Department of Communications and Networking

Empirical analysis of mobile interpersonal communication service usage

Juuso Karikoski

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 119/2013

Empirical analysis of mobile interpersonal communication service usage

Juuso Karikoski

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall S1 of the school on the 6th of September 2013 at 12.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Network Economics

Supervising professor
Prof. Heikki Hämmäinen

Thesis advisor
Prof. Heikki Hämmäinen

Preliminary examiners
Prof. Anders Henten, Aalborg University, Denmark
Prof. Sune Lehmann Jørgensen, Technical University of Denmark, Denmark

Opponent
Prof. Harry Bouwman, Delft University of Technology, The Netherlands

Aalto University publication series
DOCTORAL DISSERTATIONS 119/2013

ISBN 978-952-60-5266-3 (printed)
ISBN 978-952-60-5267-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5267-0

http://lib.tkk.fi/Diss/

Unigrafia Oy
Helsinki 2013

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Juuso Karikoski

Name of the doctoral dissertation
Empirical analysis of mobile interpersonal communication service usage

Publisher School of Electrical Engineering

Unit Department of Communications and Networking

Series Aalto University publication series DOCTORAL DISSERTATIONS 119/2013

Field of research Network Economics

Manuscript submitted 19 February 2013 **Date of the defence** 6 September 2013

Permission to publish granted (date) 16 May 2013 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

The ways in which people communicate and keep in touch with each other are changing. Our relationships are shaped by the wide variety of communication possibilities available, and the overall increase in mediated communication. On the other hand, the relationships also shape the use of different means of communication. Since the proliferation of mobile phones, mobile voice calls and SMSs have become the world's most ubiquitous form of mediated communication. More recently also newer mobile communication services have emerged, which are not necessarily provided by mobile operators, but based on the Internet instead. The merger of the mobile and Internet domains is further accelerating the changes in how we socialize.

The purpose of this thesis is to empirically analyze the use of mobile interpersonal communication services by using a quantitatively driven mixed method design, with handset-based measurements as the core data collection method. The aim is to fill in the research gap of sparse objective user-level research on mobile communication service usage. Handset-based measurements enable collecting actual mobile phone usage data from users who have opted in to participate and install a research application to their devices. Because of the personal and social "always on" nature of the data collection device, we are able to collect rich location-based and context-sensitive data about how mobile communication services are used. The main data analyzed are collected from up to 200 participants from a longitudinal research panel with a span of more than a year. In addition to handset-based measurements, surveys and semi-structured interviews are used as complementary methods.

In this thesis mobile communication services are found to be used differently based on a number of factors. Mobile phone use context and the strength of the tie between the persons communicating are the most salient ones. For example, SMSs are used more with strong ties when compared to voice calls. Moreover, the lengthiest and the briefest voice calls are made in the Home and Office contexts, respectively. Consequently, the usage interrelationships of mobile communication services are discussed from a mobile operator perspective. The use of multiple and dynamic datasets in social network analysis is also encouraged, as the analysis of different communication channels may or may not lead to different views of the social network under analysis. Thus, the thesis and the results of the research can be valuable for practitioners and academics in the area of mobile ecosystems, particularly mobile operators. The thesis also contributes to the emerging fields of communication and social network research, namely computational social science and social computing.

Keywords handset-based measurements, mobile services, interpersonal communication, social networks, usage analysis, mixed method

ISBN (printed) 978-952-60-5266-3 **ISBN (pdf)** 978-952-60-5267-0

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2013

Pages 188 **urn** http://urn.fi/URN:ISBN:978-952-60-5267-0

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Juuso Karikoski

Väitöskirjan nimi
Mobiilien ihmistenvälisten viestintäpalveluiden käytön empiirinen analyysi

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Tietoliikenne- ja tietoverkkotekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 119/2013

Tutkimusala Tietoverkkotalous

Käsikirjoituksen pvm 19.02.2013 **Väitöspäivä** 06.09.2013

Julkaisuluvan myöntämispäivä 16.05.2013 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Tavat, joilla ihmiset viestivät ja pitävät yhteyttä toisiinsa ovat muuttumassa. Ihmissuhteisiimme vaikuttaa viestintämahdollisuuksien laaja kirjo, sekä yleinen välitteisen viestinnän lisääntyminen. Toisaalta ihmissuhteemme muokkaavat myös sitä, kuinka eri viestintätapoja käytämme. Matkapuhelinten lisääntymisen myötä matkapuheluista ja tekstiviesteistä on tullut maailman käytetyin välitteisen viestinnän muoto. Lisäksi viime aikoina on ilmestynyt uusia mobiileja viestintäpalveluita, jotka eivät välttämättä ole mobiilioperaattoreiden tarjoamia, vaan Internet-pohjaisia. Mobiili- ja Internet-alojen yhteensulautuminen kiihdyttää muutoksia viestinnässämme entisestään.

Tämän väitöskirjan tarkoitus on empiirisesti analysoida mobiilien viestintäpalveluiden käyttöä. Pääasiallisena tutkimusmenetelmänä käytetään kvantitatiivisia päätelaitepohjaisia mittauksia. Tutkimuksen tavoitteena on tuottaa lisää tietoa mobiilien viestintäpalveluiden käytön objektiivisista käyttäjätason tutkimuksista. Päätelaitepohjaiset tutkimukset mahdollistavat todellisen käyttötiedon keräämisen matkapuhelinten käyttäjiltä, jotka ovat päättäneet osallistua tutkimukseen ja asentavat tutkimussovelluksen matkapuhelimiinsa. Matkapuhelin soveltuu erittäin hyvin mobiilien viestintäpalveluiden käytön tutkimiseen henkilökohtaisena, sosiaalisena ja aina päällä pidettävänä laitteena, koska sen avulla voidaan kerätä rikasta, paikkakohtaista ja kontekstiherkkää tietoa. Analysoitava pääaineisto on kerätty jopa kahdensadan osallistujan pitkittäisestä, yli vuoden kestävästä, paneelitutkimuksesta vuosien 2009 ja 2010 aikana. Päätelaitepohjaisten mittausten lisäksi kyselytutkimuksia ja puolistrukturoituja haastatteluita käytetään täydentävinä tutkimusmenetelminä.

Tässä väitöskirjassa on havaittu mobiilien viestintäpalveluiden käytön riippuvan monesta eri tekijästä. Matkapuhelimen käyttökonteksti (esim. aika ja paikka) ja viestivien henkilöiden välisen suhteen vahvuus ovat näistä keskeisimmät. Tekstiviestejä käytetään esimerkiksi enemmän vahvojen suhteiden kanssa, kuin matkapuheluita. Toisaalta pisimmät matkapuhelut tehdään kodin tyyppisissä, ja lyhimmät toimiston tyyppisissä konteksteissa. Väitöskirjassa pohditaan myös mobiilien viestintäpalveluiden käytön välisiä riippuvuussuhteita mobiilioperaattorin näkökulmasta. Lisäksi tutkimus kehottaa käyttämään useita aikasarjaan perustuvia aineistoja sosiaalisten verkostojen analyysissä, koska eri viestintäkanavien analyysi voi johtaa erilaisiin näkökulmiin analysoitavasta sosiaalisesta verkostosta. Täten väitöskirja ja sen tulokset voivat olla erityisen arvokkaita mobiilioperaattoreiden alalla toimiville ammatinharjoittajille ja akateemikoille. Väitöskirja edistää myös viestinnän ja sosiaalisten verkostojen tutkimuksen uusimpia aloja, kuten laskennallista yhteiskuntatiedettä.

Avainsanat päätelaitepohjaiset mittaukset, mobiilipalvelut, ihmistenvälinen viestintä, sosiaaliset verkostot, käyttöanalyysi, sekametodi

ISBN (painettu) 978-952-60-5266-3 **ISBN (pdf)** 978-952-60-5267-0

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2013

Sivumäärä 188 **urn** http://urn.fi/URN:ISBN:978-952-60-5267-0

Preface

"Telefoonin tarkoitus ja tehtävä on *lausutun sanan* koneellisesti ohjaaminen pitemmälle matkalle kuuluvaksi"

Maier et al. (1900)

I want to express my gratitude to a number of people without whom I would not have been able to successfully complete this project. First of all I want to thank Professor Heikki Hämmäinen for supervising the thesis and providing me guidance during tough times. Heikki has made it possible for me to conduct research in different research projects, and provided the necessary resources. Heikki has also given me the freedom to find my own path of research for which I am most indebted. I am grateful for the insightful comments I received from the preliminary examiners Professors Anders Henten and Sune Lehmann Jørgensen. I am also honored and obliged that Professor Harry Bouwman has devoted a lot of his time to act as my opponent. My co-authors Kalevi Kilkki, Sakari Luukkainen, Olli Mäkinen, Matti Nelimarkka and Tapio Soikkeli deserve special thanks for the effort they have put in to our joint research activities.

The thesis research has been conducted in the Network Economics research team of the Aalto University Department of Communications and Networking. I have thoroughly enjoyed my time in the team mainly because of my wonderful current and former colleagues: Antti Riikonen, Timo Smura, Thomas Casey, Pekka Kekolahti, Michail Katsigiannis, Arturo Basaure, Tapio Levä, Henna Suomi, Benjamin Finley, Nan Zhang, Juuso Töyli, Antero Kivi and Mikko Heikkinen. Thanks to all of you especially for the fruitful discussions on coffee breaks and 11.20 lunches, and for creating such a lively working atmosphere.

This thesis has been funded by the OtaSizzle and MoMI/MoMI II/MoMIE research projects. I want to thank Nokia and Elisa for their sponsoring in the OtaSizzle project, as well as Aalto University's MIDE program and Helsinki University of Technology TKK's 'Technology for Life' campaign

donations from private companies and communities. I also wish to thank TEKES and the corporate partners of the MoMI projects for their support.

I am thankful to the management of the OtaSizzle project, Professor Martti Mäntylä and Dr. Olli Pitkänen, for giving me the possibility to conduct my own research in the project. Also the colleagues in the OtaSizzle project, especially Airi Lampinen, Vilma Lehtinen, Kimmo Karhu, Antti Virolainen and Juho Makkonen deserve a big thank you for all the collaboration directly and indirectly related to the thesis during the project. Regarding the handset-based measurements conducted in the thesis, I greatly appreciate the support from and collaboration with Dr. Hannu Verkasalo, Tero Lindberg and Jarno Rautakorpi from MobiTrack Ltd. / Zokem / Arbitron Mobile Oy.

I have also been lucky for receiving generous support from a number of foundations. Thus, I want to express my gratitude to the HPY Research Foundation, the Research and training foundation of TeliaSonera Finland Oyj, the Nokia Foundation, the Foundation of The Finnish Society of Electronics Engineers, and the Walter Ahlström Foundation for the grants awarded. My compliments to the Future Internet Graduate School FIGS as well, for awarding me with a two-year funded doctoral school position.

Having said all this, my deepest gratitude belongs to my friends and family who have always supported me in my studies. Fishing, playing football and playing the bass guitar with friends have been an excellent counterbalance for the research work. I am especially thankful for my dear friend Antti Meriluoto for designing the cover picture of the thesis. Finally, I want to thank my parents Anne and Jukka, for always being there for me in all aspects of life, and my beloved wife Essi for love and support.

Espoo, June 6th 2013
Juuso Karikoski

Contents

List of publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I. Karikoski, J., 2012. Handset-based data collection process and participant attitudes. *International Journal of Handheld Computing Research*, 3 (4), pp.1-21.

II. Karikoski, J. & Soikkeli, T., 2013. Contextual usage patterns in smartphone communication services. *Personal and Ubiquitous Computing*, 17 (3), pp.491-502.

III. Karikoski, J. & Kilkki, K., in press. Building social capital with mobile communication services. *International Journal of Electronic Finance*.

IV. Karikoski, J. & Luukkainen, S., 2011. Substitution in smartphone communication services. In: *Proceedings of 15th International Conference on Intelligence in Next Generation Networks*. Berlin, Germany 4-7 October 2011, pp.313-318.

V. Karikoski, J. & Mäkinen, O., 2012. Mobile social phonebooks. In: *Proceedings of 16th International Conference on Intelligence in Next Generation Networks*. Berlin, Germany 8-11 October 2012, pp.157-164.

VI. Karikoski, J. & Nelimarkka, M., 2011. Measuring social relations with multiple datasets. *International Journal of Social Computing and Cyber-Physical Systems*, 1 (1), pp.98-113.

Author's contribution

Publication I: "Handset-based data collection process and participant attitudes"

Karikoski was the sole author of this publication.

Publication II: "Contextual usage patterns in smartphone communication services"

The idea for this publication was formed by Karikoski. Karikoski was the head author of the publication and wrote the manuscript. Karikoski collected and analyzed the data, and Soikkeli developed the context detection algorithm used in the analysis of the data. Karikoski and Soikkeli reviewed and edited the publication together.

Publication III: "Building social capital with mobile communication services"

The idea for this publication was formed by Karikoski. Karikoski was the head author of the publication and wrote the majority of the manuscript. Karikoski collected and analyzed the majority of the data. Kilkki reviewed the manuscript and conducted the long-tail analysis in the results chapter. Karikoski and Kilkki reviewed and edited the publication together.

Publication IV: "Substitution in smartphone communication services"

The idea for this publication was formed jointly with Luukkainen. Karikoski was the head author of the publication, wrote most of the manuscript, and collected and analyzed the data. Luukkainen provided comments and contributed text to the chapters two and three. Karikoski and Luukkainen reviewed and edited the publication together.

Publication V: "Mobile social phonebooks"

The idea for this publication was formed jointly with Mäkinen. Karikoski was the head author of the publication and wrote the manuscript, except for chapter three which was written by Mäkinen. Karikoski and Mäkinen designed and conducted the interview study, and reviewed and edited the publication together.

Publication VI: "Measuring social relations with multiple datasets"

The idea for this publication was formed jointly with Nelimarkka. Karikoski was the head author of the publication and wrote most of it. Karikoski collected the handset-based data, and Nelimarkka the OtaSizzle service data. Karikoski and Nelimarkka analyzed the data, and reviewed and edited the publication together.

List of figures

List of tables

List of abbreviations

3G	Third Generation
4G	Fourth Generation
AP	Access Point
ARPU	Average Revenue per User
CAB	Converged Address Book
CDR	Call Data Record
CID	Mobile Network Cell ID
CMC	Computer-Mediated Communication
ESM	Experience Sampling Method
ICT	Information and Communications Technology
IM	Instant Messaging
IMS	IP Multimedia Subsystem
M2M	Machine-to-Machine
MAC	Media Access Control
MCC	Mobile Country Code
MMS	Multimedia Messaging Service
MSP	Mobile Social Phonebook
NFC	Near Field Communication
OMA	Open Mobile Alliance
OTT	Over-the-Top
PDA	Personal Digital Assistant
RCS	Rich Communication Suite
RP	Research Problem

RQ	Research Question
SMS	Short Message Service
SPSS	Statistical Package for the Social Sciences
URL	Uniform Resource Locator
U&G	Uses & Gratifications
VoIP	Voice over Internet Protocol
WLAN	Wireless Local Area Network

1. Introduction

1.1 Background and motivation

The recent developments in Information and Communications Technology (ICT), the new interpersonal communication services, and the overall increase in mediated communication are changing the ways we are social (see, e.g., Licoppe & Smoreda, 2005). Especially with the emergence of the smartphone and third (3G) and fourth generation (4G) mobile networks many new mobile interpersonal communication services have appeared on the market. Social media services, which have already gained momentum in the Internet, have shown their potential as a communication channel in the mobile domain as well. Other emerging (and already established) Internet-based (or Over-the-Top (OTT)) communication services include Instant Messaging (IM) and Voice over Internet Protocol (VoIP). The mobile Internet is becoming a user-driven environment, where interpersonal interaction and social networks are essential. At the same time, mobile operators are trying to compete against and in OTT communication services and decreasing Average Revenue per User (ARPU) figures by establishing their own network-based solutions, such as the Rich Communication Suite (RCS). After all, most of the mobile operators' revenues are still being generated by Short Message Services (SMS) and voice calls.

Usage studies on communication services and related social networks have been traditionally conducted with self-report methods, such as surveys and interviews. Moreover, market-level data have been used in some cases to study how the usages of different services are interrelated. However, drawbacks of the two presented approaches exist, as self-reports of service usage might not correlate well with actual behavior (see, e.g., Gerpott, 2011b; Boase & Ling, 2013). Furthermore, market-level analysis does not take user-level differences into account, and thus the overall usage analysis might be affected by the diverse usage patterns found in different user segments (Gerpott et al, 2012). Recently, however, new methods of data collection have arisen that enable studying objectively the actual usage of communication services on user-level. As a result, a field of research called computational social science (see, e.g., Lazer et al., 2009) is emerging,

where the new capacities to collect and analyze digital trace data are leveraged. Handset-based measurements are one type of a method that can be utilized in analyzing objective user-level data on communication service usage. With this method rich data can be collected from mobile handsets by installing data collection software on the devices. The pioneers of the method have been the researchers of University of Helsinki (Raento et al., 2005) and Massachusetts Institute of Technology (Eagle & Pentland, 2006). Handset-based research has also been previously conducted in the Aalto University by Verkasalo (2009b). Because of the novelty of the method, however, relatively little related research exists.

The location-based and context-sensitive information that are characteristic for mobile phones can be utilized to study the usage of mobile interpersonal communication services. Mobile phones are personal and social "always on" type of devices that are well suited for studying these services and the related social networks, as they provide possibilities for strong network effects (or externalities, see, e.g., Courcoubetis & Weber, 2003). Fixed line phones, on the other hand, do not support social network analysis as well as mobile phones, because they are not as personal, not carried with the user, and not programmable. Furthermore, the mobile phone, and especially the smartphone, can be seen as a hub of different communication channels for the user, whereas fixed line phones only provide audio communication. Thus, we can study the communication habits and social networks of users in multiple channels, instead of focusing just on a single channel. Consequently, the smartphone has been identified as a useful emerging tool for social scientists (Raento et al., 2009) and psychologists (Miller, 2012). Studying how different communication services are used is important both from human behavior and mobile operator perspectives. For example, the mobile operators' marketing strategy and service offering decisions are affected by the usage interrelationships of the services. From a human behavior viewpoint it is essential to study the various relations between people that these different services and communication channels support and create. Studying the underlying social networks and the usage of communication services is important as it has been identified that the social network indeed affects service usage (Szabo & Barabasi, 2006).

This thesis draws on handset-based measurements as a novel method for conducting empirical research on mobile interpersonal communication service usage. The aim is to fill in the research gap of sparse objective user-level research on communication service usage identified and motivated above. Moreover, this thesis contributes to the emerging fields of computational social science and social computing. The thesis and the

results of the research may be especially significant for academics and practitioners in the areas of mobile communications ecosystems (particularly mobile operators) and social sciences.

1.2 Objectives and scope

This thesis aims to address a larger research problem (RP) to which the publications give partial solutions. The research problem is as follows:

RP: How can one empirically analyze the use of mobile interpersonal communication services utilizing handset-based measurements and complementary methods?

As the research problem indicates, the core method used in the thesis is handset-based measurements. However, complementary methods such as surveys and interviews are also used to supplement the results acquired with the core method. Thus, the methodological research approach of this thesis is a quantitatively driven mixed method design, as described later in the methods and data chapter.

The research questions (RQ) of the thesis are as follows:

RQ1: How can behavioral data be collected with handset-based measurements, and what motivational and attitudinal factors need to be considered?

RQ2: How are mobile interpersonal communication services used depending on: 1) the use context of the mobile phone, and 2) the strength of the tie between the persons communicating?

RQ3: How are the usages of mobile interpersonal communication services interrelated, and how do the mobile social phonebooks affect the interrelationships?

RQ4: How do the social networks derived from distinct communication services differ?

The empirical results based on the research questions are discussed from mobile operator's and social network researcher's perspectives. This is done by measuring and analyzing the usage interrelationships of communication services, and by analyzing social networks acquired from multiple communication services.

