr>
<td>&gt; 6001 sFr/month</td>
<td>0.91**</td>
<td>-0.16</td>
<td>-0.31*</td>
<td>0.02</td>
</tr>
<tr>
<td>Car always available</td>
<td>0.40*</td>
<td>0.07</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>National season Ticket</td>
<td>-0.49**</td>
<td>-0.11</td>
<td>-0.05</td>
<td>-0.07</td>
</tr>
<tr>
<td>Age &lt; 30 years</td>
<td>0.28</td>
<td>0.31</td>
<td>0.30</td>
<td>0.62</td>
</tr>
<tr>
<td>Age 30 to 44 years</td>
<td>-0.05</td>
<td>0.24</td>
<td>0.32</td>
<td>-0.02</td>
</tr>
<tr>
<td>Age &gt; 60 years</td>
<td>-0.44*</td>
<td>0.27</td>
<td>-0.31*</td>
<td>-0.52**</td>
</tr>
<tr>
<td>R²</td>
<td>0.33</td>
<td>0.16</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>N (Clusters)</td>
<td>277</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significance level < 0.05; ** significance level < 0.01
goodness of fit measures. The impact of the duration of contact between alter and ego varies by the type of contact. It has a significant positive impact for phone contacts between friends. The distance effect also turns positive for email messages between friends, while it otherwise consistently negative. The relative strength of the distance decay remains otherwise unchanged.

On Level 2, the socio-demographic effects are generally as expected: age reduces the frequency of contacts, but for the phone interactions. The number of previous move increases it, as the number of alters increases with the number of moves. High income increases the number of face-to-face interactions, while reducing email contact. Regular car availability increases contact, as it lowers the marginal cost of travel. The negative sign for the ownership of the national season ticket comes as a surprise in this context, but for the fact that it is also indicates regular above-average distance commuting and therefore less available time, which none of the other variables capture. To illustrate the impact of the distance and the interaction of the different modes, a short quantitative example is given for the distance bands 0.1, 1, 10, 100 and 1000 km, which shows the importance of the distance. For an average of 56 telephone calls per year constant over distance the number of face-to-face contacts are estimated the following (in brackets are just the within values): at a distance of 0.1 km 175 (15), at 1 km 112 (14), at 10 km 97 (14) at 100 km 72 (14) which then drops to 47 (13). The strong impacts of the type of alter indicate, that the effects of the other variables might vary by type. The same model formulation was estimated for the different subsets of the contacts.

Since structural equation models and path model have been often criticized by statisticians, in the following section a multivariate linear regression model is estimated and then compared to the path model. The disadvantages of structural path models are partly rooted in the modelling technique, where path models fit the sample covariance instead of the sample values themselves, as the statistical theory underlying SEM pertains to covariance, rather than correlation, matrices (Cudeck and Henly [1991]). When correlation matrices are analysed, standard errors of parameter estimates are usually inaccurate.

Another problem here is the phenomenon of propagation of specification errors.
The maximum likelihood estimator used in SEM uses all available information in the covariance matrix of the observed variables to generate parameter estimates. This allows the effects of a misspecified parameter to be propagated through the modelling structure (Kaplan, 1989). Thus, for example, an omitted path can bias estimates of other structural or measurement parameters. Further, it is very likely that the optimization only finds a local optimum, due to the complexity to minimize the objective function in SEM.

**Multivariate multilevel linear regression model**

In the case which is used in this paper we model the frequency of social contact by different modes, which is e.g. how often one person (ego) interacts with another person (alter) to whom the ego has a relationship (tie) by different modes (face-to-face, phone, short message service or email). The ego is in the data we model the interviewee, whereas the information about the alters are less detailed. The hierarchical structure implies the egos as Level 3 units \( i \) with belonging alters (Level 2 units) \( j \), with four variables \( Y_1, \ldots, Y_4 \), measuring the frequency of contact by mode 1, ..., 4 for each alter, which are the Level 1 unit. The dependent variables is denoted as \( Y_{hij} \), which is the measurement of the \( h \)'th variable for alter \( i \)'s relationship to ego \( j \).

Following the model formulations in Snijders and Bosker (1999) we use \( Y_h \) for the vector of dependent variables, where \( h \) denotes the index for the dependent variable, \( x \) for the explanatory variable at the individual level, \( z \) for the explanatory variable at the group level. \( j \) is the group index, \( i \) the index for the individuals. \( \gamma \) represents regression coefficients, \( R \) is the residual term and \( U \) is here the group dependent variation of the intercept. A random intercept model with a mean intercept of \( \gamma_{00h} \) and a group-dependent deviation \( U_{hj} \) can be denoted as:

\[
Y_{hij} = \gamma_{00h} + \gamma_{10h}x_{1ij} + \gamma_{20h}x_{2ij} + \ldots + \gamma_{p0h}x_{pij} + \gamma_{01h}z_{1j} + \ldots + \gamma_{0ph}z_{pj} + U_{hj} + R_{hij},
\]

where \( R_{ij} \) and \( U_{hj} \) are residual covariance matrices: \( \sum = \text{cov}(R_{ij}) \) and \( T = \text{cov}(U_j) \). The covariance matrix of the complete observations, conditional on the explanatory variables, is the sum of these, \( \text{var}(Y^c) = \sum + T \).
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

Because of the skewness of the contact frequencies between egos and alters, outliers and extreme values are in the dataset. A residual maximum likelihood estimator (REML) is therefore preferred over the least squares method (OLS) as OLS estimation is known for having disadvantages when dealing with skewed distributions including multiple outliers. Although both REML and Maximum Likelihood (ML) estimators fit parameters to the overall dataset, REML is preferred since it is more appropriate when estimating mixed models including fixed as well as random effects (Snijders and Bosker 1999).

The following model was estimated on data from 307 interviewees (Level 3) and 3,302 reported relationships (Level 2). The dependent variable (Level 1) is the annual frequencies of face-to-face, phone, short message service and email contacts between egos and alters. The independent variables at Level 3 are characteristics of the ego and on Level 2 the relationships characteristics.

The frequency of contacts by mode is measured as reported counts of days with a social contact within a year. This makes the Poisson distribution (rather than the Normal) appropriate since the Poisson mean is $> 0$. So the logarithm of the response variable is linked to a linear function of explanatory variables such that the $\log(Y) = \beta_0 + \beta_1X_1 + \beta_2X_2...$ etc. or $Y = e^{\beta_0}e^{\beta_1X_1}e^{\beta_2X_2}...$ etc. The contact frequencies are plotted against the distances in Figure 5.3.3 with a logarithmic x and y scale. The logarithmic transformation seemed to give homoscedastic results if we consider a simple linear model of which the result is the plotted red line for each contact. The plot already shows the dependence of the contact frequency on the physical distance between the ego and alter. The influence on face-to-face contacts is the strongest and decreases from phone to SMS and Email.

For modelling the frequency of contact we start with an empty model and include stepwise variables. First only fixed effects and random intercept models are considered, where first variables on Level 3 are included and then as well on Level 2. Further random slopes and also interaction effects within and between levels could be considered if significant and meaningful. The models are then compared by ANOVA to evaluate legitimate the additional variables. Finally, the final model will be interpreted and compared to the multilevel path analysis model from the previous paragraph.
As the Poisson model assumes that the covariates have variance equal to the mean, we test the model for this assumption on under- or over-dispersion with Pearson residuals. Where we check for if residual deviance is much different compared to the degrees of freedom. This is with 0.01 strongly the case, so we have to deal with under-dispersion. To deal with this, there are different strategies. Quasi-Poisson methods could be used, but these are not to prefer, as there are no likelihood functions available for them. Implemented approaches for multivariate multilevel models in software were not reliable. To treat over-dispersion one can introduce an observation level random effect, which is the same as using a log-normal mixing distribution, where the log-link is used for the Poisson distribution and the marginal distribution of the observation-level effects are assumed to be normal and models the excess variation (see e.g. Hinde, 1982). One can unfortunately
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

Figure 5.3.4: Trellis plot of the residuals of the Poisson multivariate multilevel linear regression model

not model under-dispersion the same way, as one would need to allow a negative “variance” estimate for the observation level random effects, as variance can not be negative. We examine the residuals of the Poisson multivariate multilevel linear regression model to test the model assumptions further. The model’s residuals are plotted in Figure 5.3.4 by the different contact means. This residual analysis shows strong heteroscedasticity in the low frequency, why we use instead of the Poisson model, a variance stabilizing transformation for count data with less rare observations (Box et al. 1978).

This makes the model linear and therefore easier for interpretation. As a rule of thumb, the results from a transformed linear model for count data have unbiased estimator if the mean number of observations is larger than 8 (Snijders and Bosker)
With a mean of 28, this criterion should be no problem in this case. The variance stabilizing transformation should be used to reach normality and equal variance, neither of which hold in most of the cases perfectly. However, simple power transformations often improve the situation considerably. The Box-Cox transformation \cite{box1964} is a scale and location shifted version of the power transformation, given by \( f_\lambda(x) = \frac{x^{\lambda} - 1}{\lambda} \) for \( \lambda \neq 0 \), with \( \log x, \lambda = 0 \). In this case, the optimal power is computed as \( \lambda = 0.09 \). This value is very close to the more simple to interpret log transformation, where \( \lambda = 0.1 \). (see Figure 5.3.5). We can visually confirm the success of this transformation using a \( Q - Q \) plot of the transformed values, shown in Figure 5.3.6. Figure 5.3.5 shows the optimal power for the model presented in Table 5.11 with the 95%-confidence interval for only the face-to-face contact frequency. Figure 5.3.6 shows the \( Q - Q \) plot for the optimal power transformation and the \( \log_{10} \) transformation. The two transformation results are very similar and only marginal different. Because of the interpretation advantage of the \( \log_{10} \) transformation, it is preferred and used for the following model prediction. To avoid zeros in the log transformation, the zero values are set to their corresponding minimum in each contact mean.