Figure 1 depicts how the publications are related. The methodological foundation of the thesis is formed by presenting the handset-based data

collection process utilized in the thesis and by surveying the participants' motivations and attitudes towards the measurements, and mobile phones and services in general (Publication I). Then, empirical analysis on mobile interpersonal communication service usage is performed from contextual, social, and interrelationships perspectives (Publications II-VI). Finally, the results are contemplated especially from a mobile operator's perspective (Publications IV and V), as well as from a social network researcher's perspective (Publication VI).

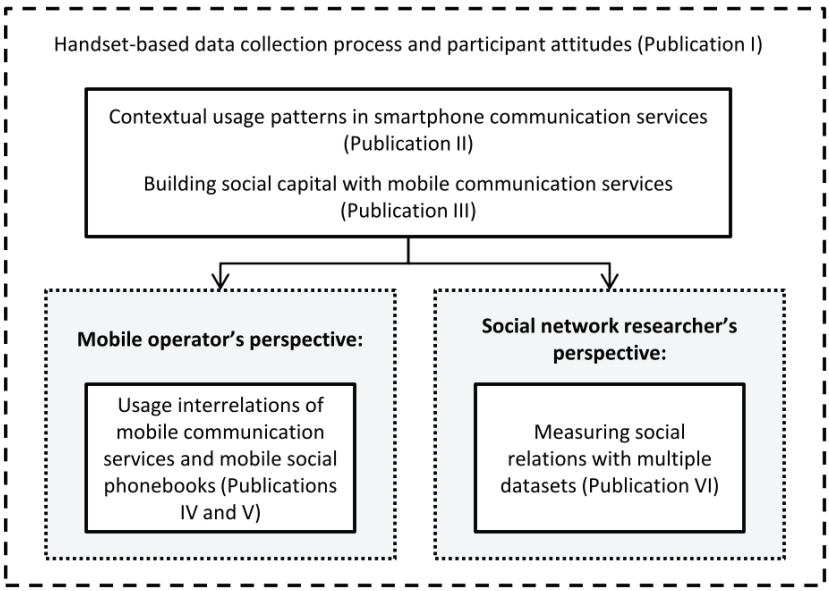

Figure 1 Relations of the publications

The key terms used in this thesis are defined rather broadly because of the dynamic nature of the target of analysis. First of all, *communication* refers simply to '*social interaction through messages*' (Fiske, 1982, p.2). *Mobile interpersonal communication services* in general refer to services or applications that can be used by a mobile device or a handset, and enable interpersonal communication, i.e., exchanges of calls and messages between persons. The services are further divided to traditional operator-provided services, such as SMS, Multimedia Messaging Service (MMS) and cellular voice calls, and to mobile Internet or OTT services, which contain all Internet-based operator-independent services, such as email, IM, VoIP and social media. The term *handset-based measurements* implies that data are collected from handsets. In this thesis especially mobile phones and smartphones are considered as handsets. *Mobile phones* and smartphones both have cellular connectivity, and are not restricted to, for example, Wireless Local Area Network (WLAN) connections, as is the case with some

mobile devices such as tablets or Personal Digital Assistants (PDA). Furthermore, *smartphones* differ from mobile phones, because they are programmable and encompass the ability to install third party application software. With the term *usage*, we refer mainly to post-adoption usage. For example, we analyze usage interrelationships only with those users who have already adopted all the services under analysis (Publication IV). Finally, *empirical analysis* refers to research approaches seeking '*to gain knowledge of the world, that is, of the reality in which we live*', as de Groot (1969, p.1) has defined. Moreover, actual usage data are collected directly from individuals' mobile phones with help of a mobile research application. The general research approach of the thesis will be discussed in more detail in the next subchapter.

1.3 Research approach

The thesis is multi-disciplinary in nature, as it combines technology and sociology. Furthermore, the results are discussed from an economic perspective. In the context of information systems and paradigms of science, Hevner et al. (2004) have claimed that behavioral science seeks to develop and justify theories that explain or predict organizational and human phenomena. Design science, on the other hand, seeks to create innovations for problem-solving purposes. Hevner et al. (2004) also state that methodologies related to behavioral science are usually rooted in data collection and empirical analysis, whereas design science methods are mainly used to assess the quality and effectiveness of artifacts.

Järvinen (2004) has distinguished between different research approaches (**Figure 2**) based on the work of Hevner et al. (2004) and March & Smith (1995). *Mathematical* approaches are utilized in studies where, for instance, a given theorem, lemma or assertion is proven. The mathematical notations do not have any direct connection with reality. In the other main branch, *conceptual-analytical* research approaches do not require empirical research to be conducted – instead we are interested in what existing theories, models and frameworks consider as part of reality. *Theory-testing* and *theory-creating* approaches, on the other hand, belong to the category of empirical research, where raw data are collected from reality. Theories, models and frameworks can be tested, for example, by performing experiments, field studies, and case studies. When a proper theory does not exist, it can be created by generalizing from the researcher's own observations and from those of previous empirical studies. The research approaches that stress the utility of innovations can be divided to *innovation-building* and *innovation-evaluating* approaches, which both belong to the paradigm of design science. The former approach asks if a

certain innovation can be built to perform a given task, whereas the latter tries to determine how well the innovation performs the task.

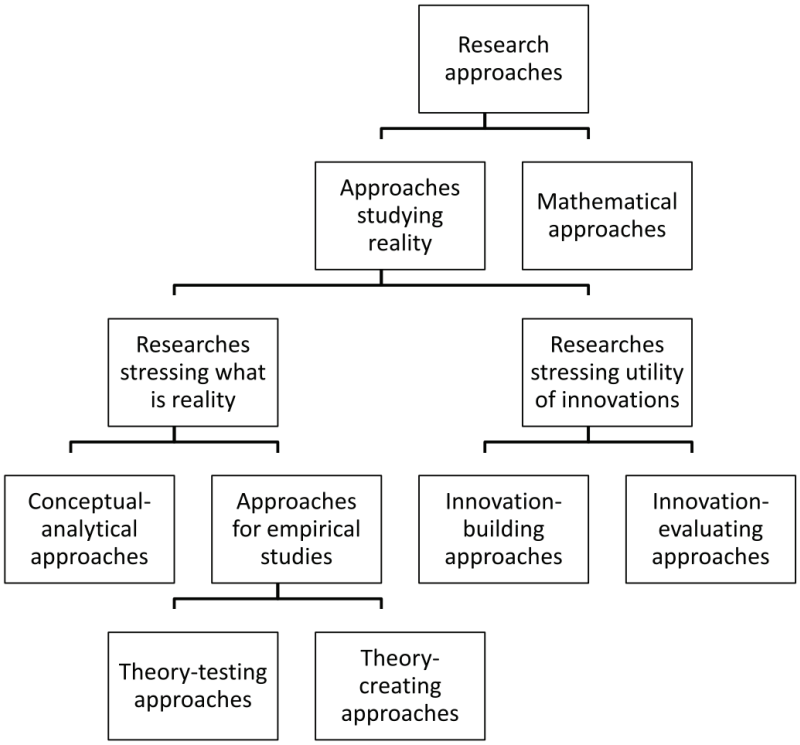

Figure 2 Taxonomy of research approaches (adapted from Järvinen, 2004)

The traditional way of conducting research is *theory-driven* or *hypothesis-driven*, where assumptions of the real world are tested (so called deductive research). Recently, however, a paradigm building on the long tradition of inductive research has emerged. In *data-driven* science (see, e.g., Schweitzer & Vespignani, 2012), one of the targets is to identify new empirical laws emerging from massive datasets that are being collected at an unprecedented rate. These two paradigms should not be treated as substitutes, however, as they do not only differ in terms of quantity of data, but also in the nature of data. The formulation of hypotheses plays a key role in identifying to which paradigm the research belongs to. As the hypotheses and research questions of this thesis mainly emerged after exploring the collected data, rather than formulating the hypothesis beforehand and then designing the research and data collection, this research fits better to the paradigm of data-driven science than to that of hypothesis-driven science. Furthermore, the definitions of some concepts used in this thesis are highly dependent on the type of data collected.

Based on the different points of view presented above, the general research approach of the thesis can be identified. The research paradigm of this thesis is data-driven behavioral science, and the research approaches are mainly empirical studying reality. The empirical research results are discussed from the viewpoint of communication technology affordances, but no theories were created based on the research conducted in this thesis. Thus, the research approach of the thesis is more related to theory-testing, than to theory-creating. However, the data in this thesis are more used for illustrating different theories, rather than testing a single theory or a set of coherent theories. One can also claim that the overall research problem of the thesis (especially Publication I) could be categorized as design science stressing the utility of innovations. However, it is used more as the methodological foundation of the thesis, on top of which the actual research is conducted.

1.4 Thesis structure

After this introduction chapter the theoretical background of the research is discussed, including communication research and social network analysis approaches relevant to this thesis. In the third chapter related empirical research on mobile interpersonal communication services conducted by other researchers is reviewed. The methods and data chapter describes the mixed method design of the thesis, with handset-based measurements as the core method, and surveys and semi-structured interviews as complementary methods. The answers to the research questions are provided in the results chapter, and they address the research problem together. Then, the implications of the results are discussed from mobile operator's and social network researcher's perspectives in the discussion chapter. Finally, after methodological implications and privacy, the limitations of the research, as well as a number of future research items are discussed.

2. Theoretical background

The theoretical background of this thesis is divided to two parts – communication research and social network research. In the first part, communication research in general and related theories and approaches utilized in this thesis are reviewed. Communication channels and interpersonal communication result in social networks, which will then be discussed in the second part of the chapter.

2.1 Communication research

Fiske (1982, p.2) has defined communication simply as '*social interaction through messages*'. However, also more detailed definitions have been suggested, such as '*a process by which information is exchanged between individuals through a common system of symbols, signs, or behavior*' by Kilkki (2012, p. 314) and Merriam-Webster Dictionary (2012). Communication studies come in a wide variety of forms and traditionally include those based on face-to-face communication. McQuail (1994, p.9) has summarized the topics of communication research to five categories: 1) who communicates to whom, 2) why communicate, 3) how does communication take place, 4) what about, and 5) what are the consequences of communication. Computer-Mediated Communication (CMC), as the name implies, refers to communication that is happening via or with the help of computers. In this thesis the mobile phone and the communication channels that it encompasses can be thought of as the common (computer-mediated) system through which information is exchanged between individuals. Thus, the mobile interpersonal communication services studied in this thesis can be placed under the umbrella term of CMC. This chapter discusses the theoretical background of the thesis in terms of communication channel usage and selection, and the related theories and approaches. The chapter is focused on the approaches and theories utilized and relevant to this thesis.

2.1.1 Communication channels

In this thesis mobile interpersonal communication services are seen as communication channels - not as a technical term, however, but rather as a

sociological term as in diffusion research by, for instance, Rogers (2003). Rogers (2003) has defined a communication channel as: '*the means by which messages get from one individual to another*'. He has also distinguished between mass media channels and interpersonal channels. Mass media channels enable one or a few individuals to reach an audience of many while interpersonal channels involve a face-to-face exchange between two or more individuals. Interpersonal channels are the focus in this thesis instead of mass media channels. However, we also consider other types of communication channels than just face-to-face, including technology or computer mediated communication channels (Van Cleemput, 2012). The terms *communication channel* and *communication media* are used interchangeably in this thesis, mainly because the terms are not used in a consistent way in the related literature. As Haddon (2005) has discussed, we should not only ask why we use a certain channel (or in this case service), but why we choose it from all the other possibilities. Furthermore, if we use a given channel, what are the implications of that to the other channels, i.e., what are the usage interrelationships between the channels?

People tend to use multiple communication channels together as bundles to stay socially connected. Building on Granovetter's (1973) work, Haythornthwaite (2005) has discovered that people with stronger ties tend to use more means of communication to stay socially connected, than those with weaker ties. She has termed this phenomenon *media multiplexity*. Similar results have been obtained more recently by, for instance, Van Cleemput (2012) and Boase et al. (2006). Boase (2008) has coined a term called *personal communication system*, which refers to the way people draw on multiple communication channels to stay connected to their personal networks, instead of using separate channels to connect to separate social worlds. The key question presented by Boase (2008) is how and to which extent these channels are used together. The affordances of the communication channels and the composition (size and diversity) of the personal network both affect the use of the channels. Haddon (2005), on the other hand, talks about a *communication repertoire*, i.e., how people manage communications and make choices of using the channels based on a variety of factors. First of all, qualities of communication media (technical, economical and symbolical) favor certain choices. These mean that the communication choices are affected by the technological forms and capabilities of the channels (e.g., display size and input capabilities of a device), financial costs, and perceptions of the channels. Social factors also favor certain communication choices. These include the purpose and content of the communication, social relationship to the other

communicator, physical proximity of communications, social location of the communicators (or context as defined by Bouwman & Van De Wijngaert (2002)) and communication norms. Furthermore, de Bailliencourt et al. (2011) and Bouwman & Van De Wijngaert (2002) have identified a number of other factors affecting the choice of communication channels, including user's characteristics (own customs and preferences), possibilities and specificities of the channels, availability of communicators, and characteristics of the task. Specifically in the mobile domain, Gerpott (2011a) has also claimed that the pricing scheme in which the mobile services are charged from the users affects the gratifications acquired from those services and, consequently, the selection and usage of a given service.

Although these kinds of checklists for channel usage and selection are always simplifications, they map the landscape of factors affecting the communication channel choices. These *systems* or *repertoires* naturally evolve over time, and thus one needs to consider which usage patterns are relatively settled and which might evolve in the long, or even in the short term, as Haddon (2005) points out. As Livingstone (1999) has well stated, '*Research on 'new media' involves studying a moving target*'. Next, two communication research approaches relevant to this thesis are reviewed in more detail, namely communication technology affordances and Uses & Gratifications (U&G). Also other communication theories from the areas of organizational and computer-mediated communication are shortly introduced.

2.1.2 Communication technology affordances

Hutchby (2001) has claimed that social processes and the properties of technologies are interrelated. Thus, he has proposed an approach to study technologies and social life, which involves seeing technologies in terms of their affordances. Technology affordances can be interpreted as the possibilities that the technology offers for action. Moreover, Whittaker (2003) has claimed that the '*fundamental goal of mediated communication theories has been to explain the relationship between the affordances…of different mediated technologies and the communication that results from using those technologies*'. This approach has been applied by a number of scholars to study the affordances of (mobile) communication technologies. For instance, Van Cleemput (2012) has distinguished three criteria based on which the affordances of communication technologies differ. First of all, communication technologies differ in terms of *time and place controllability*. For example, SMSs afford sending messages anywhere anytime, while voice calls do not. Second, some communication technologies, such as SMSs, afford high control of messages. Consequently, *message controllability* is another

criteria based on which communication technologies differ. Finally, *audience controllability*, i.e., what type of audience can be reached with a given communication technology, is the third criteria suggested by Van Cleemput (2012). Furthermore, Rettie (2009) has claimed that *synchrony* (i.e., whether mediated communication is synchronous or asynchronous) affects the affordances of communication technologies. A related approach is that of social affordances presented by Boase (2008). This approach supposes that communication media are chosen to be used when the opportunities that the medium provides, and the characteristics of the ties with whom the medium is used to communicate are congruent. As an example Boase (2008) suggests that the telephone affords communication with distant ties that would be troublesome to reach otherwise. The concept of communication technology affordances is salient especially in mediated communication research. Based on a review of communication theories relevant to mediated communication, Whittaker (2003) has claimed that the theories all take the general form of: 1) characterizing how technologies differ in relation to their communication affordances, 2) describing how the affordances differ from those of face-to-face communication, and 3) explaining how the affordances create differences between face-to-face and mediated communication.

Uses and Gratifications (U&G) approach is similar to that of technology affordances, but it has its foundations in mass communication research. According to Katz et al. (1974) individuals seek gratification from using certain technologies based on their own individual needs and motivations. The mobile phone related U&G research has focused mostly on the mobile phone as a single entity, and considered gratifications such as fashion or status, and entertainment (Peters & ben Allouch, 2005) as well as mobility, immediacy, and instrumentality (Leung & Wei, 2000). However, Van Cleemput (2012) has stated that U&G is more used in media content studies, while technology affordances are used more for studying interpersonal communication via communication technologies.

In addition to the U&G and technology affordances approaches, also a number of other communication theories mainly based on organizational and computer-mediated communication exist, such as media richness theory and channel expansion theory. Carlson & Zmud (1999) have categorized organizational communication theories to three general conceptual areas. First of all theories that explain the enablers and motivators of selecting and using a particular channel exist. The second category of theories explains the nature of the use itself. Third, theories that explain the perceptions derived from using the channel also exist.

In the light of this review on communication research approaches and theories, communication technology affordances seems as the most suitable approach for this thesis. As said, technology affordances are salient in mediated interpersonal communication research. Helles (2013) has also claimed that the affordances approach and the interplay between the affordances of different media are especially salient in the case of mobile communication. U&G as a theory used mainly in media content studies, as well as organizational communication theories in general, are seen as less suitable approaches in the scope of this thesis.

2.2 Social network analysis

Interpersonal communication channels discussed above are used to create ties between pairs of individuals. A tie can be maintained by one or more types of social (network) relations (i.e., exchanges or interactions) and thus interpersonal communication builds into social networks (Haythornthwaite, 2005). Furthermore, as Licoppe & Smoreda (2005) have stated: '*Social ties are the basic units of social networks*'. Wasserman & Faust (1994, p. 9) have defined a social network as '*the set of actors and the ties among them*', where the definitions of actors and ties are not restricted to people and social relationships, but instead they include actors such as organizations, and trade relations as ties. In general social networks can be analyzed either on personal (or ego-centered) network level or full network level. In the ego-centered approach, ties from a focal actor (i.e., ego) to its alters are measured (Wasserman & Faust, 1994, p. 42). Full network data, on the other hand, show how the alters also interact with each other, and not just how the ego interacts with its alters (Van Cleemput, 2012). Full networks can be analyzed in terms of, for instance, cohesion and centrality measures. Measures of cohesion include density, diameter, and transitivity, whereas centrality is measured, for instance, with degree and betweenness (see, e.g., Wasserman & Faust, 1994). Degree measures the number of links a node has to other nodes in the network. Nodes that have higher betweenness feature on many shortest paths between other nodes in the network. Density measures the ratio between the number of existing links and the maximum number of links in a network. Diameter is the length of the longest shortest path between any pair of nodes in a network. Finally, a triad of nodes X, Y, and Z is transitive if 1) X directs a link to Y, 2) Y directs a link to Z, and 3) X directs a link to Z.

Social networks can, however, be analyzed in multiple levels. By analyzing different communication channels among individuals, you may or may not end up with different views of the social network under analysis. After all, people select and use different channels based on a variety of factors as introduced in the previous chapter. McPherson et al. (2001) concluded

already more than a decade ago that the priority for future social network researchers should be to gather dynamic data on multiple social ties. More recently, Banford et al. (2010) have also claimed that using multiple proxies for tie strength increases the accuracy of social network analysis.

Social network analysis has increased and gained a lot of interest lately, mainly because of the massive amount of data that are being collected in online social networking services, such as Facebook (see, e.g., Golder et al., 2007) and Twitter (see, e.g., Huberman et al., 2009; Mislove et al., 2011). As a result, a field of research called computational social science (Lazer et al., 2009) is emerging, where the new capacities to collect and analyze digital trace data are leveraged. Furthermore, Licoppe & Smoreda (2005) have claimed that the new communication services and the overall increase in mediated communication are changing the ways we communicate and keep in touch with each other. In the following subchapter, some social network analysis approaches and concepts relevant to this thesis are introduced.

2.2.1 Strength of ties and social capital

Granovetter's (1973) theory on the strength of weak ties is one of the fundamental issues in sociology. According to him interpersonal ties can be classified to *strong*, *weak* and *absent* ties depending on the degree of overlap of the friendship networks of the individuals. The strength of the tie is affected by the amount of time invested in it, the emotional intensity and intimacy of the tie, as well as the reciprocal services (i.e., favors) characterizing the tie. Granovetter (1973) claims that weak ties are indispensable to the opportunities of individuals and their integration into communities, whereas strong ties breed local cohesion and lead to overall fragmentation. Moreover, he argues that a bridge between micro-level interactions and macro-level phenomena can be established through analyzing interpersonal networks.

Social capital is a concept related to the strength of ties. Numerous definitions for social capital exist, but as Wilken (2011) has pointed out, the works of Coleman (1988) and Putnam (2000) have been foundational in the way social capital research has developed in the past decade. Coleman (1988) sees social capital as an individual resource acquired from relationships, whereas Putnam (2000) sees it as a collective resource of social networks. Putnam (2000) has distinguished between two forms of social capital – bonding and bridging. Weak ties form the foundation for *bridging social capital*, which is outward looking social capital that networks of people across diverse social backgrounds encompass. Strong ties on the other hand form the foundation for *bonding social capital*, which refers to inward looking social capital reinforcing exclusive identities

and homogenous groups. Thus, bridging social capital refers to linking between groups from different social circles, forming new relationships and information diffusion, whereas bonding social capital refers to already established relationships, where the same social circles are shared.

2.2.2 Circles of acquaintanceship

A framework of social groups called the circles of acquaintanceship has been proposed by Hill & Dunbar (2003). According to this framework and the extended social brain hypothesis (Dunbar, 1998), personal social networks consist of a series of layers with roughly 5, 15, 50 and 150 members. The inner circles are more emotionally close (than the outer ones), meaning that they are characterized with strong ties and bonding social capital. The upper limit for the social group size in humans is 150 (so called Dunbar number), because of the limited information processing capacity of the human brain, and the limited time that individuals have available for social interaction.

Recently, the social brain hypothesis has been confirmed by Saramäki et al. (2012). By means of a personal network survey and mobile voice call records, Saramäki et al. (2012) show that individuals have a "social signature", where they tend to allocate approximately the same amount of time to their alters based on their rank, but independent of who they are. Consequently, old relationships are downgraded, when new ones are acquired.

3. Related empirical research

Empirical research related to mobile interpersonal communication service usage utilizing data collected with or from mobile phones can be roughly divided to two categories. In a bottom-up approach, data are collected directly from a mobile phone by installing data collection software on the device (as in this thesis with handset-based measurements). On the other hand, a top-down approach exists, where mobile operators' Call Data Records (CDR) or other "big data" from service providers are utilized to collect information on a large number of individuals – usually millions. In the following subchapter, empirical studies from these two perspectives relevant to this thesis are reviewed. Then, studies related to the usage interrelationships of communication services are reviewed. In the reviews we have purposefully excluded usage studies where self-report measures (e.g., surveys and interviews) or market-level data analysis have been used as a primary method. A possibility exists that self-reports of service usage might not correlate well with observed behavior, and that market-level data might mask diverging user-level change patterns in different user segments, as Gerpott et al. (2012) state. However, we acknowledge that these kinds of subjective and market-level methods should be used in conjunction with objective behavioral data collection to complement the results.

3.1 Service usage and social networks

As introduced above, mobile interpersonal communication service usage and social networks can be studied from two perspectives – bottom-up or top-down. In this subchapter, we have included the most relevant studies from these two approaches. **Table 1** and **Table 2** summarize the literature reviewed in the scope of this thesis. For more information on using mobile phones as sensors for social research, see Eagle (2011).

3.1.1 Bottom-up research

The bottom-up approach enables collecting rich datasets that have a large amount of information or "throughput" as Aharony et al. (2011) put it. However, the sample has been restricted to a maximum of thousands of individuals so far (with sample sizes depending on the software distribution

strategy as discussed by Miller (2012)), and the devices enabling this kind of data collection are not ubiquitous yet. Moreover, this approach is novel, because data collection has only been possible since the invention of the smartphone, and its ability to install third party application software. This approach also provides a possibility to collect longitudinal data, and thus the duration of observation period in these kinds of studies usually spans months. This approach is utilized in this thesis as a core research method, and more information about it can be found from the methods and data chapter of this thesis.

Table 1 Related empirical research from the bottom-up perspective

Project / Software	**Research topic**	**Services analyzed**	**Sample**	**Obs. period**	**Citation**
Proprietary Symbian software	Place dependent communication propensity	Voice calls, SMSs, MMSs	24 individuals	Six months	Licoppe et al. (2008)
Reality Mining (ContextPhone)	Inferring friendship network structure by using mobile phone data	Voice calls	94 individuals	Nine months	Eagle et al. (2009)
Reality Mining (ContextPhone)	Inferring individuals' community affiliations and friendship from behavioral data	Voice calls	100 individuals	Nine months	Eagle & Pentland (2009)
Nokia SmartPhone360	Contextual usage patterns in mobile service usage	Voice call, SMS, Email, IM	324 individuals	Three months	Verkasalo (2009a)
Nokia Research Center	Large-Scale Analysis of Applications and Context	Voice calls, SMSs, Email, Voice chat	77 individuals	Nine months	Do et al. (2011)
Communication Explorer	Bonding and bridging with multimedia mobile phones	Voice calls, SMSs, Emails	233 individuals	Two months	Boase & Kobayashi (2012)
LiveLab	Characterizing Communications in Social Networks	Voice calls, SMSs	42k calls, 346k SMSs	Seven months	Tossell et al. (2012)
OtaSizzle	Spatiotemporal correlations of handset-based service usages	Voice calls, SMSs, Emails	124 individuals	16 months	Jo et al. (2012)
Chimps Lab	Mining Smartphone Data to Classify Life-Facets of Social Relationships	Voice calls, SMSs	40 individuals	Four months	Min et al. (2013)

Most of the bottom-up research analyzes mobile phone or service usage in general, such as Verkasalo (2009b) and Falaki et al. (2010a; 2010b). Only a handful of studies concentrating specifically on the use of mobile interpersonal communication services exist, and most of those studies have considered only voice calls and SMSs. We have made an effort to include

specifically that literature to the review, where services, which fit into the scope of this thesis, have been analyzed. In some cases, however, where mobile services in general have been analyzed, we have included the communication service specific parts of the research to the review. Consequently, the related empirical research on mobile interpersonal communication service usage and social networks from the bottom-up perspective are presented in **Table 1**.

Among the most cited mobile phone related social network analysis from the bottom-up perspective is the research conducted by the researchers from the Reality Mining group of Massachusetts Institute of Technology. Eagle et al. (2009) reported accurately predicting 95% of self reported friendships based only on objective measurements of behavior. They claim that inferred friendships might contain more information, than what is possible to capture with surveys. Another study performed with the Reality Mining dataset has been conducted by Eagle & Pentland (2009). They identified community affiliations of individuals with 96% accuracy by using behavioral similarity measures. Moreover, friendship can be estimated with the difference between individuals' common behavior. The Reality Mining dataset has also been used in the area of mobile interpersonal communication services by a number of researchers not directly affiliated with the group, such as Hossain et al. (2007) and Banford et al. (2010).

Both Verkasalo (2009a) and Do et al. (2011) studied the contextual usage patterns of mobile (communication) services. In the studies, the term context is used as a highly place-related term, but time based heuristics and user given labels are used in the studies to give semantic meaning to the places. Furthermore, Do et al. (2011) used Bluetooth to detect the social context of the users. Verkasalo (2009a) observed that voice calls are used more in *Home* context, where the voice calls tend to be also significantly longer, compared to the *Office* and *On the move* contexts. The most usage of email and IM services, however, happens in the *On the move* context. Do et al. (2011), on the other hand, observed that voice calls are associated with moving (or nomadic) contexts such as bus stops, parks or shopping centers. Moreover, SMSs are associated with many stationary indoor contexts such as home and work. They claim that the preference for synchronous communication and micro-coordination are reasons for choosing voice calls over SMSs in nomadic contexts. Regarding the social context, Do et al. (2011) observed that the use of voice calls and SMSs is higher in contexts with a high number of nearby Bluetooth devices. In a related study, Licoppe et al. (2008) also analyzed the propensity of users making a voice call in different places. They observed that the propensity to call from home with the mobile phone is low compared to other places. Furthermore, during

mobility periods the propensities to call are the highest. They claim that the availability of communication resources in different places plays a key role in the propensities to call.

More recently, Tossell et al. (2012) studied a longitudinal dataset of voice calls and SMSs and concluded that individuals contacted with both voice calls and SMSs are more likely strong ties. This relates to the media multiplexity theory by Haythornthwaite (2005) introduced previously. A similar result was also obtained by Boase & Kobayashi (2012), who report that voice calls and SMSs are used together for bonding social capital (i.e., communication with strong ties). Moreover, they observe that voice calls are used more for bridging social capital (i.e., communication with weak ties), whereas SMSs are used more for bonding social capital.

In a study closely related to this thesis, Jo et al. (2012) compared the communication network of users constructed from voice calls and SMS data, and the behavioral overlap network based on clusters of mobile service usage. The results indicate that an overall positive correlation between the networks exists, meaning that connected users having many common neighbors have also similar weekly service usage patterns. This finding is highly related to the principle of *homophily* in sociology (McPherson, 2001), which states that similarity breeds connection, or in other words, that birds of a feather flock together.

Min et al. (2012) used voice call and SMS logs to classify mobile phone contacts to three relationship types, namely *family*, *work* and *social*. They use different measures to classify the contacts, such as intensity, regularity and temporal tendency of communication. By comparing classification results to ground truth data collected from the participants, they report achieving 90% accuracy. For family type of relationships call intensity was the strongest factor, whereas for work type of relationships intensity was negatively correlated. In social type of relationships communications on weekends and SMS usage were the strongest factors.

According to this literature review, the empirical research utilizing the bottom-up approach and studying OTT communication services is sparse. However, these services are becoming more common, and as more and more handset-based data collection efforts are established and smartphones diffuse more among the majority, we are foreseeing an increase in the related literature. Moreover, data collected from the bottom-up perspective have recently been openly released for academic research purposes as part of the Nokia Mobile Data Challenge 2012 (Laurila et al., 2012).

3.1.2 Top-down research

The data collected from the top-down approach have a lower "throughput" (Aharony et al., 2011), than the bottom-up datasets. The contextual information of the individuals in these kinds of datasets is low (or non-existent), usually restricted to, e.g., location in terms of mobile network cell IDs (CID) in the case of CDRs. However, CIDs are only recorded when the phone is used for communication, because the phone does not send signals alerting the mobile network base stations of its location due to power constraints. In comparison, handset-based measurements capture also the offline usage of the mobile phone, enabling data collection also when the device is not connected to the network at all. The datasets collected from the top-down approach are usually also constrained to a specific domain (or communication service), such as voice calls or SMSs. Whereas bottom-up data can be collected only from devices which support installation of third party application software, the top-down approach collects data about all devices (also so called dumb phones) connected to the mobile network. **Table 2** summarizes the most relevant related literature from the top-down perspective.

Table 2 Related empirical research from the top-down perspective

Dataset source	Research topic	Services analyzed	Sample	Obs. period	Citation
Mobile operator	Social network effects on mobile service usage	Voice calls, SMSs, IM, Email	Six million subscribers	14 months	Szabo & Barabasi (2006)
Mobile operator	Structure and tie strengths in mobile communication networks	Voice calls	Seven million users	18 weeks	Onnela et al. (2007)
Mobile operator	The dynamics of a mobile phone network	Voice calls	Two million people	One year	Hidalgo & Rodriguez-Sickert (2008)
Mobile operator	Reciprocity of mobile voice calls	Voice calls	5,3 million customers	18 weeks	Kovanen et al. (2011)
European mobile operator	Interplay between Telecommunications and Face-to-Face Interactions	Voice calls	One million customers	One year	Calabrese et al. (2011)
European mobile operator	Correlated Dynamics in Egocentric Communication Networks	Voice calls, SMSs	6,2 million individuals	182 days	Karsai et al. (2012)
European mobile operator	Close relationships: A study of mobile communication records	Voice calls, SMSs	3,2 million subscribers	Seven months	Palchykov et al. (2013)

Most of the top-down research studies voice calls and/or SMSs, and the mobility data that are collected as a by-product. We have only come across one study, where other services than just voice calls and/or SMSs have been analyzed from a top-down perspective. Szabo & Barabasi (2006) utilized data collected from a mobile operator, and compared different communication services from a social network perspective. They have identified that the underlying social network of individuals (identified by voice calls and SMSs) affects the usage of other communication services, such as IM and email. For example, email is used uniformly across communities while IM displays a strong community-based segregation. This means that email usage does not apparently correlate with the underlying social network, whereas IM usage does. Regarding network effects, the authors also state that IM usage is driven by network externalities, while email is not. This is probably because mobile email is just an interface to email in general, whereas the IM service analyzed is an entirely mobile phone based service. Szabo & Barabasi (2006) have also identified that if a user uses a specific service regularly it is more likely (than what is expected based on a random adoption model) that his or her phone-based contacts use the service also. This effect is not limited to the adoption phase, but includes day-to-day usage as well. Again, this relates to the concept of homophily in sociology (McPherson, 2001).

The other studies in **Table 2** present examples of top-down studies based on voice calls and SMSs. The difference to Szabo & Barabasi's (2006) study is that in these other studies communication services are usually not compared to each other. Onnela et al. (2007) were the first to confirm the weak tie hypothesis on societal level. They analyzed a mobile operator dataset of voice calls and observed that the majority of strong ties are found within communities, whereas weak ties tend to connect communities to each other. Furthermore, the stronger the tie between two users, the more friends they have in common. When analyzing the removal of ties, Onnela et al. (2007) also observed that the whole network collapses if weak ties are removed. If strong ties are removed, however, the network only gradually shrinks but does not collapse. These results indicate that weak ties are crucial for maintaining a network's structural integrity, whereas strong ties are crucial for maintaining local communities. Hidalgo & Rodriguez-Sickert (2008) and Kovanen et al. (2011) studied reciprocity (i.e., the degree of mutuality in a relationship) of mobile voice calls. Hidalgo & Rodriguez-Sickert (2008) observed that reciprocity is the strongest predictor of tie persistence (i.e., stability of ties across time). The other measures considered were degree and clustering. Kovanen et al. (2011) also showed that ties, where one participant is much more active than the other (i.e.,

lopsided relations) are common in a voice call network. Furthermore, Calabrese et al. (2011) studied the interplay between voice calls and face-to-face interactions by utilizing CID location information. They claim that voice calls are used for coordination calls just before face-to-face meetings.

Two recent studies by Karsai et al. (2012) and Palchykov et al. (2013) have considered both voice calls and SMSs in their analyses. Karsai et al. (2012) compared voice calls and SMSs in terms of overall communication balance of links (similar to reciprocity discussed above), and observed strong differences. SMS links are strongly balanced, whereas voice call links are strongly unbalanced. They have suggested that SMS is technologically determined and requires mutuality, whereas voice calls do not. Palchykov et al. (2013) studied the closeness of relationships by identifying "best friends" of egos using a number of measures. They concluded among other things that voice calls and SMSs have a comparatively low level of overlap when identifying best friends, probably because the services serve different functions in human communication.

In addition to mobile operator based top-down research, also studies that utilize service provider data, such as Backstrom et al. (2011) study on Facebook or Kikas et al. (2012) study on Skype exist. However, we have not come across service provider based studies, where specifically the mobile usage of a given service has been analyzed. In a similar manner as with the Nokia Mobile Data Challenge 2012 (Laurila et al., 2012) in the bottom-up perspective, top-down data have also been recently openly released to the scientific community as part of the Data for Development challenge by Orange (Blondel et al., 2012).

3.2 Service usage interrelationships

The usage interrelationships between communication services can be divided to three general types – *complementary*, *substitutional* and *independent*. In the economics literature the interrelationships are usually defined in terms of cross elasticity measures (see, e.g., Courcoubetis & Weber, 2003). Marketing scholars, on the other hand, use terms such as multiproduct interactions (see, e.g., Bayus et al., 2000). This chapter reviews related empirical research on usage interrelationships, restricting the review to studies where objective user-level communication behavior is measured.

Kim et al. (2010) and Gerpott (2010a) have discovered that the usage interrelationship literature related to mobile interpersonal communication services is sparse or almost silent. According to Kim et al. (2010) the lack of objective user-level data and the relatively recent nature of the phenomenon are the main reasons for the low amount of related work. In most of the interrelationship studies, only voice calls and SMSs have been

studied. Only a handful of studies have been conducted, where the use of mobile Internet has also been considered. Moreover, we have not come across any studies where the OTT communication services in particular have been studied. Instead, the analysis has been conducted with more coarse-grained measures such as mobile Internet traffic usage. **Table 3** summarizes the literature related to the usage interrelationships of mobile interpersonal communication services based on the restrictions introduced above.

Table 3 Related empirical research on usage interrelationships

Dataset source	Usage interrelationships	Sample	Obs. period	Citation
Handset-based measurements	Mobile Internet services and voice calls and SMSs are independent.	659 mobile phone users	Three months	Verkasalo (2007)
Portuguese mobile telephone bills	Voice calls and SMSs are complements.	Representative sample of Portuguese consumers	11 months	Grzybowski & Pereira (2008)
German mobile operator	SMSs and mobile Internet use are substitutes.	304 subscribers	Three months	Gerpott (2010a)
German mobile operator	SMSs, voice calls and mobile Internet use are independent.	443 subscribers	Three months	Gerpott (2010b)
Asian mobile operator	SMSs and voice calls are substitutes.	6847 subscribers	Nine months	Kim et al. (2010)
South Korean mobile operator	Voice calls and mobile Internet use are complements.	180000 subscribers	Three months	Ghose & Han (2011)
German mobile operator	Mobile Internet use is a complement to voice calls and SMSs.	8312 subscribers	Three months	Gerpott et al. (2012)

Some empirical evidence exists in favor of a complementary interrelationship between SMSs and voice calls. Grzybowski & Pereira (2008) report complementarity as an average of the sample, and acknowledge that the services might be substitutes for some users and complements to others. A complementary interrelationship has also been observed between mobile Internet usage intensity and voice call consumption (Ghose & Han, 2011). However, mobile Internet usage was measured in terms of content downloading and uploading. In a similar manner to Grzybowski & Pereira (2008), Ghose & Han (2011) also acknowledge that the user characteristics affect the interrelationships on user-level. Gerpott et al. (2012) have observed that mobile Internet use intensity is a complement to voice calls and SMSs. The complementary interrelationship held, even if the users were charged with flat rate from mobile Internet access, and on usage basis from SMSs and voice calls (i.e.,

they had a strong economic incentive for substitution). However, the practical relevance of the results is small, with, e.g., demographics of the users and handset types strongly influencing the usage.

Some evidence of small substitutional interrelationships can also be found, but as reported by the authors, the magnitude or practical relevance of the studies' results has been rather low. A substitutional interrelationship between SMSs and voice calls has been observed by Kim et al. (2010), whereas Gerpott (2010a) has reported of a substitutional interrelationship between SMS and mobile Internet usage intensity. Kim et al. (2010) analyzed different demographic segments and concluded that the substitutional interrelationship between SMSs and voice calls holds for all of them. However, the substitution effect is stronger the younger the users are, and among females. Gerpott's (2010a) results on substitution hold for both incoming and outgoing SMSs. However, the operationalization of mobile Internet use intensity was rather broad, measured on the basis of mobile Internet traffic.

Empirical evidence in favor of an independent interrelationship between mobile interpersonal communication services is rather limited as well. Verkasalo (2007) has reported of an independent interrelationship between mobile Internet service usage, and SMSs sent or frequency and duration of voice calls made. However, Verkasalo (2007) analyzed a dataset from 2005 and 2006 and thus it was rather early to analyze the usage of mobile Internet services. Moreover, OTT services in particular were not studied. In another study by Gerpott (2010b), no significant correlation between mobile Internet use intensity and mobile voice communication or SMS activity levels was observed. Gerpott's (2010b) results hold both for the whole sample, as well as for subsamples, where the effects of heavy communication users were excluded from analysis.