Based on the transformation analysis for the further models a logarithmic transformation is used. For the following results of the parameter estimates the regression coefficient can be interpreted as for a one unit change in the predictor variable, The difference in the logs of expected counts is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant.

In Table 5.11 all significant variables are shown and it is denoted from here on as the full model. The overall model goodness of fit statistics are highly significant. The bulk of the explanatory power is due to the Level 2 differences. The intraclass correlation \( \rho_I \) coefficient, defined as \( \rho_I = \frac{\gamma^2}{\gamma^2 + \sigma^2} \) where \( \gamma^2 \) is the Level 3 variance and \( \sigma^2 \) the variance at Level 2. It is the proportion of the variance that is accounted for by the group level. There is more variance on the Level 2 than on the Level 3 (\( \rho_{IF-to-F} = 0.39, \rho_{IFPhone} = 0.32, \rho_{IFEmail} = 0.43, \rho_{SMS} = 0.56 \)) for face-to-face contacts, Email and SMS. This means, that people in general are more homogeneous in their behaviour of how frequent they meet friends and relatives,
Figure 5.3.5.: Power transformation analysis
Figure 5.3.6.: $Q - Q$ plot of the optimal power transformation and the $\log_{10}$ transformation.
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

but very heterogeneous among their different alters.

The correlation between the different contact modes are as expected, where especially phone and SMS contact frequencies are positive correlated with face-to-face contact frequencies, as they are used to coordinate meetings and are immediate means of contact. Email is negatively correlated with face-to-face contacts and seems to be rather complementary to face-to-face contacts. Probably also because of its use for longer distances where time difference and language barrier can be influencing. But also as Email it is a different type of contact mode, much less personal and immediate, so is the correlation also very low.

Contact frequencies are decreasing significantly with an increasing network size. This confirms the idea of limited network capacities as each social relationship needs time to be maintained (Dunbar, 2003). In the models, only the variables which are significant at least on the 0.1 level on one of the contact modes are kept.

Still, all of the equations contribute significantly. At both levels the various modes of contact are complementary, except for E-mail. The socio-demographic effects are generally as expected: age is not significant for face-to-face contacts, but reduces the frequency of contacts for the phone in higher ages and especially for Email and SMS contact. This is expected, as older people did not grow up with the technology of these contact modes. Education doesn’t show a clear picture and isn’t significant. High income increases the number of face-to-face interactions, while reducing phone contact. Regular car availability increases face-to-face contact, as it lowers the marginal cost of travel. The number of previous move decreases the face-to-face contacts, as the spatial social network tends to be more scattered, as the number of moves increases. On the other hand the contacts with the phone increases slightly with the number of moves.

Table 5.12 shows the ANOVA between all the estimated models, which are presented in this paper. All the models are nested within each other and are pairwise compared, so that each following column presents its model fit and the comparison statistics of the model in the column above. The empty model and the model with fixed effects on level three in the first two columns. The model fit does improve. Whether the difference between model two and three is significant can be tested with a likelihood-ratio test: $D = -2(\log\text{Liknullmodel} - \log\text{Likalternativemodel})$
### Table 5.11.: Parameter estimates for full multivariate linear log-model with interactions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Face-to-Face</th>
<th>Phone</th>
<th>Email</th>
<th>SMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.30**</td>
<td>0.20</td>
<td>1.73**</td>
<td>0.33</td>
</tr>
<tr>
<td>Age_31-41</td>
<td>-0.17</td>
<td>0.17</td>
<td>-0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>Age_42-58</td>
<td>-0.12</td>
<td>0.18</td>
<td>-0.59*</td>
<td>0.30</td>
</tr>
<tr>
<td>Age_59-70</td>
<td>-0.21</td>
<td>0.19</td>
<td>-0.87**</td>
<td>0.31</td>
</tr>
<tr>
<td>Age_71+</td>
<td>-0.20</td>
<td>0.18</td>
<td>-0.58*</td>
<td>0.30</td>
</tr>
<tr>
<td>College</td>
<td>-0.05</td>
<td>0.15</td>
<td>0.30</td>
<td>0.24</td>
</tr>
<tr>
<td>&lt; 1999 sFr/month</td>
<td>0.06</td>
<td>0.18</td>
<td>-0.05</td>
<td>0.29</td>
</tr>
<tr>
<td>&gt; 6001 sFr/month</td>
<td>0.25*</td>
<td>0.14</td>
<td>-0.34</td>
<td>0.23</td>
</tr>
<tr>
<td>Number of relocations</td>
<td>-0.04*</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>GA Yes</td>
<td>-0.10</td>
<td>0.14</td>
<td>0.30</td>
<td>0.23</td>
</tr>
<tr>
<td>Car Available Yes</td>
<td>0.21*</td>
<td>0.12</td>
<td>0.26</td>
<td>0.20</td>
</tr>
<tr>
<td>degree</td>
<td>-0.02*</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Family Yes</td>
<td>0.29*</td>
<td>0.12</td>
<td>1.23**</td>
<td>0.19</td>
</tr>
<tr>
<td>ln distance</td>
<td>-0.38**</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Work Yes</td>
<td>-0.28**</td>
<td>0.10</td>
<td>0.47**</td>
<td>0.16</td>
</tr>
<tr>
<td>Relationship duration</td>
<td>-0.01**</td>
<td>0.00</td>
<td>0.01**</td>
<td>0.00</td>
</tr>
<tr>
<td>Family:ln_dist</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Work:ln_dist</td>
<td>0.11**</td>
<td>0.03</td>
<td>-0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Loglikelihood: -26968.33

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_h=Ego:Alter$</td>
<td>FtoF</td>
<td>0.93</td>
<td>0.96</td>
<td>$(\rho(R_{1ij}, R_{2ij}))$</td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td>3.93</td>
<td>1.98</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Email</td>
<td>3.93</td>
<td>1.98</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>SMS</td>
<td>3.33</td>
<td>1.83</td>
<td>0.25</td>
</tr>
<tr>
<td>$\gamma^2_h=Ego$</td>
<td>FtoF</td>
<td>0.66</td>
<td>0.81</td>
<td>$(\rho(U_{1j}, U_{2j}))$</td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td>1.85</td>
<td>1.36</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Email</td>
<td>1.66</td>
<td>1.29</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>SMS</td>
<td>2.35</td>
<td>1.53</td>
<td>0.21</td>
</tr>
</tbody>
</table>
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

Table 5.12.: ANOVA of all linear models

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>BIC</th>
<th>logLik</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>P(&gt;Chisq)</th>
<th>Pseudo R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base model</td>
<td>25</td>
<td>59,330</td>
<td>-29,546</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Only Level 3 variables</td>
<td>45</td>
<td>54,567</td>
<td>-27,071</td>
<td>4,950</td>
<td>20</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Full model w/o interaction effects</td>
<td>85</td>
<td>54,778</td>
<td>-26,988</td>
<td>167</td>
<td>40</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Full model</td>
<td>93</td>
<td>54,814</td>
<td>-26,968</td>
<td>39</td>
<td>8</td>
<td>0.00</td>
<td>0.09</td>
</tr>
</tbody>
</table>

The empirical value for \( D \) has to be compared to the theoretical value of the Chi-square distribution for \((Df1-Df2)\) degrees of freedom. All model improvements discussed above are significantly improving the overall fit of the model.

We examine the residuals of the full model to test the model assumptions further. The histogram for the model’s residuals are plotted in Figure 5.3.7. The residuals seem to be roughly normal distributed. Figure 5.3.8 shows the fitted versus the predicted residuals. The variance seems to be more or less stable. The line like shape of the residuals is coming from the count like data.

To visualize the effects of the model the model predictions are printed against the distance for different contact means. The predicted values assume average values for the not plotted variables. The confidence intervals are drawn based on the per-point variances. and the limits are based on ±1.96\( σ \). The predicted values are shown in Figure 5.3.9. It is clearly visible how strong the distance impact is on the face-to-face contact frequency, also the 95%-confidence band is very tight, which indicates the overall good model fit for physical contacts. SMS contact frequency is overall not very strong, but decreases nearly parallel with the phone contact frequency. Whereas the face-to-face contact is the dominant form in terms of frequency for shorter distances, the phone contacts become dominant at around 60 km. Note here also, how similar the model predicts the different curves compared to Figure 5.3.1. Email is the only contact form which increases with distance in absolute terms. It surpasses face-to-face frequency at around 1,000 km, but stays behind phone contact frequency.
Figure 5.3.7: Histogram of the residuals of the model presented in Table 5.11, for all contact modes and trellis histogram by contact mode.
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

Figure 5.3.8.: Trellis plot of the residuals of the full model presented in Table 5.11
Figure 5.3.9: Predicted contact frequency per year plus the 95% confidence interval against distance from the model presented in Table 5.11. The left part of the Figure shows the contact frequency for ties within 1 km, whereas the whole distance spectrum is shown in the right part.
Comparison between the multivariate multilevel linear regression model and the multilevel path model

Table 5.13 compares the result from the multivariate multilevel model with a similar specified structural equation model. The model results are overall fairly consistent in the sign of the effect. But the standard errors and the resulting significance is often different, which can be a problem with path models (see the discussion above).

5.3.3. Conclusions

This chapter analysed three ego-centric network properties with uni- and multivariate empirical analyses techniques. In the first detailed analysis the ego-centric social network degree was analysed. With regard to the number of relationships it is shown that younger people maintain more social relationships. Also, people with public transit passes maintain more social contacts. This also applies to persons with frequent residence changes, higher education, full-time employees and persons with children in their household.