Building on this literature review, we can conclude that the mobile interpersonal communication service usage interrelationship literature based on objective user-level communication behavior is scarce, and so far the results are differing. Furthermore, usage interrelationship analysis on OTT services specifically has been non-existent, and the analysis has been conducted on mobile Internet usage in general. The usage interrelationship literature is related to the communication research presented in the previous chapter. As introduced, many user-level (and other) factors affecting the communication channel choices exist, which, consequently, also affect the usage interrelationships of the channels. Thus, more detailed research in this area is definitely needed.

4. Methods and data

This chapter first presents the overall methodological research approach of the thesis. After that, the individual research methods used in the thesis (i.e., handset-based measurements, surveys and semi-structured interviews) are depicted, together with the characteristics of the data collected from those methods.

4.1 Methodological research approach

Bryman (2006) has distinguished three different methodological research approaches: *qualitative*, *quantitative* and *mixed methods research* (also called multi-methods, multi-strategy or mixed methodology research). In its simplest form qualitative and quantitative research methods are integrated (or combined) in the latter approach. However, as Creswell (2010) states, the definition of mixed methods research has varied considerably during the last twenty years, with definitions differing, for instance, in terms of what is being mixed. Morse (2010) has distinguished between mixed method design and multiple method design. Mixed method design refers to research, where a single complete core method and one or more complementary methods exist. By means of the complementary methods, other relevant areas of the research problem can be accessed, which are not accessible with the core method. Multiple method design, on the other hand, refers to research, where two or more studies addressing the same research problem are approached with different, but equal, methods.

Following Morse's (2010) definition, the methodological research approach of this thesis is a *quantitatively driven mixed method design*, where the core method is handset-based measurements and the complementary methods are surveys and semi-structured interviews. A similar approach has been recently utilized by, for example, Bouwman et al. (2013). With this design we are able to enhance and clarify the core quantitative results obtained with handset-based measurements, by administering surveys and semi-structured interviews to the population under study. Surveys are used in a qualitative manner to develop the

handset-based measurements and to characterize the sample used. Semi-structured interviews, on the other hand, are used to explore new phenomena (ideas and concepts) that are not yet accessible otherwise, and to initiate new research questions. According to Bryman's (2006) content analysis and Robson's (2011) research, semi-structured interviews are the most common qualitative method used in combination with quantitative methods. Morse's (2010) design was also selected, because it entails a diagramming technique with which the design can be illustrated in a simple way (**Figure 3**).

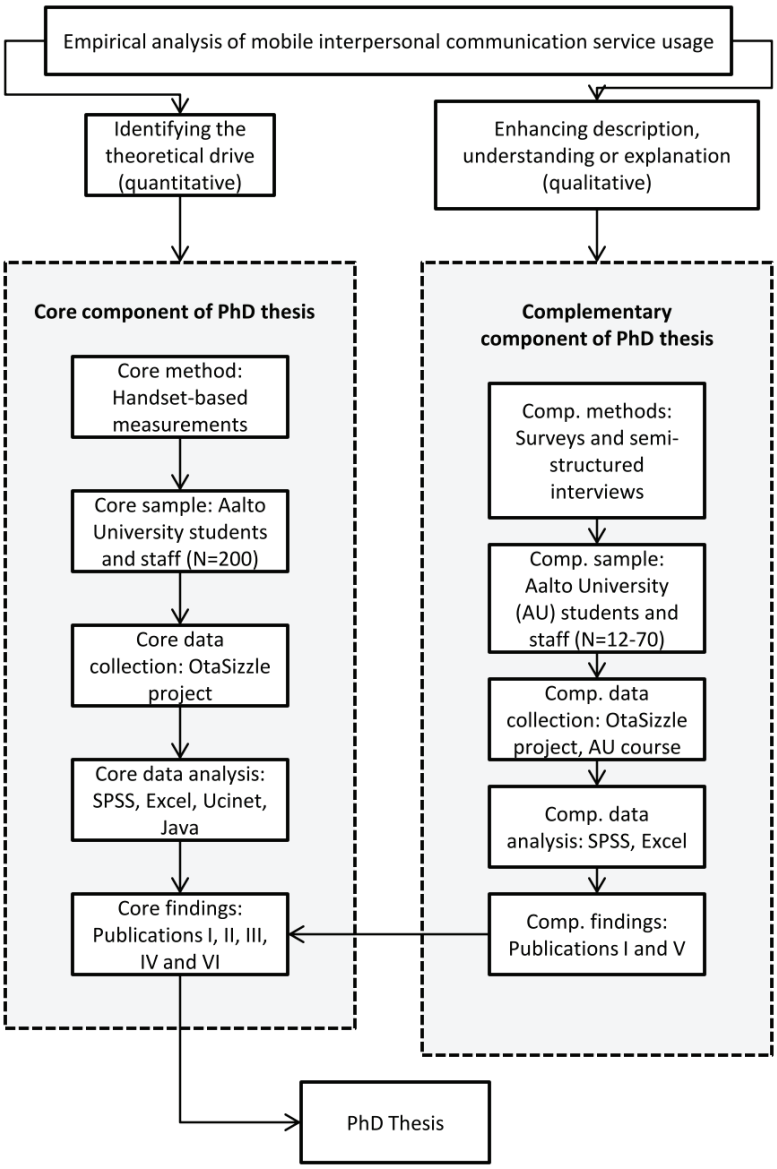

Figure 3 Quantitatively driven mixed method design of the thesis (adapted from Morse, 2010)

The research approach of this thesis is implemented as a panel study. With the term panel, we refer to collecting data with different methods from the same sample of individuals over long time periods. In this thesis, handset-based panels have a span of more than a year, and surveys are administered to the panelists throughout the course of the panel. Thus, the panel study as a whole concerns all the steps from designing and setting up the measurements and surveys, recruiting the participants, actual measuring, as well as data preprocessing and analysis. Because this methodological approach concerns the collection of a variety of personal data from the panel participants, informed consent (i.e., opt in) is required from all participants. The World Economic Forum (2011) has defined personal data as '*data relating to an identified or identifiable person or persons*', and divided it to three forms – volunteered, observed and inferred data. *Volunteered data* refer to personal information that is shared explicitly by individuals. *Observed data*, on the other hand, are captured by logging activities of individuals. Finally, *inferred data* are based on the analysis of personal data. In our measurements volunteered data are collected when the individuals opt-in to participate in the measurements and answer the surveys about, for example, demographics. Observed data are collected by means of the software installed in the participants' mobile phones in terms of, for instance, communication service usage and location. Finally, the context detection algorithm utilized in this thesis enables inferring contextual information of the participants, which belongs to the category of inferred data. Next, we will go through the individual methods that are part of the quantitatively driven mixed method design of this thesis.

4.2 Handset-based measurements

As discussed in the methodological research approach chapter above, handset-based measurements are the core method used in this thesis and are depicted in detail in Publication I, and previously by Verkasalo (2009b). With this method, quantitative behavioral data can be collected from mobile phone users by installing a research application to their devices. The data collected are location-based and context-sensitive as the mobile phones are always carried with the users, personal and "always on" type of devices. Moreover, even though a user might have multiple mobile phones, the underlying assumption is that nobody else is using them.

Handset-based measurements have only been possible to conduct on a large-scale after the introduction of the smartphone (with its ability to install third party software on it) and 3G/4G mobile networks. Especially in social sciences, the mobile phone (or smartphone) has been identified as an

emerging and useful research tool (Raento et al., 2009). Moreover, researchers in psychology have recently argued that mobile phones should be taken more seriously as research tools as their capabilities encompass and surpass most existing tools (Miller, 2012). Even if the method itself is not widely used yet, some related work in industry and academia exist, which will be reviewed shortly. The Context project of Helsinki Institute for Information Technology HIIT was one of the earliest handset-based data collection efforts. They wanted to provide researchers a robust and reliable tool for collecting handset-based data and thus developed the ContextLogger software on the ContextPhone platform (Raento et al., 2005). This software has been used, for instance, to study mobility patterns and collect data on Bluetooth device proximity. The ContextLogger software was adopted by the Reality Mining project of Massachusetts Institute of Technology, where they used a version of it, along with other self-developed software to collect handset-based data (Eagle & Pentland, 2006). The publications from this project are among the most cited handset-based studies, including identifying structure in daily behavior (Eagle & Pentland, 2009) and inferring social network structure (Eagle et al., 2009). The work conducted in the Reality Mining project has been continued by Aharony et al. (2011), who have recently developed an open source software framework for handset-based data collection called Funf. Another more recent academic handset-based data collection effort has been conducted by Falaki et al. (2011). They have developed a tool called SystemSens for monitoring contextual smartphone usage and published results related to the traffic on smartphones (Falaki et al., 2010a) and diversity in smartphone usage (Falaki et al., 2010b). In the industry sector, Nokia Research Center has been collecting handset-based data in Lausanne, Switzerland, and recently they openly released the data for academic research purposes as part of the Nokia Mobile Data Challenge 2012 (Laurila et al., 2012). For a more detailed review of related work conducted with handset-based measurements, see Publication I.

The data collection software used in this thesis was initially selected based on a number of attributes, and the software selection process followed simple steps from identification, screening and evaluation to selection (Publication I). Initially the software was called MobiTrack and it was provided by MobiTrack Innovations Ltd. Later on the software was called MyLife and provided by Zokem and most recently by Arbitron Mobile Oy. The software measures real-life user behavior, usage of devices, mobile services and various technical parameters. The data include application usage, application installations, processes, battery levels and charging, Bluetooth and WLAN entries, voice calls, SMSs, MMSs, Uniform Resource

Locator (URL) entries, network sessions and location based on mobile network CIDs. The monitoring software has been available as a Symbian, Google Android, Windows Mobile, BlackBerry and Apple iOS application but most of the data utilized in this thesis are collected from Symbian devices. For more information about the software, see Publication I, and Verkasalo (2010, 2012).

The handset-based measurements utilized in this thesis were conducted in the OtaSizzle project of Aalto University during 2009 and 2012. The measurement participants are all students and staff of Aalto University, and the sample is overrepresented with innovators or early adopters of mobile phones and services when compared to the general population (Publication I). In addition to taking part in the handset-based measurements, the participants responded to surveys in the recruitment phase of the research panel, and also later in the panel (surveys as a research method will be discussed in more detail later in this chapter). The panel recruitment process utilized in this thesis is depicted in **Figure 4**. First of all, the prospective participant (user in **Figure 4**) becomes aware of the measurements via, for example, email advertising or word-of-mouth. In this phase usually some sorts of incentives (e.g., vouchers and device lotteries) for participation are included. Then, the participant fills in an initial survey on a web page designed specifically for this purpose. If the participant does not agree with the terms of use of the survey (i.e., opt in), he or she will not be able to participate in the research. Otherwise, the participant also indicates if he or she has a suitable mobile device for the handset-based measurements. Those participants, who do not have suitable devices, can only participate in the survey part of the research. Others navigate to the measurement software registration page via a link received with SMS or email. Those participants, who successfully fill in the registration details, agree with the research agreements and privacy policy of the software, and install the software on their devices, become active data producers in the measurements. Others will participate in the survey part of the research. However, it is also possible for the participants to navigate directly to the software registration page and skip the survey part. This approach has been successful in increasing the number of participants in the handset-based part of the research. During the measurements the data are pre-processed, stored locally in the device and sent encrypted and compressed to servers once a day. Then the data are aggregated from all the participants and the usage data are exported to the researchers.

Calculating the hit rate (the percent of people targeted for recruitment actually joining the panel) is hard to measure, because people also joined the panel through word-of-mouth during the course of the panel. In case of

email marketing we sent an email invitation in spring 2010 to all of the Aalto University student union members (N~17000), and as a result we got 26 new panelists in four days to the handset-based measurements. So the hit rate of email marketing was approximately 1.5 per mil. A more successful approach in terms of hit rate was to utilize a more personal recruitment strategy, where the panel was promoted in, for example, student events or courses.

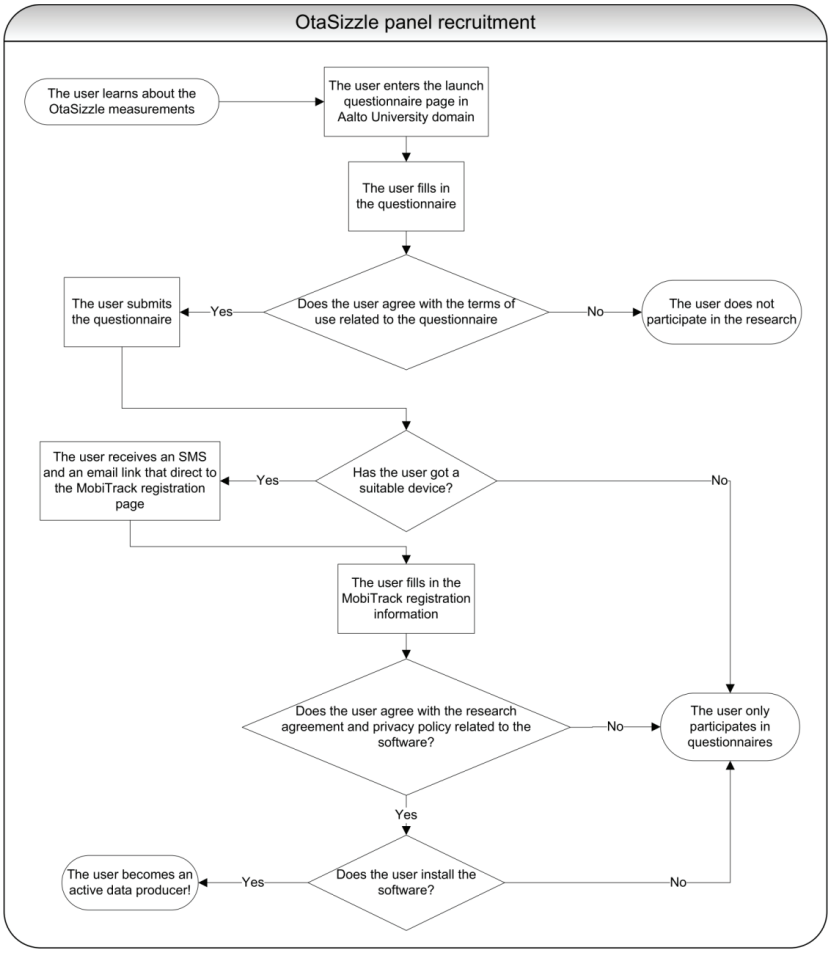

Figure 4 Panel recruitment process[1]

Mobile phones as data collection devices are highly sensitive to privacy issues, because they are personal and always carried. They are also sensing information of the surrounding people and not just the participants themselves, which creates further challenges for privacy and anonymity.

[1] Published in Publication I, p. 13. Copyright 2012 IGI Global. Reprinted with permission.

Moreover, it is a question of debate what personal data constitute, although classifications are emerging such as the one by The World Economic Forum (2011) discussed previously. The solutions regarding privacy in handset-based studies are still to a large extent case-specific, and no standards exist for setting up these kinds of measurements and dealing with behavioral data in general. Our approach to preserve the privacy of the participants is to separate the data handling roles between a limited number of authorized researchers, so that the possible leakage of personal information is reduced to a single responsible researcher. As **Figure 5** portrays, a dedicated person (person N) combines and anonymizes the collected data utilizing numerical identifiers. This person then hands the anonymized data together with the aggregate level demographics of the sample to person X, who performs the aggregate level analysis of the data. Thus, the data handled by person X are only quasi-identifiable and cannot be easily de-anonymized.

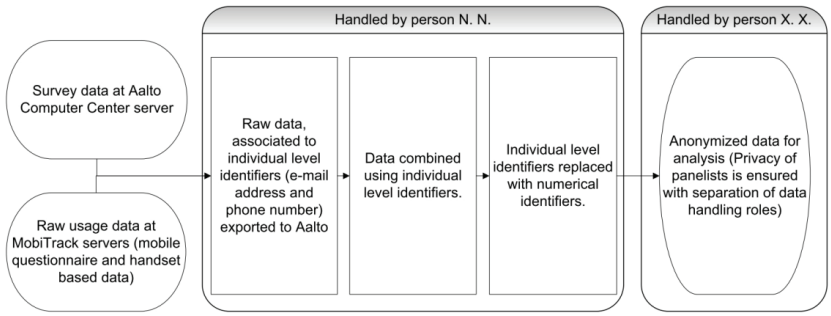

Figure 5 Data anonymization process[2]

These kinds of measurements also have a number of other challenges which are specified in more detail in Publication I. The possible biases that are present in the measurements include conditioning effects, attrition bias and panel selection bias (Lohse et al., 2000), as well as sample selection bias and follow up bias (Wunsch et al., 2010). Moreover, the mapping of mobile services to categories introduces a possible bias to the results, but can be mitigated by using service categorization frameworks such as the one by Smura et al. (2009) used in this thesis.

After the data are collected and before the actual usage analysis can begin, preprocessing of data needs to be conducted. This is a laborious process including, e.g., data cleaning and mapping of services, which only can be automatized to some degree. For example, during the course of the panel,

[2] Published in Publication I, p. 10. Copyright 2012 IGI Global. Reprinted with permission.

more than 500 different mobile services from multiple mobile platforms were manually harmonized and mapped to more holistic service categories. In the scope of this thesis preprocessing and data analysis in general have required both thorough domain knowledge of mobile devices and services, as well as programming skills and statistical know-how. The data utilized in Publications II, III, IV and VI were collected with MobiTrack software during 2009 and 2010 from a longitudinal panel of 200 participants. In addition to that, experience sampling data were collected with Zokem software in 2011 and 2012 from a panel of 31 participants – these data were utilized in Publication III. Furthermore, additional handset-based data previously collected in the Aalto University during 2008 were used in this thesis. These data were collected with the Nokia Smartphone360 software (Verkasalo, 2009b) and utilized in Publication IV. The data are mainly analyzed with statistical tools such as the Statistical Package for the Social Sciences (SPSS) and Microsoft Excel. Ucinet software (Borgatti et al., 2002) is used for social network analysis, and the context detection algorithm presented next is implemented in Java programming language.

4.2.1 Context detection

The context detection algorithm utilized in this thesis (Publication II) has been developed first by Jiménez (2008) and Verkasalo (2009a), and most recently by Soikkeli (2011). The algorithm consists of two parts that work also independently – one is based on CID data and the other on WLAN data. Mobile network CID data are collected from all users and thus the CID based context detection is performed first. After that the WLAN based context detection is performed, as it is possible that an individual user lacks WLAN data. The WLAN based context detection verifies and gives more accuracy to the results extracted from the CID data. The detected place-related contexts include *Home*, *Office*, *Other meaningful*, *Elsewhere* and *Abroad*. *Home* refers to a context, where the majority of nighttime is spent, and *Office* refers to a context where the majority of normal office hours are spent. A considerable amount of time is spent in the *Other meaningful* context, but it does not have the characteristics of *Home* or *Office* context. For example, parents' or spouse's apartment can be an *Other meaningful* context. Places that are not frequently visited, and being on the move, belong to the context *Elsewhere*, which also refers to contexts which are not that meaningful for a person. Finally, *Abroad* refers to a context outside the country, where the majority of the person's time is spent.

The CIDs are collected to a time stamped data log every time the CID changes. The CID gives information under which mobile network base station the user resides, and thus the algorithm recognizes how many times a given cell is visited, when it is visited and how much time is spent there.

This information, together with some time based heuristics, can be used to infer the importance of the location (or cell) to a user. The time based heuristics used are based on everyday observations of people's daily routines and statistical data on how Europeans (Eurostat, 2004) and Finns in particular (OSF, 2011) spend their time. For instance, people tend to be at work or school during daytime and at home during nighttime on weekdays. The first step in the CID based context detection is to detect if the user is abroad by using the Mobile Country Codes (MCC). The MCC with the most CIDs is assumed as the user's country of residence and the rest are treated as abroad. For the foreign CIDs, the context *Abroad* is assigned and no further detection is performed. In the second step adjacent cells are clustered into a cell cluster, if the algorithm detects that the mobile device has jumped back and forth between overlapping cells (even a stationary device can do this). Finally, the CID based context (*Abroad*, *Home*, *Office*, *Other meaningful* or *Elsewhere*) is detected and assigned to the cells or cell clusters.

In the WLAN based context detection the sensed Access Points (APs) within range of the mobile phone are used to verify and give more accuracy to the results obtained in the CID based detection. The APs have unique Media Access Control (MAC) addresses which are logged every 30 minutes or whenever the WLAN is used by the user. The signal strengths are also logged and used together with the MAC addresses to recognize WLAN fingerprints pertaining to a particular location. If the fingerprints are similar enough in terms of common APs and the Spearman rank-order correlation coefficient of the ordered relative signal strengths, they are assigned to the same location. The contexts are assigned to the locations in a similar heuristic based manner as in the CID based detection. However, as the resolution of the WLAN data is lower, no *Elsewhere* context can be detected with the WLAN based method. Thus, it is only used to increase the accuracy of the CID based context detection of the more static or significant contexts such as *Home*, *Office* or *Other meaningful*.

The final step of the algorithm combines the two parts. The context information acquired with the WLAN based detection overrides the CID based contexts, and thus, two significant places under a single cell (or cell cluster), for instance, can be separated from each other. We evaluated the accuracy of the context detection algorithm by utilizing subjective ground truth data (i.e., user given labels for contexts) collected from data collection participants. According to Soikkeli et al. (2012) the average accuracy of the algorithm, measured over all users as the time spent in accurately labeled contexts per the time spent in all contexts, is slightly below 90%. *Home* and *Office* contexts are detected quite reliably, whereas *Other meaningful* and

Elsewhere contexts are more difficult to detect accurately. For more information about the context detection algorithm, please see Soikkeli (2011) and Soikkeli et al. (2012), where it is documented in detail. Furthermore, other related studies where a version of the context detection algorithm has been utilized, have been conducted by Soikkeli et al. (2011) and Smura (2008).

4.2.2 Experience sampling

The handset-based data collection software enables experience sampling from the participants. This variant of the Experience Sampling Method (ESM) (Hektner et al., 2007) is used in this thesis (Publication III) to sample the participants' experiences related to the use of mobile interpersonal communication services. ESM has been used to study the experience of individuals in the natural contexts of everyday life since the early 1970s already. Experience refers to any of the contents of consciousness of an individual (Hektner et al., 2007). In ESM questions are asked from individuals at random points of the day by using signaling devices that prompt them to respond. Beepers have been used as signaling devices traditionally and pen and paper have been used for documentation. However, since the introduction of computerized ESM (Feldmann Barrett & Barrett, 2001), more developed devices and methods have been used for signaling and documentation.

In our version of the ESM, we used mobile phones as signaling devices and collected the ESM response data in conjunction with other handset-based mobile phone usage data. The questions can be triggered based on certain events, in a time-based manner or a combination of these. For instance, the communication service related questions were asked after voice calls were disconnected or ended by a user, and after a message had been sent to or received by a user (with a minimum delay of 12 hours between consecutive questions) (Publication III). Our version differs from traditional ESM, since the sampling is not purely random. However, if an individual's experience related to a certain event needs to be collected, the sampling needs to happen instantly. Otherwise the response quality might get deteriorated, for example, because of lapse of memory. Previously, the ESM has been used for similar purposes as in this thesis by, for example, Froehlich et al. (2007) and Kubey et al. (1996).

4.3 Surveys

Surveys are used in this thesis as a complementary method for handset-based measurements. As discussed in the previous subchapter, surveys were used, for example, when recruiting the participants to the research panel. Moreover, additional surveys were targeted to the participants as the panel research advanced. Surveys are used in general to collect data from

and about people, and they provide a straightforward approach to study, for instance, attitudes, values, beliefs, and motives of individuals. Although they come in a variety of forms, Robson (2011) has tried to identify the typical features of a survey. According to him a survey typically has a fixed design and the data are collected in a standardized form from a large group of people with samples of individuals representing known populations. Surveys are not well suited for exploratory research (Robson, 2011), contrary to semi-structured interviews which are discussed in the next subchapter. With online survey tools, such as LimeSurvey (2012), surveys can be set up quickly and with low costs. Furthermore, more complex survey structures can be implemented with online surveys than with traditional paper-based surveys, and visual aids can be used to support the survey. Survey data are affected by the characteristics of the respondents (and non-respondents) and a number of response biases (e.g., social desirability response bias). Respondents might also misunderstand the survey questions and not treat the survey seriously. The reliability of a survey can be assessed based on its internal and external validity (Robson, 2011).

In this thesis (Publication I) an online survey was deployed in a qualitative manner to complement the data acquired with handset-based measurements. Rather than conducting extensive statistical analysis and trying to generalize the results, the survey data were used to segment the participants of the measurements, and to develop the data collection process. The survey was purposefully administered to all handset-based data collection participants of the OtaSizzle project during the fall of 2010, after the participants had been producing data for about a year. In total 70 usable responses were received, totaling a 35% response rate from the participants. The respondents are biased towards technology-oriented Finnish male students in their mid-twenties.

4.4 Semi-structured interviews

When researching phenomena that are not yet well understood or defined, it is generally agreed that qualitative research is the only approach to have (see, e.g., Ritchie & Lewis, 2003). Moreover, the generative nature of qualitative research provides rich and highly illuminating data that can be used to explore further, towards more conclusive research opportunities (Malhotra et al., 2012; Robson, 2011). Thus, semi-structured interviews were used in this thesis (Publication V) for explorative research purposes, and as a complementary method to the handset-based measurements. The goal of the semi-structured interviews was to explore mobile phone users' perceptions towards new service concepts and ideas that were thought of beforehand.

In a semi-structured interview, the interviewer has an interview guide, which consists of the topics to be covered and a default wording and order of the questions to be asked. However, the wording and the order of the questions can vary based on the flow of the interview and also additional unplanned questions can be asked. Probe questions (or probing) can also be used to get the interviewee to expand or elaborate a response more (Robson, 2011; Malhotra et al., 2012). A semi-structured interview was developed in this thesis based on research on prior art in industry and academia (Publication V). Furthermore, comments from colleagues in social sciences and mobile industry experts were used to iterate the interview structure and probes. The interviews were carried out individually instead of focus-group interviews, so that there would not be any social pressure to conform to the group's general opinions. The goal of the research was not to prove a theory in a statistical way.

The interview sample consists of 12 technology-oriented master level students of Aalto University biased towards Finnish male students in their mid-twenties. To aim for diversity in the data, students using different mobile operating systems and devices were purposefully recruited as interview participants. The interviews were carried out in January and February 2012. This kind of a sample is good for studying and exploring an emerging new service concept, for which rich data might not be accessible with a sample representing the general population.

5. Results

This chapter presents the results of the thesis following the research questions presented in the introduction chapter. The summary of the results is presented and contemplated from the perspective of the research problem at the end of the chapter. The first subchapter contributes to the body of knowledge regarding measurements conducted with mobile phones, and functions as a foundation of the thesis research. The following two subchapters present results of empirical usage studies and interviews on mobile interpersonal communication services. Finally, the last subchapter before the summary presents the results of the social network analysis conducted on multiple communication services.

5.1 Measurement development

Handset-based measurements as an emerging data collection method are not widely documented in detail in academia. Thus, one of the purposes of this thesis is to shed light on how these kinds of measurements are conducted, by depicting the data collection process used in the thesis in the most detailed manner possible. Furthermore, participants' motivations and attitudes towards the measurements and mobile phones and services in general are surveyed, in addition to collecting basic demographic information (Publication I).

The handset-based data collection process steps include data collection software selection, data acquisition and panel recruitment (Publication I) as discussed in the methods and data chapter of the thesis. Moreover, several challenges are identified related to the collection of handset-based data. These encompass legal agreements, the lack of standards related to behavioral data and the method as a whole, as well as the privacy and anonymity of the participants. The major lessons learnt based on the measurements conducted in this thesis include the following: 1) The data collection process is iterative, as mobile phones are equipped with more and more sensors, 2) not all sensor data can be collected – the data collection must be carried out keeping the device limitations in mind, and 3) the participants cannot be burdened with additional tasks without careful forethought, instead they need to be provided with strong enough incentives for participation. The process description also provides

guidelines for other researchers interested in following a similar data collection approach. As said, the process is also depicted in more detail than reported previously in other data collection efforts.

The survey results related to the data collection process (Publication I) can be divided into three parts. First of all, the measurement participants (N=70) were asked for their valuations of different incentives (i.e., motivation) related to the data collection. Then, the innovativeness and general privacy concern levels of the participants (i.e., attitudes) were measured. The results indicate that the participants value the device and voucher lotteries as the most important incentives for data collection (**Figure 6**). However, the participants also seem to be interested in research activities in general. Innovators and early adopters of mobile phones and services are over-represented in the sample, when compared to the general population (**Figure 7**). Finally, the majority of the participants are generally trustful towards organizations collecting their personal information (*the Unconcerned*) (**Figure 8**). However, the participants are quite divided between the extremes, with one third being generally distrustful towards organizations that ask for their personal information (*the Privacy Fundamentalists*). The basic demographic data collected from the surveys indicate that the participants of the data collection are biased towards technology-oriented Finnish male university students in their early twenties.

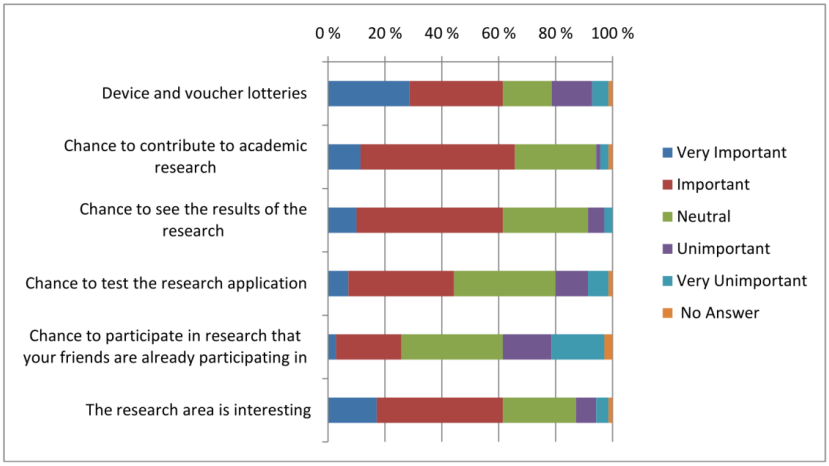

Figure 6 Responses to incentive survey[3]

[3] Published in Publication I, p. 14. Copyright 2012 IGI Global. Reprinted with permission.

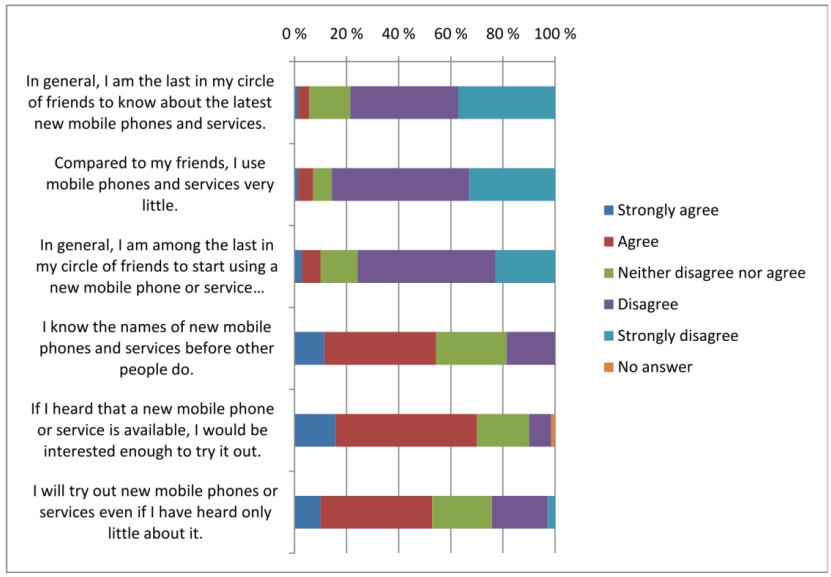

Figure 7 Responses to innovativeness survey[4]

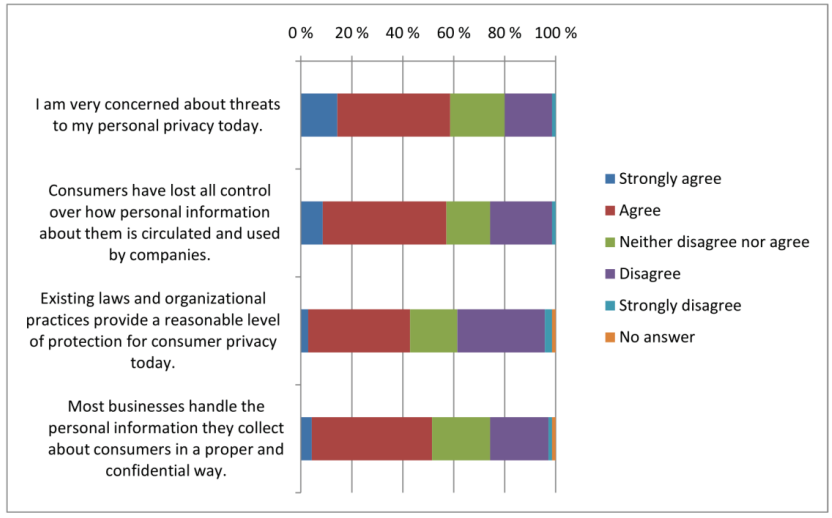

Figure 8 Responses to privacy concern level survey[5]

5.2 Contextual and social usage patterns

Another purpose of the thesis is to study contextual and social usage patterns of mobile interpersonal communication services empirically. First, the context of mobile phone use is identified by means of mobile network CID and WLAN AP data as described in the methods and data chapter of

[4] Published in Publication I, p. 15. Copyright 2012 IGI Global. Reprinted with permission.
[5] Published in Publication I, p. 16. Copyright 2012 IGI Global. Reprinted with permission.

the thesis. Then, the context is used as a variable affecting the use of different communication services (Publication II). Third, it is studied how social capital is built with communication services, namely SMSs and voice calls (Publication III). Usually these types of studies focus on the mobile phone as a single entity and specific communication services are not analyzed. Thus, it is analyzed in this thesis how people bond (communicate with strong ties) or bridge (communicate with weak ties) social capital using these services.

The handset-based data utilized in the context detection (Publication II) include the time stamped (one second resolution) CID of the mobile network base station that the handset is connected to, and the surrounding WLAN APs detected by the device (scanning happens every 30 minutes and whenever the user connects to an AP). Furthermore, the data utilized in the service usage analysis include application foreground data, actual voice calls made and actual SMS, MMS and email messages sent. Application foreground data are collected when an application is visible to the user in the foreground of the device's screen, and not, for instance, when it is running in the background as a process.

The identified place-related contexts include *Home*, *Office*, *Other meaningful, Elsewhere* and *Abroad*. *Home* refers to a context, where the majority of nighttime is spent, and *Office* refers to a context where the majority of normal office hours are spent. A considerable amount of time is spent in *Other meaningful* contexts, but they do not have the characteristics of *Home* or *Office* context. For example, parents' or spouse's apartment can be an *Other meaningful* context. Places that are not frequently visited, and being on the move, belong to the context *Elsewhere*, which refers to contexts which are not that meaningful for a person. Finally *Abroad* refers to a context outside the person's homeland; however, it is not utilized in this thesis. The contexts are identified by comparing WLAN AP fingerprints and (clusters of) CIDs. Time-based heuristics are used to label the places inferred from these location-based data. These heuristics are based on everyday observations and common impressions of people's daily routines. **Figure 9** depicts the shares of identified contexts per hour of day.

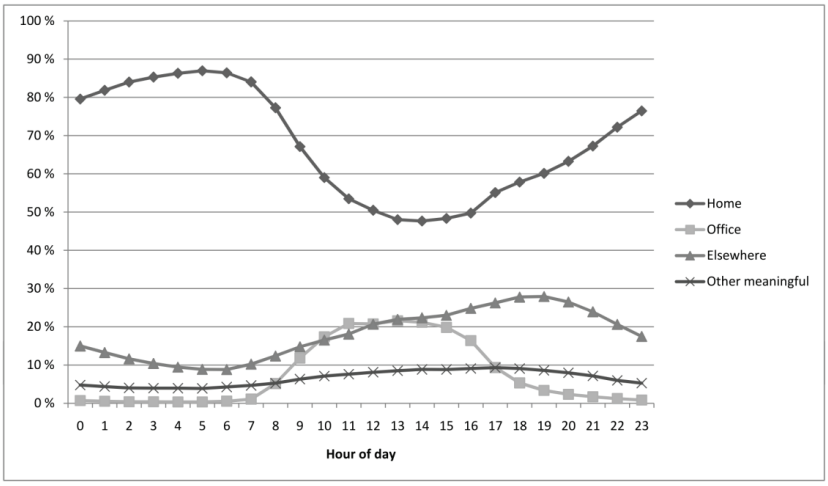

Figure 9 Shares of identified contexts per hour of day[6]

The service usage analysis is conducted on user-level with intensity measures for a total of 140 Symbian mobile phone users. Thus, actions of a user (i.e., application launches, voice calls made or messages sent) are divided by the user's mobile phone interaction time in a given context. This gives us a good measure for how intensively the users have used different services when interacting with the mobile phone in a given context. Non-usage and the length of the voice calls are also taken into account in the analysis. To classify the mobile services to different categories, the framework of Smura et al. (2009) is used, with the exception of combining messaging and calling categories to a single category called communication.

The results of the analysis indicate that people use communication services differently depending on the context of mobile phone use. Communication category services (compared to services of the other categories, such as multimedia or business / productivity) are used most intensively in all contexts. This suggests that the mobile phone is still used mainly for communication purposes, regardless of the context. Overall, services of the communication category are used most intensively in the *Office* context. The lengthiest voice calls are made in the *Home* context, where the voice calls are used least intensively. The *Office* context is characterized by the most intensive usage of email and SMS. Furthermore, voice calls are the briefest in the *Office* context. Finally, social media and mobile IM and VoIP are used most intensively in *Elsewhere* and *Other meaningful* contexts. These results together indicate that the use context is a factor affecting the choice of communication service used.

[6] Published in Publication II, p. 496. Copyright 2011 Springer-Verlag London Limited. Reprinted with permission.

When studying the social usage patterns (Publication III), the main analysis is based on actual SMS and voice call communication logs collected from handset-based measurements from 200 mobile phone users. Contrary to the contextual analysis, we do not only focus on outbound communication, but consider also inbound communication in the analysis, since the operationalization used in the research is based on distributions of contacts and communication activity in general. Furthermore, multipart SMSs (i.e., messages consisting of more than 160 characters) are considered as single messages in our analysis, and the Christmas and New Year's Eves are excluded from analysis. We also assume in the analysis that a single individual has only one phone number.

The operationalization can be divided into two parts: 1) the larger the number of unique contacts with whom SMSs (voice calls) are exchanged per day, the larger the level of bridging activity of the SMS (voice call) service, and 2) the larger the number of SMS (voice call) events per unique contact per day, the larger the level of bonding activity of the SMS (voice call) service. In other words, 1) the larger the service specific degree of a node, the larger the level of bridging activity of that service, and 2) the larger the weight of the service specific edge, the larger the level of bonding activity of that service. Moreover, it was hypothesized that people use more media to communicate to strong ties than to weak ties (so called media multiplexity). Consequently, if both voice calls and SMSs are used to communicate to a given contact, then the level of bonding activity is greater.