Further, the spatial properties of ego-centric social networks were analysed. In terms of network geography shows that young people, a high level of education, car ownership, educational and occupational change of residence are associated with a geographically more expansive networks.

Geographical distance influences how people interact today, as it did in the past (Mok and Wellman, 2007). The effort involved in the different modes of communication (face-to-face, phone, email and SMS) matches the strength of distance decay in the frequency of those interactions. The patterns of the shares follow this differential distance decay. The frequency of face-to-face visits and their share falls quickest with distance.

While email frequency is unaffected by distance, its share rises fastest with it. The reverse is true for SMS messaging, which stays stable in terms of share, while only slowly falling with distance in terms of frequency. The high frequency of phoning at short distances translates into growing shares, even if the absolute frequencies fall
Table 5.13.: Model comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Multilevel Mixed Model</th>
<th>Multilevel path model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Mode:</td>
<td>Face-to-face</td>
<td>Phone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Email</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face-to-face</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Email</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SMS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.18**</td>
<td>1.79**</td>
<td>-0.35</td>
<td>1.74**</td>
<td>3.30**</td>
<td>2.13**</td>
<td>0.74*</td>
<td>1.10**</td>
</tr>
<tr>
<td>duration</td>
<td>-0.01**</td>
<td>0.01**</td>
<td>-0.01**</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01**</td>
<td>0.00*</td>
<td>0.00*</td>
</tr>
<tr>
<td>WorkmateYes</td>
<td>-0.03</td>
<td>0.45**</td>
<td>0.71**</td>
<td>0.44**</td>
<td>-0.14*</td>
<td>-0.10*</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>FamilyYes</td>
<td>0.37**</td>
<td>1.04**</td>
<td>0.30*</td>
<td>0.46**</td>
<td>-0.07</td>
<td>0.16**</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>ln_dist</td>
<td>-0.35**</td>
<td>-0.03</td>
<td>0.15**</td>
<td>-0.11**</td>
<td>-0.11**</td>
<td>-0.04**</td>
<td>0.00</td>
<td>-0.03**</td>
</tr>
<tr>
<td>Gender Male</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.30*</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Moves</td>
<td>-0.04*</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.09*</td>
<td>0.06*</td>
<td>-0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td>College degree</td>
<td>0.14</td>
<td>0.24</td>
<td>-0.04</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highschool</td>
<td>0.42</td>
<td>0.56</td>
<td>1.24*</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary School</td>
<td>-0.18</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC1999Yes</td>
<td>0.06</td>
<td>-0.05</td>
<td>-0.35</td>
<td>0.13</td>
<td>0.25</td>
<td>-0.11</td>
<td>-0.46*</td>
<td>0.08</td>
</tr>
<tr>
<td>INC6001Yes</td>
<td>0.26*</td>
<td>-0.33</td>
<td>-0.13</td>
<td>-0.25</td>
<td>0.91**</td>
<td>-0.16</td>
<td>-0.31*</td>
<td>0.02</td>
</tr>
<tr>
<td>Car Avail.Yes</td>
<td>0.22*</td>
<td>0.26</td>
<td>0.08</td>
<td>-0.13</td>
<td>0.40*</td>
<td>0.07</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td>GAYes</td>
<td>-0.09</td>
<td>0.30</td>
<td>0.46*</td>
<td>0.05</td>
<td>-0.49**</td>
<td>-0.11</td>
<td>-0.05</td>
<td>-0.07</td>
</tr>
<tr>
<td>Age_31-41</td>
<td>-0.16</td>
<td>-0.08</td>
<td>-0.61*</td>
<td>-1.28**</td>
<td>0.28</td>
<td>0.31</td>
<td>0.30</td>
<td>0.62</td>
</tr>
<tr>
<td>Age_42-58</td>
<td>-0.11</td>
<td>-0.58*</td>
<td>-1.05**</td>
<td>-1.96**</td>
<td>-0.05</td>
<td>0.24</td>
<td>0.32</td>
<td>-0.02</td>
</tr>
<tr>
<td>Age_59-70</td>
<td>-0.19</td>
<td>-0.87**</td>
<td>-1.46**</td>
<td>-2.43**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age_71+</td>
<td>-0.19</td>
<td>-0.58*</td>
<td>-2.17**</td>
<td>-3.23**</td>
<td>-0.44*</td>
<td>0.27</td>
<td>-0.31*</td>
<td>-0.52**</td>
</tr>
</tbody>
</table>

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_{\lambda}=$  Ego:Alter</td>
<td>FtoF</td>
<td>0.34</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Phone</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Email</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>SMS</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.37</td>
<td>0.20</td>
</tr>
</tbody>
</table>

| $\gamma^2_{\lambda}=$ Ego     | FtoF       | 0.11      | 0.14      |
|                               | Phone      |           |           |
|                               | Email      | -0.10     | 0.19      |
|                               | SMS        | 0.20      | 0.32      |
|                               |            | 0.32      | 0.14      |
|                               |            | 0.46      | 0.53      |
|                               |            | 0.53      | 0.33      |

136
5.3. SOCIAL CONTACTS IN A SHRUNKEN WORLD

with distance. While other characteristics of the respondent-contact dyad influence frequency and shares (duration of the relationship, context of the initial acquaintance), these are less influential in absolute terms than the socio-demographics of the person, especially income, age and the number of moves stand out among them. The impacts of the characteristics of the ego-alter dyad varies systematically with type of relationship, which reflect the underlying social expectations for relatives, friends and work mates.

Further was shown, that the interaction of the different contact modes and Face-to-Face meetings are complementary.
6. Extrapolation and experiments of geographical embedded social networks

This chapter is with permission of the co-authors an updated and extended version of parts of Frei and Axhausen (2011); Frei (2011); Frei and Axhausen.

This chapter explores collective decision-making in social networks applied to a model of activity location choice for leisure activities and the participation in joint activities. Social networks pertain to the decision of resource allocation in planning and executing leisure activities, the manifestation of such decisions in the travel patterns of daily lives and the less frequent journeys outside the daily activity space and therefore social network structures have an important influence on a realistic representation of travel behaviour. A real practical experiment to use the spatial attributes of social networks in demand or supply models is based on the available data in this thesis not possible. Nevertheless, the empirical work presented in the previous chapters is a further step into practical applications of spatial and other discussed attributes of ego-centric social networks. How the empirical data could be of use in further studies is already discussed in the framework presented in Chapter 2. Anyhow, this chapter builds on the framework idea and uses some of the explored social network properties to build experiments to show, what the effect of including social network data into demand models could be and how they could improve those models, as well as what current modelling practices are missing with not including social network effects.
6.1. Models of social networks

The driving force behind social network structure is the homophily, which has been shown in many empirical studies (see McPherson et al., 2001). For example a big part of friendships’ origin is work or family, the so called “inbreed homophilie”. Born into a family, we share automatically many characteristics, such as race and religion. With our friends from work we often share characteristics such as social class, education. All those characteristics are not exclusive characteristics, which means that they do not limit our active choice. Many models of social networks are based on these characteristics (e.g., van Duijn et al., 2004; Newman, 2003).

But much more important than the “inbreed homophily” is the homophily which actually limits directly the pool of potential ties by geographic space. Many studies show a strong relationship between physical distance and social structure Merton (e.g., 1948); Festinger et al. (e.g., 1950); Caplow and Forman (e.g., 1950); Whyte (e.g., 1956); Sommer (e.g., 1969); Snow et al. (e.g., 1981); Latane et al. (e.g., 1995); Wellman (e.g., 1996); Wellman et al. (e.g., 2001).

Previous research on the detailed spatial distribution of social networks is very sparse, and the spatial properties of social network as discussed in the previous chapter is to the authors knowledge the most precise and complete till today.

Research on models of spatially embedded networks is even more sparse than literature on the analysis of the spatial dimension of social contacts. Hackney and Marchal (2009) propose a model which is based on the idea that there is a certain probability that people become friends if they remain at the same place in an overlapping time interval. To extract the trajectories of peoples’ mobility patterns, they use a microscopic traffic simulation, which simulates daily traffic in the greater area of Zurich, Switzerland. Hackney and Marchal find that distance distribution of social contacts is heavily right-skewed, but they provide no detailed analysis on the shape of the distribution.

Besides the model of Hackney and Marchal, there are several social network models, which are based on probability, but which are not especially used to model geographically embedded social networks. However, Toivonen et al. (2009) classifies these social network models in two main categories: those were the link
6.1. MODELS OF SOCIAL NETWORKS

updates depends on local network structures (network evolution models, NEMs) and those were the update of the links is dependent on the attributes of the vertices or nodes (nodal attribute models, NAMs). A third family of model, which can be interpreted as NEM or NAM, is the exponential random graph model (ERMG). The difference between NEMs and ERMGs are that the ERMGs do not include a link update step based on probabilities. The ERGMs link existence is defined by the probability distribution. Markov Chain Monte Carlo (MCMC) sampling methods, which are used in ERMGs, can also be used to model the evolution of social network, as proposed by Snijders (1996).

All of these models generate non bipartite, social networks, which are not directed.

Since the spatial dimension of a social network is defined as an attribute of each link and not as a network structure, a recent NAM (Boguna et al. 2004) (BPDA) and an ERGM (Wong et al. 2006) (ERGM) are studied following in more depth. Both approaches, the BPDA and ERGM will be tested to generate spatial social networks, to see, which approach fits the needs of this research better in terms of reproduction of spatial patterns obtained in a the Zurich survey and in general how they produce typical social network structures. These typical network structures, which are observed in social networks are:

- Large clustering $c_i = \frac{2e_i}{k_i(k_i - 1)}$, where $e_i$ represents the number of triangle connected to vertex $c_i$ and $k_i$ is $c_i$’s degree. The denominator represents the possible triangles.