The results indicate that although both SMSs and voice calls are used for bonding and bridging social capital, SMSs are used more for bonding purposes than voice calls. Furthermore, media multiplexity is more associated with bonding than bridging activity. This means that voice calls are more likely exchanged with those contacts that SMSs are also exchanged. On the other hand, SMSs are not necessarily exchanged with those contacts that voice calls are exchanged. In addition to the SMS and voice call analysis, a pilot study utilizing experience sampling to study social capital in the context of emerging communication services, such as mobile Internet or OTT services is presented. This pilot study provides a framework for how to conduct social capital related communication service usage research in a more detailed manner in the future. Based on these results, it can be claimed that the strength of the tie is another factor affecting the decision to use or select a given communication service.

5.3 Service usage interrelationships

Studying the usage interrelationships of mobile interpersonal communication services and the effect of Mobile Social Phonebooks (MSP) on these interrelationships is another purpose of the thesis. First, handset-

based data are utilized to study if direct substitution can be found between the traditional operator-provided communication services (i.e., voice calls, SMSs and MMSs) and the mobile Internet or OTT services, such as IM, VoIP, email and social media services (Publication IV). This was accomplished by analyzing the user-level service usage correlations with users that have experience in using these services. This method enables more fine-grained analysis than is possible with market-level data analysis as discussed in the introduction chapter. Then, more usage interrelationships are identified and semi-structured interviews are carried out to study mobile phone user perceptions towards MSPs and their effects on the usage interrelations (Publication V). As these phonebooks are rather new and emerging in the mass consumer markets, a semi-structured interview was selected as the research approach, as it provides rich data that can be used to explore further.

The user-level correlation analysis (Publication IV) is conducted for 183 Symbian mobile phone users. In a similar manner as in the contextual analysis, we consider user actions as application launches, voice calls made or messages sent. Usage intensity is considered, meaning that actions are calculated on user-level on a daily basis. However, we also consider the duration of service usage for all services, whereas in the contextual analysis we only considered it for voice calls. The framework of Smura et al. (2009) is again used in overall categorization of the services, and mobile Internet services are further divided to three categories, namely email, IM/VoIP and social media (e.g., Facebook). IM and VoIP are classified to the same category, because most of the services in that category contain both features (e.g., Skype).

The results indicate that no clear evidence can be found in favor of substitution between the categories of traditional and mobile Internet (or OTT) communication services (Publication IV). Email usage was also analyzed separately, but no substitution was observed between it and the use of voice calls and SMSs. Consequently, the services are considered to be independent services, which are possibly used for different purposes. These results are also supported by similar studies conducted in academia.

The term MSP is coined in this thesis. The term refers to the integration of contacts from mobile Internet (or OTT) services to the mobile device's phonebook. The term is characterized as follows: '*To distinguish the term from other non-mobile phone based contact lists and to emphasize the fact that we are dealing with mobile phones and not with other mobile devices such as tablets, we use the terms mobile and phonebook. The term social on the other hand refers to the fact that this integration of multiple social networks on the device's phonebook makes the phonebook more social for*

the user' (Publication V). Three general types of usage interrelationships are identified in the thesis, namely complementary, independent and substitutional.

The mobile phone user perception results indicate that the interviewees see communication services as different in nature, and thus use them differently depending on a variety of factors. The identified factors include the type of relationship, own use context (physical, social), presence of others (location, availability), costs and subscription type, communication type (interpersonal vs. group, leisure vs. work or school) and device status (e.g., battery level). However, the interviewees still see mobile communication as phonebook-centric, although communication can be initiated from, e.g., the communication log application of the device, or directly from the different communication services (e.g., Facebook) themselves.

As another result based on the literature review and the interviews, we claim that the MSPs will affect the usage interrelationships of communication services, and that the future analysis of these interrelationships needs to happen in a more detailed manner than what has been achieved so far. This means that usage interrelationships need to be studied on user-level, for example, between specific services such as IM and SMS or MMS, or between VoIP and voice calls. Regarding the general usage interrelations, the MSP can enable more enriched connections and result in complementation, provide easier access to substituting services and result in substitution, or result in independence, if the services have different affordances.

5.4 Social network analysis

The thesis also utilizes mobile phone communication data gathered with handset-based measurements, and data from an online social media service developed in the OtaSizzle project of Aalto University with the purpose of studying how the social networks derived from distinct communication services differ (Publication VI). First, the SMS and voice call logs gathered from handset-based measurements are utilized to identify the social network under analysis. This network of identified users (N=20) is then analyzed from the online social media service and compared to the mobile phone based social networks.

Handset-based data were collected from 113 participants in the fall of 2009. After excluding those participants, who 1) did not communicate to each other during the data collection period, 2) had produced data for less than a week, and 3) had not been members of the online social media service for a minimum of three months, we were left with a group of 20 participants with whom the social network analysis was performed. The

analysis is performed both on full network and node level, and the measures used can be divided to centrality and cohesion measures. Centrality measures include degree, betweenness and network centralization index. Cohesion measures, on the other hand, include density, diameter and transitivity.

The social network structure analysis indicates that the online social media service network (**Figure 10**) is denser and has a larger ego network for a user on average than the mobile phone networks (**Figure 11** & **Figure 12**). The node sizes are set by the degrees of the nodes and the thin red lines indicate reciprocal ties, whereas the thick blue lines indicate non-reciprocal ties. The weights of the nodes are multiplied by ten for illustrational purposes and indicate the numbers of messages sent or voice calls made per an active day of mobile phone usage. Furthermore, the node degree correlation results indicate that the correlations between the online social media service network and the mobile phone networks are not significant, meaning that the networks may not be correlated at all. These results indicate that the online social media service has a different kind of friendship structure than those based on mobile phone communication.

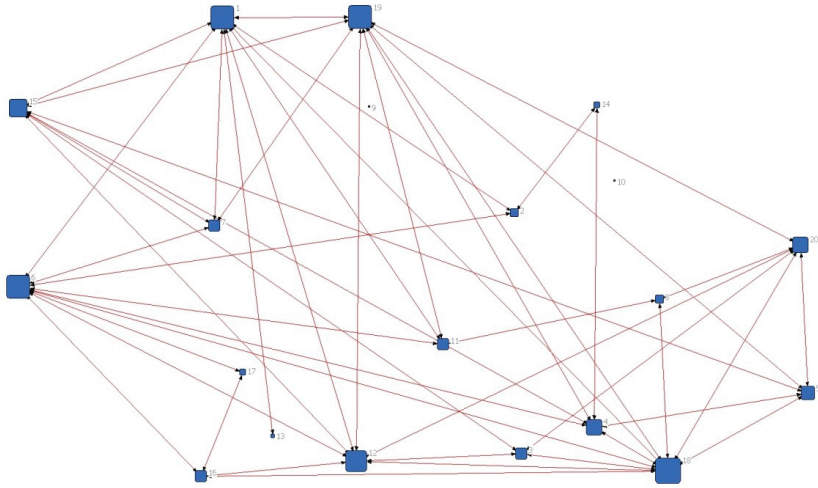

Figure 10 Online social media service network[7]

[7] Published in Publication VI, p. 107. Copyright 2011 Inderscience Enterprises Ltd. Reprinted with permission.

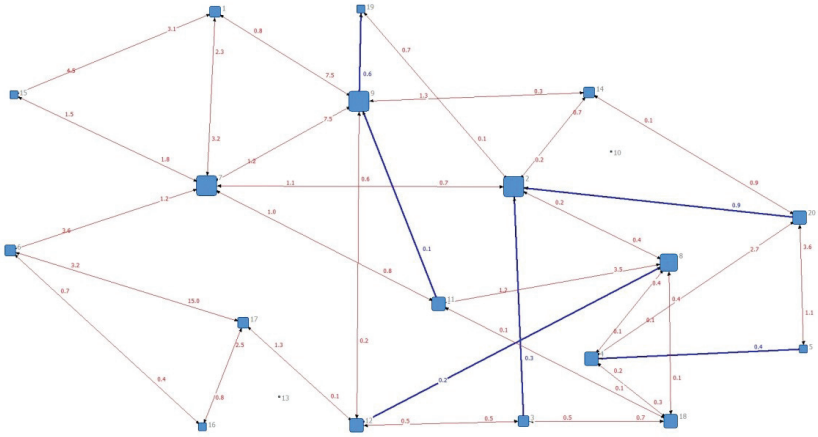

Figure 11 SMS communication network[8]

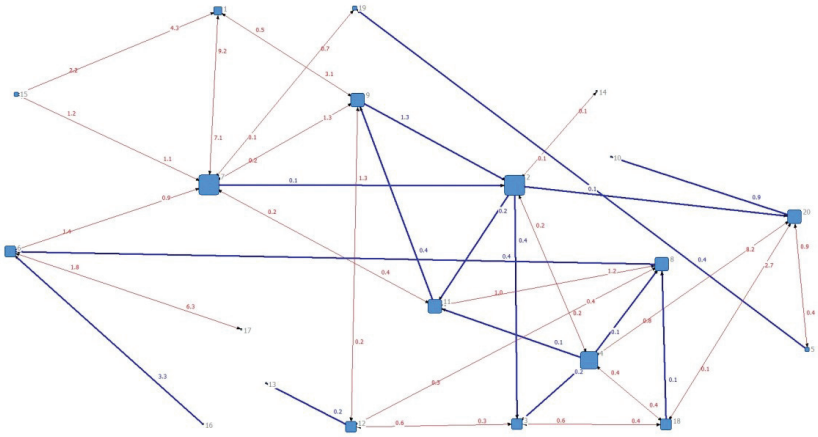

Figure 12 Mobile voice call communication network[9]

5.5 Results summary

The results of this thesis are derived from semi-structured interviews, surveys, and handset-based usage measurements of mobile interpersonal communication services. **Table 4** summarizes the publications, research questions, methods, and main results of this thesis. These results together answer the research problem, namely '*How can one empirically analyze the use of mobile interpersonal communication services utilizing handset-based measurements and complementary methods?*'

Handset-based measurements provide the foundation for the research, because the actual usage data analyzed are collected from them. Thus, the

[8] Published in Publication VI, p. 106. Copyright 2011 Inderscience Enterprises Ltd. Reprinted with permission.
[9] Published in Publication VI, p. 106. Copyright 2011 Inderscience Enterprises Ltd. Reprinted with permission.

handset-based data collection process used in this thesis is presented as a guideline for other researchers interested in following a similar data collection approach. However, complementary methods need to be used to identify the motivations, attitudes, and demographics of the sample being used, and to research services that are not yet accessible with handset-based measurements because of low or non-existent use. The results of the handset-based usage studies demonstrate how the rich data collected can be used in a number of ways to empirically analyze the use of mobile interpersonal communication services. First, it is identified that the service usage varies in different mobile phone use contexts. For instance, voice calls are the briefest in the *Office* context, whereas the lengthiest voice calls are made in the *Home* context. Second, it is observed that the strength of the tie between the persons communicating affects the use of communication services. For example, SMSs are used more to communicate with strong ties than voice calls. Third, as a result of the latter two studies and a usage interrelationship study, it is identified that no direct substitution between traditional and OTT services exists. Finally, when comparing mobile phone communication data with data collected from an online social media service, it is observed that the social networks derived from distinct communication services are different in nature, in terms of, e.g., node degrees.

According to the surveys conducted in this thesis, the sample used in the handset-based measurements is biased towards early adopters or innovators of mobile phones and services, who are also interested in research activities in general. This kind of a sample is well suited for studying newer services that are not yet adopted by the general population. The sample can be more of a layman nature for analyzing the use of traditional services, as those services are being used by practically all mobile phone users. Semi-structured interviews were used to explore a new service concept, which was coined as mobile social phonebook in this thesis. Three different usage interrelationships were identified in the interviews, namely complementary, independent, and substitutional. It was also observed that a need for more detailed usage interrelationship analysis exists. In addition to the mobile phone use context and strength of the tie between the persons communicating discussed above, a number of other factors affecting the use of communication services were identified in the interviews. These include presence of others (location, availability), costs and subscription type, communication type (interpersonal vs. group, leisure vs. work or school), and device status (e.g., battery level). Moreover, it was observed that the interviewees still see mobile communication as phonebook-centric.

As discussed, this thesis utilizes a mixed method approach, and depicts different usage studies on communication services to answer the research problem and the individual research questions, respectively. Thus, the contributions of this thesis are both methodological and empirical, spanning chapters four and five. Next, the empirical results are contemplated from a mobile operator's, and communication or social network researcher's perspectives. Methodological and privacy implications are also discussed. The discussion in the next chapter builds on the contributions presented in the last two chapters, and reflects the results on the theoretical background and related empirical research presented in chapters two and three, respectively.

Table 4 Publications, research questions, methods, and main results of this thesis

Publication	I	II	III	IV	V	VI
Research question	RQ1: How can behavioral data be collected with handset-based measurements, and what motivational and attitudinal factors need to be considered?	RQ2: How are mobile interpersonal communication services used depending on: 1) the use context of the mobile phone, and 2) the strength of the tie between the persons communicating?		RQ3: How are the usages of mobile interpersonal communication services interrelated, and how do the mobile social phonebooks affect the interrelationships?		RQ4: How do the social networks derived from distinct communication services differ?
Methods	Handset-based measurements, Surveys	Handset-based measurements, Context detection	Handset-based measurements	Handset-based measurements	Semi-structured interviews	Handset-based measurements, Online social media service data analysis
Main results	The handset-based data collection process used in this thesis presented as a guideline for other researchers interested in following a similar data collection approach. Moreover, participants of handset-based measurements were identified as early adopters or innovators of mobile phones and services, who are interested in research activities in general.	Mobile interpersonal communication service usage varies in different use contexts (*Home, Office, Other meaningful* and *Elsewhere*).	SMSs are used more to communicate with strong ties than voice calls. Media multiplexity is also more associated with strong than weak ties.	No direct substitution between traditional and mobile Internet (or OTT) communication services exists.	Term Mobile Social Phonebook was coined and three different usage interrelationships were analyzed. Furthermore, a number of factors affecting the use of communication services were identified, and the need for more detailed usage interrelationship analysis was observed.	Social networks based on an online social media service and mobile phone communication are different in nature, in terms of, e.g., node degrees.

6. Discussion

This chapter builds on the results introduced in the last chapter, and discusses how the results can be interpreted and discussed from a number of perspectives. Although the results and discussion can be utilized by a number of actors in the mobile communications ecosystem, they may be especially useful for communication and social network researchers, and mobile operators. Moreover, methodological and privacy implications are discussed, before concluding the thesis to discussion on research limitations and future research.

6.1 Communication and social network research implications

The empirical observations presented in this thesis suggest that mobile interpersonal communication services are used for different purposes, and that many factors affect the decision to use or select a given service, including mobile phone use context and strength of tie between the persons communicating (Publications II and III). It might be that a need for discreet personal communication in the *Office* context exists and thus messaging services are used most intensively there. Furthermore, more (less) restricted time resources in the *Office* (*Home*) context might exist, which results in shorter (longer) voice calls there. Verkasalo (2009a) also observed that voice calls tend to be significantly longer in the *Home* context, whereas Do et al. (2011) observed that voice calls and SMSs are associated with moving (or nomadic), and stationary indoor contexts, respectively. Regarding the strength of the tie, we speculate that SMS might be perceived as a more personal form of communication than voice calls, and thus it is used more for bonding social capital (i.e., communication with strong ties). Similar results have also been reported by, for instance, Boase & Kobayashi (2012). In the future, the newer communication services might be used more for bonding than bridging social capital purposes, as '*technology may slow relationship decay rate, but be poor for creating new ones*' (Dunbar, 2010). On the other hand, de Bailliencourt et al. (2011) have predicted that the convergence of communication services will iron out the current differences between various service usage rates.

Based on the interview study (Publication V), mobile voice calls are still the primary means of communication, because they are effortless in nature and they entail the ability to get an immediate response. Messaging, however, is used in contexts where the use of voice calls might be disturbing, such as public transport or lectures. IM seems to be preferred for group communication because it is more interactive than the traditional services, and photos, for example, can be shared within the group. Email in the mobile domain is perceived as a work or school related communication service, which is more used to communicate with weak than strong ties. VoIP is used for international calls by those mobile phone users who do not have a usage-based mobile data subscription, mainly because of the high costs of international circuit-switched voice calls. Presence information is perceived as more important with OTT than with the traditional communication services. The reason for this might be that with voice calls and SMSs one can always be sure that the message reaches the other party if their devices are on. With IM and VoIP you usually have to have the application running on the device to be able to communicate. This is now changing, however, with the emergence of platform-specific communication services such as Apple's iMessage. With the majority of the cross-platform services, such as WhatsApp, you still need to have the application running on the device. Finally, mobile social media services differ from the other communication services, because they are not necessarily used only for interpersonal communication, but also for time killing purposes, by, for instance, browsing through friends' status updates. It was also observed in this thesis that social media services are more free-time oriented, being used the most intensively in other contexts than *Home* and *Office* (Publication II). When considering mobile devices and services, one cannot exclude the effect of battery life on service usage. Thus, communication services in general are used less when battery is running low and especially synchronization of emails is discontinued.

Based on these empirical observations and the theory of communication technology affordances, we speculate that:

- Voice calls afford effortless communication regardless of who is being contacted
- SMS affords more personal and discreet communication
- IM affords interactive group communication with multimedia content
- VoIP affords voice communication internationally
- Email affords work or school related communication
- Social media services afford time killing purposes

Along the lines of Boase (2013), we believe that these affordances are general enough, so that they exist on multiple mobile platforms and are thus not greatly affected by the specifics of different software and hardware configurations.

As previously mentioned, people use multiple communication services to stay socially connected to each other. We claim that this personal communication system (Boase, 2008) consists of services that have different affordances. Consequently, the services each serve a distinct purpose, and direct displacement between the services in the system is not likely. If a service outside the system is to displace an existing service, then it needs to offer at least the same affordances that the existing service is offering. Adopting a system view of the different communication services and their affordances is especially salient in the case of mobile communication, as Helles (2013) points out. This thesis presents empirical observations, based on which communication services are used and selected, and consequently, a personal mobile communication system can be constructed. These systems vary from individual to individual, and the interplay between the affordances of different services needs to be understood (Helles, 2013).

Since some communication services are used specifically for some purposes, the resulting social networks derived from distinct services can also differ. This is why we claim that multiple datasets should be used when analyzing the social dynamics of a given population (Publication VI). Furthermore, because of media multiplexity (Haythornthwaite, 2005) the use of multiple datasets is even more important when ties get stronger. In addition to using multiple datasets, the analysis should also be dynamic (i.e., based on time-series data) in an ideal situation. This idea is not novel, however – already a decade ago McPherson et al. (2001) suggested that the priority for future social network researchers should be to gather data on multiple social ties, and one should also strive for dynamic, not static, social network data. This thesis builds on that suggestion.

Another point worth discussing is the validity of measurements, when considering social ties. In general it can be claimed that the quantitative method used in this thesis (i.e., handset-based measurements) is very well suited for studying the usage of communication channels, while the qualitative methods (i.e., surveys and interviews) are more suited towards qualitative examination of the contents of communication. One could also claim that the reliability of data collection on social ties has increased, because of the automatic logging of data, for example, in online services and CDRs of mobile operators. However, it is not always clear what a social

tie represents. For instance, what does it mean to be a friend in Facebook, compared to being a friend in the real world? Furthermore, since we cannot access the content of communication or the identity of the persons communicating in studies utilizing CDRs, how can we make a decision on what kind of social ties we are dealing with? This is why models, such as the one by Nelimarkka & Karikoski (2012), have been presented, that help categorizing social ties. With the help of these models, the nature of different social ties can be understood, and researchers can discuss the domain and validity of their observations better. Thus, we see the validity of measurements as an important question that has not been addressed enough, especially in recent SNA studies based on digital trace data.

6.2 Practical implications for mobile operators

Continuing on the thought that communication services afford different things and are thus used for different purposes, we can discuss the results from the perspective of a mobile operator. In general the usage interrelationships of the services can be substitutive, complementary or independent. The observations of this thesis indicate that mobile interpersonal communication services are not direct substitutes to each other, or at least the possible substitution happens on a more detailed level than what has been analyzed so far. Most of the related research conducted on objective user-level behavioral data (**Table 3**) also suggest that no clear substitutive relation between the services exists. Small substitutive relations between SMS and mobile Internet use as well as between SMS and voice calls have been observed by Gerpott (2010a) and Kim et al. (2010), respectively. However, the authors report that the magnitude (or practical relevance) of the relations is rather low. Talking about media substitution in general, Lin (2011) claims that in order for a new technology to completely substitute an existing technology, at least three criteria need to be met – content superiority, technical superiority, and cost efficiency – otherwise the interrelationship will be complementary or independent. Thus, in terms of technology affordances, the services need to be similar for complete substitution to happen. As Tan & Henten (2006) have discussed, the strategy that the companies pursue is also a key factor affecting whether a new technology will be complementary or substitutive, and thus pave the way for sustaining market change or market disruption. If we look at the current market-level data and the diminishing revenues of the mobile operators, however, possible substitution between different services definitely needs to be taken into account.

The usage interrelationships of the services will evolve in the future, and we claim that the emerging Mobile Social Phonebooks (MSP) will have an effect on the interrelationships. Consequently, time-series data would be

the most valuable data to analyze, similar to using dynamic datasets in social network analysis. Regarding mobile Internet or OTT communication services, for instance, we are still in the early phase of the adoption curve (Rogers, 2003) and network externalities (Courcoubetis & Weber, 2003) will play a key role in them being adopted on a larger scale in the future. Time-series data would enable dynamic analysis where analyzing shifts from one service (category) to another would be possible. One possible switch could be the change from traditional operator-provided services to OTT services, aided by the MSPs. However, one could speculate that the MSPs can also have a complementary effect on the usage of services by enabling enriched social connections for the operator's customers, and thus leading to less churn. Users can also learn to use the existing services in new ways because of the newer services and the MSPs, which would also lead to complementary usage interrelationships between the services.

For some time now, mobile operators have offered their services in bundled packages to their customers. Gerpott et al. (2012) have stated that operators should consider offering several different types of mobile service bundles, for example, differing based on the amount of "free" megabytes of mobile data, SMS messages or voice call minutes included in the monthly charge. They also conclude that so called "fair use" tariff schemes should be used for mobile data. These schemes would reduce the data transmission speed, or impose use-dependent charges to the customers if too much mobile data are used. Kim et al. (2010) also claim that if multiple services are to be bundled, a careful understanding of how the services interact is required. Especially understanding how a new service would interact with the existing services that the firm is offering is critical. We claim that even more complex service bundles could be offered by the operators. The bundles can be more sophisticatedly tailored to different customer segments, based, for instance, on the different personal mobile communication systems or general usage profiles of the customers. In this tailoring process, the usage interrelationships of the services play a key role, and they also affect the pricing strategies of the services as Grzybowski & Pereira (2008) have noted. Operators could utilize more complex pricing strategies for their services as well, such as application-aware pricing or quality-based pricing. However, complex bundles and tariff schemes make it hard for the operator to interpret how changes in them impact the demand and profitability of the company, as pointed out by Kim et al. (2010). Moreover, in the battle against decreasing revenues, operators could even revert completely back to usage-based pricing instead of the nowadays popular flat-rate pricing schemes. Although users are willing to

pay more for flat-rates, these pricing schemes also substantially stimulate usage (Odlyzko, 2009).

In addition to bundling their services and coming up with innovative pricing strategies, operators can also build new innovative service offerings themselves. They could provide the MSP, and thus get even richer data on their customers, which can be utilized to be more competitive in the market and decrease customer churn. Gerpott et al. (2012) have also suggested that operators should develop services, where conversations can seamlessly continue from one service (or communication channel) to another – similar as in MSPs. In fact, mobile operators have already acknowledged this and are developing their own network-based solutions (usually IP Multimedia Subsystem (IMS) based technologies) to compete against the OTT providers. With the RCS, operators are aiming to offer better services with the same functionalities as the OTT communication services. Furthermore, network-based social phonebooks, such as Open Mobile Alliance's (OMA) Converged Address Book (CAB) have been developed offering similar functionalities as the MSP across various devices and services. However, the challenge in these network-based solutions is that they should be interoperable across all operators at least on national level, to enable widespread adoption by the customers. This requires tight collaboration among the operators in a field, which has been traditionally characterized by fierce competition.

Despite challenges that the operators are facing regarding the rapidly changing mobile communications ecosystem, they still have many unique assets based on which they can compete against and in the OTT service providers, and avoid ending up as a sole bit-pipe provider. First and foremost, they have already established billing relationships with the customers and developers, which give them an upper hand in the competition. Moreover, the operators are in possession of rich usage data from their customers, which can be utilized to serve the customers better. One of the key assets is also the telephone number that forms the basis of their relationship with the customers. However, the telephone number can also be utilized by the OTT providers when connecting users, as is the case with, for example, WhatsApp. With regard to the network-based solutions, operators can provide better reliability, security and privacy than the OTT providers, as discussed by Nikou et al. (2012). Moreover, mobile operators were perceived as trustworthy service providers in the interview study of this thesis (Publication V). Finally, mobile operators can collaborate by making international roaming agreements, so that their customers can use the services also abroad without fears of excessive roaming costs.

As Grzybowski & Pereira (2008) and Gerpott et al. (2012) have discussed, also regulators must take the usage interrelationships of different communication services into account when regulating termination, and especially, international roaming prices for these services. The Net Neutrality debates and laws considering economic efficiency vs. fairness (see, e.g., Odlyzko, 2009) will have a direct effect on the mobile operators as well. In fact, some operators are already charging separately for mobile VoIP traffic in Europe, or have even restricted or blocked the use of VoIP in some of their subscriptions (BEREC, 2012). However, in the Netherlands, for instance, a Net Neutrality law is in place, which prohibits such restrictions. This has lead to widespread adoption of OTT communication services, especially WhatsApp, in the Netherlands.

The mobile communications ecosystem is rapidly evolving, and many actors are involved in providing the services, as well as regulating them. It remains to be seen what the future looks like and how the models of collaboration and competition evolve. However, according to De Reuver et al. (2010), a hybrid model, where OTT providers and mobile operators collaborate, would be the most beneficial scenario for both actors. Thus, the actors could leverage their resources and capabilities together. In fact, this scenario has already taken place in some regions. Odlyzko (2009) has claimed that from a societal viewpoint it would be better to enforce Net Neutrality and thus stimulate innovation in the mobile network by others (e.g., OTT providers). This would not be a final solution, however, as new actors would emerge as choke points of innovation and neutrality rules would have to be imposed again. As a final note we would like to stress that concurrently to adjusting to the changes in the mobile communications ecosystem, mobile operators should also seek for the next areas of growth. These can include mobile payments, entertainment and content services, Near Field Communication (NFC) and Machine-to-Machine (M2M) applications, digital homes and mobile identity.

6.3 Methodological implications and privacy

The core method used in this thesis, i.e., handset-based measurements, is highly sensitive to privacy issues, because of the personal nature of the data collection device – the mobile phone. We have made an effort to depict the data collection process utilized in this thesis (Publication I) in more detail, than what has been achieved in other similar efforts conducted in academia and industry before. We believe that more transparency in data collection is needed in the future as signs of an approaching open data era are visible (see, e.g., Laurila et al., 2012; Blondel et al., 2012), and as the users' awareness of different measurements happening in the digital domain is increasing. Regarding the privacy of the participants, we have made a best

effort to fulfill the privacy principles of Langheinrich (2001) in our measurements. However, the types of personal data (volunteered, observed and inferred (World Economic Forum, 2011)) evolve constantly, and the term personal data also has several meanings in the digital world. At the moment we are in a phase, where it usually is determined case by case what personal data constitute. Privacy itself is a complex construct, and includes technical, sociological and legal domains, all of which are intertwined and cannot solve privacy issues solely by themselves. Thus, research ethics play a critical role in these kinds of measurements, and it is essential that the participants know and understand how personal data are processed in the measurements (Kosta & Dumortier, 2008). Boase (2013) has also noted that as the participants of handset-based data collection may not understand all the technical details of how data are collected and processed, researchers should use clear and simple language when describing them. Furthermore, although the data anonymization process used in the thesis (**Figure 5**) is adequate in a university setting where the participants have opted in, there are always concerns about de-anonymization (Lane et al., 2012), re-identification (Beach et al., 2010) or deductive disclosure (Eagle, 2009) of the participants that need to be taken into account.

The data collection process presented in this thesis is a case study of how these kinds of measurements can be conducted and naturally has several things in need of development. As long as no overarching solutions for how to conduct these kinds of measurements and preserve the privacy of the participants exist, these case-specific solutions provide guidelines for other researchers interested in following a similar approach. In the future, the participants' control over the data that are being collected from them could be increased, so that they can choose what data are collected, and delete data if they wish. However, the more control the participants have on the data, the more it might restrict the possibilities of research, if the same data are not collected from all participants. The ownership of personal data has also been a hot topic in the academia lately. Pentland (2009) has suggested that individuals should own the digital data they have produced, and be able to access and remove their data from databases at any point in time. Consequently, Moiso & Minerva (2012) have suggested that a new actor is needed in the personal data ecosystem, who would be a provider of personal data management services, in a similar manner as commercial banks are providers of money management services.

6.4 Limitations

Mobile interpersonal communication services, especially those based on the Internet, can be thought of as a form of new media. As Livingstone (1999) has stated, this kind of research is based on a moving target, because new

media evolve all the time. Thus, we have made an effort to understand the underlying motivations of use and the affordances of communication technologies in our analysis, in addition to describing usage patterns empirically. We want to stress, however, that the results should be treated as descriptive observations rather than the truth, as limitations of the research exist which will be discussed next.

As previously discussed in this thesis, the sample used in the handset-based measurements is biased towards early adopters of mobile phones and services (Finnish university students in their early twenties). This has been the case also in other similar data collection campaigns, such as the one by Nokia Research Center. Kiukkonen et al. (2010) observed that high involvement participants, who are more eager to replace their mobile phones and use the services and functions of their mobile devices, were overrepresented in their sample. Furthermore, as the majority of the participants are students, the contextual results in **Figure 9**, for example, can be different compared to a situation where a panel of full time office workers would have been used as participants. The main challenge still remains that the users of advanced handsets (from whom data collection with the methods used in this thesis is possible) tend to be more technology-savvy than the general population. Although this kind of a sample together with its relatively small size (especially in the social network analysis conducted in Publication VI), yield challenges related to the external validity of the results, we see it as a possibility as well. The innovativeness of the participants enables studying new innovative communication services (such as OTT services in Publication IV), that have not yet been adopted by the general population. The usage patterns of the early adopters might also reflect those of the majority in the future. Moreover, the networked nature of the sample enables studying the social networks among the participants (as in Publication VI), which would not be possible with a random sample of the same size. The handset-based observations are mostly restricted to the Symbian platform. However, in the case of voice calls and SMSs, we claim that the platform plays no significant role in the analysis, since these services are available for all mobile phones. In the future, the practical limitations that different mobile OSs pose to the data collection need to be studied in more detail. In this thesis, for example, we were not able to collect messaging data in iOS devices, and email was only accessible in Symbian devices.

The definitions of certain concepts utilized in this thesis also create challenges to the external validity of the results. For instance, our definition of context (Publication II) relies heavily on location and the type of data collected. Furthermore, assuming that individuals only have a single phone

number (Publication III) is somewhat limited, as people might have several phone numbers that they use actively and simultaneously. In the social network analysis (Publication VI) we were limited by the type of data we were able to get from the online services. For instance, the assumption that all online relations are reciprocal is a potential threat to the external validity. Moreover, the social network analysis is mainly based on full network measures that might conflate individual differences. Thus, the analysis could be enhanced by comparing single node properties across the various communication services. Finally, determining which mobile services belong to which category inevitably brings some subjectivity to the observations. However, we have tried to minimize this subjectivity by using the service classification framework of Smura et al. (2009) consistently throughout the thesis. In general, one could ask what the value of these kinds of empirical analyses is, where a snapshot of a given situation in the service usage patterns of users is provided, and which patterns are relatively settled in the long or even short term (Haddon, 2005). The consistent usage of the abovementioned framework helps in this case as well, because the observations get more stability by analyzing holistic service categories instead of single fad-type services. On the other hand, the usage interrelationships of individual services should also be studied in more detail.

6.5 Future research

Related to mobile interpersonal communication service usage in general, we believe that in the future more emphasis should be placed on understanding how people use multiple services or communication channels together as a bundle, and which part(s) of the bundle result in the decision to adopt. After all, people use a variety of communication services to stay connected to their personal networks (Boase, 2008; Haddon, 2005), and thus the various services an individual is using probably do not overlap much in terms of their affordances. The usage interrelationships need to be analyzed in more detail, than what has been achieved so far. For instance, the usage interrelationships between SMS or MMS and IM, as well as between voice calls and VoIP are interesting topics for the future. Quantitative analysis related to the use of mobile social phonebooks is also needed in the future, as they become more common among the majority of mobile phone users.

In the top-down domain, we hope to see more comparative research conducted on big datasets of communication services. Future research in the area of social network analysis should strive for analyzing dynamic and multiple datasets, as discussed. Especially interesting would be to study how the social networks from different communication services conform to

the circles of acquaintanceship framework presented in the theoretical background chapter of this thesis. For instance, SMS networks are more skewed and concentrated than voice call networks (Publication III), but how do they compare to the other services? More emphasis is also needed on the validity of the research, especially in terms of what kind of a social tie is being measured. Continuing the development and especially empirical validation (or invalidation) of the social tie categorization model by Nelimarkka & Karikoski (2012) is also foreseen in the future. This can be done, for example, in a similar manner as Min et al. (2013) have empirically classified mobile phone contacts to different types of relationships based on mobile phone usage data.

On the methodological side, forthcoming data collection efforts should strive for bigger sample sizes preferably from different cultures. On the other hand, smaller samples such as the one used in this thesis are also foreseen in the future, as long as the sample is naturally networked so that social network analysis is possible. Gathering data on multiple mobile platforms is challenging, and many practical challenges exist that need to be tackled. However, cross-platform development tools, for instance, can be used to facilitate the data collection efforts. Related to the contextual part of this thesis, the context detection (Publication II) will be developed in the future by adding data from more sensors (e.g., acceleration sensor) to the input data of the algorithm. Moreover, instead of the current one-size-fits-all approach, the algorithm needs to be developed so that it is tailored to each individual user's spatiotemporal trajectories to enable more accurate context detection.

In general, the paradigms in different disciplines (e.g., between social and computer science) need to become more receptive in the future, so that collaboration across disciplines can be developed. As Willinger et al. (2009) have noted, a great danger exists in analyzing available data without a deeper understanding of the domain where the measurements were conducted, even if the analysis is mathematically rigorous. Thus, especially in the case of big and open data, those with the mathematical skills need to be accompanied with domain experts so that possible pitfalls can be avoided. Robust collaboration models and data sharing agreements between academia and corporations are also needed in the future. Thus, large-scale analysis of human behavior could become more common and not restricted to privileged researchers whose research and results cannot be critiqued or replicated (Lazer et al., 2009). In addition to the future research items presented above, which have mainly emerged based on the work conducted in the scope of this thesis, Eagle (2011) has outlined a new set of social research questions that can be tackled by using mobile phones

as sensors. Finally, in the brink of the open data era, we are hoping to see more datasets being openly published in the academia, and enabling groundbreaking work across research disciplines.

References

Aharony, N., Pan, W., Ip, C., Khayal, I. & Pentland, A., 2011. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7 (6), pp.643-659.

Banford, J., McDiarmid, A. & Irvine, J., 2010. Estimating the Strength of Ties in Communication Networks with a Small Number of Users. In: *The 6th International Conference on Wireless and Mobile Communications (ICWMC2010)*. Valencia, Spain 20-25 September 2010.

Bayus, B.L., Kim, N. & Shocker, A.D., 2000. Growth models for multi-product interactions: Current status and new directions. In: V. Mahajan, E. Muller & J. Wind, eds. *New Product Diffusion Models*. Kluwer, Boston, USA, pp.141-163.

Beach, A., Gartrell, M. & Han, R., 2010. q-Anon: Rethinking anonymity for social networks. In: *The Second IEEE International Conference on Social Computing (Socialcom2010)*. Minneapolis, Minnesota, USA 20-22 August 2010.

BEREC (Body of European Regulators for Electronic Communications), 2012. *A view of traffic management and other practices resulting in restrictions to the open Internet in Europe*. [Online] Available at: http://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/45-berec-findings-on-traffic-management-pra_0.pdf [Accessed 16 May 2013].

Blondel, V.D., Esch, M., Chan, C., Clerot, F., Deville, P., Huens, E., Morlot, F., Smoreda, Z. & Ziemlicki, C., 2012. Data for Development: the D4D Challenge on Mobile Phone Data. *Arxiv preprint physics*: arXiv:1210.0137v1.

Boase, J., Horrigan, J.B., Wellman, B. & Rainie, L., 2006. *The strength of Internet ties*. Washington, USA: Pew Internet and American Life Study.

Boase, J., 2008. Personal Networks and the Personal Communication System. *Information, Communication & Society*, 11 (4), pp.490-508.

Boase, J. & Kobayashi, T., 2012. Bonding and bridging with multimedia mobile phones: A study using the Communication Explorer smartphone application. In: *International Communication Association Annual Conference (ICA 2012)*. Phoenix, USA 24-28 May 2012.

Boase, J. & Ling, R., 2013. Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*, doi: 10.1111/jcc4.12021.

Boase, J., 2013. Implications of software-based mobile media for social research. *Mobile Media & Communication*, 1 (1), pp.57-62.

Borgatti, S.P., Everett, M.G. & Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Bouwman, H. & Van De Wijngaert, L., 2002. Content and context: an exploration of the basic characteristics of information needs. *New Media & Society*, 4 (3), pp.329-353.

Bouwman, H., de Reuver, M., Heerschap, N. & Verkasalo, H., 2013. Opportunities and problems with automated data collection via smartphones. *Mobile Media & Communication,* 1 (1), pp.63-68.

Bryman, A., 2006. Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6 (1), pp.97-113.

Calabrese, F., Smoreda, Z., Blondel, V.D. & Ratti, C., 2011. Interplay between telecommunications and face-to-face interactions: a study using mobile phone data. *PLoS ONE*, 6 (7), e20814.

Carlson, J.R. & Zmud, R.W., 1999. Channel Expansion Theory and the Experiential Nature of Media Richness Perceptions. *The Academy of Management Journal*, 42 (2), pp.153-170.

Coleman, J.S., 1988. Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94, pp.95-120.

Courcoubetis, C. & Weber, R., 2003. *Pricing communication networks - Economics, technology and modeling*. New York, USA: Wiley.

Creswell, J.W., 2010. Mapping the developing landscape of mixed methods research. In: A. Tashakkori & C. Teddlie, eds. *Sage handbook of mixed methods in social & behavioral research*. Sage Publications, Thousand Oaks, California, USA, pp.45-68.

de Bailliencourt, T., Beauvisage, T., Granjon, F. & Smoreda, Z., 2011. Extended Sociability and Relational Capital Management: Interweaving ICTs and Social Relations. In: R. Ling & S.W. Campbell, eds. *Mobile Communication: Bringing Us Together and Tearing Us Apart*. Transaction Publishers, New Brunswick, New Jersey, USA, pp.151-179.

de Groot, A.D., 1969. *Methodology – Foundations of inference and research in the behavioural sciences*. The Hague, Paris: Mouton & Co.

de Reuver, M., Visser, A. & Bouwman, H., 2010. The emerging ecosystem for future mobile communication services. In: *Proceedings of the 18th Biennial Conference of International Telecommunications Society (ITS 2010)*. Tokyo, Japan 27-30 June 2010.