- Positive degree correlation, which means that the degree $k_n$ is positively correlated with the average degree of the nearest neighbours of $n$: $\text{cor} [k_n, \overline{k_{nn}}(k)] > 0$.

- The existence of communities; In many cases it was shown that clustering does not occur evenly over the entire network: Subgroups of actors who are highly connected within themselves are loosely connected to other subgroups, which are themselves highly connected.
CHAPTER 6. EXTRAPOLATION AND EXPERIMENTS OF GEOGRAPHICAL EMBEDDED SOCIAL NETWORKS

6.2. Model of geographically embedded social networks

6.2.1. Basic assumptions

The simulated social networks represent always loop-less undirected social networks \( G = (N; H) \) with a known vertex set \( N \) and a set of links \( H \). \( G \) is spatially embedded, in the sense that a space \( S \) is defined and the set \( N = \{n_1, ..., n_N\} \) such that \( N \subset S \). The distance function, \( d \), on \( S \), is the great circle distance in kilometres. The models’ specification include \( x \) and \( y \) coordinates for each Node \( n_i \), which may cause boundary problems, which will not be analysed in this paper. A less critical model could be generated by assigning the space in one dimension by just modelling the distance between each node, instead of assigning each node \( x \) and \( y \) coordinates. But for further usage of the model and its network structure implications on activity/location choice, the model calculates distances by assigning coordinates to each node in a physical space. For simplicity a population of size \( N = 750 \) is placed uniformly within a square region of size \( l \times l = 100 \text{km} \). Such a model may be thought of as a first approximation to a very sparse population distribution in physical space over large areas. The models are configured to generate social networks in physical space, with characteristics to represent the Zurich data [Frei and Axhausen, 2007], with a mean degree of 12.36 and a edge length distribution flowing an exponential distribution \( f(x; \lambda) = \lambda e^{-\lambda x}, x \geq 0 \) and \( 0, x < 0 \) with \( \lambda = 0.06 \).

6.2.2. Two dimensional BPDA

The BPDA model rules the establishment of friendship in a way that individuals with short distance have a large probability of being connected, and vice versa for long distances. Each Node \( i \) has a location \( h_i \). Connection between each pair of individuals, \( h_i \) and \( h_j \) is given by: \( r(h_i, h_j) = \sum_{n=1}^{N} \omega_n r_n(h_i^n, h_j^n) \), where \( \omega_n \) is a normalized weight factor and \( r_n \) is the probability for a friendship given by:
6.2. MODEL OF GEOGRAPHICALLY EMBEDDED SOCIAL NETWORKS

\[ r_n(h^n_i, h^n_j) = \frac{1}{1 + [\beta_n^{-1} d_n(h^n_i, h^n_j)]^{\alpha_n}} \], where \( \beta_n \) is a characteristic length scale (that controls eventually the average degree) and \( \alpha_n > 1 \) is a homophily measure, the tendency people connect to similar people. \( n \) is the number of possible links: \( i \ast (i - 1) \).

To generate the social network two parameters alpha and beta have to be adjusted to reproduce the empirical Zurich data. For optimizing alpha and beta to reproduce the right-skewed distribution and the degree distribution, an incremental search is used to minimize the deviance of the generated and empirical link distance distribution within six bins, weighted by cases, as well as the mean degree of social contacts. The resulting, chosen social network model configuration is given by \( \beta = 2.25 \) and \( \alpha = 1.99 \). The resulting distance distribution are visualized in Figure 6.2.1. The target distribution can be very adequately reproduced.

The simulated network meets also the the required criteria for a social network. The network has a large clustering coefficient of 0.42 with a low number of total ties (9450 of 561,750). The degree of correlation is positive and relative high (0.42).

6.2.3. ERGM

The ERGM model describe the local selection forces that shape the global structure of a network. The probability of a tie is described as

\[ P(Y = y | X) = \frac{\exp[\theta^T g(y, X)]}{\kappa(\theta)}, \]

where \( Y \) is random set of relations (edges and non-edges), \( y \) is a particular given set of relations, \( X \) is the matrix of attributed for the vertices, \( g(y, X) \) the vector of network statistics, \( \theta \) vector of coefficients and \( \kappa(\theta) \) normalizing constant. The log-odds of any given edge given the current state of the rest of the network is described bye the logit function:

\[ \text{logit}(Y_{ij} = 1) = \theta^T \delta[g(y, X)]_{ij}, \]

where \( Y_{ij} \) is an actor pair in \( Y \), and \( \delta[g(y, X)]_{ij} \) is the change in \( g(y, X) \) when the value of \( y_{ij} \) is toggled from 0 to 1. This is in this simple case the fraction of possible edges which are realized. Since we use for this paper the edge distance as the local selection force, \( g(y, X) \) contains the distance of the edges and the \( \theta \) vector of coefficients is given by the parameters for the exponential distribution. To simulate a network, the Gibbs-sampling method is used, where from a constant initial set of edges \( G \) of all possible edges \( N \) are chosen, to generate the desired degree.
Further is in each step one edge of $H$ and one edge of $G - H$ is randomly drawn and if $\frac{p(G-H)}{p(G-H) + p(H)} \geq U(0, 1)$ the set of $H$ is updated. As the probability to chose an edge with a certain distance is not uniform, the probability to chose a certain edge is therefore weighted by the inverse kernel density estimation for each edge: $\frac{\text{max(density(distance))}}{\text{density(distance)}}$. Candidates of edges to include in $N$ are $10^7$ times sampled. Which is for the size of the Network enough long, as the distance distribution did no more change after about $10^6$ samples. The resulting distance distribution is visualized in Figure 6.2.1. The target distribution can be very adequately reproduced, but deviates a little bit from the exponential distribution, which is targeted, due to the weighting procedure, which is not 100% accurate, as the weighting is done in bins of finite length of 0.5 km. Nevertheless, if the edge length distribution of the initial whole network were uniform, the simulated network would follow exactly the desired exponential distribution.

The simulated network meets also the the required criteria for a social network. The network has a large clustering coefficient of 0.38 with a low number of total ties (9446 of 561,750). The degree of correlation is positive and relative high (0.38).

Figure 6.2.1.: Empirical data of the edge length distribution, exponential distribution and simulated distribution
6.2. MODEL OF GEOGRAPHICALLY EMBEDDED SOCIAL NETWORKS

6.2.4. Results and discussion

This Section presents two stochastic models for spatially embedded social networks. Empirical data show that the edge probability does not follow a simple distribution, because of geographical reasons and the tendency of people to cluster in cities. However, an exponential distribution based on the edge length with a parameter of 0.060 can approximate the edge distance distribution. The paper shows that the observed exponential distribution can be explained with a relatively simple homophily model where the probability of an edge can be modelled as a function of the distance between two persons. With a simulation on generated vertices, we demonstrate that we can reproduce the observed edge length distribution with the two proposed models. The simulated network shows a right-skewed degree distribution and high degree correlation, which are both typical properties for social networks. Where as both models perform good on reproducing the edge distance distribution, the ERGM approach has to be favoured, as it allows to incorporate further explanatory variables.

In the context of transportation research, the great circle distance is probably not the appropriate measurement for the spatial effects between two persons. Especially regarding topology e.g. mountains etc. have a strong influence on the selection process. To include the real cost, which one has to be overcome, to maintain an edge, travel time should be used as a measurement instead of distance. However, distance could easily be replaced by travel time in both of the suggested models.

For a variety of reasons, it may be interesting to have a complete model of all social contacts of a region or a country, as it is suggested in this paper. Possible applications are the location choice of leisure traffic or visiting friends and relatives. Dynamic models such as the ones by Hackney (Hackney and Marchal, 2009) are able to generate similar results. But statistical measures such as distance, clustering coefficients etc. are very difficult to control. Statistical modelling to generate social networks can produce robust static results, which can then be used for further models, which represent short to mid time decisions, which do not affect the social network immediately.
CHAPTER 6. EXTRAPOLATION AND EXPERIMENTS OF GEOGRAPHICAL EMBEDDED SOCIAL NETWORKS

6.3. Experiments

Activities and travel involving multiple persons, particularly in leisure travel, result from a collective decision process that requires its participants to find a way to allocate and arrange their individual needs along with those of the other person(s). On an abstract level, this means finding an appropriate modelling structure to represent the utility derived by each individual from participating in leisure activities and the collective effects of these individual utilities on joint decisions, such as the participation in a sports training with your favourite training partner(s) further away or participating in a training with less preferred training partner(s) at a closer location.

At a more basic level, this research involves explaining travel behaviour patterns which are only understandable when collective decision making is used as a base process. For example, the leisure trip distance distribution in Switzerland has a mean of 11.1 km [ARE/BfS 2007b]. The nearest leisure activity location of each leisure category has a mean of 1.1 km in a spatially detailed agent based micro simulation of the Swiss overall work day [Horni et al. 2009]. Even though the categories of leisure activities are aggregated and do not consider the differences of those locations, the difference between the nearest leisure activity location and the travel patterns which evolve from the actual activity location cannot be explained by considering only the choice of each single person independent from one another. The consideration of a collective location choice between individuals participating in an activity together and considering the distance distribution of the social contacts (Within 120 km the distance distribution between egos and their alters has a mean of 16.5 km [Frei et al. 2009] could shed light on this difference.

This Experiment aims to model the process which leads to a joint location choice for leisure activities. To do so three different steps have to be accomplished - a spatial social network structure, a process to describe the organization within this network of smaller activity-groups and a joint location choice model.

The first research objective is accomplished by generating a static geographical embedded social network. To do so, an exponential random graph model is used to reproduce the edge distance distribution observed in an empirical study in Zurich,
6.3. EXPERIMENTS

Switzerland. For further details of the method used to generate the social network, see the previous Section.

The second research objective is the social organization within the social network. A joint activity participation model will be used to model the participation of each individual in a certain activity with others and to build an activity group. A randomly selected group of individual can find or invent a focus around which to combine various others with whom they are tied in an activity. The focus gets smaller (larger), with decreasing (increasing) number of participants (or individuals in the social network of interest). The foci are substitutes for each other.

The third research objective is to model the collective or joint location choice. To choose a location to perform a specific activity, a utility aggregation function will be introduced (preference function). The utility for location \( a \), \( U_a = f(\alpha_i, U_i) \), is an aggregated function over all \( i \) participants in an activity. \( \alpha_i \) represents a weight on the individuals’ utility \( U_i \), which are among others’ network attributes.

**6.3.1. Joint activity participation organization model**

To be able to compare the model of a joint activity participation model, empirical data would have to be needed, which shows at a minimum the group size distribution. Further, the composition of these groups would be interesting, e.g. how many of the participants are friends with each other. Such data is to the knowledge of the authors not yet available. But at least the participant numbers should be higher than the number of accompanying travellers in Figure 6.3.1.
Modelling of joint decision structures has been done in the past mostly by interactions between household heads. Fujii et al. (1999) modelled the allocation of time to in-home and out-of-home activities with other family members, with non-family members and alone using a production function paradigm. Golob and McNally (1997), Simma and Axhausen (2001) and Carrasco and Miller (2006) modelled interactions in a structural equation format, simultaneously accounting for the allocation of time to activities. Zhang et al. (2002) expressed household utility as an aggregation of the linear utilities of individual household members, weighted by the gender of the decision maker, using simultaneously unrelated regression (SUR) to derive the optimal level of time allocation to in-home, out-of-home indi-
6.3. EXPERIMENTS

individual and out-of-home shared activities. A similar and more extended approach is used by Gliebe and Koppelman (2000). Based on the idea of Townsend (1987) and Bhat and Koppelman (1993), they propose a utility-based model to explain the decision to allocate time to both joint and independent activity types. They offered an empirical model, structured as a two-level nested share model, in which two household heads make an 28 upper-level choice between joint maintenance, joint leisure or independent activity time allocation (see also Gliebe and Koppelman, 2002). The independent activity is partitioned at the lower level into choices between at-home, subsistence, independent maintenance and independent leisure time allocation. The composite utilities of the independent activities of the individual decision makers are weighted using a parametrization based on their attributes, such as employment level.

Townsend (1987) forms the basis for the approach taken in this research. He developed a concept for the analysis activity patterns of households and their members. Household utility maximization is combined in an aggregation function of each individual’s preferences. In his empirical research he split the utility in three components related to interpersonal and personal parts to trade-off in activity and travel patterns: Efficiency, companionship and power/altruism. Efficiency is linked to the specialization and the efficiency gained through it. Companionship is the need of people to engage in activities jointly, and power is illustrated through e.g. higher income and therefore a higher influence in decision making. But power is also illustrated as a form of altruism, where a person engaged in an activity with a less powerful individual by helping them or fulfill other altruistic tasks in the free time to gain esteem.

Gliebe and Koppelman (2000, 2002) extended the approach of Townsend (1987) with the concept of Bhat and Koppelman (1993) to generate household activities with three different types of activities: consumption, work, household (maintenance) and leisure (discretionary) activities. The total utility of a household is defined as the joint utility of the household members e.g. $U_T = f(\alpha_iU_i + \ldots + \alpha_jU_j)$, where the $\alpha$’s are weights of each individuals $i$ to $j$. The utility of each household member has two components. The first component is a combination of $V_i = \beta_{bi} + \beta_{bi}C_i + \beta_{b2i}S_{bi} + \beta_{b3i}S_{wi} + \beta_{b4i}S_{hi} + \beta_{b5i}S_{lij} + \beta_{b6i}S_{hij}$, where $C_i$ denotes
the consumption and $S_{xi}$ denotes the satisfaction of each leisure, work and household. The joint components reflect the degree of positive utility each individual gains from joint activity participation and are modelled as taste variations across individuals. The second component, the altruism effect, describes the total utility together with component one as follows: $U_i = V_i + \beta_7 V_{j,...,k-i}$, where $V_{j,...,k-i}$ describes the utility of all others participating in the joint activity. The satisfaction terms are modelled as a function of time spent, e.g. $S = nT_x - T_x^2$ where $S$ convex. and the total time is is limited by $T_0 \leq T_w + T_h + T_l$.

This research concentrates on leisure activities only, but the context of joint household activities will be extended to general joint leisure activities. The influence factor $\alpha$ in the aggregation function will be extended and the influence which is described by [Townsend (1987)] and [Gliebe and Koppelman (2000)] as a power relative to income and gets replaced with the approach described by [Carrasco and Miller (2006)], where the structural influence will be tested with different approaches. Different structural measurements are reviewed in the following paragraph.

**Network Structural Centrality Measurements**

Network structural centrality measurements are described e.g. in [Wasserman and Faust (1994)], [Hanneman and Riddle (2005)] or in [Monge and Contractor (2003)]. The basic idea is to measure the relative importance of each individual in a social network. There are four different approaches which are mainly used to measure the centrality, and each different measurement has a different meaning.

**Degree Centrality** The degree centrality is the simplest measurement and describes the number of total direct neighbours of each individual relative to the total network size: $C_D(v) = \frac{\text{deg}(v)}{n-1}$. The degree centrality is interpreted as the intermediate probability of an individual to receive something which is flowing through the network, e.g. the risk to catch a virus, or the probability to receive certain information.
Closeness Centrality  Degree centrality measures do only account for the immediate neighbours. A critical example of the degree centrality approach could be, that an individual might have a high degree, but all those individuals he is tied to are isolated from the rest of the network and so he is not very central even though he has a high degree. There are several measurements for closeness centrality. The most common two measurements are the mean path distance and the eigenvector of geodesic distance. The mean path distance measures the mean shortest distance $d_G$ in social space between individual $v$ and all other members of network $V$:

$$C_C(v) = \frac{\sum_{a,t\in V\backslash v} d_G(v,t)}{n-1}.$$  But similar to the degree centrality, the mean path measurement can be somewhat misleading in larger networks, as e.g. an individual with a short path to many members of a disconnected part of the network and long paths to the rest of the networks, can have a similar measurement as e.g. an individual with a medium distance to all members. It can be argued that the second individual described is more central in the sense, that he can reach more other individuals with the same amount of effort. The eigenvector is an approach to find a measurement on the overall network structure with paying less attention to local structures (e.g. Google’s rank page uses some kind of eigenvector centrality). The eigenvalue score of $v_i$ is $x_i$, where $A_{ij}$ is the adjacency matrix. The centrality score for $v_i$ is proportional to the sum of the scores of all his neighbours:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x_j,$$

$N$ is the total number of individuals in the network. There might be different eigenvalues $\lambda$ for which a solution to this equation exists. As a centrality measurement, we are only interested in the greatest eigenvalue.

Betweenness Centrality  The betweenness centrality measures positional advantages, or power of an individual $v$, such that the shortest path from an other individual to an other individual goes through individual $v$. The idea is that individuals who are between others, might be able to use this as some form of power.

The betweenness centrality is the measurement of how many of the shortest paths $\sigma$ between other individual of the network $V$ go through individual $v$: $C_B(v) =$
6.3.2. Collective choice model

The implementation of a joint activity and location choice model is grouped in three hierarchically dependent models: A social network model \((a)\), a joint activity participation model \((b)\) and a joint location model \((c)\). The dependence is static and hierarchical in the sense of a fixed and timely order, where \((b)\) depends on \((a)\) and \((c)\) depends on \((a)\) and \((b)\): \[(a) \rightarrow (b) \downarrow \downarrow \quad \rightarrow (c)\]. The timely, hierarchical dependence is not dynamic and will therefore not be true for certain situations, where these three structures/decision situations have dynamic interactions among each other on different levels.

For example, a certain activity will be uniquely located on a short or mid-term horizon, and the activity participation is therefore a function of the location. On a longer horizon, the social network might be influenced by the location choice (e.g. when people decide to move). In general this model structure assumes that the social network is relatively stable over time. The mean duration for social contacts is estimated based on a survival model, which is modelled with a hazard function describing the exponential failure density of the Zurich data (based on a small sample, see [Ohnmacht and Axhausen, 2005]) to be around 8 years (see Figure 6.3.2) and therefore influence of short term decisions should be negligible. The dependence of the joint location model on the activity participation model, and not vice versa, is chosen because the pool of possible locations for a certain activity is normally much bigger than the pool for possible activities for a certain location. For example, the choice to go to a film gives us a relatively high flexibility in location choice (in bigger cities), whereas if we chose to go to a film theatre, the activity options are very limited. The argument here is that the model process should reflect the choice process while maintaining high flexibility.
6.3. EXPERIMENTS

Figure 6.3.2.: Survival function of the empirical duration of social contacts

Source: Own calculation based on a small sample of respondents in Zurich (see Ohnmacht and Axhausen 2005)

Basic assumptions

The utility function for the joint activity participation model and the collective choice model is a modification of the utility function proposed by Gliebe and Koppelman (2002) and Townsend (1987). $U_T = f(\alpha_iV_i + \ldots + \alpha_jV_j)$, where the $\alpha$'s are weights of each individual $i$ to $j$. The utility of each person is $V_i = \beta_{i1}S_{i1} + \beta_{2i}S_{i1, for all H=1} - \beta_{3ij}S_{i1, for all H=0}$, where $S_{i1}$ denotes the satisfaction of each leisure activity and $H = 1$ means that person $i$ is connected to person $j$ and
vice versa, if $H = 0$. The joint component $S_{ij}$ reflects the degree of positive utility each individual gains/loses from joint activity participation.

**Joint Activity Participation Organization Model**

The static social network generated by the model described in Frei and Axhausen (2011) is used to organize the population in activities.

The idea of the joint activity participation model is that a joint activity $A_i$ is more satisfying than an activity performed alone. A joint activity is defined as $A_i\{n_i, \ldots, n_j\}$, whereas as the $\sum H \in A_i \geq 1$. Each $n_k \in A_i$ with $H = 0$ has a negative impact on the activity utility. This is chosen, because of the static social network, where it is assumed, that the probability of $n_i$ knows $n_k$ is high, but they are not maintaining a friendship.

The joint activity participation choice cannot be evaluated directly, because the simultaneously decision of each person affects the decision of the others. To get a stable result from the organizing process within the network to participate in a certain activity an iterative approach is chosen as follows:

1. 50% of the population are randomly picked to initiate activities.
2. Each person which has a direct connection to one of the persons who initiated an activity will obtain the information about this activity.
3. An initial choice for participating in one of the activities is made based on the physical distance to the person who initiated the activity.
4. Iterate: In each iteration 10% of the population is randomly chosen to make a new choice based on all of the available information, which includes the possibility a choice based on an other person participating in an activity, who then changes the activity.

The iteration process will be continued till the overall utility stabilizes. The utility for each person is calculated by $V_i = \beta_0i + \beta_1S_{i} + \beta_2S_{ij}, for all H=1 - \beta_3iS_{ij}, for all H=0$. 

154
6.3. EXPERIMENTS

Choice Model

The collective choice model is based on the generated social network and the activity participation model. Each leisure activity $A$ can be performed at each location $L = \{l_1, \ldots, l_i\}$, as there is no further categorization of leisure activities in this model. Each location $L$ is spatially fixed within $S$. The location choice for an activity $A$ is defined by the group utility $U_T = f(\alpha_i U_i + \ldots + \alpha_j U_j)$, where $U_i = V_i + \beta_4 V_{j_{k-i}}$ and $V_i = \beta_1 S_{lj} + \beta_2 S_{l_{ij}, for all H=1} - \beta_3 S_{l_{ij}, for all H=0}$. The weights $\alpha$ are network structural measurements as described on page 150. The choice set of activity locations is created by including each in the activity participating persons three closest locations. This can account for the effects that not all information about locations is available for everyone, but that the groups information is based on its members.

Experiment results: Distance comparison

The stop criterion for the joint activity participation model was set to 2.5% relative change in the sum of the utility of the population $\sum V_i$. It took on average 273 iterations per run to reach the criterion. As there are no group size distributions for leisure activities known to the author, the resulting mean, median, std. deviation of the group size and mean distance between members of a group are calculated for different proportions of the parameters of the utility function. The utility function $V_i = \beta_0 + \beta_1 S_{li} + \beta_2 S_{l_{ij}, for all H=1} - \beta_3 S_{l_{ij}, for all H=0}$ consists of three parameters, where the intercept $\beta_0$ is set to zero. The satisfaction terms $S_{1i,2i,3i}$ are set inverse proportional to the distance to the activity initiator and multiplied by ten. The parameter $\beta_1$ is set to one. The parameters $\beta_2$ and $\beta_3$ are set so that the sum of each pair equals to one: $\beta_2 + \beta_3 = 1$. The results for different parameter proportions are presented in Table 6.1. It is not surprising, that the group size depends very strongly on the proportion of $\beta_2$ and $\beta_3$, as they affect the total utility each gets from the participation of others in the activity. But it also shows, that through the effect of a negative utility each one gets from an $H = 0$, the group sizes are substantially smaller than the mean degree of the underlying social network. For group sizes equal to one, the distance was set to 1.1, to make it
comparable to the results in Table 6.2.

Table 6.1.: Resulting group organization by parameter proportion

<table>
<thead>
<tr>
<th>$\beta_3 / \beta_2$</th>
<th>Group size</th>
<th>Mean [-]</th>
<th>Median [-]</th>
<th>Std. dev. [-]</th>
<th>Mean distance within group members [km]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8.6</td>
<td>5.6</td>
<td>3.4</td>
<td>19.8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6.6</td>
<td>5.2</td>
<td>3.0</td>
<td>16.6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.8</td>
<td>5.1</td>
<td>2.2</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.2</td>
<td>3.8</td>
<td>0.5</td>
<td>11.8</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2.3</td>
<td>2.6</td>
<td>0.2</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>0.33</td>
<td>2.1</td>
<td>2.4</td>
<td>0.2</td>
<td>8.0</td>
<td></td>
</tr>
</tbody>
</table>

For the choice model, a location set was randomly generated and placed uniformly within the square region of size $l \times l = 100 \text{km}$, so that the mean physical distance for each vertex to the closest location is 1.1 km. The group’s utility for the location choice is calculated as $U_T = f(\alpha_i V_i + \ldots + \alpha_j V_j)$, where the $\alpha$’s are normalized weights represented by the network structure centrality measurements and the $V_i$’s are calculated in the same way as in the joint activity participation model. The resulting descriptive statistics for the proportion of $\beta_2$ and $\beta_3$, and the different $\alpha$’s used are available in Table 6.2. The resulting distances to the chosen locations are close to the distances which are observed by Horni et al. (2009) for larger groups, but still are substantially larger than the mean distance to the closest locations for smaller more realistic mean group sizes. The different network structure measurements have an impact. The degree centrality measurement as the $\alpha$ results in shorter distances than the closeness centrality, which makes sense as a high degree centrality is a measurement for people which share people around themselves, whereas people with a high closeness centrality are people with rather fewer contacts but in between two groups, which makes the average distance larger. The betweenness centrality does not measure much of the spatial impact, as there are no information passed further than one degree.
Table 6.2.: Descriptive statistics of the distance of each vertex to the chosen activity location in km

<table>
<thead>
<tr>
<th>$\beta_{2L}/\beta_{3L}$</th>
<th>Degree centrality</th>
<th></th>
<th>Closeness centrality</th>
<th></th>
<th>Betweenness centrality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>4</td>
<td>10.4</td>
<td>6.2</td>
<td>12.6</td>
<td>6.3</td>
<td>11.1</td>
<td>6.1</td>
</tr>
<tr>
<td>3</td>
<td>8.6</td>
<td>5.8</td>
<td>9.8</td>
<td>5.9</td>
<td>9.0</td>
<td>5.8</td>
</tr>
<tr>
<td>2</td>
<td>7.0</td>
<td>5.4</td>
<td>7.6</td>
<td>5.5</td>
<td>7.2</td>
<td>5.3</td>
</tr>
<tr>
<td>1</td>
<td>5.6</td>
<td>5.3</td>
<td>6.4</td>
<td>5.3</td>
<td>6.1</td>
<td>5.2</td>
</tr>
<tr>
<td>0.5</td>
<td>4.6</td>
<td>4.3</td>
<td>5.1</td>
<td>4.4</td>
<td>4.9</td>
<td>4.3</td>
</tr>
<tr>
<td>0.33</td>
<td>4.4</td>
<td>4.3</td>
<td>4.6</td>
<td>4.3</td>
<td>4.5</td>
<td>4.2</td>
</tr>
</tbody>
</table>
7. Conclusion and closing views

The goal of the final chapter is on the one hand to summarize the empirical findings considering the hypotheses and research gaps drawn in Section 4.2. These findings are then further reflected in the context of the theoretical discussion in Chapter 3. Here, the contribution of this work to transportation planning and social networks are illustrated including policy implications. Further, the contribution of this work towards social capital is discussed. Finally, new research perspectives are presented.

7.1. Empirical evidences and hypothesis

Following, an overview of the empirical findings of Chapter 5 compared with the hypothesis and the results from previous research is given. In general, the empirical analysis from Chapter 5 follows the hypothesis and the previous research findings. The social network geography, tie distances and network degree, as well as the contact frequencies are connected to a variety of explanatory dimensions. A summary of the empirical findings are presented in Table 7.1 for an overview.
## Table 7.1: Summary of the empirical results

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Network degree</th>
<th>Tie distance</th>
<th>Network geography</th>
<th>Physical contact Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic influences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Female</td>
<td>n.s.</td>
<td>-</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>High income</td>
<td>n.s.</td>
<td>n.s.</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Age</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.s.</td>
</tr>
<tr>
<td>Higher education</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>n.s.</td>
</tr>
<tr>
<td>One person household</td>
<td>n.s.</td>
<td>+</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Swiss citizenship</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Part time employed</td>
<td>-</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Work Status: Retiree</td>
<td>-</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Children in household</td>
<td>+</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Living with partner</td>
<td>n.s.</td>
<td>-</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Network degree</td>
<td>n.m.</td>
<td>n.s.</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Mobility biographical influences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. of relocations</td>
<td>+</td>
<td>-</td>
<td>n.s.</td>
<td>-</td>
</tr>
<tr>
<td>N. of educ./work changes</td>
<td>n.s.</td>
<td>n.s.</td>
<td>+</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Mobility tool ownership influences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car availability always</td>
<td>n.s.</td>
<td>n.s.</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Season ticket</td>
<td>+</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Tie characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship duration</td>
<td>n.m.</td>
<td>-</td>
<td>n.m.</td>
<td>-</td>
</tr>
<tr>
<td>Immediate Family</td>
<td>n.m.</td>
<td>-</td>
<td>n.m.</td>
<td>+</td>
</tr>
<tr>
<td>Extended Family</td>
<td>n.m.</td>
<td>+</td>
<td>n.m.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Work mate</td>
<td>n.m.</td>
<td>n.s.</td>
<td>n.m.</td>
<td>-</td>
</tr>
<tr>
<td>Very close tie</td>
<td>n.m.</td>
<td>-</td>
<td>n.m.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Distance</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>ICT use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet access</td>
<td>n.s.</td>
<td>+</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Telephone Frequency</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.m.</td>
<td>+</td>
</tr>
<tr>
<td>SMS Frequency</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.m.</td>
<td>+</td>
</tr>
<tr>
<td>Email Frequency</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.m.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

+/- Significant (< 0.1) variable and the sign of influence
n.s. Not significant variable
n.m. Not included in the model
7.1.1. Sociodemographic influences

In general, nearly no gender effects are observed, except for the empirical finding that women tend to maintain shorter distant ties. This effect can not be confirmed by the social network geography analysis, which hints, that women tend to maintain shorter distant relationships compared to men, but less accumulated compared to men. Surprisingly, the income effects are also not strong, and in the case of the social network geographies higher income has even a negative impact, which could mean that people with higher income have less time to maintain a more geographic extended network. But also, that people with higher income do not have to migrate in comparison to poorer people, which might be forced to migrate based on their financial situation. This result is in contrast to the results from Carrasco and Miller (2006), where the income has a positive impact on the social network geography. But income has a positive impact on the face-to-face contact frequency (and negative on the other contact means), which indicates that high income people can visit those that matter as often as they want physically, where less well situated people have to use cheaper contact means. There is empirical evidence that tertiary education leads to higher degree networks, spatial more expansive networks and longer distance ties. In contrast, Flamm and Kaufmann (2006) found no positive influence of education on social network geographies Younger persons (<29 years) have the tendency to maintain higher degree and spatially more expansive networks as well as longer tie distances than older people. It might be speculated whether this is a cohort effect, which can not be empirically investigated in this thesis. This result is consistent with the results of Beige and Axhausen (2006), which show that young people are the most mobile persons. Persons with immigrant background can not be distinguished by either of the dependent ego-centric social network measurements. The impact of work status measured by the variables part time employee and Retiree. Only an negative impact on the network degree can be verified, whereas there is no significant effects on the network geography, tie distance nor face-to-face interactions. The smaller network degree is consistent with the results from Páez and Scott (2007), where they show, that a work environment strongly affects the social inclusion. Children in the household have no effect on the network geography, but have a
positive effect on the degree. Living with a partner has a negative impact on the
tie distance, suggesting that those who care for others primarily maintain small
distance relationships.

7.1.2. Mobility-biographical influences

The different analyses show that the mobility biography has a strong influence
on all social network measurements, except tie distance. As more eventful the
mobility biography of an actor is, the more spatially disperse the social network
is designed, as mobility biographies are increasingly the subject of conception of
people’s lifes, which is consistent with the findings of [Ohnmacht and Axhausen
(2005)].

7.1.3. Mobility tool ownership influences

The findings of [Ohnmacht and Axhausen (2005)] that reduced or impeded access
to mobility tools lead to social exclusion can be confirmed. Groups of people with
access to a car facilitate the maintenance of longer distant ties. Public transport
season tickets seem to support rather local contact and shows a positive significant
influence on the network degree. This result is consistent with the analyses of [Viry
et al. (2009)].

7.1.4. Contacts in a shrunken world and the influence of ICT
use

Geographical distance influences how people interact today, as it did in the past
[Mok and Wellman (2007)]. This is confirmed by the strong decay of face-to-
face contact frequency with distance. The effort involved in the different modes
of communication (face-to-face, phone, email and SMS) matches the strength of
distance decay in the frequency of those interactions.

While email frequency is unaffected by distance, its share rises fastest with it. The
reverse is true for SMS messaging, which stays stable in terms of share, while only
slowly falling with distance in terms of frequency. The high frequency of phoning at short distances translates into growing shares, even if the absolute frequencies fall with distance. While other characteristics of the respondent-contact dyad influence frequency and shares (duration of the relationship, context of the initial acquaintance), these are less influential in absolute terms than the socio-demographics of the person, especially income, age and the number of moves stand out among them. The impacts of the characteristics of the ego-alter dyad varies systematically with type of relationship, which reflect the underlying social expectations for relatives, friends and work mates.

The survey reported here provides a measure of the process of the globalisation of personal social networks. People mix local, regional and international contacts in their social networks distributing their social capital widely. They maintain relationships via all the modes at their disposal. Still, face-to-face meetings favour the regional scale, but at a distance beyond the walking distances of the past. The scale ties the respondents to their cars or season tickets to be able to keep the travel times involved acceptable. And even though modern telecommunication techniques allow maintaining contacts living far away, maintaining a large distance contact is still more costly than maintaining contacts to direct neighbours. And those who live closer are better able to provide help and companionship.

7.2. Relevance and contributions

Manski (1993) warned long ago, that recognizing the social network influence is not possible without knowing it’s present and without knowing the underlying social structures and their influence strengths. This Thesis reinforces this explicitly for demand modelling in transportation. Activity timing, location choice and trip generation effects can not be understood without understanding the underlying social network processes. Although the causality of the measured effects in this thesis cannot be known for certain, the social dynamics in leisure travel are highly probable under the assumption that travel is derived demand of activities. Trying to explain such behaviours without the data about socializing components of the behaviour presented in this thesis, researcher had to use latent variables to
CHAPTER 7. CONCLUSION AND CLOSING VIEWS

capture these components. However, with the data and the experimental scenario presented in this thesis, we are a step closer to be able to fully include socializing components in transportation models.

7.2.1. Transportation, policy and social networks

The interaction of the different contact modes and face-to-face meetings are complementary. As the costs for the use of ICT as e.g. telephone costs are still decreasing and decreased in the past centuries much more than the costs for travelling, the impact of decreasing telecommunication costs on socially motivated travel is not to be neglected. Based on the change of telephone costs between the 1990 and 2001 by a factor of 5.8 (FCC 2001) and a price elasticity of telephone costs of -0.34 (Nadiri and Nandi 1997), the number of telephone contacts in this period have increased by a factor of 1.97, which increased the number of face-to-face contacts, assuming no transportation price changes in travelling costs, by a factor of 1.20. This is not unneglectable in forecasting travelled distances, as 45 percent of the distances travelled in Switzerland (95 billion passenger kilometres) is due to recreational activities, of which around 80 percent is socially motivated. Accounting for this, similar drops in telecommunication costs would account for an additional 5.84 billion passenger kilometres in Switzerland. Growths in leisure travel similar to those since the 1970 until now by 60 percent (Grotrian 2007) can be expected.

The distance distribution shows even for local contacts a strong decay with a mean distance between the egos and their social contacts within 100 km of 14.5 km. The practical application in Chapter 6 shows the impact of a joint location choice with a theoretical simulation. Compared with the leisure trip distance distribution in Switzerland of 11.1 km (ARE/BfS 2007a) and a nearest leisure activity location of 1.1 km (Horni et al. 2009), which is the outcome of the trip distance gap, based on location choice models not accounting for social influences or quality heterogeneity of the locations. The gap between the distances of the average closest leisure location and observed trip distances based on location choices can be explained to some extend with spatial social networks.

Further the experimental results from this thesis show how the collected data
7.2. RELEVANCE AND CONTRIBUTIONS

and analysis performed in this thesis can be implemented in to activity based transportation models to improve their precision.

7.2.2. Social capital

Social capital is associated with the physical social space. As discussed in Chapter 2 about social capital and space, researchers suggest that long distances relationships can result in significant loss in terms of social support. On the other hand, Axhausen (2007) defines the social capital as the stock of joint abilities, shared histories, understandings and commitments enabling the skilled performance of joint activity, even at a distance, of a pair or larger number of persons. And argues that increasing accessibilities enables a larger pool of possible social contacts to select from and hence increases the social capital. Also discussed in Ohnmacht et al. (2008) is the involvement of a shift away from local socializing reference points toward more expansive spatial areas of common interest in the advanced modernisation of societies, which would follow the idea of increased social capital based on increased accessibility. The findings in the case of Zurich are supporting both views. But in interpreting the results, one has to keep in mind that while the sample is large, it is drawn from Zurich, a very specific, highly international environment and one of the second-rank global cities (Taylor, 2004; Sassen, 2001). Expected influences of income towards more spatial extended social networks to support the more specialized selection of social contacts enabled trough higher monetary travel budgets can not be found. In contrary, the income has a slightly negative influence towards shorter distant ties. This indicates that the overcoming of space is also bounded and limited by time. On the other hand, higher education has a positive influence throughout the social network measurements presented in this thesis, which supports the view, that highly skilled people make use of the increased accessibility to social capital, which also means that modern societies require attributes such as high education, mobility and flexibility from people to be successful. On the other hand, the importance of ICT and ownership of mobility tools in the models underline, that reduced or impeded access to mobility tools is a dimension of social exclusion. The large measured social network geographies require additional mobility for meetings and visits to maintain contact, which is
also visible in the growth of leisure travel. Although the literature speaks of a “death of distance” \cite{cairncross2001}, the spatial distances effect on social interactions is enormously, as the decay of tie probability and contact frequency with distance has been shown in the previous chapters. These findings are consistent with the findings of \cite{mok2007}. An other empirical evidence for the effects of how mobility affects the sociality, is that the relationship duration has a negative impact on the tie distance, which means that the effort to overcome physical space burdens long-distance relationships over-proportionally in the long-term. The results of this thesis show that people interact more frequently with local close contacts, as that they feel the need to spend more time with emotionally close contacts and family.

7.3. Research Perspectives and Outlook

It would be highly desirable to replicate the study both nationally and internationally across a range of urban and rural environments, which is partly already ongoing with the excellent work done in Switzerland, the Netherlands and Conception \cite{kowald2009, vandenbergetal2009, carrasco2012}. Although the empirical difficulties are enormous, it would be worthwhile to construct benchmarks for decades past, so as to be able to assess the dynamics of this personal globalisation (see \cite{rae2003} for a starting point).

The analyses could be improved technically. The magnitude of proportions of the different beta parameters used in the experiments are unknown and would need further survey work. Also the impact of other utility functions has not yet been tested. E.g. the satisfaction terms $S_{ij}$ could be modelled as a function of the number of persons involved, e.g. $S_{ij} = nS_{ij},...,ik-1, - S_{ij},...,ik-1$ where $S_{ij}$ is convex. The $S_{li}$ could also be modelled as taste variations across individuals.

Further there should have been socio-demographical variables incorporated in the surveys to make the tie probability models dependent on such variables. This thesis’ research analyses ego-centric social networks empirically, whereas the overall network could shed more insight in the underlying social processes in travel. Together with a travel diary, much of the data not available, but needed to integrate
the experiments in chapter 6 are currently surveyed in a follow up project on this thesis (see Kowald et al., 2009a). This project uses snowball sampling to collect social networks in Switzerland including network structures. In addition, the social network questionnaire (similar to the one presented here) is accompanied by a travel diary including a social interaction diary, which allows to get data about activity group-sizes, location and, time of social meetings.

The complex interplay between space, society and transportation was demonstrated by theoretical and empirical work in this thesis. The empirical data collection and analysis results provide an important contribution to current research of transportation planning, space sociology and social capital research.
Bibliography


Axhausen, K. W., A. Frei and T. Ohnmacht (2006a) Networks, biographies and travel: First empirical and methodological results, in proceedings of the 11th International Conference on Travel Behaviour Research, Kyoto.


Bernard, H., E. Johnsen, P. Killworth, C. McCarty, G. Shelley and S. Robinson
Bibliography


Bibliography


Frei, A. (2007a) Distance, mode and frequency of contact in egocentric social networks, in *paper presentation at the Cosmobilities Network Meeting 2007: Mobilities, space, and inequality*, Basle.


Granovetter, M. (1973) The strength of weak ties, American Journal Of Sociology, 78 (6), 1360–1380.


Hanneman, R. A. and M. Riddle (2005) *Introduction to social network methods*, University of California, Riverside, CA.


Jackson, M. O. (2005) The economics of social networks, in *proceedings of the 9th World Congress of the Econometric Society*.


Lauterbach, W. and K. Pillemer (1996) Familien in späten Lebensphasen: Zerrissene Familienbande durch räumliche Trennung?, Technischer Bericht 23, Univer-


Bibliography


Bibliography


Wellman, B., A. Haase, J. Witte and K. Hampton (2001) Does the internet in-


A. Questionnaire
### Part 1: Personal information

**When were you born?**

<table>
<thead>
<tr>
<th>Month</th>
<th>Year</th>
</tr>
</thead>
</table>

**Are you ...?**

- [ ] Male
- [ ] Female

**Are you...?**

- [ ] A Swiss national
- [ ] An other nationality, please specify: 
- [ ] Naturalised, if yes when (year):

**Are you mainly ...?**

- [ ] In education or apprenticeship
- [ ] Full-time employed
- [ ] Part-time employed
- [ ] Seeking employment
- [ ] Homemaker
- [ ] Retired

**Where is your current place of education or employment located?**

**Post code and municipality**: 

**Which qualifications have you acquired? (Tick all that apply)**

- [ ] Primary school
- [ ] Secondary school
- [ ] Baccalauréat
- [ ] Apprenticeship
- [ ] University of applied science degree
- [ ] University degree/ETH degree

**What type of education or employment are you mainly engaged in?**

**Education/Employment:**
If you are employed, where is your workplace in the main?

- [ ] At home
- [ ] Varying places
- [ ] At a fixed place of work outside the home

Do you own a car driving licence?  
If yes, since when?

- [ ] No
- [ ] Yes  
  Year of acquisition: 

How often is a car available to you?

- [ ] Always
- [ ] Frequently
- [ ] Infrequently
- [ ] Never

Do you own one or several of the following public transport tickets?

- [ ] None
- [ ] Regional annual or monthly ticket
- [ ] National annual ticket
- [ ] Point-to-point season ticket
- [ ] Half-fare discount ticket
- [ ] Other: 
Part 2: Mobility Biography

Please fill in the following tables chronologically.

Please note:
- At the top of every table is an example.
- The questions concerning the different places of residence request the exact address, preferably including the post code, municipality, street and house number. If you can not remember the address exactly, please enter the name of the municipality and in case of bigger towns the name of the neighbourhood.
- If you did live abroad, we are also interested in information about your residences there.

Life events and places of residence

Please assign important life events to specific periods:

<table>
<thead>
<tr>
<th>Life events</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., birth of brother</td>
<td>1981</td>
<td>1981</td>
</tr>
</tbody>
</table>

(and 6 more lines)

Please indicate your places of residence since your year of birth:

<table>
<thead>
<tr>
<th>Place of residence</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., Bahnhofstrasse 1, 8001 Zürich</td>
<td>1982</td>
<td>1985</td>
</tr>
</tbody>
</table>

(and 8 more lines....)
### Employment and Income

Please indicate your places of work:

<table>
<thead>
<tr>
<th>Place of work</th>
<th>From</th>
<th>To</th>
<th>Main means of transport to and from work</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., Wolfsburg</td>
<td>1978</td>
<td>1999</td>
<td>Car</td>
</tr>
</tbody>
</table>

(and ten more lines)

Please indicate your personal gross income per month:

<table>
<thead>
<tr>
<th>Income class</th>
<th>From</th>
<th>To</th>
<th>Income class</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 0 to 999 Fr.</td>
<td></td>
<td></td>
<td>(3) 2 000 to 5 999 Fr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) 1 000 to 1 999 Fr.</td>
<td></td>
<td></td>
<td>(4) 6 000 to 9 999 Fr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) 10 000 to 13 999 Fr.</td>
<td></td>
<td></td>
<td>(6) 14 000 Fr. and more</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income class</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., 3</td>
<td>1979</td>
<td>1999</td>
</tr>
</tbody>
</table>

(and 8 more lines)

Mode of transport, ownership of public transport tickets and car availability

Please indicate your main mode of transport:

<table>
<thead>
<tr>
<th>Mode of transport</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., 1</td>
<td>1979</td>
<td>1999</td>
</tr>
</tbody>
</table>

(and 5 more lines)
Please indicate your ownership of public transport tickets:

<table>
<thead>
<tr>
<th>Public transport ticket</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., 3</td>
<td>1979</td>
<td>1999</td>
</tr>
</tbody>
</table>

(and 4 more lines)

Please indicate your car availability:

<table>
<thead>
<tr>
<th>Availability</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., 2</td>
<td>1979</td>
<td>1999</td>
</tr>
</tbody>
</table>

(and 4 more lines)
Memberships in groups that meets periodically for a shared activity

Please indicate your memberships, including the frequency of meetings in groups that meet regularly for a shared activity:

Types of groups which meet periodically for a shared activity:
  religious, sports, politics, social, nature, etc.

<table>
<thead>
<tr>
<th>Type of group</th>
<th>Your frequency of attendance</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., nature</td>
<td>2 times per month</td>
<td>1981</td>
<td>1990</td>
</tr>
</tbody>
</table>

(and 4 more lines)

Additions and comments

Would you like to make any additions or comments?

Thank you very much for your assistance!
Part 3: Social networks

In this part of the questionnaire we have some questions about the people with whom you interact privately. These persons can be family members, friends or acquaintances. Please note:

- At the top of the lists is always one example.
- The questionnaire has two parts. Please start with the personal-data section and then fill in the contact-description section. The persons section and the contact-description section are each divided into two parts, part A and part B.
- Please name as many relevant persons as you can think of in this section.
- The contact-description section is located in a separate part of the questionnaire to make it easier for you to assign a number to each name.
- Photo albums or address books may help you remember relevant persons.
- You may also enter nicknames instead of full names.
- The questions concerning the places of residence of your contacts request their exact addresses, preferably including the post code, municipality, street and house number. If you can not remember the address exactly, please enter the name of the municipality and in case of bigger towns the name of the neighbourhood.

Persons

Please indicate persons with whom you discuss important problems, with whom you stay in regular contact or whom you can ask for help:

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Name</th>
<th>Nr.</th>
<th>Name</th>
<th>Nr.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>A</td>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>e.g., Tom</td>
<td>6</td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>7</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>8</td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>9</td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>10</td>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>11</td>
<td></td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>
Please indicate additional people with whom you undertake leisure activities:

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Name</th>
<th>Nr.</th>
<th>Name</th>
<th>Nr.</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td>B</td>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>e.g., Jenny</td>
<td>11</td>
<td></td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>B</td>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>12</td>
<td></td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>B</td>
<td></td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>14</td>
<td></td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>15</td>
<td></td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>17</td>
<td></td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>19</td>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>21</td>
<td></td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>22</td>
<td></td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>23</td>
<td></td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>24</td>
<td></td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

Holidays and acquaintances

Please specify holidays you have gone on with acquaintances:

<table>
<thead>
<tr>
<th>Holiday place</th>
<th>Year</th>
<th>With whom?</th>
<th>Means of transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., New York/USA</td>
<td>2001</td>
<td>Friend from the USA, Nr. A2</td>
<td>Airplane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Contact description - A

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Acquaintance from/of the...</th>
<th>Through family</th>
<th>Association, others...</th>
<th>Year</th>
<th>Face-to-face</th>
<th>By telephone</th>
<th>By email</th>
<th>By SMS</th>
<th>Where did you meet this person the last time?</th>
<th>Contact's place of residence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Contact description - B

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Acquaintance from/of the...</th>
<th>Through family</th>
<th>Association, others...</th>
<th>Year</th>
<th>Face-to-face</th>
<th>By telephone</th>
<th>By email</th>
<th>By SMS</th>
<th>Where did you meet this person the last time?</th>
<th>Contact's place of residence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>