Do, T.M.T., Blom, J. & Gatica-Perez, D., 2011. Smartphone usage in the wild: A large-scale analysis of applications and context. In: *Proceedings of the 13th International Conference on Multimodal Interaction (ICMI'11)*. Alicante, Spain 14–18 November 2011.

Dunbar, R.I.M., 1998. The social brain hypothesis. *Evolutionary Anthropology*, 6, pp.178-190.

Dunbar, R., 2010. Phoning home: Keeping in Touch in Social Networks. In: *ICTeCollective Symposium - Social Networks from science to technology*. [Online] Espoo, Finland 9 September 2010. Available at: http://becs.aalto.fi/ictecollective/docs/Dunbar-Oxford.pdf [Accessed 16 May 2013].

Eagle, N., 2009. Engineering a common good: Fair use of aggregated, anonymized behavioral data. In: *IEEE Engaging Data: First International Forum on the Application and Management of Personal Electronic Information (EDF)*. Cambridge, MA, USA 12-13 October 2009.

Eagle, N. & Pentland, A., 2006. Reality Mining: Sensing Complex Social Systems. *Personal and Ubiquitous Computing*, 10 (4), pp.255-268.

Eagle, N. & Pentland, A., 2009. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63 (7), pp.1057-1066.

Eagle, N., Pentland, A. & Lazer, D., 2009. Inferring Social Network Structure using Mobile Phone Data. *Proceedings of the National Academy of Sciences (PNAS)*, 106 (36), pp.15274-15278.

Eagle, N., 2011. Mobile Phones as Sensors for Social Research. In: S.N. Hesse-Biber, ed. *The Handbook of Emergent Technologies in Social Research*. Oxford University Press, New York, USA, pp.492-521.

Eurostat, 2004. *How Europeans spend their time: Everyday life of women and men*. [Online] Luxembourg: Office for Official Publications of the European Communities. Available at: http://www.unece.org/stats/gender/publications/Multi-Country/EUROSTAT/HowEuropeansSpendTheirTime.pdf [Accessed 16 May 2013].

Falaki, H., Lymberopoulos, D., Mahajan, R., Kandula, S. & Estrin, D., 2010a. A first look at traffic on smartphones. In: *Proceedings of the 2010 Internet Measurement Conference (IMC2010)*. Melbourne, Australia 1-3 November 2010.

Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R. & Estrin, D., 2010b. Diversity in Smartphone Usage. In: *Proceedings of the 8th International Conference on Mobile systems, applications, and services (MobiSys '10)*. San Francisco, USA 15-18 June 2010.

Falaki, H., Mahajan, R. & Estrin, D., 2011. SystemSens: A Tool for Monitoring Usage in Smartphone Research Deployments. In: *Proceedings of The 6th ACM International Workshop on Mobility in the Evolving Internet Architecture (ACM MobiArch 2011)*. Washington, USA 28 June 2011.

Feldman Barrett, L. & Barrett, D.J., 2001. An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review*, 19 (2), pp.175-185.

Fiske, J., 1982. *Introduction to Communication Studies*. London, UK: Methuen & Co.

Froehlich, J., Chen, M.Y., Consolvo, S., Harrison, B. & Landay, J.A., 2007. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In: *Proceedings of the 5th international Conference on Mobile Systems, Applications and Services (MobiSys2007)*. San Juan, Puerto Rico 11-14 June 2007.

Gerpott, T.J., 2010a. Impacts of mobile Internet use intensity on the demand for SMS and voice services of mobile network operators: An empirical multi-method study of German mobile Internet customers. *Telecommunications Policy*, 34 (8), pp.430-443.

Gerpott, T.J., 2010b. Communication behaviors and perceptions of mobile internet adopters. *Info*, 12 (4), pp.54-73.

Gerpott, T.J., 2011a. Tariff-type suitability's impact on satisfaction of Mobile Internet customers in Germany. *International Journal of Mobile Communications*, 9 (3), pp.243-260.

Gerpott, T.J., 2011b. Determinants of self-report and system-captured measures of mobile Internet use intensity. *Information Systems Frontiers*, 13 (4), 561–578.

Gerpott, T.J., Thomas, S. & Weichert, M., 2012. Usage of established and novel mobile communication services: Substitutional, independent or complementary? *Information Systems Frontiers,* Online First Article.

Ghose, A. & Han, S.P., 2011. An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Science*, 57 (9), pp.1671-1691.

Golder, S.A., Wilkinson, D. & Huberman, B.A., 2007. Rhythms of social interaction: messaging within a massive online network. In: *Proceedings of the 3rd International Conference on Communities and Technologies (C&T2007)*. East Lansing, Michigan, USA 28-30 June 2007.

Granovetter, M.S., 1973. The strength of weak ties. *American Journal of Sociology*, 78 (6), pp.1360-1380.

Grzybowski, L. & Pereira, P., 2008. The complementarity between calls and messages in mobile telephony. *Information Economics and Policy*, 20, pp.279-287.

Haddon, L., 2005. Research questions for the evolving communications landscape. In: R. Ling and P.E. Pedersen, eds. *Mobile Communications: Re-negotiation of the Social Sphere*. Springer, Surrey, UK, pp.7-22.

Haythornthwaite, C., 2005. Social networks and Internet connectivity effects. *Information, Communication & Society*, 8 (2), pp.125-147.

Hektner, J.M., Schmidt, J.A. & Csikszentmihalyi, M., 2007. *Experience sampling method: measuring the quality of everyday life*. California: SAGE Publications.

Helles, R., 2013. Mobile communication and intermediality. *Mobile Media & Communication*, 1 (1), pp.14-19.

Hevner, A.R., March, S.T., Park, J. & Ram, S., 2004. Design science in information systems research. *MIS Quarterly*, 28 (1), pp.75-105.

Hidalgo, C.A. & Rodriguez-Sickert, C., 2008. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387 (12), pp.3017-3024.

Hill, R.A. & Dunbar, R.I.M., 2003. Social network size in humans. *Human Nature*, 14 (1), pp.53-72.

Hossain, L., Chung, K. & Murshed, S., 2007. Exploring temporal communication through social networks. In: C. Baranauskas et al., eds. *Human-Computer Interaction – INTERACT 2007, Lecture Notes in Computer Science, 4662*. Springer, Berlin/Heidelberg, pp.19-30.

Huberman, B., Romero, D.M. & Wu, F., 2009. Social networks that matter: Twitter under microscope. *First Monday*, 14 (1-5).

Hutchby, I., 2001. Technologies, Texts and Affordances. *Sociology*, 35 (2), pp.441-456.

Jiménez, B., 2008. *Modeling of mobile end-user context*. M. Sc. Espoo, Finland: Helsinki University of Technology.

Jo, H.-H., Karsai, M., Karikoski, J. & Kaski, K., 2012. Spatiotemporal correlations of handset-based service usages. *EPJ Data Science 2012*, 1 (10).

Järvinen, P., 2004. *On Research Methods*. Tampere, Finland: Opinpajan kirja.

Karsai, M., Kaski, K. & Kertesz, J., 2012. Correlated Dynamics in Egocentric Communication Networks. *PLoS ONE*, 7 (7), e40612.

Katz, E., Blumler, J.G. & Gurevitch, M., 1974. Uses and Gratifications Research. *The Public Opinion Quarterly*, 37 (4), pp.509-523.

Kilkki, K., 2012. *An introduction to Communications Ecosystems*. Charleston, SC, USA: CreateSpace Independent Publishing Platform.

Kim, Y., Telang, R., Vogt, W. & Krishnan, R., 2010. An empirical analysis of mobile voice service and SMS: a structural model. *Management Science*, 56 (2), pp.234-252.

Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D. & Laurila, J., 2010. Towards rich mobile phone datasets: Lausanne data collection campaign. In: *Proceedings of the 7th ACM International Conference on Pervasive Services (ICPS2010)*. Berlin, Germany 13-16 July 2010.

Kosta, E. & Dumortier, J., 2008. Searching the man behind the tag: privacy implications of RFID technology. *International Journal of Intellectual Property Management*, 2(3), pp.276-288.

Kovanen, L., Saramäki, J. & Kaski, K., 2011. Reciprocity of mobile phone calls. *Dynamics of Socio-Economic Systems*, 2 (2), pp.138-151.

Kubey, R., Larson, R. & Csikszentmihalyi, M., 1996. Experience sampling method applications to communication research questions. *Journal of Communication*, 46 (2), pp.99-120.

Langheinrich, M., 2001. Privacy by design – principles of privacy-aware ubiquitous systems. *Lecture Notes in Computer Science*, 2201, pp.273-291.

Lane, N.D., Xie, J., Moscibroda, T. & Zhao, F., 2012. On the feasibility of user de-anonymization from shared mobile sensor data. In: *The Third International Workshop on Sensing Applications on Mobile Phones (PhoneSense 2012)*. Toronto, Canada 6 November 2012.

Laurila, J.K., Gatica-Perez, D., Aad, I., Trinh-Minh-Tri, D., Blom, J., Bornet, O., Dousse, O., Eberle, J. & Miettinen, M., 2012. The mobile data challenge: Big data for mobile computing research. In: *Proceedings of the Tenth International Conference on Pervasive Computing (Pervasive2012)*. Newcastle, UK 18-22 June 2012.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. & Van Alstyne, M., 2009. Computational Social Science. *Science*, 323 (5915), pp.721-723.

Leung, L. & Wei, R., 2000. More than just talk on the move: Uses and gratifications of the cellular phone. *Journalism & Mass Communication Quarterly*, 77 (2), pp.308-320.

Licoppe, C. & Smoreda, Z., 2005. Are social networks technologically embedded? How networks are changing today with changes in communication technology. *Social Networks*, 27 (4), pp.317-335.

Licoppe, C., Diminescu, D., Smoreda, Z. & Ziemlicki, C., 2008. Using mobile phone geolocalisation for 'socio-geographical' analysis of co-ordination, urban mobilities, and social integration patterns. *Tijdschrift voor economische en sociale geografie*, 99 (5), pp.584-601.

LimeSurvey, 2012. *LimeSurvey – the free & open source survey software tool*. [Online] Available at: http://www.limesurvey.org/ [Accessed 16 May 2013].

Lin, C.A., 2011. Media Substitution – Supplementation, Complementarity, or Displacement? In: A. Viswanath & G.A. Barnett, eds. *The Diffusion of Innovations – A Communication Science Perspective*. Peter Lang, New York, USA, pp.171-191.

Livingstone, S., 1999. Young People and the New Media: On learning lessons from TV to apply to the PC. *Réseaux*, 7 (1), pp.59-81.

Lohse, G.L., Bellman, S. & Johnson, E.J., 2000. Consumer Buying Behavior on the Internet: Findings from Panel Data. *Journal of Interactive Marketing*, 14 (1), pp.15-29.

Maier, J., Preece, W.H., Nyström, C.A. & Lyytinen, M., 1900. *Sähkö II - Telefooni ja sen käyttäminen*. Tampereen Sanomain Osakeyhtiön kirjapaino.

March, S.T. & Smith, G.F., 1995. Design and natural science research on information technology. *Decision Support Systems*, 15 (4), pp.251-266.

Malhotra, H.K., Birks, D.F. & Wills, P., 2012. *Marketing Research: An Applied Approach*. 4th ed. Essex, UK: Pearson Education Limited.

McPherson, M., Smith-Lovin, L. & Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, pp.415-444.

McQuail, D., 1994. Mass Communication Theory. 3rd ed. London, UK: Sage Publications.

Merriam-Webster Dictionary, 2012. *Dictionary and Thesaurus - Merriam-Webster Online*. Available at: http://www.merriam-webster.com/dictionary/communication [Accessed 16 May 2013].

Miller, G., 2012. The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*, 7 (3), pp.221-237.

Min, J.-K., Wiese, J., Hong, J.I. & Zimmerman, J., 2013. Mining Smartphone Data to Classify Life-Facets of Social Relationships. In: *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing*. San Antonio, Texas, USA 23-27 February 2013.

Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P. & Rosenquist, J.N., 2011. Understanding the demographics of twitter users. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11)*. Barcelona, Spain 17-21 July 2011.

Moiso, C. & Minerva, R., 2012. Towards a User-Centric Personal Data Ecosystem - The Role of the Bank of Individuals' Data. In: *Proceedings of 16th International Conference on Intelligence in Next Generation Networks*. Berlin, Germany 8-11 October 2012.

Morse, J., 2010. Procedures and Practice of Mixed Method Design: Maintaining Control, Rigor, and Complexity. In: A. Tashakkori & C. Teddlie, eds. *Sage handbook of mixed methods in social & behavioral research*. Sage Publications, Thousand Oaks, California, USA, pp.339-352.

Nelimarkka, M. & Karikoski, J., 2012. Categorizing and measuring social ties. In: *The RC33 Eighth International Conference on Social Science Methodology*. Sydney, Australia 9-13 July 2012.

Nikou, S., Bouwman, H. & de Reuver, M., 2012. The potential of converged mobile telecommunication services: a conjoint analysis. *Info*, 14(5), pp.21-35.

Odlyzko, A., 2009. Network Neutrality, Search Neutrality, and the Never-ending Conflict between Efficiency and Fairness in Markets. *Review of Network Economics*, 8 (1), pp.40-60.

Onnela, J.P., Saramäki, J., Hyvönen, J., Szabo, G., Lazer, D., Kaski, K., Kertesz, J. & Barabasi, A.L., 2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences (PNAS)*, 104 (18), pp.7332-7336.

OSF (Official Statistics of Finland), 2011. *Time Use Survey 2009*. [Online] Helsinki: Statistics Finland. Available at: http://www.stat.fi/til/akay/2009/02/akay_2009_02_2011-02-17_tie_001_en.html [Accessed 16 May 2013].

Palchykov, V., Kertész, J., Dunbar, R.I.M. & Kaski, K., 2013. Close relationships: A study of mobile communication records. *Journal of Statistical Physics*, 151 (3-4), pp.735-744.

Pentland, A., 2009. Reality mining of mobile communications: Toward a new deal on data. In: S. Dutta & I. Mia, eds. *The Global Information Technology Report 2008-2009*. World Economic Forum, Geneva, Switzerland, pp.75-80.

Peters, O. & ben Allouch, S., 2005. Always connected: a longitudinal field study of mobile communication. *Telematics and Informatics*, 22 (3), pp.239-256.

Putnam, R.D., 2000. *Bowling Alone*. New York, USA: Simon & Schuster.

Raento, M., Oulasvirta, A., Petit, R. & Toivonen, H., 2005. ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. *IEEE Pervasive Computing*, 4 (2), pp.51-59.

Raento, M., Oulasvirta, A. & Eagle, N., 2009. Smartphones: An Emerging Tool for Social Scientists. *Sociological Methods & Research*, 37 (3), pp.426-454.

Rettie, R., 2009. Mobile Phone Communication: Extending Goffman to Mediated Interaction. *Sociology*, 43 (3), pp.421-438.

Ritchie, J. & Lewis, J., 2003. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London, UK: Sage Publications.

Robson, C., 2011. *Real world research: a resource for users of social research methods in applied settings*. 3rd ed. West Sussex, UK: John Wiley & Sons Ltd.

Rogers, E.M., 2003. *Diffusion of Innovations*. 5th ed. New York, USA: Free Press.

Saramäki, J., Leicht, E.A., Lopez, E., Roberts, S.G.B, Reed-Tsochas, F. & Dunbar, R.I.M., 2012. The persistence of social signatures in human communication. *Arxiv preprint physics*: arXiv:1204.5602.

Schweitzer, F. & Vespignani, A., 2012. Editorial. *EPJ Data Science 2012*, 1 (1).

Smura, T., 2008. Access alternatives to mobile services and content: analysis of handset-based smartphone usage data. In: *The 17th Biennial International Telecommunications Society Conference (ITS Montréal 2008)*. Montréal, Canada 24-27 June 2008.

Smura, T., Kivi, A. & Töyli, J., 2009. A framework for analysing the usage of mobile services. *Info*, 11 (4), pp.53-67.

Soikkeli , T., 2011. *The effect of context on smartphone usage sessions*. M. Sc. Espoo, Finland: Aalto University.

Soikkeli, T., Karikoski, J. & Hämmäinen, H., 2011. Diversity and end user context in smartphone usage sessions. In: *Proceedings of the 5th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST2011)*. Cardiff, Wales, UK 14-16 September 2011.

Soikkeli, T., Karikoski, J. & Hämmäinen, H., 2012. An End User Context Framework for Handset-based Studies. In: *International Telecommunications Society 19th Biennial Conference*. Bangkok, Thailand 18-21 November 2012.

Szabo, G. & Barabasi, A.L., 2006. Network effects in service usage. *Arxiv preprint physics*: arXiv:physics/0611177v1.

Tan, S-E. & Henten, A., 2006. New Mobile Technologies: Sustaining or Disruptive? In: *Proceedings of the International Conference on Mobile Business (ICMB'06)*. Copenhagen, Denmark 26-27 June 2006.

Tossell, C.C., Kortum, P., Shepard, C.W., Rahmati, A. & Zhong, L., 2012. Getting Real: A Naturalistic Methodology for Using Smartphones to Collect Mediated Communications. *Advances in Human-Computer Interaction*, 2012 (815972), pp.1-10.

Van Cleemput, K., 2012. *The Everyday Use of Communication Technologies in Adolescent Peer Groups. A Social Network Approach*. Ph. D. The Netherlands: University of Antwerp.

Verkasalo, H., 2007. Handset-based measurement of smartphone service evolution in Finland. *Journal of Targeting, Measurement and Analysis for Marketing*, 16 (1), pp.7-25.

Verkasalo, H., 2009a. Contextual patterns in mobile service usage. *Personal and Ubiquitous Computing*, 13 (5), pp.331-342.

Verkasalo, H., 2009b. *Handset-based analysis of mobile service usage*. Ph. D. Espoo, Finland: Helsinki University of Technology.

Verkasalo, H., 2010. Analysis of Smartphone User Behavior. In: *Proceedings of the 9th International Conference on Mobile Business and 2010 Ninth Global Mobility Roundtable (ICMB-GMR 2010)*. Athens, Greece 13-15 June 2010.

Verkasalo, H., 2012. On-Device Metering in Measuring the Biggest Mass Medium – Mobile. In: *Audience Measurement 7.0*. New York, USA 11-12 June 2012.

Wasserman, S. & Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Whittaker, S., 2003. Theories and Methods in Mediated Communication. In: A.C. Graesser et al., eds. *Handbook of Discourse Processes*. Routledge, USA.

Wilken, R., 2011. Bonds and Bridges: Mobile Phone Use and Social Capital Debates. In: R. Ling & S.W. Campbell, eds. *Mobile Communication: Bringing Us Together and Tearing Us Apart*. Transaction Publishers, New Brunswick, New Jersey, USA, pp.127-149.

Willinger, W., Alderson, D. & Doyle, J.C., 2009. Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. *Notices of the American Mathematical Society*, 56 (5), pp.586-599.

World Economic Forum, 2011. *Personal Data: The Emergence of a New Asset Class*. [Online] Available at: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf [Accessed 16 May 2013].

Wunsch, G., Russo, F. & Mouchart, M., 2010. Do We Necessarily Need Longitudinal Data to Infer Causal Relations? *Bulletin de Methodologie Sociologique*, 106, pp.5-18.

The overall increase in mediated interpersonal communication and the merger of the mobile and Internet domains are accelerating the changes in how we socialize. At the same time the data collected by and from mobile phones offer unprecedented possibilities for empirically studying how we communicate and keep in touch with each other. Thus, this thesis aims to empirically analyze the usage of mobile interpersonal communication services by utilizing a quantitatively driven mixed method design, with handset-based measurements as the core data collection method.

Aalto-DD 119/2013

ISBN 978-952-60-5266-3
ISBN 978-952-60-5267-0 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Communications and Networking
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS