Thesis

Analysing smartphone users “inner-self”: the perception of intimacy and smartphone usage changes

GUSTARINI, Mattia

Abstract

Smart mobile services and applications use users’ context. However, we never investigate how users perceive this context and how to leverage this perception for even smarter services. We represent the perception of the context of users as their intimacy, their familiarity with their current place, the number, and kind of people around them. The adjective ‘intimate’ describes the context as familiar, being private and comfortable. First, we validate the intimacy concept. We establish that users use mobile services differently in different intimacy situations, and we create a first theoretical model to estimate their intimacy. Second, we investigate the intimacy predictability in practice (limitations and solutions). Finally, we show how we can leverage intimacy for studies on users’ context or deploy our intimacy model to help apps developer. Advertisers can use it to deliver their content, and it can support the innovative projects, as Google Project Tango, to get smarter.

Reference

GUSTARINI, Mattia. *Analysing smartphone users “inner-self”: the perception of intimacy and smartphone usage changes*. Thèse de doctorat : Univ. Genève, 2016, no. GSEM 26

URN : urn:nbn:ch:unige-850316
DOI : 10.13097/archive-ouverte/unige:85031

Analysing smartphone users "inner-self": the perception of intimacy and smartphone usage changes

THÈSE

présentée à la Faculté d'Economie et de Management
de l'Université de Genève par

Mattia Gustarini

sous la codirection de

Assoc. Prof. Katarzyna Wac

Prof. Dimitri Konstantas

pour l'obtention du grade de

Docteur ès Économie et Management
mention Systèmes d'Information

Membres du jury de thèse:

Prof. Gilles Falquet (Président du jury)
Assoc. Prof. Katarzyna Wac (Co-directrice de thèse)
Prof. Dimitri Konstantas (Co-directeur de thèse)
Prof. François Grey, University of Geneva
Prof. Anind K. Dey, Carnegie Mellon University
Prof. Effie Lai-Chong Law, University of Leicester

Thèse N° 26
ISBN: 978-2-88903-044-6
Genève, le 26 avril 2016

La Faculté d'Economie et de Management, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 26 avril 2016

La Doyenne
Maria-Pia VICTORIA FESER

Impression d'après le manuscrit de l'auteur

Table of Contents

Résumé vii
Abstract ix
Sommario xi
Acknowledgements xiii
Publications Related to this Ph.D. xv
List of Figures xvii
List of Tables xxi
Acronyms xxii
1. Introduction 1
2. Why Intimacy? 5
2.1 Context and Smartphone Usage 5
2.2 Defining Intimacy 6
2.3 Intimacy Implications 7
2.4 Leveraging Intimacy 9
3. Feasibility Study: Can we Assess Intimacy “in the wild”? 11
3.1 From MDC Raw Data to Intimacy 11
3.1.1 Raw Data Exploration 11
3.1.2 Features Selection 12
3.1.3 Intimacy Estimation Algorithm 13
3.2 Results 15
3.2.1 Observers and Safe Places 15
3.2.2 Intimacy Levels 16
3.2.3 Demographic Analysis 18
3.3 Discussion 19
3.3.1 Study Assumptions & Limitations 19
3.3.2 Ground Truth and Experience Sampling 20
3.3.3 Where to Apply Intimacy? 21
3.4 Conclusion 21
4. The mHUMAC Method and User Studies 23
4.1 The mHUMAC Method 23
4.2 The User Studies 24
4.2.1 User Study One: Validation and Modelling of Intimacy 24
4.2.2 User Study Two: Evaluation of the Preliminary Intimacy Prediction Model 28
4.2.3 User Study One and Two Summary Table 28
4.2.4 Study Limitations 29
5. Validating the Intimacy Definition 31
5.1 General Statistics of ESM 31
5.2 ESM Data Preparation 32
5.3 Intimacy vs. Place, Kind and Number of People Around the User 32
5.3.1 Semantic Place 33
5.3.2 Number of People Around the User 34

5.3.3 Kind of People Around the User 34
5.4 Intimacy and Mood Components: Valence and Arousal 35
5.5 Conclusion 36

6. Intimacy and Differences in Smartphone Usage 39
6.1 Experiment 1: Data Preparation 39
6.1.1 Usage Session and Usage Window 39
6.1.2 High-Level Smartphone Usage Features 40
6.1.3 Aggregating ESM Beeps to High-Level Smartphone Usage Features 41
6.2 Experiment 1: Results 41
6.3 Experiment 1: Discussion 44
6.4 Experiment 2: Data Preparation 45
6.4.1 Valid Transitions 45
6.4.2 Applications Switches and Screen Touches 46
6.4.3 Intimacy ESM and Intimacy Models 46
6.4.4 Data Cleaning 47
6.5 Experiment 2: Results 49
6.5.1 Intimacy in Time 50
6.5.2 PRESENT-OFF Transition Intimacy Models 50
6.5.3 Applications Used Intimacy Models 52
6.5.4 Communication Category Intimacy Models 53
6.6 Experiment 2: Discussion 54
6.7 Conclusions 56

7. Predictive Modeling of Intimacy 59
7.1 Data Preparation 59
7.1.1 Creating Neighborhoods of Cell IDs 59
7.1.2 The Neighborhoods Cell IDs Data 60
7.2 Features Extraction 61
7.3 Transforming Intimacy ESM Data: From Classes to Ranks 62
7.4 Generating Data Sets for RPC Classification 63
7.5 Testing and Validating the Intimacy Model 64
7.5.1 Experiment 1: Per Subject Modeling 64
7.5.2 Experiment 2: Per Subject Modeling vs. Other Subject Data 66
7.5.3 Experiment 3: User Study 'in the Wild' 69
7.6 Discussion 75
7.7 Conclusions 76

8. The Role of Intimacy in Anonymous Smartphone Data Collection 77
8.1 People-Centric Sensing and Mobile Crowd Sensing 77
8.2 Investigating Users' Perception of Anonymity 81
8.3 Public Sharing of Anonymous vs. Identifiable Data 81
8.4 Data Anonymisation 82
8.5 Analysis of Factors Influencing Anonymity 83
8.5.1 User's Deletion Choices: Public on Facebook vs. Anonymously on a Public Server 83
8.5.2 People Factor: Users Privacy Clusters 84
8.5.3 Data Kind Factor: User's Deletion Choices and Content Type 85
8.5.4 Context Factor: User's Deletion Choices and Intimacy 87
8.5.5 Summary of Quantitative Results 88

8.6 DRM: Participants Interviews 89
8.6.1 Identifiable Public Sharing on Facebook 89
8.6.2 Anonymous Public Sharing on a Server 91
8.7 Discussion 92
8.7.1 Anonymous vs. Identifiable Sharing 92
8.7.2 Factors Influencing Anonymous vs. Identifiable Sharing 93
8.7.3 Understanding Users' Perception of Anonymity 94
8.8 Conclusions 94
9. Future Work and Conclusions 97
9.1 Leveraging Intimacy Further in the UnCrowd TPG Mobile Application 97
9.1.1 Users Involved 97
9.1.2 Study Methods 98
9.1.3 Variables Involved 99
9.2 Conclusions and Open Research Questions About Intimacy 100
Bibliography 103

Résumé

Les services et applications mobiles utilisent le contexte de leurs utilisateurs pour leur fournir des nouvelles fonctionnalités plus intelligentes. La plupart de ces services utilisent la position de l'utilisateur et d'autres contextes comme, par exemple, l'activité de l'utilisateur. A l'avenir, il sera possible d'acquérir de nouvelles informations sur le contexte de l'utilisateur. Par conséquent, les services mobiles l'exploiteront encore plus. Il est impératif que nous comprenons comment les différentes parties du contexte vont ensemble et comment l'utilisateur les perçoit. Plus important encore, nous avons besoin de savoir si la perception du contexte par les utilisateurs modifie la façon dont ils utilisent les services mobiles. Comment pouvons-nous tirer parti de cette perception pour fournir des services encore plus intelligents ?

Pour représenter la perception du contexte chez les utilisateurs, nous définissons l'intimité des utilisateurs comme leur familiarité avec leur lieu actuel, le nombre de personnes et la nature des gens autour d'eux. Pour résumer avec un exemple, nous faisons l'hypothèse suivante : « quand les gens sont dans leur maison dînant et discutant avec leurs familles, ils se sentent intime, alors que quand ils sont dans un bus avec des inconnus, ils ne se sentent pas intime. » La définition prend en compte aussi des aspects spécifiques qui sont notoirement associées au concept classique de l'intimité que nous ne considérons pas comme une partie de celui-ci, comme la vie privée et la qualité des relations entre les gens.

Une première étude avec des utilisateurs, nous permet de valider le concept d'intimité, et nous confirmons que les variables de contexte : lieu, nombre et genre des personnes proche des utilisateurs reflète leur perception d'intimité (la familiarité de leur contexte). Avec la même étude, nous évaluons la corrélation entre l'intimité et l'utilisation du smartphone. Des interactions plus courtes, plus fréquentes, et moins privilégiés ont lieu lorsque l'intimité est plus faible. Lorsque l'intimité est plus élevée nous enregistrons des interactions plus long, moins fréquentes, et plus privilégié. Enfin, avec les données recueillies lors de cette étude, nous avons également définit un modèle théorique d'intimité pour prédire l'intimité des utilisateurs de smartphone.

Nous étudions la prévisibilité de l'intimité en pratique avec une seconde étude. Nous avons créé un logiciel qui peut prédire l'intimité et nous l'avons utilisé dans un environnement avec des utilisateurs. Nous avons découvert que la localisation de l'utilisateur et le temps passé dans un endroit sont prédictifs de l'intimité. D'autres variables extrait avec les smartphones améliore la précision de la prédiction. Nous analysons également les problèmes de notre modèle d'intimité et fournissons des solutions pour améliorer ses capacités de prévisions.

Enfin, tout au long de notre travail, nous examinons comment nous pouvons tirer parti des conclusions sur l'intimité et quelles sont ses implications pratiques dans le monde des applications et des services mobiles. Nous pouvons utiliser notre modèle pour des études sur le contexte ou le déployer pour aider les développeur d'applications mobiles (pour automatiser les services ou créer une meilleure interface). Les annonceurs peuvent l'utiliser pour livrer leur contenu au

bon moment, et il peut aussi soutenir les projets innovants, comme le Google Project Tango.

Abstract

Mobile services and applications use their users' context to provide them new and more intelligent features. Most of these services use the users' location and another context like, for example, the user activity. In future, it will be possible to sense even more users' context. Therefore, mobile services will leverage it even more. It is imperative that we understand how different pieces of context go together and how the end-user perceives them. More importantly, we need to know if the users' perception of their context changes the way they use mobile services. How can we leverage this perception to provide to users even more intelligent services?

To represent the perception of the context of users, we define users' intimacy as their familiarity with their current place, the number of people and kind of people around them. To summarize with an example, we make the following assumption: "when people are at home having a dinner with their families discussing family matters, they feel intimate, while when they are on a bus with strangers, they do not feel intimate." The definition takes into accounts also some specific aspects that are notoriously associate with the classical concept of intimacy that we do not consider to be a part of it, such as privacy and the quality of relations between people.

With a first user study, we validate the intimacy concept, and we confirm that the context variables: place, the number and kind of people around the users represent the users perception of intimacy (familiarity of their context). With the same user study, we evaluate the intimacy correlation to smartphone usage features. We obtain that shorter, more frequent, and less engaging interactions take place when intimacy is lower, while longer, less frequent, and engaging interactions when intimacy is higher. Finally, with the data collected in user study one, we also theoretically define an intimacy model to predict the users' intimacy using the users' smartphones.

With a second user study, we investigate the intimacy predictability in practice. We create a software package that can predict intimacy and we study it in the real users' environment. We discover that location-time features are predictive for the intimacy, and other smartphone-based features can improve the intimacy prediction accuracy. We also analyze the problems of our intimacy model and provide solutions to improve further the intimacy predictions capabilities in future iterations.

Finally, all along our work, we discuss how we can leverage the findings about intimacy and which are its practical implications in the world of mobile applications and services. We can use it for studies on users' context or deploy our intimacy model to help apps developer (to automate services or create better UIX). Advertisers can use it to deliver their content at the right moment, and it can support the innovative projects, as Google Project Tango, to get even better.

Sommario

I servizi e le applicazioni per smartphone utilizzano il contesto dei propri utilizzatori per fornire nuove funzioni sempre più intelligenti. La maggior parte di questi servizi sfruttano la posizione dell'utilizzatore o altre informazioni come la sua attività corrente. In futuro, le capacità di estrarre informazioni dal contesto dell'utilizzatore aumenteranno e i servizi mobili ne faranno ancor più largo uso. Quindi è estremamente importante che comprendiamo come l'utilizzatore percepisce le informazioni estratte dal suo contesto e come queste informazioni sono connesse fra loro. Altrettanto importante è definire se la percezione di queste informazioni da parte degli utilizzatori di smartphone influenzi come essi lo utilizzano. Possiamo sfruttare questa percezione per fornire servizi sempre più intelligenti?

Per rappresentare la percezione del contesto dell'utilizzatore, definiamo l'intimità degli utilizzatori come la loro familiarità con il luogo in cui si trovano, e il numero e il tipo di persone che si trovano attorno a loro mentre usano i loro smartphones. Per esempio, assumiamo che "quando le persone sono a casa loro seduti a tavola mangiando con la propria famiglia, esse si sentano intime, mentre quando si trovano in un bus pieno di sconosciuti, esse non si sentano intime. La nostra definizione di intimità prende in considerazione anche aspetti che sono notoriamente associati con la nozione d'intimità piu' comune, come le relazioni tra persone e la preservazione della propria vita privata e le esclude da essa coscienziosamente.

Con un primo studio degli utilizzatori, validiamo il nostro concetto d'intimità, e confermiamo che: luogo, numero e tipo di persone del contesto dell'utilizzatore sono variabili della sua intimità (sono parte della familiarità del suo contesto). Con lo stesso studio, definiamo il rapporto tra l'intimità dell'utilizzatore e differenti variabili di utilizzazione degli smartphones. Abbiamo verificato che l'utilizzatore esegue interazioni corte, frequenti, e meno impegnate quando si trova in uno stato di bassa intimità. Mentre interazioni lunghe, meno frequenti, e impegnate possono indicare alta intimità. Per concludere, con i dati acquisiti in questo primo studio, abbiamo anche definito un modello computazionale teorico capace di stimare l'intimità degli utilizzatori di smarthpones.

Con un secondo studio, abbiamo investigato, in pratica, la possibilità di predire l'intimità degli utilizzatori di smartphones. Abbiamo creato un'applicazione per smartphone capace di stimare l'intimità nel contesto reale degli utilizzatori basata sul nostro modello teorico. Abbiamo appurato che le variabili tempo e posizione dell'utilizzatore combinate sono componenti importanti per essere in grado di stimare l'intimità degli utilizzatori e scoperto variabili addizionali, tutte derivabili dagli smartphones, in grado di accrescere la precisione della stima. Questo esperimento ci ha anche permesso di identificare i limiti del nostro modello teorico e identificare possibili soluzioni per renderlo più performante in future iterazioni del suo sviluppo.

Per concludere, proponiamo come possiamo sfruttare i risultati e le scoperte di questa ricerca e analisi (uso degli smartphones in differenti stati di intimità) per soluzioni pratiche nel mondo dei servizi mobili. Per esempio, possiamo applicare le nostre scoperte in altre ricerche sul contesto degli utilizzatori di smartphones,

o rendere il nostro modello di stima dell'intimità disponibile come servizio per gli sviluppatori di applicazioni mobili (per esempio, per automatizzare altri servizi o creare una migliore esperienza per l'utilizzatore). Il nostro modello può essere utilizzato dai pubblicitari per indirizzare le pubblicità e promozioni all'utilizzatore di servizi mobili gratuiti al momento opportuno (per esempio, quando l'utilizzatore è più ricettivo), o supportare progetti innovativi, come Google Project Tango, concentrati sullo sviluppo di tecnologie per captare precisamente il contesto di utilizzazione degli smartphones.

Acknowledgements

First of all, I would like to thank my family, particularly my parents for their guidance and support efforts during my whole life. Without them, I would not be able to reach my goals and therefore, there would not be this thesis.

Then, I would like to thank my supervisor, Assoc. Prof. Katarzyna Wac, for her help and guidance, particularly for putting at my disposal her vast network of contacts, any resource I needed, and create all the necessary conditions to complete this work successfully. I would also like to thank my co-supervisor Prof. Dimitri Konstantas always at disposal to solve any issue and for the many pieces of advice given, sometimes going beyond the Ph.D. itself.

Moreover, I would like to thank everyone part of the Information Science Institute (ISI), other Ph.D. students, scientific collaborators, professors, administration, etc. for the help whenever it was needed or for spontaneous collaborations and initiatives. Among these people, there are some that became my friends, which I thank very much for making my thesis path much more entertainment, enjoyable, and part of bigger scope than just scientific research.

Furthermore, I would like to thank the co-founders of our company MobileThinking SARL: Carlos Ballester Lafuente, Jody Hausmann, Jérôme Marchanoff, and Kevin Salvi, which are helping me to smoothly transition from to my future after this thesis.

In addition, I would like to thank all my jury members: Prof. Anind K. Dey (Carnegie Mellon University), Prof. Gilles Falquet (jury president, University of Geneva), Prof. François Grey (University of Geneva), and Prof. Effie Lai-Chong Law (University of Leicester) for their time and valuable feedback while they conscientiously evaluated my work.

A particular note, I am also grateful to every subject that participated in the studies carried in this work, particularly all the people from the Pittsburgh (USA) area and the participants to our mQoL living lab from the Geneva (Switzerland) area.

Finally, I would like to thank all the funding institutions that financially supported this work: Swiss National Science Foundation Project "Enabling People-Centric Sensing by Overcoming the privacy BarriEr" (PCS-OBEY) (SNSF-149591) (2013-2015), European Ambient Assisted Living project "A Pervasive Guardian for Elderly with Mild Cognitive Impairments" (AAL-MyGuardian) (AAL-2011-4-027) (2012-2015), European Ambient Assisted Living project "Way Finding Seniors" (AAL-WayFiS) (AAL-2010-3-014) (2011-2013).

Publications Related to this Ph.D.

As First Author:

1. Mattia Gustarini, Katarzyna Wac, **Estimating People Perception of Intimacy in Daily Life From Context Data Collected With Their Mobile Phones**, Mobile Data Challenge by Nokia Workshop, co-located with the PERVASIVE Conference, Newcastle, UK, June 2012.
2. Mattia Gustarini, Katarzyna Wac, **Ubiquitous Inference of Mobility State of Human Custodian in People-Centric Context Sensing**, International Workshop on Ubiquitous Mobile Instrumentation (UbiMi), co-located with the UBICOMP Conference, Pittsburgh, PA, USA, September 2012.
3. Mattia Gustarini, Selim Ickin, Katarzyna Wac, **Evaluation of Challenges in Human Subject Studies "In-the-Wild" Using Subjects' Personal Smartphones**, International Workshop on Ubiquitous Mobile Instrumentation (UbiMI), co-located with the UBICOMP conference, Zurich, Switzerland, September 2013.
4. Mattia Gustarini, Katarzyna Wac, **Smartphone Interactions Change for Different Intimacy Contexts**, Fifth International Conference on Mobile Computing, Applications and Services (MobiCASE), G. Memmi and U. Blanke (Eds.), Springer LNICST 130, pp. 72-89, Paris, France, November 2013.
5. Mattia Gustarini, Jerome Marchanoff, Marios Fanourakis, Christiana Tsiourti, Katarzyna Wac, **UnCrowdTPG: Assuring the Experience of Public Transportation Users**, 2nd IEEE International Workshop on Smart City and Ubiquitous Computing Applications (SCUCA'14), co-located with the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (IEEE WiMob 2014), Larnaca, Cyprus, October 2014.
6. Mattia Gustarini, Katarzyna Wac, Anind K. Dey, **Anonymous Smartphone Data Collection: Factors Influencing the Users' Acceptance in Mobile Crowd Sensing**, Personal and Ubiquitous Computing, Springer. Impact Factor: 1.518.
7. Mattia Gustarini, Marcello Paolo Scipioni, Marios Fanourakis, Katarzyna Wac, **Differences in Smartphone Usage: Validating, Evaluating, and Predicting Mobile User Intimacy**, Pervasive and Mobile Computing, Elsevier. Impact Factor: 2.325.

As Co-Author:

1. Katarzyna Wac, Gerardo Pinar, Mattia Gustarini, Jerome Marchanoff, **Smartphone Users' Mobile Network's Quality Provision and VoLTE Intend: Six-months Field Study**, IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (IEEE WoWMoM 2015), Boston, USA, June 2015.
2. Katarzyna Wac, Maddalena Fiordelli, Mattia Gustarini, Homero Rivas, **Quality of Life Technologies: Experiences from the Field and**

Key Research Challenges, IEEE Internet Computing, Special Issue: Personalized Digital Health, July/August 2015.

3. Katarzyna Wac, Mattia Gustarini, Jerome Marchanoff, Marios Fanourakis, Christiana Tsiourti, Matteo Ciman, Jody Hausmann, Gerardo Pinar, **mQoL: Experiences of the 'Mobile Communications and Computing for Quality of Life' Living Lab**, International Workshop on the Living Lab Approach for Successful Design of Services and Systems in eHealth (LivingLab'15), co-located with the IEEE HealthCom 2015, Boston, USA, October 2015.
4. Katarzyna Wac, Gerardo Pinar, Mattia Gustarini, Jerome Marchanoff, **More Mobile & Not so Well-connected yet: Users' Mobility Inference Model and 6 Month Field Study**, 7th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2015), Brno, Czech Republic, October 2015.

List of Figures

Figure 1: Schematic representation of the five top objectives of this work. Semi-circles are the actions needed to fulfil the related goals and rectangles are representing their outcomes. The dashed part refers to on-going and future work. 4
Figure 2: Smartphone's ring status of participant P9. 11
Figure 3: The three main greedy intimacy algorithm steps. In step 1 features get assigned a score from 1 to 6 depending on their values, in step 2 we average these values into one score for 'observers' and 'safe places', and finally, in step 3 their average defines the current 'intimacy state' of the user. 15
Figure 4: Observers most frequent intimacy levels for participant P26 per time interval (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy. 16
Figure 5: Safe places most frequent intimacy levels for participant P26 per time interval (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy. 16
Figure 6: Most frequent intimacy levels (average of intimacy levels of 'observers' and 'safe places') for participant P26 per interval of time (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy... 17
Figure 7: Probability of participant P26 to be in a particular intimacy level (considering missing data as well) for the seven days of the week. Legend: 1 (purple or dark, on the left of the graph) means low intimacy - 6 (green or grey, on the right of the chart) means high intimacy, and NA (blue or black, at the very end right of the chart) means not available data. 18
Figure 8: Probability for all the participants to be in a given intimacy level (considering missing data as well) for the whole dataset. Legend: 1 (purple or dark, on the left of the graph) means low intimacy - 6 (green or grey, on the right of the graph) means high intimacy, and NA (blue or black, at the very end right of the chart) means not available data. 18
Figure 9: Data acquisition methods (subjective and objective) and their ideal granularity regarding frequencies (from minute acquisition to month/year intervals). 24
Figure 10: The ESM question about the perception of intimacy of the user at the moment of the answer. 27
Figure 11: The ESM question about the place in which the user was at the time of the answer. 27
Figure 12: The ESM question about the number of people around the user at the moment of the answer. 27
Figure 13: The ESM question about the kind of people around the user at the moment of the answer. 27
Figure 14: The ESM question about the valence of the user at the time of the answer. 27
Figure 15: The ESM question about the arousal of the user at the moment of the answer. 27
Figure 16: The ESM question about the deletion of a content type (here activity) posted publicly on Facebook. 27

Figure 17: The ESM question about the deletion of a content type (here photo) posted publicly but anonymously on a Server. 27
Figure 18: Number of users answers to ESM questions. Some users are contributing more than others creating a different number of contributions. 31
Figure 19: The different intimacy scores mean per each study participant (1 „completely" intimate - 6 „not at all" intimate) [error bars 95% CI]. 31
Figure 20: Mean intimacy z-score for places. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of places. 33
Figure 21: Mean intimacy z-score for number of people. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of people around. 34
Figure 22: Mean intimacy z-score for kind of people. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of kind of people. 35
Figure 23: Mean intimacy z-score for valence. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. 36
Figure 24: Mean intimacy z-score for arousal. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. 36
Figure 25: An example of usage window containing the DURING session (including the user answer to the ESM - beep), the BEFORE, and AFTER sessions. The usage sessions labeled as OUT are in usage windows without ESM answer. 40
Figure 26: Estimated mean z-score of context intimacy for the short, medium, and long *session_duration* for each *timeout*. 42
Figure 27: Estimated mean z-score of context intimacy for the few, several, and many *number_of_sessions* for each *timeout*. 43
Figure 28: Estimated mean z-score of context intimacy for the glance, review, and engage *session_kind* for each *timeout*. 44
Figure 29: State machine that represents valid transitions between the *screen* and *presence* events. 45
Figure 30: The transition model presents the variables considered in our research, especially in the PRESENT-OFF transition. 46
Figure 31: Frequency for each intimacy state for each hour and day. 48
Figure 32: Probability (left Y axis) for each intimacy state depending on the *hour* (bottom X axis), *day* (right Y axis), and *app_switches* (top Y axis). 52
Figure 33: Probability (left Y axis) for each intimacy state (colored lines) depending on *touches* (bottom X axis), *sub_category* (right Y-axis), and a *day* (top Y axis). 54
Figure 34: Distribution of the intimacy states over all the users. 63
Figure 35: Distribution of intimacy ranks after transformation from intimacy scores. 63
Figure 36: Mean Kendall's Tau across all the users for the datasets and the two RPC classifiers, ZeroR (baseline) and J48 for the 10CV test of experiment 1 [error bars 95% CI]. 65
Figure 37: Mean Kendall's tau of the ten 10CV runs of RPC ZeroR (baseline) and RPC J48 for each user for the population_clean dataset [error bars 95% CI]. 66

Figure 38: The mean Kendall's Tau for each user tested against the models of other users over all the datasets [error bars 95% CI]. 67
Figure 39: The mean Kendall's Tau for each user tested against the models of other users for each dataset. The dashed black is the population_clean dataset that performs better than the others [error bars 95% CI]. 67
Figure 40: The mean Kendall's Tau for all the datasets over all the run of users against other users model. The dataset population_clean is significantly better than the others [error bars 95% CI]. 68
Figure 41: The notification we presented to the users during experiment 3. Notification A is the countdown of the seven days of data acquisition. With notification B users could provide their current intimacy state. We were displaying B each time the intimacy model was computing a new prediction. Notification C was presenting the users input vs. our prediction (in here two examples where our prediction was matching the user intimacy). 70
Figure 42: The number of users answers to the notifications of experiment 3. As for the first user study, some users contributed more than others. 70
Figure 43: The number of predictions that our model did for each user involved in experiment 3. Some had a higher number of predictions probably due their higher mobility. 71
Figure 44: Mean Kendall's Tau heat map for all the 12 datasets (c1 datasets are generated by users cluster one and c2 by users cluster two) and majority vote (from the 12 datasets outcome). A positive mean Kendall's Tau means we correctly predicted the intimacy ranks; a zero means there is no correlation (wrong prediction), and a negative one represents an inverse ranking prediction. 72
Figure 45: Mean Kendall"s Tau over all the users for all the datasets. No datasets seem to be significantly better than the others [error bars 95% CI]. .. 73
Figure 46: The heat map is showing the counts in percent of the users activities in the different intimacy rankings. We normalized the counts per rank and user. Users tend to be less mobile in high intimacy than in low intimacy. 73
Figure 47: The heat map is showing the counts in percent of the users WiFi in the different intimacy rankings. We normalized the counts per rank and user. Users tend to be connected to a frequent WiFi when in high intimacy and not connected to WiFi when in low intimacy. 74
Figure 48: The heat map is showing the mean light (in Lumen) for all the users in the different intimacy rankings. We normalized the light values with z-scores for each user. Light seems to be overall more intense when users are in low intimacy (probably outdoor) than when they are in high intimacy. 75
Figure 49: Mean of users' mean delete for sharing on Facebook (not anonymously) and on Server (anonymously) ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI). 84
Figure 50: The users divided into three privacy clusters using Jenks Natural breaks, 14 fundamentalists Del_m = [1.19, 2.18], 16 pragmatists Del_m = (2.18, 3.26], and 12 unconcerned Del_m = (3.26, 4.31] ('1' very likely to delete the content, '5' being the opposite). 84
Figure 51: Mean users' mean delete for sharing on Facebook (not anonymously) and on Server (anonymously) separated in the three privacy clusters ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI). 85

Figure 52: Estimated mean users' mean delete for sharing a different kind of data, to highlight how different content is treated depending on the sharing format Facebook (not anonymously) and on Server (anonymously) ('1' very likely to delete the content, '5' being the opposite).....86
Figure 53: Mean users' mean delete for sharing different kind of data in Facebook (not anonymously) and on Server (anonymously) with the groups of similarly shared contents: A easily shared, B more protected depending on information that is conveyed, C and D highly protected contents ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI). 86
Figure 54: Mean users' mean delete for sharing different kind of data in Facebook (not anonymously) and on Server (anonymously) and intimacy states (delete: '1' very likely to delete the content, '5' being the opposite, intimacy: '1' completely intimate – '6' not intimate, error bars: 95% CI).....88

List of Tables

Table 1: The six different scenarios, one for each intimacy state, with their corresponding number of observers, kind of observers and possible places. Observers are (a) strangers, (b) co-workers and classmates, (c) friends, (d) girl/boyfriend, and (e) family members. 13
Table 2: The US1 and US2 with their data collection methods, the data collected, and where we used the data in this work. 29
Table 3: Extracted phone usage data statistics about usage windows and usage sessions. 41
Table 4: LMM analysis and pairwise comparisons results for *session_duration* for each *timeout*. 42
Table 5: LMM analysis and pairwise comparisons results for *number_of_sessions* for each *timeout*. 43
Table 6: LMM analysis and pairwise comparisons results for *session_kind* for each *timeout*. 44
Table 7: Basic statistics for transitions subset, A=ON-OFF, B=ON-PRESENT, C=PRESENT-OFF, D=OFF-ON. 47
Table 8: Basic statistics for the application used. 48
Table 9: Basic statistic of *Most_Sign_Tr* model data after cleaning. 51
Table 10: Most_Sign_Tr model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '), confidence intervals (CI). 51
Table 11: *Most_Sign_App* model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ') and confidence intervals (CI). 52
Table 12: Basic statistics of *Most_Sign_Com* model data after cleaning. 53
Table 13: *Most_Sign_Com* model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ') and confidence intervals (CI). 53
Table 14: Original continuous features and their corresponding binary binned features. 61
Table 15: The datasets generated from the six different cleaning techniques with their percentages of kept and removed instances. 64
Table 16: The separation of users in clusters for each of the six datasets with cluster centroid and the average Silhouette value for cluster quality (> 0.5 represents a good cluster separation). 68
Table 17: From sensor networks to people-centric sensing campaigns – research challenges 80
Table 18: Comparison of different studies performed to understand differences in the users' privacy concerns depending on diverse sharing aspects (*only public sharing). 82
Table 19: Distribution of content types for the random sharing survey question. 85

Acronyms

Acronym	Full text
CELLID	Cellular Cell Unique Identifier
DRM	Day Reconstruction Method
EMA	Ecological Momentary Assessment
ESM	Experience Sampling Method
FB	Facebook
MCS	Mobile Crowd Sensing
MDC	Mobile Data Challenge
SAM	Self Assessment Manikin Scale
SE	Server
mHUMAC	Understand Human Aspects in Mobile Computing
PCS	People Centric Sensing
US	User Study
US1	User Study 1
US2	User Study 2

1. Introduction

It is now widely accepted that researchers and developers leverage the context of mobile users to provision them with mobile services. Smartphones follow us everywhere in our everyday life. People have become accustomed to using smartphones at home as well as at work, in public transportation or while walking in the city, sitting in a pub or on the go, with the family or alone, on vacation, and in many more situations. Recent study reports that people use their smartphones more than 200 times a day [1]. However, people do not use their smartphones in the same way in every situation. While sometimes a quick glance at the phone is enough to check notifications status, in other cases people engage in longer sessions. For example, to text with a friend, answer to emails, buy food, browse Facebook, search for a place on a map, or check the balance on their credit card [2].

Currently objective context elements, such as location, are the most used and exploited for mobile services provision. Most of the subjective context are still technically hard to sense (e.g., users' mood [3], [4], kind and people around users [5]–[7]). It is tough for researchers and developers to test their assumptions and model such context. As to summer 2015, only tech giants as Google are starting to introduce the necessary tech tools to revolutionize research and development around inferred context elements (Project Tango[1] and Nearby API[2] are examples). However, the missing technology does not prevent researcher as us to already experiment and test new subjective context. We can be a step ahead when technology is fully available. To enrich the potential that subjective context can bring to smartphone users, we defined a new subjective context variable and researched six inter-related objectives. We present them here with their main contributions that are the novelties of this work.

Objective 1: the definition of a new subjective context variable that we called *intimacy*.

Contribution 1: we identified and fused together into *intimacy* three objective context elements, namely the place where a user is (*e.g*. at home, at work, or on the bus), the number of people around the user, and the kind of people (*e.g*. friends, family or strangers). We provide the novel and extended definition of *intimacy* in Chapter 2.

Our primary goal is to define theoretically, a social computational model and then practically operationalize *intimacy* as a subjective context metric that we can leverage for user-centered mobile service deployment. We will show that the proposed *intimacy* concept is a suitable candidate for being a new relevant context information to refine further certain mobile application design aspects. For example, the design of interfaces and patterns of interaction, or to alleviate intrusiveness (reduce notifications), to evaluate real users' needs, and more.

---

[1] https://www.google.com/atap/project-tango/
[2] https://developers.google.com/nearby/

After defining intimacy, we focused on the feasibility of designing an intimacy model and on the intimacy definition validation itself. Then we investigated the potential of intimacy related to smartphone usage and as context element influencing the users perception of anonymity in Mobile Crowd Sensing (MCS). Finally, to operationalize intimacy, we study the design and development of computational intimacy prediction algorithm able to estimate users' intimacy perception.

Objective 2: understand if the data collected from a smartphone is suitable to greedily model and estimate the defined perception of *intimacy*. This step includes the definition of what data we could leverage and the design of a greedy algorithm able to estimate intimacy.

Contribution 2: in Chapter 3 we use the Mobile Data Challenge (MDC) Nokia dataset to provide some assumptions about relevant data to identify intimacy (e.g., Cell IDs, Bluetooth scans). Based on these assumptions we defined a greedy algorithm able to estimate the intimacy of users. The data analysis results suggest that we can leverage smartphone's data to estimate the intimacy of their users.

We explored the following objectives by designing and executing two ad-hoc user studies. We present the details about these studies in Chapter 4. However, given the technological limitations cited above we can anticipate that we needed to execute the studies "in the wild" using real end users smartphone and not controlled "in lab experiments". We applied techniques such as Experience Sampling Method (ESM). For example, to collect the *intimacy* ground truth and other user context elements, such as the place, number and kind of people around. Additional data automatically logged from users' smartphone (to collect app used, screen touches, and more), and in lab interviews, Day Reconstruction Method (DRM) (to refine the *intimacy* definition).

Objective 3: understand if there is a relation between the three objective context elements we considered: place, number of people and kind of people around the individual and the defined *intimacy*.

Contribution 3: in Chapter 5 with the help of data we collected in a user study we prove that place, number and kind of people strongly relate to users' intimacy perception. When users are in familiar places, like home, they feel significantly more intimate (i.e., attached to their place), than when they are in public places like a bar or in the street. Also, the number of people and the kind of people around are significantly influencing the perception of *intimacy* positively or negatively depending on the situation. Additionally, we found that subjective context elements such as mood can greatly influence *intimacy*, proving how this metric correlates to several context elements together (detailed results in Section 5.3 and Section 5.4).

Objective 4: we investigate if the proposed *intimacy* metric is relevant for mobile services: are there significant differences in the smartphone usage and interaction when users are in different *intimacy* states?

Contribution 4: we analyzed the data about user smartphone usage (Chapter 6) that we automatically logged from their devices (*e.g.*, screen ON/OFF events). We evaluate that intimacy influences significantly the length of smartphone

usage sessions, the number of sessions in a given window of time, and the kind of session executed (characterized by the operations and apps used). At a very low level of analysis, we discovered that (1) when users switch between many applications (most used apps) in a single smartphone usage session they are more likely to be intimate. (2) When users use some applications actively for a long time, they are more likely to be intimate, and (3) users use messaging applications mostly when intimate and prefer to browse the internet when in lower intimacy. Depending on the *intimacy* perception, users are using their smartphone differently, which shows the relevance of considering such metric for mobile services provision.

Objective 5: proved the relevance of *intimacy*, we must research the most suitable modeling approach that allows estimating automatically users' *intimacy* in real-time, unobtrusively, and accurately.

Contribution 5: once again thanks to the intimacy ground truth we collected in the users studies we modeled intimacy and tested the preliminary intimacy model we created and operationalized "in the wild" with real end-users. In Chapter 7 we show that it is feasible to predict the users' intimacy perception using machine learning techniques such as Ranking by Pairwise Comparison (RPC) and (to start) simple location and time-related features. We also present some new variables that can help further improve the prediction accuracy and we are sure that the new recent development in ubiquitous computing technologies can help further the process.

This contribution is different from contribution 2 because there we were applying a greedy algorithm on a dataset that was not targeting intimacy directly and there was no ground truth to compare the algorithm performance. Furthermore, the greedy algorithm of contribution 2 was not able to perform in real-time.

Objective 6: the implications of understanding the user's intimacy for users themselves, and mobile applications and services providers.

Contribution 6: as the objective the contribution of this point span over several Chapters (i.e., 5, 6, and 8) we wrap up about the intimacy importance in the different discussions we open along this work. To summarize, We may leverage the *intimacy* model for better designs of pervasive systems. (1) Service providers can leverage the attention of the users and provide them relevant information or notify them about particular events when the users are more likely to spend more time on their device, and we capture their attention (*i.e.*, they are in a *high intimacy*) (smartphone usage and intimacy case study Chapter 6). (2) Particular applications can adapt their interface to offer easy tasks or shortcuts when users are in a *low intimacy* or *high intimacy* (we present the design of a study about intimacy in a real published app in Section 9.1). (3) Services can reduce their intrusiveness when the users are in *high intimacy*, for example by disabling the collection of personal context information that may be more privacy sensitive in this context (Mobile Crowd Sensing and intimacy case study Chapter 8). (4) The smartphone itself can adapt its functionalities never to lock the device or change the way notifications are delivered (*e.g.*, sound, vibration and light) in *high intimacy*, or always automatically lock the phone and change the notification settings in *low intimacy*. (5) Application designers may use *intimacy* to analyze users' behaviors unexplored so far.

To conclude we also discuss the overall contribution of this work with a future work section (Chapter 9) in which we design a final study about intimacy that we will execute in a real and fully deployed mobile application. We will test the developed *intimacy* module for its accuracy, speed, and dependability in the real environment and applied to a real use case (existing mobile service). We will evaluate the theory of the concept of *intimacy* and the effectiveness of changes in service provision, as well as user experience with this service, based on this metric (*i.e.*, evaluate the outcome of this intervention). Additionally, this future work will allow us to operationalize the *intimacy* model as a software-based algorithm to enable mobile services to leverage *intimacy* as a subjective user variable for provision service personalization. We conclude the analysis with a section (Section 9.2) dedicated to new open research questions that we need to answer for the future of subjective context in general and intimacy in particular.

To draw a complete picture of what we present in this work, in Figure 1 we show the research approach enabling to achieve the above objectives. All the objectives are dependent, and they required an ordered execution, which implied a high complexity for the data collection (ad-hoc users studies) and analysis procedure (e.g., handle and model subjective data).

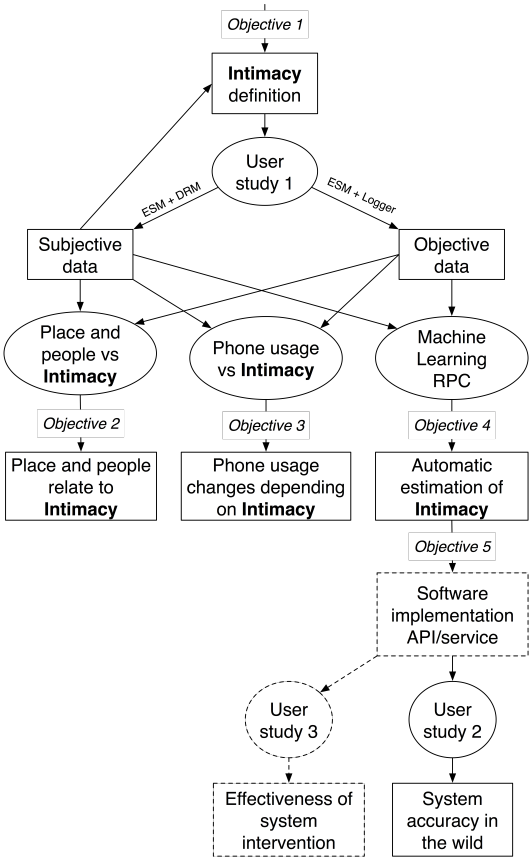

Figure 1: Schematic representation of the five top objectives of this work. Semi-circles are the actions needed to fulfil the related goals and rectangles are representing their outcomes. The dashed part refers to on-going and future work.

2. Why Intimacy?

2.1 Context and Smartphone Usage

Past and ongoing research have focused on the understanding of the user's context, *i.e.*, what are the context elements, context modeling methods, context discovery, and so on. Comprehensive survey papers on context are those by Chen and Kotz [8], Bolchini *et al*. [9], Bettini *et al*. [10], Ye *et al*. [11], enumerating the existing context definitions, *e.g.,* the one of A. K. Dey [12]: "*Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*" They also investigate in depth their field of interest related to context, *e.g.*, application design, situation identification, and context modeling.

Many works focused on understanding the context of users and how this influences the use of (mobile) devices. Eagle and Pentland [5], [13] used mobile phone Bluetooth capability to identify the users' social context and study how they were accessing and consuming information in different social situations. Adams *et al*. [6] focused on the users' colocation to establish social ties, using mobile phone GPS and Bluetooth. They analyzed how social interactions were influencing the use of two applications: a social photo and video album and a blog auto-generating content from social information. Zheng and Ni [7] presented an innovative system to detect social circles using smartphone Bluetooth and envisioned the use of these circles in combination with social network apps usage.

Researchers also investigated the subjective nature of smartphone usage. Ryan and Gonsalves [14] present how a context information such as location can be necessary to influence the use of mobile applications. They empirically compared the usability of a restaurant booking application in four domains: PC application, Web-based PC application, Web-based mobile application, and native mobile application. They found that the native mobile application - the only one that was able to make use of the location context information, can perform better than the others regarding subjective usability measurements and time to carry out the task for which Ryan and Gonsalves designed it. With the possibility to use the current location of the users, the mobile application was reducing by an average of 10 seconds the duration of the task to find a restaurant nearby and book a table. They showed how just the location context information has the potential to improve performance in information retrieval tasks, and they claim that is important to investigate the influence of context on application usability. Shin *et al*. [15] widely analyze context factors and the use of mobile apps. They performed a user study where they collected GPS and cellular network location (for the location), app open/close events (for the time), and battery charging patterns, and so on. They used the data to predict the next app used, thus be able to adapt the apps menu. Böhmer *et al*. [16] did an analysis of mobile apps' usage. They collected data from over 4100 users. Users used 22'626 different apps and researchers collected 4.92 million events of app usages (*e.g.*, install, opened, closed). They analyzed apps' usages

depending on location, time, and chain of app usages context information. They presented how the analysis of this information could be used to improve the design of mobile apps. Ickin *et al.* [17] also investigated mobile apps' usages regarding the location (semantic place), time of the day, and connectivity (Quality of Service analysis). When researchers interviewed participants, they admitted that they learn how to maximize their experience based on their previous apps usage experience, connectivity options and app needs at hand. There is a need to provide to users automated mechanism to adapt the use of their smartphones and mobile apps. Floch *et al.* [18] presented the European research project MUSIC describing typical context (*e.g.*, noise, light, network, location, users interactions) and adaptation features (*e.g.*, different user interaction and provided functionalities) that are relevant to develop self-adaptive mobile apps. They propose a framework to help developers to make use of contextual information to reduce the development complexity of self-adaptive mobile apps. Falaki *et al*. [19] showed how individual usage patterns can be exploited to personalize battery consumption prediction. Soikkeli *et al*. [20] focused on how smartphone usage changes depending on the user's location.

Differently from most of the previous works, we focus on smartphone usage patterns in a wider domain than the location, also considering the number and kind of people surrounding the user. In the next section, we motivate the choice of these three objective context elements and provide the definition of a new subjective context variable that we called intimacy.

2.2 Defining Intimacy

The following definition of intimacy is the result of several iterations and feedback from user studies analysis and papers submitted to conferences and journal along the whole thesis. Its core meaning never changed from the beginning to the end of the thesis period, but we added several specifications to reduce confusion and misinterpretation of our intimacy concept.

The definition of intimacy as a new subjective context element closely relates to three objective contexts that are part of the smartphone users' daily live: place, number and kind of people around them. We fused them together in a single subjective element because they are objectively identifiable by anyone, but their perception as a whole is subjective and depends on the relation of users with a given place, and the people around them.

As reported by Prager [21], the concept of 'intimacy' is complex, and it has several definitions. The word 'intimacy', from the Latin 'intimare', means 'impress, make familiar', (from 'intimus', *i.e*. 'inmost') [22]. According to the dictionary, the adjective 'intimate' describes the context (place or setting) as familiar, being private and comfortable [23]. In our study, we refer to intimacy with respect to a location/place or an environmental situation ("a cosy and private or relaxed atmosphere" [22]).

The place is one of the elements that connects with the perception of intimacy. Manzo shows [24] how being at home or in another location can influence several traits of the people personality, such as emotions and perception of the environment. Seamon and Sowers build upon that by researching report on the work of Relph [25] researching about the deep meaning of people's attachment

to a location/place. Relph [25] indicates 'home' as the most intimate place. Thus, we consider the location/place as a potentially predictive variable for the individual's intimacy.

Besides the place, another component that influences intimacy and describes the current setting is the presence of other people and an individual's relationship with them. Saegert [26] explains how the environment of people influences their cognitive perception of it. The more people are physically around, the more complicated the situation becomes cognitively, and the person tends to be distracted more. Diener and Seligman [27] point out how the presence of closest people contributes to the well-being of people. Miller and Lefcourt [28] consider these relations as social intimacy. In their study, they interviewed undergraduate students about their relationships with friends, acquaintances, and family members to define the characteristics of the relationships that they consider intimate.

Based on these considerations, for this research, we focus solely on (1) the place, (2) the number and (3) the kind of people physically around the user as the three top contextual elements bound to users' perception of intimacy. We are aware that the context of users does not include only the three components we include in our intimacy definition. However, we made the choice to keep intimacy simple and focused, and iteratively enrich it. Once we ground its core elements, we can enlarge the intimacy domain. Our future goal is to include new context elements that we can prove are a fit for the definition with the help of further large scale user studies. Already at this stage we show the example of mood components, valence and arousal, as good candidates to enrich the intimacy definition.

Furthermore, we are not investigating the relation of intimacy with other concepts such as love, sexuality, self-disclosure, or privacy [29], [30]. While there is a relation between intimacy and privacy [31], in this definition we only refer to intimacy as a perception of *attachment to and familiarity with* the three context elements: place, number and kind of people analyzed in our study. Such perception does not aim at analyzing communication with others or controlling the disclosure of information about oneself, or at capturing the quality of users' relationships as defined, for example, by Westin [30], but focuses solely on a subjective evaluation of one's current situation.

Based on this interpretation of intimacy, we hypothesize that when people are at home having a dinner with their families discussing family matters, they feel intimate, while when they are on a bus with strangers, they do not feel intimate.

2.3 Intimacy Implications

The main implications for intimacy are closely related to mobile applications developers. Developers are designing mobile applications that users can customize to have the best experience. Usually, users can customize applications behaviours through a set of settings that offer the possibility to define some static cause-effect actions. A simple example can be a messaging application that provides the possibility for the users to customize how to get notifications about new messages. For instance, via a full-screen pop-up, with message preview and phone screen ON, simple notification icon, with or without

message preview, any possible sound, vibration and LED light colors combination, or different combinations of all the above, depending on the contact or group of people, and so on. All these settings may overwhelm the users that may just do not explore all the functionalities of the application [32]. Users may not be able to exploit the potential of these apps fully, either because they are not aware of how to set up the application to enable different behaviors, or they just agree with the default settings, because the set up seems to be cumbersome. Also, usually these settings are static, so recalling the example, if the user decides that she/he wants a given new message notification type, the application will always notify the user in that way, in any situation. That can irritate the user if this setting is not appropriate for some situations. The developers become aware of that and to address it, on a growing scale they develop new kinds of application-adaptations techniques to meet the users' needs and expectations [33]–[35]. Namely, developers are starting to use contextual information about the users to not only provide new functionality on their mobile applications but also to adapt the applications functionalities and the way they provide them. There are mobile applications that after a very basic setup are automatically adapting to the user context, to the extreme example by *Google Now*, that is not even requiring a minimal configuration, but acts autonomously from the beginning and "settings and preferences" of the application derive are from the user's actions. However, the number of these mobile applications is still tiny compared to the total available ones, and we still need to explore diverse types of user's contextual information [36]. We need to research new contextual clues, in particular, the ones that are considering the user as a human being with a set of motivational and attitude traits and exploit these [37]. Towards this end, in our research, we focus on one of such traits, and at this moment, we investigate if the *intimacy* of the users is relevant to mobile applications' adaptation.

We want to contribute to enriching the information about users context that is a disposal to mobile applications and services developers. Recalling our primary objectives presented in Chapter 1, this study about the intimacy of users and smartphone usage aims to answer five dependent questions:

1. **Can we somehow derive users' intimacy from their smartphone? Can we at least greedily estimate it? (Chapter 3)**
2. **Is intimacy, as we defined it, a representative for users context in term of place, number and kind of people around them? (Chapter 5)**
3. **How the interaction of users with their smartphone may change depending on their perception of intimacy? (Chapter 6)**
4. **Can we model and predict intimacy with a smartphone? (Chapter 7)**
5. **What are the implications of understanding the user's intimacy? (Chapter 6, Chapter 8, and Chapter 9)**

Answering the questions above can bring new possibilities for developers that are striving to make their applications more usable and useful for their users, in different circumstances. If we consider our simple example about the messaging application, we may think how the notification of a new message may automatically adapt to the user intimacy. For example, if the user is in an intimate situation (at home, possibly with her/his family) the notification for a new message may be presented as a full screen pop up, with the content of the message displayed in it (given the assumption that there are fewer privacy issues when at home with people the user trust). The notification may switch the smartphone screen ON and emit a longer notification sound with a LED blinking (to attract the user attention probably focused elsewhere). Instead, when the user is in a less intimate context (in a bus full of strangers), the notification may need to be more discreet without showing the message content in a pop-up (so people around cannot read it). The notification may be using the normal notification system of the smartphone, and without emitting a very long and loud sound or just vibrating. Of course, we may reverse or refine these settings following the user needs and/or behavior patterns.

2.4 Leveraging Intimacy

We leverage intimacy along this study by focusing on its relations with smartphone usage, and additionally we provide a use case in which we study intimacy and Mobile Crowd Sensing (MCS).

As in this case with intimacy, other researchers considered the importance of subjective context elements, for example: detecting users' mood [4] or stress [38] states, and relate them to different smartphone usage patterns. Our goal is to do a step further than the works we highlighted in Section 2.1 about how users use smartphones and the context in which they use them. We based our research on smartphone usage mainly on the work of Banovic *et al.* [2]. They researched how long users are interacting with smartphones by labeling the smartphone usage sessions in three categories: 'glance', 'review', and 'engage'. They proved how these categories of interactions are able to represent the users' general behavior while emphasizing that each user keep her/his unique interaction pattern. We build upon this research by employing the proposed categorization of interaction sessions to carry out our analysis of smartphone usage. This part of our work wants to understand if there is a real opportunity to use intimacy as new and relevant context element to develop more intelligent mobile applications and services.

For a more concrete and less general use of intimacy than just the smartphone usage, we chose to put intimacy in the context of MCS. Smartphones are devices that can be exploited to collect valuable information ranging from users context to more personal smartphone interaction such as applications used, and number of touches performed. Intimacy is yet another information that we may be able to capture from users' smartphone. With this intimacy use case, we want to explore if intimacy has a role in MCS. Particularly, we want to understand how users involved in MCS perceive data anonymity? Is intimacy playing a role in this anonymity perception? Does anonymity influences how data sharing decisions and which is the role of intimacy on the users decisions? In Chapter 8,

we present our analysis of anonymity in MCS and how intimacy and other factors impact the perception of anonymity from the point of view of users directly participating as MCS contributors.

3. Feasibility Study: Can we Assess Intimacy “in the wild”?

In this chapter, we present a first study that explores the Mobile Data Challenge (MDC) Nokia dataset to investigate the feasibility of detecting intimacy of mobile users from the data collected from their smartphones. We propose a greedy algorithm based on simple assumptions that make use of the available data to output the intimacy level of users. At the end of this chapter, it will be clearer if there is hope for intimacy and how it can be applied more concretely to real use cases.

3.1 From MDC Raw Data to Intimacy

3.1.1 Raw Data Exploration

In [39] Laurila *et al*. provide a detailed description of the MDC dataset and the data collection procedure. We could access approximately one year of data of 38 participants that the organizer of the MDC selected out of a total 185 contributors. We started our research with a literature study and a high-level analysis of the raw MDC data to understand which features were best suited to describe the intimacy perception of the users (*c.f.*, Section 3.1.2). For example, firstly we confirmed that Bluetooth data (periodical scans of surrounding Bluetooth devices) is a good indicator to have an estimation of the crowd around the user (*i.e.*, ~10m circle) and the possible relations between people [5], [6], [13]. The results show that only two participants of the study may strongly relate to each other (*e.g.*, be a couple), and that the majority of them have some frequently encountered, but unknown to the study, devices logged in their scan data. The second example of critical data is the ring phone status (*e.g.*, normal, ascending, silent). For each user, we computed the overall percentage of all phone’s ring status during each hour of the day. We found out that all users follow a precise similar ring pattern for every hour. In Figure 2 we show the percentage of each ring state of user P9, for the full day over the whole study (note that ‘ringonce’ option is present, but no user uses it). All the other users follow the same behavior, but with a different distribution of ring status. This result can suggest that users react to particular situations by changing their ring status and that these situations are almost uniformly distributed over the whole time of the data collection.

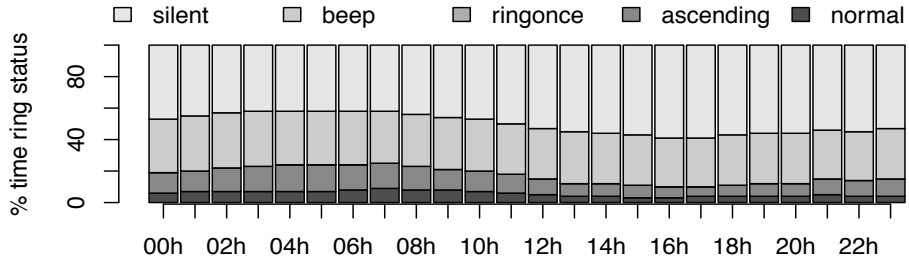

Figure 2: Smartphone’s ring status of participant P9.

A third example of raw data we used for deriving the intimacy level of mobile users is the analysis of the phone's charging status (*e.g.*, no charging, charging, and full). As Ferreira *et al.* present in [40] from the resulting charging patterns we noticed that users have a predictable behavior on charging their phones. In particular, they charge more often their phone during the night, and when the phone completes its charge, it stays for a long time attached to the power adapter. During the day, charging and fully charged times are shorter. In addition to these three particular raw data examples, we also investigated another kind of data. Calls and SMS logs (essentially durations and relations) [5], [6], [13], GPS and WIFI traces (for a greedy indoor/outdoor recognition: available GPS implies outdoor while WIFI implies indoor). We split all the data we analyzed above in working days and weekends to see if there were differences in the resulting patterns.

3.1.2 Features Selection

Given the result of the first raw data analysis and based on the definition of intimacy (*c.f.*, Section 2.2) we decided to select some specific features and split them into two categories: *observers* (for people) and *safe places* (for locations). For the *observers* category, we chose all the characteristics that can help us to identify if some people surround the users and what is their relation to them. In the *safe places,* we selected only features that can give us an indication of the users trust in the place they are in at a given moment (*i.e.*, if users feel secure).

For each feature, we devised some assumptions upon raw data that helped us to decide which of them to use and in which category they are supposed to be. In the category of **observers** we have:

Bluetooth: the number of the devices around the user can reveal the (minimum[3]) number of people observing him. Also by using the overall appearance frequency of devices we can also derive the relation of the user with these observers.

Ring status: can represent the willingness of the user to share the events of the device with others. A silent status may indicate that people surround the user, and he does not want to disturb them or that they are not supposed to know that he received a message, call, or similar. A normal status can represent the opposite.

Outgoing call: the duration of a call made by the user and the relation with the called person (based on the overall frequency of call exchanged between them) can give us a hint about how the user feels about speaking on the phone at that moment. If the user trusts the observers (or he is alone), he may feel more relaxed to call a family member or a friend and speak for a long time.

Outgoing SMS: the concept is the same as for calls, but we reversed it. If a user is exchanging many SMS with a family member or a friend, it may indicate that he is in the company of people that are not supposed to know the content of the conversation or even that he is communicating.

---

[3] Assuming each device correspond to a person. There can be more people around than discovered devices.

Furthermore, for the **safe places** we have the followings:

Charging status: can reveal if a user is in a trusted place. If the phone is charging it can indicate that the user is currently at home, office or in his car. Also, the fully charged status for a long time could tell us that the phone is left attached to the charger for long and confirm that the user trusts that place.

Ring status: is the inverse of the ring status in the observers' category. This time is related to “how much” the user wants to be disturbed by external events. A silent status may indicate that the user is in a safe place and does not want any other to enter that place in any way, for example with a call.

Indoor/Outdoor: there is a high probability that if the user is outdoor, he may not be in a safe place.

3.1.3 Intimacy Estimation Algorithm

We want to evaluate the intimacy level of a user. After the selection of the features, we attempt to combine them to obtain a single score representing the intimacy. Our primary idea is not to have an accurate way to devise the intimacy level, but to have an estimation of it. Since we do not have the ground truth to evaluate the accuracy of our algorithm, we base its construction on the assumptions we made. For this reason, we chose to create a greedy algorithm that using a fixed score system combines all the features to obtain the final estimation of the intimacy level. The first step we made to assess the intimacy level was to divide the raw data of each user in intervals of 10 minutes (in this way we can estimate the intimacy status six times per hour, and have enough raw data to process in each interval). Then we decided to fix the intimacy scale from 1 to 6. We chose this scale to have enough distinct intimacy levels and at the same time fit the scales of the single features presented before. The intimacy level 1 represents the *not at all intimate*, 2 *not intimate*, 3 *more not intimate than intimate*, 4 *more intimate than not intimate*, 5 *intimate*, or 6 - *completely intimate* level. These levels relate to six different scenarios we present in Table 1.

Intimacy	# Observers	Observers	Places
1	50+	a,b,c,d,e	Big structures, open air, ...
2	15-30+	a,b,c,d,e	Cinema, shopping center, pub, ...
3	10-20+	a,b,c,d,e	Train, bus, metro, traffic jam, ...
4	5-25	b,c,d,e	Work, classroom, auditorium, ...
5	1-15	c,d,e	Different kind of private places
6	0	none	Any, mostly private places

Table 1: The six different scenarios, one for each intimacy state, with their corresponding number of observers, kind of observers and possible places. Observers are (a) strangers, (b) co-workers and classmates, (c) friends, (d) girl/boyfriend, and (e) family members.

For both *observers* and *safe places*, we defined the following rules to compute the level of intimacy for each feature in each time interval. For the **observers** category:

Bluetooth: we ranked all the different devices found with Bluetooth scans by their appearance frequency. The most frequent device is the first in the rank (most known observer) and the less frequent - the last. Then to each Bluetooth device found in the considered interval, we assigned a weight between 0 and 1

depending on the position of the rank (0 for the first position). The inverted mapped sum of all these weights between 1 and 6 give us the level of intimacy for this feature, *i.e.*, the known observer is the most intimate.

Ring status: in this case we simply assigned an intimacy score between 1 and 6 to the different ring status accordingly to the assumptions made for this feature (to equally distribute the five states of this feature over the six levels of intimacy). We have 6 for *normal*, 4.32 for *ascending*, 2.66 for *beep*, and 1 for *silent*. In the case of different states in the same interval, an average of the scores is taken.

Outgoing call: we ranked all the phone numbers found in the call log depending on the number of interaction the user had with each of them. First in the rank is the most contacted number (most special person). The importance of the called weights each call's duration. For each interval, we summed all the weighted durations and mapped this sum to the interval 1 to 6 accordingly to the assumptions we made for this feature.

Outgoing SMS: we ranked the phone numbers as for the calls. Each message found in the interval has a weight depending on its position in the rank. As for Bluetooth, we summed all these weights and mapped the result on a number between 1 and 6 to obtain the intimacy level of this feature.

The **safe places** follow the same line of thinking:

Charging status: as done for ring status; in this case, we just assigned an intimacy value depending on the state. When the phone was *no charging* we have 2, when *charging* 4, and when *fully charged* 6. In the case of different charging situations in the same interval, an average of the scores is taken.

Ring status: the procedure is similar to the same feature in the category of observers, but accordingly to the assumption of this feature, in the category of safe places we inverted the scale of intimacy levels.

Indoor/Outdoor: this feature is just a simple binary 'yes'/'no' decision. If the user is *indoor* during the interval we considered, we give the score of 6, otherwise the score of 1. In the case of a mix between outdoor and indoor in the same interval, the score is 3.

As we show in Figure 3, the algorithm completes by putting together all the scores for each category. It first computes an average of the scores for each time interval for the observers and then it does the same to the safe places. We obtain the final intimacy level for a given time spot with the average of the score of the two primary categories. In this way, all the features of the categories and the two categories have the same weight to derive the intimacy level. It is important to remark that the approach we propose uses only the data of each user; we do not share data among them to perform the analysis. The motivation is that such an algorithm may be implemented directly on the user's personal smartphone, not depending on the others' data.

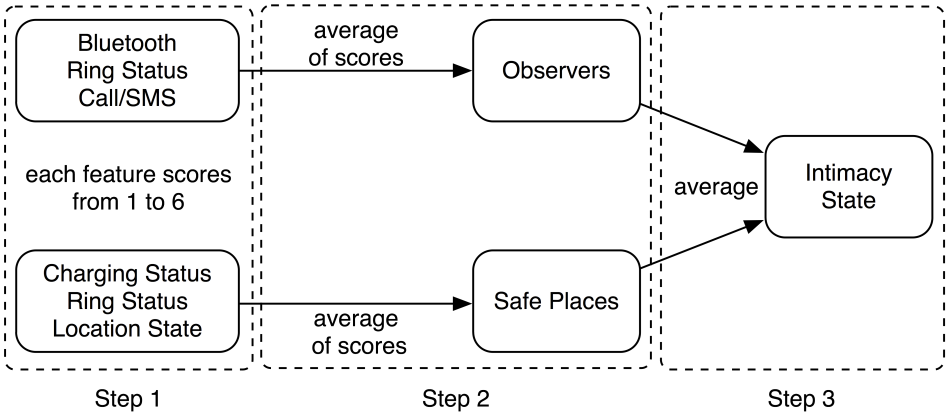

Figure 3: The three main greedy intimacy algorithm steps. In step 1 features get assigned a score from 1 to 6 depending on their values, in step 2 we average these values into one score for ‘observers’ and ‘safe places’, and finally, in step 3 their average defines the current ‘intimacy state’ of the user.

3.2 Results

To derive mobile users’ intimacy level we used the MDC data of 38 different participants with different demographic attributes such as sex, occupation (*e.g.*, students and full-time job), and age range. The data collected from the participants is not uniform. We have different starting and ending dates of the data gathering and non-uniform missing data across all the users (12% to 85% of the raw data used in this study). In total, we analysed data for 38 participants over the period of 13 Aug 2009 – 26 September 2010. As an example, we have selected a user that presents the most interesting results and who is among the ones with less missing data (P26, 15% missing data).

3.2.1 Observers and Safe Places

We analyzed the results for each feature independently, but for simplicity, we are going to present only details about their main categories. In Figure 4 we present the observers and in Figure 5 the safe places most frequent intimacy level per interval of the whole experiment for the selected participant (P26). We divided the data into the days of the week from Monday to Sunday and each day in 144 intervals of 10 minutes. For observers, we can notice a particular intimacy level pattern. From Monday to Friday the level of intimacy is always reduced during working hours (around 7 am and 5 pm) and it is higher during nights and evenings. During the weekend the pattern is different. This fact can suggest that in weekends the user is more intimate, or he tends to meet people that are more close to him as family members, and best friends. Instead, for safe places we observe a slightly different pattern. For all the weekdays the night hours and some part of working hours are more intimate than the rest of the day. Also, in this case, the pattern is a little bit different for the weekend, *i.e.*, when the person does not work. The time spent in less safe places is more frequent than during the working week, where this behavior is more present just at the end of the day. From the intimacy level of the indoor/outdoor feature during the weekend, we can also add that the user is more active outdoors, so he may be for more time

in possibly less safe places. If we look to both categories at the same time, we can say that they share some similarities that can indicate that our reasoning about intimacy may be right. Although the working place is somehow considered safer than home, the differences may not be necessary against our reasoning. For example during weekend afternoons and evenings, the person tends to be less intimate accordingly to safe places, but intimate accordingly to the observers. That can mean that she is in a not safe place (*e.g.*, a park), but she may be alone or with someone that is close to her.

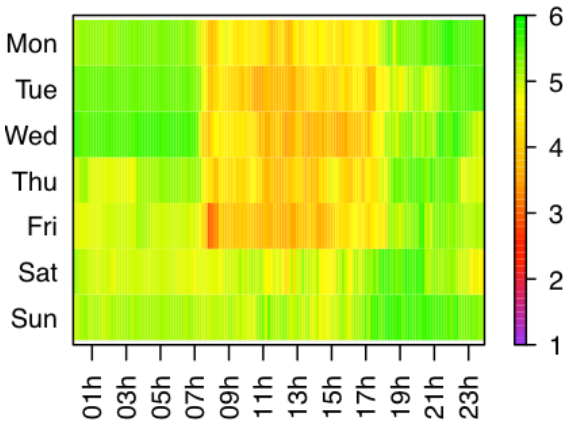

Figure 4: Observers most frequent intimacy levels for participant P26 per time interval (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy.

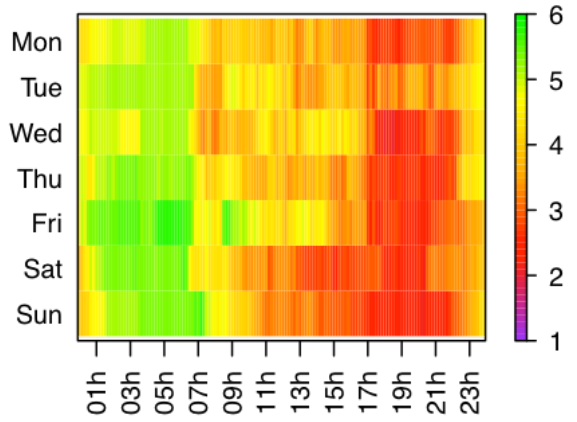

Figure 5: Safe places most frequent intimacy levels for participant P26 per time interval (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy.

3.2.2 Intimacy Levels

Always considering participant P26, in Figure 6 we show his overall most frequent intimacy level per interval (combination of observers and safe places as explained in the algorithm) during a week. Also, in this case, is possible to

recognize a pattern that reflects the ones depicted when discussing the two primary categories alone. Weekdays are similar but different from the weekend. The higher level of intimacy is always during evenings and nights except for Friday and Saturday nights, in which this level seem to shift to later times. This finding can suggest us that P26 uses to go out and be more social on those nights.

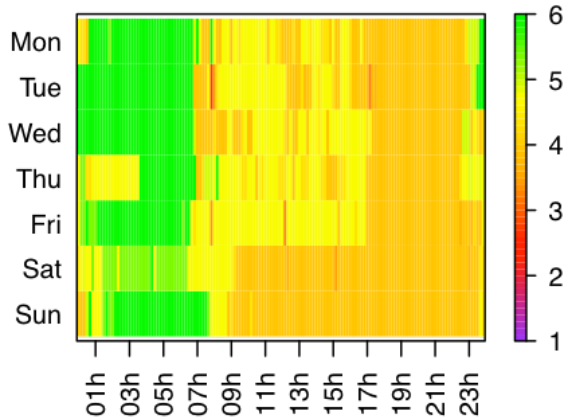

Figure 6: Most frequent intimacy levels (average of intimacy levels of ‘observers’ and ‘safe places’) for participant P26 per interval of time (10 minutes). Legend: 1 (purple or dark) means low intimacy - 6 (green or grey) means high intimacy.

In Figure 7 we present the probability of participant P26 to be in a given intimacy level (considering missing data as well) for the seven days of the week. To categorize the data in 7 distinct categories (intimacy levels plus not available data) we rounded the outcome of our algorithm to the closest integer. From the graph, we can see that this person is most of the time around the 4th and 5th level of intimacy. She tends to be more social at the end of the week (and thus less intimate) and on Sunday, she prefers to spend more time alone or with a closer person (and thus to be intimate). Based on data for Monday we make the assumption that if the majority of the missing data for the rest of the week would be present, the probability to be in intimacy level 4 may increase. Another important fact to depict from the graph is that we do not have many situations of absolute intimacy or not intimacy at all (no probability for levels 6 and 1). To have such levels of intimacy the user would need to be in extreme situations either with a lot of not known people to reach level 1 (*e.g.*, an opera) or, to get level 6, all the features at once would need to correspond to an intimate case (really unlikely to happen given our assumptions).

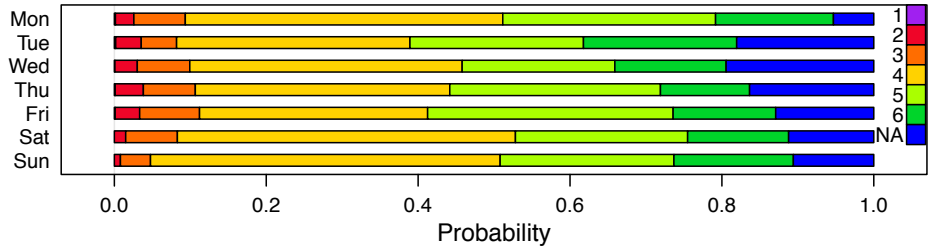

Figure 7: Probability of participant P26 to be in a particular intimacy level (considering missing data as well) for the seven days of the week. Legend: 1 (purple or dark, on the left of the graph) means low intimacy - 6 (green or grey, on the right of the chart) means high intimacy, and NA (blue or black, at the very end right of the chart) means not available data.

3.2.3 Demographic Analysis

With the help of Figure 8, we want to consider all the MDC participants. From a survey filled by 29 of them (out of 38) we have demographic information that may help us to connect specific intimacy level patterns to the population.

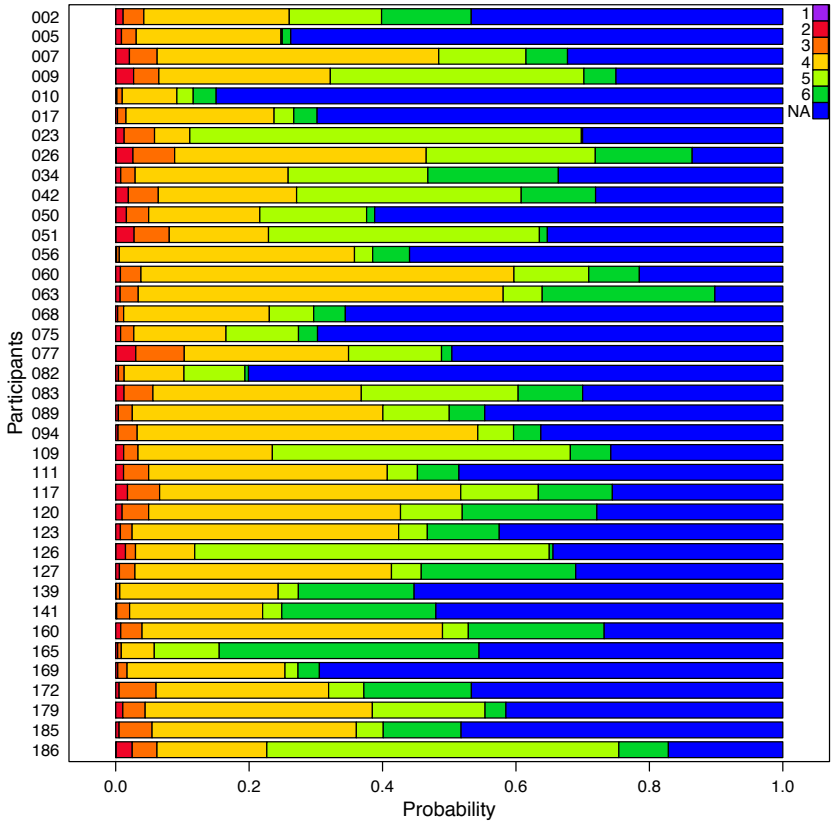

Figure 8: Probability for all the participants to be in a given intimacy level (considering missing data as well) for the whole dataset. Legend: 1 (purple or dark, on the left of the graph) means low intimacy - 6 (green or grey, on the right of the graph) means high intimacy, and NA (blue or black, at the very end right of the chart) means not available data.

In general for the majority of the people under analysis, from the graph is evident that the quantity of missing data is greater than the probability to be in a given intimacy state. Also, we can say that most of the users tend to be around the 4th and 5th level of intimacy. This fact can reveal that we tend to stay alone or with people that we trust most of the time (trend is confirmed by results presented in [27]). The users gender shows that female seem to be less intimate, but we have only eight females in the whole group. Hence, our conclusion is drawn with care. In each age range, there are different behaviors and the distribution of the people is not uniform enough to make further assumptions. The occupation of the participants does not seem to correlate with the level of intimacy. In each category (students and full/part-time workers) there is no evidence of similar intimacy patterns. We have some indications that people that use public transportation to go to work are less intimate than the one using the car, bike or walk. Also, in this case, the number of answers and the distribution are not enough to be sure about this phenomenon. We wanted to investigate more the correlation between our results and survey data about relationships and time spent with trusted people, but given the distribution of the answers seen so far, we cannot derive statistically significant results. A less general study with pre-selected participants would be more representative to have meaningful results.

3.3 Discussion

In this section, we firstly discuss the validity of the results of this study, and particularly the limitations stemming from assumptions employed in our approach. Secondly, we discuss the role of ground truth and possible ways of acquiring it. Furthermore, we discuss potential application areas of our research.

3.3.1 Study Assumptions & Limitations

To develop our approach identifying the intimacy level of the study participants, we employed several assumptions on how to interpret the data. We discuss these assumptions and ponder on possible inaccuracies that can influence the outcome of our research. We are going to list them as we present them in Section 3.1.3. We start with the category of the **observers**.

Bluetooth: it is possible that there were many people around the study participants, but they do not have the Bluetooth activated (or even did not have a smartphone at all). In this case, the quantification of the level of intimacy can be inaccurate. Our approach might show that the user is more intimate than in reality. Another problem related to this assumption is the existence of fix devices (*i.e.*, printers) with Bluetooth capabilities. We can interpret these devices as people (owners of mobile phones) that we encounter often and so simulate people that are highly intimate with the study participant. As a future work, we recommend the analysis of the MAC address of the Bluetooth devices. The prefix of the MAC address can give a hint about the kind of device we are observing. In this way, one can filter out the undesired ones before to proceed with the Bluetooth data analysis.

Ring status: the problem with this assumption is that people change their ring status accordingly to the users’ situation. We assumed that the majority of people do so, but this may not always be true. Depending on the cultural context

or just personal behavior (*i.e.*, people that always have the ring tone on vibrate) "the rules" can change and for example, it may not be considered impolite to have the phone's ringer volume set at maximum during a meeting, lecture or just in an open space office. For these people there would be no difference between being in a crowded tram (where we would hypothesize the ring is ON) and being in a meeting (the ring is OFF). This matter can result in inaccuracies in our results because we are not anymore able to distinguish these kinds of situations from each other.

Outgoing call and SMS: we based these assumptions on two events, namely performing a call and sending SMS, that for the overall considered time may be infrequent and irregular. The occurrence of these events can help us to improve the accuracy of the intimacy assessment given some specific time interval, where these events are present, but the absence of such events do not provide any information. Also, given their limited distribution, one needs to have at disposal a long trace of events, to indicate accurately, which are the most frequent numbers called or texted. This problem implies the collection of data from the user for longer periods of time to understand his behavior.

Furthermore, for the **safe places** we discuss the following assumptions.

Charging status: the assumption that when a person is charging her phone, she is in a trusted place can be inaccurate. For example, in case if she is traveling in a plane, a train or on a long bus trip, where the electricity plug is available. Nowadays all these means of transportation offer the possibility to have a power source at a disposal. In these cases, using our approach we may conclude that the user is intimate, but, in reality, he is not. Furthermore, as we show in Figure 5 (safe-places), we consider the fact of being in an office as mostly intimate, too. The intimacy perception is subjective and for example, a person working in an open space office may consider the situation somehow intimate instead another not at all. The challenge in interpretation of the data arises from the fact, that both categories of people can charge their phone at work. For the population in the second group, our assumption is wrong, and it will decrease the accuracy of our algorithm.

Ring status: this assumption closely relates to the one made for the *observers* category. We can make the same kind of observations for this category where we assume *safe places* and not *observers*. Also, in this case, we can say that a user may be at home alone, in a high intimacy state, but his phone's ringtone can be ON and phone's ringer volume set at maximum.

Indoor/Outdoor: the subjective perception plays a significant role in this assumption. We assumed that if the participant is outdoor, most likely he is not in a safe place. The challenge in interpretation of the data arises from the fact that some people may consider to be alone on a bench in a park or a tent in the forest, as being in a safe place. In this case, our algorithm recognizes an outdoor environment and therefore it can erroneously conclude that the users are not intimate.

3.3.2 Ground Truth and Experience Sampling

The validation of our assumptions and so of the accuracy of overall results of our algorithm it is not possible given that we do not have any ground truth from the

data provided for the MDC challenge. This situation posed some limits not only on the verification of the existing assumptions but also for further exploration of the data. If we make more assumptions without a check of their validity more likely is that we lead to significant inaccuracies in the final output. For these reasons, we limited the exploration space to the ones proposed in this chapter.

With experiments and a data collection targeting more precisely the given objective of estimating people intimacy, we will introduce Experience Sampling Method (ESM) to collect the ground truth we need. We will employ ESM in a form of a short questionnaire appearing automatically on a user phone along his daily life activities, and asking him to label his current context with respect to his/her feeling of intimacy. With the help of periodical and random EMS logs on the user phone, we will be able to have a better understanding of the phenomenon of intimacy in different daily life contexts and start to validate our preliminary assumptions. Once we have a solid base, we can begin to explore other categories of data that may help us to refine the algorithm.

3.3.3 Where to Apply Intimacy?

Once accurate, this concept can be applied to develop applications in several domains. They may range from another data campaign similar to the one done on the MDC data [39] where the data would be collected respecting the level of intimacy of participants. Social applications (*i.e.*, messengers and social networking applications) may share and display the status of the user and his relevant details accordingly to the level of intimacy. The development of applications that automatically control how a smartphone or other devices handle events and notifications to the users (receiving a message, a call, an email, and a request for approval). For example, assuming that when the user is intimate, the alerts shall be less intrusive (*e.g.*, just a notification without sound).

3.4 Conclusion

It is feasible to model intimacy from smartphone data and with proper experiments it has the potential to provide benefits on context research.

In this chapter, we presented our initial approach to the analysis of the concept of intimacy. We devised a simple method to derive intimacy from daily life context data acquired on a mobile phone, and we have presented its preliminary results. Although we do not have sufficient information to confirm the accuracy of results of our analysis (no ground truth available and the initial participants’ survey is not entirely applicable), we have some first hints from the patterns of intimacy levels and our personal experience. Also, we found some supporting material from the specialized literature (*e.g.*, [27]) that confirms some of our conclusions. We confirmed that there is hope for intimacy, but to obtain more significant results, we need a dedicated experiment, following the MDC approach, involving a pre-selected set of mobile Android OS participants and involving ESM deployment for ground truth availability. This first research results helped us to have a more clear view of which variables and confounding factors we need to investigate in the experiments that we present in the next chapters.

4. The mHUMAC Method and User Studies

In this chapter, we present our user study method that we applied to intimacy experiments. We will describe in details the two User Studies (US) that we designed and executed to answer our research questions about intimacy. The goal is to provide a basic understanding of the main experimental methods used to collect data during the US, the kind of data we collected, and the methods limitations. To simplify the reading, we introduce methods to analyze the data and how we generated the analysis results when needed in the dedicated results section.

4.1 The mHUMAC Method

The variety of *human aspects* plays a significant role in user perception and acceptance of mobile applications and services. These human aspects can include, but we do not limit them to mood, emotional state, intellectual and physical state, social interaction status, and so on. It is important to understand how human aspects relate to mobile computing to identify implicit and unmet users' needs to provide them novel and useful mobile applications and services. These human aspects can relate to specific phenomena. They can range from the most preferred interaction style with a mobile service (auditory, kinesthetic, visual), via the user's specific health and care needs (wellness or low radiation), to the user's particular aspects like cognitive load, physical flexibility, or momentary perception of safety, intimacy or love in a particular context. To summarize, human aspects include (i) those influencing the human-computer interaction in mobile computing (i.e., a mobile device, a smartphone in a given context), and (ii) those influencing the human interaction with her context (with the own state, the environment, and the people around). The former focus may lead to meet the unmet human needs and improvement of interaction aspects/service design aspects (in a given context) while the latter focus may result in the development of innovative mobile services supporting the implicit human needs in a particular context. In both cases, researchers should assess these human aspects *in the wild*. Specifically, in the natural environments where users use mobile devices daily.

Intimacy is part of these human aspects. To study and investigate intimacy in mobile computing we introduced in our study an interdisciplinary research methodology to Understand Human Aspects in Mobile Computing denoted as mHUMAC. The mHUMAC methodology involves users through gathering the cumulative users' opinion via *open-ended interviews and surveys*. Thus, specifically focusing on understanding the users' expectations towards a researched phenomena and current experience of this phenomena. We use surveys mostly to establish the users' baseline experience in the experiment variables and context, but also to gather general demographics of the experiment participants. The mHUMAC goal is also to collect the momentary users' opinion on some particular aspects of health behaviors, moods, feelings, social interactions, or environmental and contextual conditions via an *Experience Sampling Method (ESM)* [41]. Special momentary surveys executed multiple times per day 'in situ' (in the natural users environments). Our goal is

also to gather the episodic users' opinion on some particular aspects with semi-structured interviews based on the diary, for example by the *Day Reconstruction Method (DRM)* [42]; (d) gathering the data upon the users' daily life contexts and smartphone usage via *continuous*, automatic, unobtrusive *data collection on the users' device* through the measurements-based logger service.

Each of the methods offers different granularity of available data collection with respect to time and subjective/objective users' information (Figure 9). ESM, DRM, and classic surveys provide information on subjective variables directly related to users. ESM collects this data with an interval up to each hour (or higher frequency if needed, but it may influence negatively the users and thus jeopardize the experiment [41]). We can apply DRM weekly, and general surveys less frequently (for example at the beginning and/or the end of the experiment). The logger collects the objective variables on the users' device, and the frequency may vary. Finally, we can also use ESM to support the logger by collecting more specific users' context details and, therefore, obtain richer objective contributions.

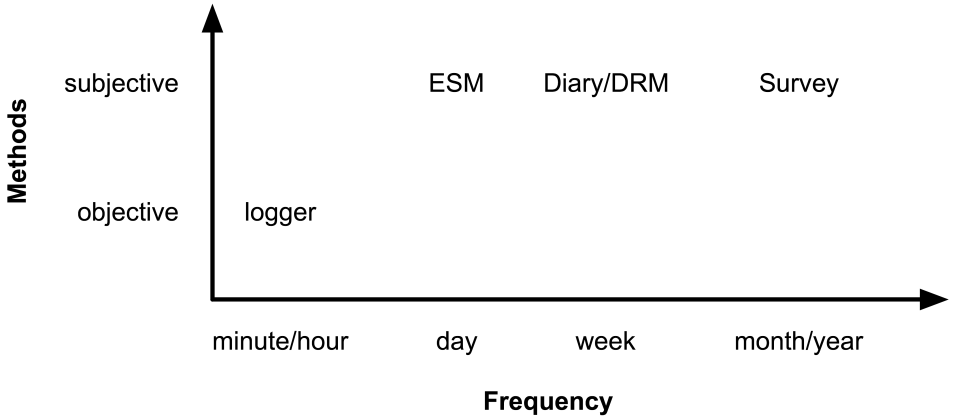

Figure 9: Data acquisition methods (subjective and objective) and their ideal granularity regarding frequencies (from minute acquisition to month/year intervals).

4.2 The User Studies

To perform the study about intimacy and its implications for mobile applications and services we designed two users studies. The first one aimed to collect the necessary data for the exploration and analysis of intimacy (ranging from the verification of the assumptions at the core of the concept to the exploration if its applicability for mobile applications and services). This study lies at the heart of the full intimacy analysis presented in this work. The second one aimed to experiment a first working prototype of the intimacy model and to collect further data that could help its refinement. For both studies, we applied the methods described in the mHUMAC methodology.

4.2.1 User Study One: Validation and Modelling of Intimacy

For User Study 1 (US1) we recruited 42 people (18 female and 24 male), aged 18 to 45 years, from Pittsburgh (PA) in the USA (22) and Geneva (GE) in

Switzerland (20). The recruitment was performed using online announcements in Craigslist in the USA and from our living lab participants in Switzerland [43]. We used all the data collected in this US1 in all of the analysis we performed, and we present in the next chapters (Chapters 5, 6, 7, and 8). We first performed the study in the USA in which we found much more collaboration from smartphone users, clearer ethical procedures, and it was easier to remunerate participants directly in cash. To perform the Geneva study, we needed to construct our living lab of end users by offering free smartphones rent in exchange of people contributions (no possible to use cash, and very low participation of people without tangible inceptives), and we could autonomously apply the same ethical protocol we used in the USA (there is no ethical committee in our department yet). Additionally, our goal was also to collect data in different demographics and cultural set up.

Entry Survey and Users Recruitment

We launched an online entry survey to assess the participant ownership of a smartphone with an experience of at least six months and to know more about the users habits on smartphone usage. We involved the participants in the study for a minimum of 22 days to a maximum of 36 (average of 27 days). They were students or employees in different working fields, using their personal Android OS-based smartphones.

Experience Sampling Method Questions

We required the study participants to answer an ESM based survey about their intimacy appearing on their smartphone screen. Each ESM survey - a so-called "*beep*" - appeared randomly, but uniformly, eight times per day, between 8 am and 11 pm. Additional *beeps* showed each time the participants plugged their smartphone for charging. The application recorded the answers and sent them over the Internet to our dedicated server. Each *beep* had several questions about the participants' current context and feelings:

1. Their subjective perception of *intimacy* via a six points Likert scale (following the answers entries order, from 1 'completely' intimate - to 6 'not at all' intimate) (Figure 10). We provided to participants the definition of intimacy at the study entry. We instructed them to indicate their perception of intimacy in their current situation, *i.e*. the location, number and kind of people around them when answering the survey;
2. Current semantic location ('bus', 'home', 'other', 'pub', 'school', 'shopping center', 'street', 'work') (Figure 11);
3. The number of people around the participant ('alone', '1', '2-10', '11-20', '21-40', '40+') (Figure 12);
4. The kind of people around ('co-workers/classmates', 'family', 'friends', 'girl/boyfriend', 'other', 'strangers') (Figure 13);
5. Users indicated their emotional state using the Self-Assessment Manikin (SAM) scale [44], with two dimensions: valence and arousal (scaled from -4 to 4). Valence options range from 'unpleasant' to 'pleasant' and arousal from 'calm' to 'highly activated' (Figure 14 and Figure 15);
6. The last question we presented to the participants is the focal point of the study we present in Chapter 8, and we have taken a particular care on how we designed and formulated it. From the work of Jensen *et al*. [45], we

know that if sharing does not occur, we can bias hypothetical data sharing questions answer (i.e., there is an attitude-behavior gap). The more recent work of Braunstein *et al*. [46] proven that is possible to reduce the bias if we ask sharing questions differently. We must never communicate to the study participants that we investigate privacy. Therefore, we designed our sharing question as suggested by Braunstein *et al*. [46]. In Figure 16 and Figure 17, we present two possible examples of this kind of question. We submitted this question once per each *beep*. It had two random variables that we changed every time we started a new survey: (a) how is the content shared and (b) which is the content shared. The first informs the participant about how we shared the content: either publicly on Facebook (thus, being identifiable) or on an 'open data' server anonymously. The second is the type of content that we proposed to share in that situation: *video*, *photo*, recorded *audio*, sport/*activity*, *location*, and *air* quality. To minimize the attitude-behavior gap, as Braunstein *et al*. [46] suggest when questioning the user, we assumed the content already shared, and we asked participants to decide if they were going to make an effort of deleting it. The participants could choose from 5 points Likert scale from 1 'very likely to delete' to 5 'very unlikely to delete'.

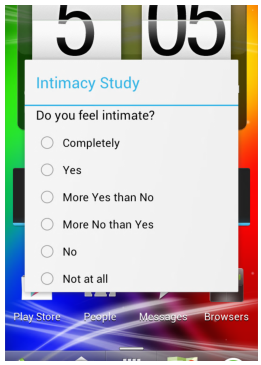

Figure 10: The ESM question about the perception of intimacy of the user at the moment of the answer.

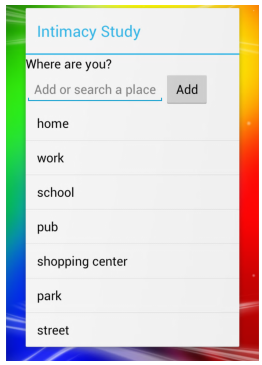

Figure 11: The ESM question about the place in which the user was at the time of the answer.

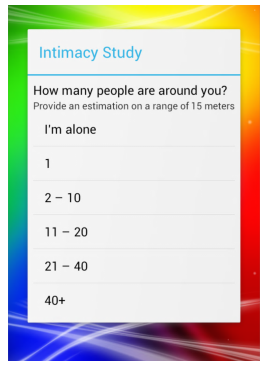

Figure 12: The ESM question about the number of people around the user at the moment of the answer.

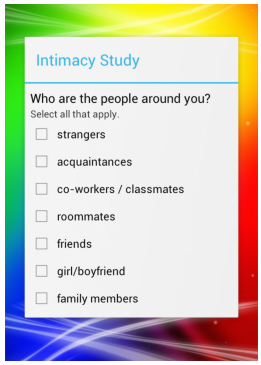

Figure 13: The ESM question about the kind of people around the user at the moment of the answer.

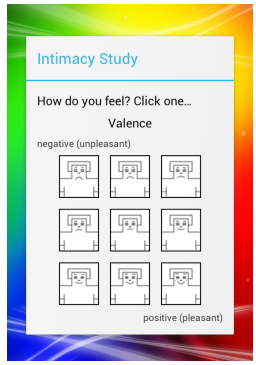

Figure 14: The ESM question about the valence of the user at the time of the answer.

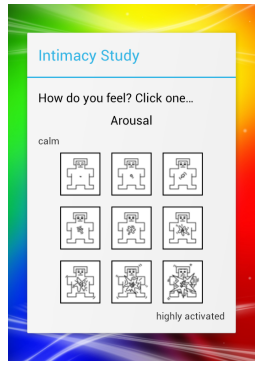

Figure 15: The ESM question about the arousal of the user at the moment of the answer.

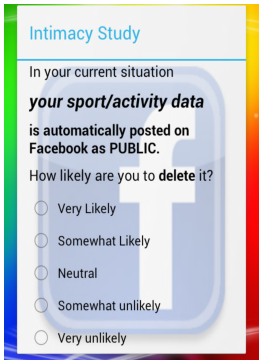

Figure 16: The ESM question about the deletion of a content type (here activity) posted publicly on Facebook.

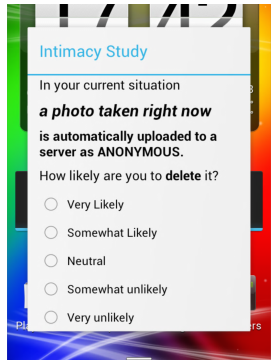

Figure 17: The ESM question about the deletion of a content type (here photo) posted publicly but anonymously on a Server.

Smartphone Logged Data
Besides issuing the ESM questionnaire, our research application was logging several system- and user-generated events as they occurred. We list here the data that we finally used for the analysis presented in the following chapters: smartphone screen events, finger touches, the apps that participants were using (from these logged events we elaborate smartphone usage variables, Section 6.1 and Section 6.4), and each minute. We also captured the unique identification number of an access point to a cellular network (CellID) to which users' smartphone were connected (we used them to define locations subsequently used for the intimacy prediction model, Section 7.1.1).

Participants Personal Thoughts Acquired Using Day Reconstruction Method
We interviewed each participant every week along the duration of the study for a total of four times each. Following the DRM [42] approach we asked the participants to explain the reasons for their intimacy choice (related to the first ESM question and the context in which they claimed to be) and of their content deletion choices (related to the last two questions of the ESM survey). We always covered mainly the events of the last 24 hours and few cases we went back to the *beep* data and selected some older answers (yet at most one week old and representing particular situations) to acquire more information from the user.

4.2.2 User Study Two: Evaluation of the Preliminary Intimacy Prediction Model

For User Study 2 (US2) we recruited 31 people (11 female and 20 male), aged 18 to 45 years, from Geneva (GE) in Switzerland. The participants were all from our living lab. The majority of them were already participants of US1. We involved all the participants for exactly one month. We used the data collected in this US2 exclusively to test and perform a first improvement iteration of the intimacy prediction model we present in Chapter 7.

The details of this US2 will be explained in Section 7.5.3, in which we describe the related experiment. Summarizing the methods used in US2, also for this experiment we used ESM to collect the intimacy ground (ESM questions and screenshots in Section 7.5.3) truth from users whenever our model was predicting the current user intimacy. For this study, this approach implies that we were running our model directly on participants' smartphone and collecting data when needed and not randomly as for US1. We also used a smartphone logger to collect further data as such the current user activity (e.g., walking, being still, being in a vehicle), the WiFi to which the smartphone was connected, and the light value sensed from the smartphone each time the user was switching on the smartphone screen.

4.2.3 User Study One and Two Summary Table

To conclude this section, we present Table 2 in which we summarize the different aspects of the two studies, from the mHUMAC methods used to the data collected and in which chapters or section of this work we use the data for our analysis.

User Study	mHUMAC methods	Data collected used in this work	Chapters in which data we used
1	ESM	User intimacy state (randomly during the day)	5 6 7 8
		User place	
		Number and kind of people around the user	
		User mood (valence and arousal)	
		User's content deletion decision	
	Logger	Screen events	
		User's finger touches	
		Apps used	
		Connected CellID	
	DRM	Personal users' thoughts about the intimacy and data delete concerns	
2	Intimacy predictor	Output of Intimacy prediction	Section 7.5.3
		Intermediate output of Intimacy model	
	ESM	User intimacy state (at prediction output)	
	Logger	User activity	
		Connected WiFi	
		Light sensed from smartphone	

Table 2: The US1 and US2 with their data collection methods, the data collected, and where we used the data in this work.

4.2.4 Study Limitations

To conclude this chapter, we report here the main limitations related to the user studies we presented (mostly related to US1). There are limitations related to ESM itself, ESM content of the questions, the dependability of our logger to the smartphone usage, and how to study privacy concerns (already extensively commented when we presented the content deleting ESM question, Section 4.2.1).

The ESM method itself is the main limitation on capturing the users' intimacy. With this approach, we may have lost some users' rare events or situations (Lathia *et al.* [47]), in particular when the users happened to be in low intimacy states, as our data indicates. Thus, in US1, some particular situations may be missing. To reduce the risk of losing critical inputs in US2, we used location-based features (related to the intimacy model prediction) to predict when to trigger ESM questions.

Another limitation stems from the choice of options we proposed in the different ESM questions (*e.g.*, for a number of people: 'alone', '2-10', '40+'). It is hard to design these in advance, especially because, to our knowledge, the literature

does not clearly present validated scales suitable for the purpose of our study. Based on this work, in subsequent works, we may refine the possible ESM questions and answers. Moreover, the six points Likert-scale we used may not be the best representative for the *intimacy* states. In US2 (as we present in Section 7.5.3) we used four points Likert-scale.

Also, the automated data logging from the users' smartphone suffers from a main limitation. For all the system related data (e.g., the current connected CellID or WiFi) we are limited to the time the users switch on their smartphone. For the data depending on the smartphone usage (e.g., screen events and finger touches) we are bounded by how much the users are active on their smartphone and the tasks they perform the most. Some users may be more active than others, and the number of the different tasks they may perform is vast.

Finally, in US1, we have not shared the content on Facebook or via an anonymous server. We may have an 'attitude-behavior gap' as mentioned earlier (Jensen *et al*. [45] and Braunstein *et al*. [46]). However, we have deployed our methods in the real user context, as well as when we have interviewed the participants. We understand that they expressed their genuine opinions about their actual content sharing in a given real context, we claim that the attitude-behavior gap may have only minimally biased our studies. Additionally, our goal is to highlight the relative differences between sharing content on Server (thus anonymously) or Facebook (thus not anonymously) and to identify the responsible factors. We are not in any way trying to define an absolute sharing attitude of users. Another possible limitation, related to this one, is that there are people that share any data publicly in running real systems like Facebook. Most probably a minority of these users are privacy unconcern, but most of them do not know how to manage privacy settings. They are not instructed about the risks of sharing data, or in the case of mobile applications they do not understand the permissions Lin *et al.* [48] and Kelley *et* al. [49].

5. Validating the Intimacy Definition

To ensure that the intimacy scores we collected via ESMs of US1 were accurately expressing the intimacy of users as we intended, we check if certain direct relations are in place. For instance, one would expect that in more familiar places, such as home, and with more familiar people, users would perceive higher intimacy (more attachment to these context elements) than in unfamiliar places or with less familiar people. Therefore, we evaluated how the variation of the (subjective) intimacy scores collected through the ESM questionnaires is bound to the variation of the objective context features we collected. We did this evaluation for place, number and kind of people around the user; as well the trending subjective context feature: mood (composed of valence and arousal variables).

5.1 General Statistics of ESM

The US1 application issued a total of 12497 ESM surveys (on average ~11 per user, per day) and we collected 6509 answers from all the users (~5.7 per user, per day). In Figure 18 we present how each user contributed to the answers. Apart from a few cases, most of the participants reached at least 100 answers (~3.7 answers per day) with some users reaching over 200. Additionally, in Figure 19 we show the mean intimacy score for all the participants over all their intimacy ESMs.

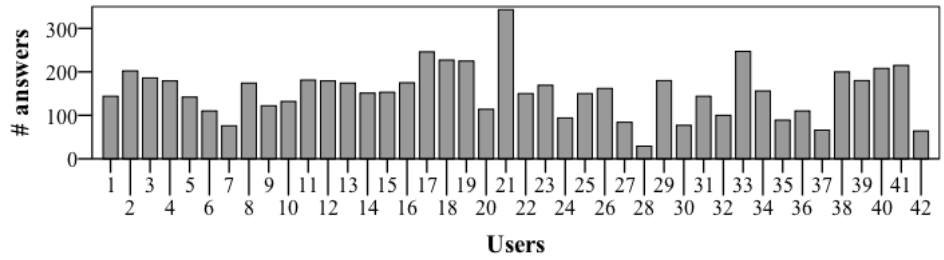

Figure 18: Number of users answers to ESM questions. Some users are contributing more than others creating a different number of contributions.

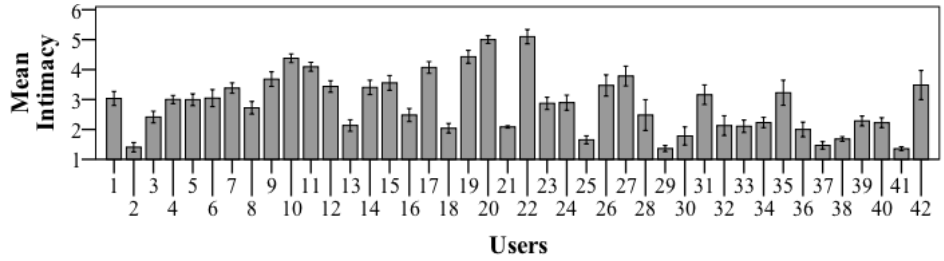

Figure 19: The different intimacy scores mean per each study participant (1 „completely" intimate - 6 „not at all" intimate) [error bars 95% CI].

5.2 ESM Data Preparation

Our data has two particularities that we need to take into account for a reliable analysis of the data. (1) As we show in Figure 19, users have a different overall perception of intimacy and tend to employ different subjective scales of intimacy; hence to be able to compare scores across the users we need to normalize these scales across all users. (2) As we present in Figure 18, users contribute differently to the dataset. For example, user 21 is a significant contributor, as well as participants 17, 18, 19 and 33. If we do not properly prepare the data for analysis, the users providing this information may bias the results. Following the method by Larson and Delespaul [50] we transformed the ESM data before the hypothesis-testing step.

Firstly, to address the individual intimacy scales problem, we standardized the users' intimacy score using z-scores [50]. To do so, we scaled the intimacy score for each user by centering her/his score *mean* to 0; representing his/her 'usual' score. This scaling results in negative values when the user is in a higher intimacy than usual, and in positive values for a lower perceived intimacy.

Secondly, we aggregated the ESM answers of each user transforming the *beeps file* into a *subjects file*. In the *beeps file*, each row represents the answers a user gave for each ESM (*i.e.*, the answers to the six questions). This file contains as many rows as the total of all the surveys answered. Instead, in the *subjects file,* we aggregated the intimacy value as the mean value of all the different situations encountered by the users (*i.e.*, the combinations of the answers to the ESM questions 'where', 'number of people', 'kind of people'). For example, as a situation, the user was on the street ('where'), with 2-10 people ('number of people'), being strangers ('kind of people'). We took all of her/his answers ('beeps') in this situation and aggregated the single intimacy scores with its mean. This *subjects file* contains as many rows as the number of different situations encountered by each user for all the users.

Differently from the beep-level analysis, by using the subject as the main unit of analysis, the assumption of independence between samples is not violated, and we avoid issues such as an inflated number of samples (*i.e.*, when a small number of individuals produce the majority of samples). For example, in our case we avoid that user 21, being a significant contributor of the data, defines a trend in the results (Figure 18). The disadvantage of using this approach is that we may not capture all the true relationships in our data [50].

5.3 Intimacy vs. Place, Kind and Number of People Around the User

In this section, we present how the intimacy concept correlates with place, number and kind of people around the user. Usually, ANOVA is used for such analysis. However, because of the nature of the ESM method, the questionnaire appeared randomly but uniformly during the day, there is no guarantee that each user happens to be in all possible situations (Lathia *et al.* [47]). Given the unbalanced nature of the data, *i.e.*, not all the situations are distributed equally and not all users fully covered them, we performed a Linear Mixed Model (LMM) analysis [51]. Differently from ANOVA, the LMM analysis can deal with missing

data without the need to discard the incomplete features and thus losing information. We performed LMM with users as subjects and the context intimacy z-score as a dependent variable. We modeled the effect of the three contextual elements as factors ('where', 'number' and 'kind of people') with a random intercept to take into account the variance of intimacy between users. Finally, for all the three context elements we tested which of their possible values were significantly different regarding the mean intimacy. In this case, we conducted pairwise post hoc tests with Bonferroni correction.

5.3.1 Semantic Place

With the result of the LMM analysis, we indicate that the place has a significant effect on the subjective perception of intimacy, $F(7, 1117) = 16.674$, $p < 0.001$. In the ESM survey, users could specify one of the following eight categories of place: 'bus', 'home', 'other' (all the rarely selected places), 'pub', 'school', 'shopping center', 'street', 'work'. Not all the users have been in all 8 places ('bus' = 19 participants, 'home' = 42, 'other' = 42, 'pub' = 30, 'school' = 31, 'shopping center' = 28, 'street' = 38, and 'work' = 37).

With pairwise comparison post hoc tests, we show that users perceive more important places, like 'home', as more intimate (more familiar) than less visited ones. The place 'home' (Figure 20, A1, labels refer to significant difference between groups of places) is significantly different from all other places belonging to group A2 $p < 0.001$. The members of group B1, 'Work' $p = 0.002$, 'school' $p = 0.034$, and 'other' $p = 0.006$ are more intimate than 'street' (B2). All other places differ amongst them, but not significantly.

The model estimated a z-score intimacy mean for 'home' of -0.02. This value is very close to 0 - representing the scaled average value of intimacy for all the users. An interpretation of this result can indicate that participants being at 'home' have their mean value level of intimacy perception, and elsewhere people feel less intimate than home (*i.e.*, increasing z-score). The intimacy perception seems to relate to the time that an individual is used to spend in a given place. Since most of our users spent most of the time at 'home', it is reasonable to notice that it has higher familiarity and attachment. Then, places like 'bus', 'shopping center', and 'street' are visited quickly and occasionally (users perceive lower intimacy and thus reduced familiarity and attachment).

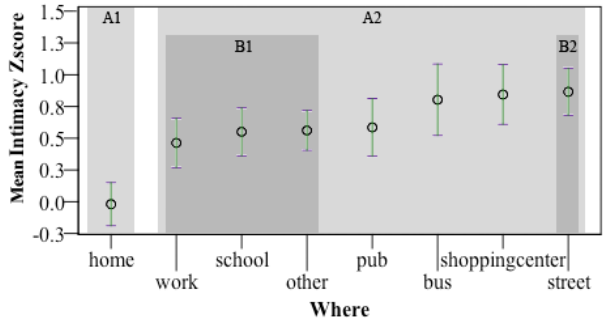

Figure 20: Mean intimacy z-score for places. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of places.

5.3.2 Number of People Around the User

With the LMM analysis result for this context element we indicate that the number of people has a significant effect on intimacy, $F(4, 1117) = 20.940$, $p < 0.001$. In the ESM survey, users could specify how many people were surrounding them by selecting one of the following six categories: 'alone', '1', '2-10', '11-20', '21-40', '40+' people. For some users not all the categories are present, in particular in the categories with many people around ('alone' = 41 users, '1' = 42, '2-10' = 42, '11-20' = 41, '21-40' = 37, '40+' = 32).

With post hoc tests we show that the fewer number of people around the user, the higher is the users' perceived intimacy. The category '0' ('alone', Figure 21, A1, labels refer to significant difference between groups of people around) is significantly different from '1' $p = 0.007$ (A2), and from the others $p < 0.001$ (A3 and A4). The category '1' (A2) is significantly different from: '2-10' $p = 0.001$ (A3), and '11-20', '21-40', and '40+' $p < 0.001$ (A4). The category '2-10' people (A3) is also significantly different from all the members of group A4 with $p < 0.001$. Finally, '11-20', '21-40', and '40+' people (A4) are not significantly different from each other, but different from all the members of groups A1, A2, and A3 with $p < 0.001$.

In particular, users perceive differently being alone or with one person (more familiar) than being with 2-10 people and with 20+. Most probably, the '2-10' category contains the threshold for which the number of people changes the perception from a high-perceived intimacy (higher familiarity and attachment) to a lower one. Having more than 20+ people around a user, results in low intimacy; for any higher number of people the intimacy level remains constantly low.

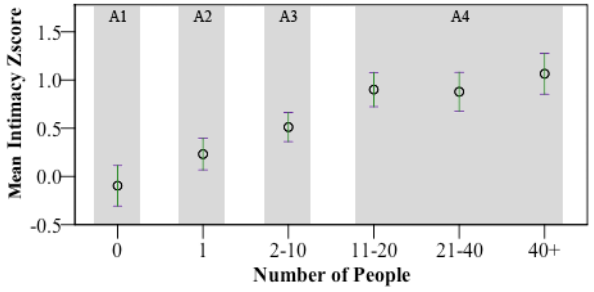

Figure 21: Mean intimacy z-score for number of people. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of people around.

5.3.3 Kind of People Around the User

The LMM results show that the factor 'kind of people' has a significant effect on the context intimacy mean, $F(5, 1117) = 17.450$, $p < 0.001$. In the ESM survey, users could specify what kind of people were surrounding them by selecting one of the following six categories: 'co-workers/classmates', 'family', 'friends', 'girl/boyfriend', 'other', 'strangers'. Also in this case, not all the users choose all the available categories ('co-workers/classmates' = 37, 'family' = 30, 'friends' = 33, 'girl/boyfriend' = 17, 'other' = 41, representing other possible combinations rarely rated, and 'strangers' = 41).

With post hoc tests, we could identify two main groups of 'kind of people': one accounting for people closer to the user ('family', 'girl/boyfriend', and 'friends' - Figure 22, A1, labels refer to significant difference between groups of kind of people), and another one with people that are less closely related ('co-workers/classmates' and 'strangers' - A2). The former contains the categories with which the users perceive high intimacy while the latter group contains the categories associated with lower perceived intimacy. The category 'other' does not belong to any of these two groups, probably due to its heterogeneous composition. The categories belonging to the group of closest people (A1) are all significantly different from the not closest one $p < 0.001$ (A2). 'Family' (B1) is also significantly different from 'other' $p = 0.046$ (B2), and there is a significant difference between 'other' (C1) and 'stranger' $p < 0.001$ (C2).

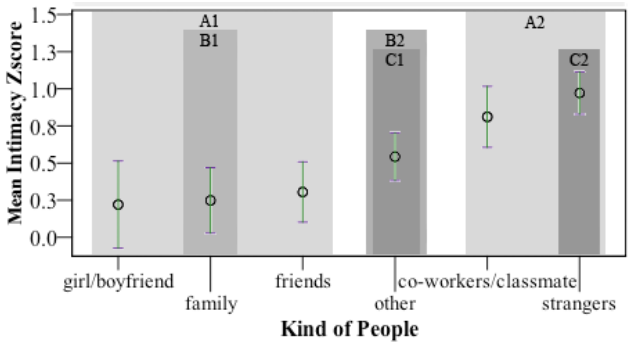

Figure 22: Mean intimacy z-score for kind of people. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI]. Labels refer to significant difference between groups of kind of people.

5.4 Intimacy and Mood Components: Valence and Arousal

Since we collected the mood components 'valence' and 'arousal' using ESM surveys, we prepared the data and analyzed it in the same way as we did in the previous Section 5.3, *i.e*. we performed an LMM analysis and post hoc tests with Bonferroni correction.

Using the self-assessment manikin (SAM) scale [44] users could specify their subjective value of 'valence' from '-4 - unpleasant' to '4 - pleasant' with the neutral value as '0'. Also in this case, not all users selected all the categories ('-4' = 12, '-3' = 15, '-2' = 22, '-1' = 28, '0' = 38, '1' = 34, '2' = 37, '3' = 37, '4' = 28).

Valence has a significant relation with intimacy, $F(8, 242) = 3.045$, $p = 0.003$. With post hoc tests we show a clear pattern (Figure 23): when users are in higher intimacy (low-negative z-score), they perceive the situation as more pleasant than when they are in lower intimacy (high-positive z-scores). We show that intimacy has a relation with valence that can be leveraged further, *i.e.*, to work with mood related tasks. After comparing the mean values of intimacy among all valence values, the only significant difference we found is between the less pleasant state of '-1' and the more pleasant states '2' $p = 0.039$ and '3' $p = 0.010$. The rest of the comparisons did not lead to significant differences (all p

> 0.05). This result is probably due to a high number of valence categories, 9, which would have required more ESM survey answers.

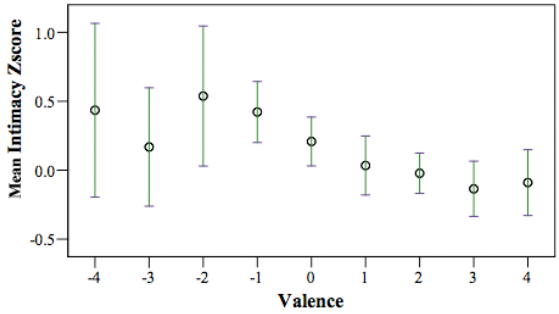

Figure 23: Mean intimacy z-score for valence. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI].

We also investigated the relation between intimacy and 'arousal' but we could not find any significant relation, $F(8, 244) = 0.411$, $p = 0.914$. As for valence users could provide their subjective value of 'arousal' from '-4 - calm' to '4 - highly activated'. Also in this case, not all the users selected all the categories ('-4' = 31, '-3' = 37, '-2' = 36, '-1' = 29, '0' = 35, '1' = 28, '2' = 23, '3' = 17, '4' = 17). Since we did not find any significant relation between intimacy and arousal, we did not perform any post hoc tests. With the plot of Figure 24, we show that arousal cannot be considered dependent on the perception of intimacy.

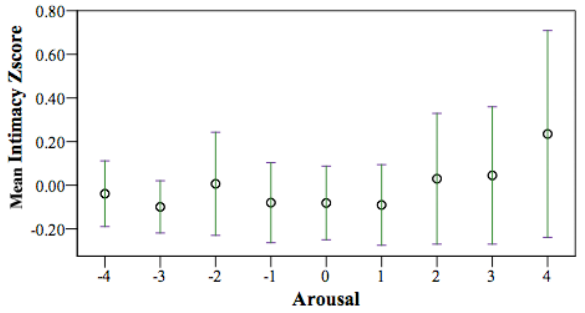

Figure 24: Mean intimacy z-score for arousal. Lower is the value of the mean z-score, higher is the user intimacy [error bars 95% CI].

5.5 Conclusion

Intimacy represents the users perception with respect to its enclosed context variables: place, number and kind of people around the users.

In the results we show that the three objective context elements: 'place', 'number of people', and 'kind of people' around correlate with users' intimacy. In particular, for the location, we have shown that users perceive 'home' differently than any other less familiar place. For the number of people, as the number of people increases, the perception of *intimacy* decreases. Finally, for the kind of people, we identified that being with the closest ones (*e.g.,* family and friends) leads to higher *intimacy* than when with non-closest ones (*e.g.*, strangers and

co-workers). The US1 results show that our definition of *intimacy* is close to what it can represent to the users in reality (considering the three objective context elements: location, number and kind of people around).

Intimacy has the potential to enclose other subject context variables, such as valence, representing users comfort in their current context.

Additionally, we explored whether the subjective context mood elements like valence and arousal relate with intimacy. Despite the low significance of the results (due to the high number of valence categories and the total number of answers from users), valence seems to be related to intimacy, while arousal does not appear to have any particular relation. The relationship with valence suggests that intimacy can enclose the users' feelings regarding the perceived comfort of the situation. However, valence is only one of the many subjective variables, and it is particularly applied to mood and stress metrics. More subjective context elements should be analyzed and compared to intimacy to achieve more accurate results.

To conclude, we can affirm that the captured intimacy of users is very close to our intimacy definition. We can now use this intimacy data to explore more in deep its relation to smartphone usage.

6. Intimacy and Differences in Smartphone Usage

Since we can reliably use the intimacy ground truth, we collected from our US1 participants we defined two experiments to analyze how the intimacy perception of users influences the use of their smartphones. In experiment 1 we focus on high-level smartphone usage variables such as the duration of smartphone usage sessions, how many sessions users perform in given window of time, and more. With experiment 2 we dig deeper, and we go to the level of applications used, number of touches users carried out on the screen, and others. We start to present experiment 1 full procedure and results followed by the same for experiment 2.

6.1 Experiment 1: Data Preparation

In this first experiment, we evaluated the collected intimacy on high-level smartphone usage variables: *session_duration*, *number_of_sessions*, and *session_kind* (details in Section 6.1.2). We used the data we collected in US1. For this experiment, we used the ESM intimacy answers, the smartphone screen events *ON*, *PRESENT*, *OFF*, and the apps the participants were using.

Since we collected intimacy through ESM questionnaires appearing randomly but uniformly during the day, we computed smartphone usage features within a time frame in which the user answered to an ESM questionnaire. This operation required the preparation of the data for analysis.

6.1.1 Usage Session and Usage Window

To create usage statistics for each user, we performed two main steps: (1) we aggregated the screen events and applications used in *usage sessions*, and (2) aggregated usage sessions in *usage windows*, defined by time thresholds (Figure 25). To each *window* (and therefore to all the contained *sessions*) we associated the intimacy value that the user submitted to the ESM questionnaire.

We delimit a *usage session* by the screen events *ON* and *OFF*. In this interval, we logged the name of the apps currently used with a sampling rate of 10 seconds (enabling us to capture micro usage, as shown by Ferreira *et al.* [52]). For each *usage session,* we noted its duration in seconds. Due to reasons bound to the implementation of the logger, we could guarantee this sampling rate only for the 20 participants from Switzerland (out of 42 study participants), thus, we only refer to these users for this analysis.

We define a *usage window* (Figure 25) as an aggregation of consecutive usage sessions that are not separated by a *timeout longer* than *X* minutes. One or many *usage sessions* compose a *usage window*. To study different usage window aggregations, we repeated the analysis with *timeout* = {2.5, 5, 7.5, 10, 15} minutes.

The majority of usage sessions lasts less than 250 seconds (4.2 minutes). After this threshold, we have a very long tail of very few longer sessions, so we decided to consider only usage sessions with a maximum length of 600s (10 minutes) [19], [53], [54]. This procedure prevents the creation of outliers in further steps of our analysis.

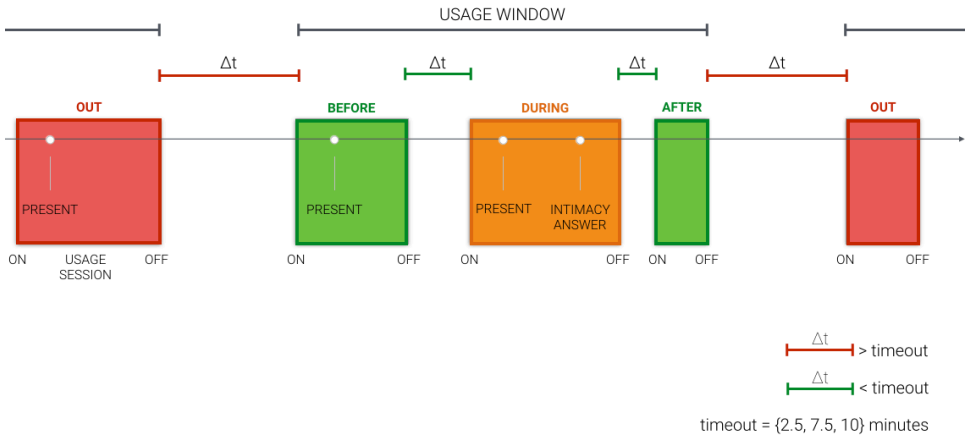

Figure 25: An example of usage window containing the DURING session (including the user answer to the ESM - beep), the BEFORE, and AFTER sessions. The usage sessions labeled as OUT are in usage windows without ESM answer.

6.1.2 High-Level Smartphone Usage Features

From the above-described base aggregation of *usage sessions* in *usage windows* we extracted three usage features:

1. *session_duration*, *i.e.*, the duration of a single interaction with the smartphone;
2. *number_of_sessions*, *i.e.*, the number of interactions within a given time window;
3. *session_kind*, describing whether users were glancing at the smartphone screen or actively using one or more applications.

Since each user has her/his way to interact with the phone [19], the features are *unique to each user.*

To define the first feature, *session_duration*, we grouped the session duration in three bins: ‘short’, ‘medium’, ‘long’. We leveraged the distribution of all the session durations and found the breaks for the three classes employing a quantile method. At the end of the procedure, each *usage session* has a label corresponding to its *session_duration* bin.

For the *number_of_sessions* per window, we grouped the counts of usage sessions per usage windows in three bins: ‘few’, ‘several’, ‘many’. As for *session_duration*, we leveraged the distribution of the variable. In this case, each *usage window* has one of the *number_of_sessions* bins as a label.

For the *session_kind* feature, we considered the kind of interaction of users with their smartphones, using Banovic *et al.* [2] approach. We considered all the sessions from each user and divided these sessions into three categories: ‘lock screen only’ sessions, ‘launcher only’ sessions, and ‘applications and phone’ sessions. We then computed the distribution of session durations of the three categories. By finding the best separations between these three distributions with a maximum likelihood approach, we split the session duration into 3 bins. We called them: ‘glance’ (sessions in which the users look at the phone lock

screen), 'review' (sessions in which the users unlock the phone and mainly interact with the home screen), and 'engage' (in which the users unlock the phone and use one or several apps or perform/receive a phone call) [2].

6.1.3 Aggregating ESM Beeps to High-Level Smartphone Usage Features

Finally, we assigned the intimacy z-scores to the corresponding *usage sessions*. We first located the *usage window* containing a beep answer. We labeled the *usage session* of the located *usage window* in which the users provided their input as *DURING* (Figure 25) while we labeled the sessions in the same *usage window* executed before and after the ESM answer, as *BEFORE* and *AFTER* (Figure 25). Then, we associated the intimacy z-score with all the *sessions* of the located *usage window*. We did not consider all the sessions outside labeled *usage windows* (denoted as *OUT* in Figure 25). We also excluded from our analysis the *DURING* sessions, since we recorded them when the users were answering a new ESM questionnaire.

In Table 3 we show the number of total *usage window* for all the considered values of a *timeout*. The decrease in the total number of usage sessions is due to the deletion of windows containing one *usage session* of more than 600 seconds (10 min), as described above. In Table 3, we present the window and session durations; for increasing *timeout* values the window duration increases. Moreover, in Table 3 we also show the percentage of intimacy labeled *AFTER* + *BEFORE* sessions (the ratio *between AFTER* and *BEFORE* always remains between 40% and 45%). We excluded from our analysis the windows generated with timeout values 10 and 15 (last two rows in Table 3) since it was unreasonable to associate a single intimacy value to windows having such a long average duration.

	Windows				**Sessions**			
timeout (min)	**Total number**	**# Avg sessions per window**	**Avg duration** (min)	**SD duration** (min)	**Total number**	**Avg duration** (min)	**SD duration** (min)	**Intimacy labeled AFTER + BEFORE** (%)
2.5	18748	25.5	12.4	10	33327	0.89	1.41	7
5	13484	37.6	26.3	19.3	31669	0.85	1.40	13
7.5	10716	48.8	43.8	33.3	29982	0.80	1.32	20
10	8786	61.4	73.3	59.7	28449	0.76	1.28	27
15	6555	84.9	151.7	135.2	25284	0.74	1.26	36

Table 3: Extracted phone usage data statistics about usage windows and usage sessions.

6.2 Experiment 1: Results

For this analysis, we used the same techniques applied before Linear Mixed Model (LMM) [51], and Bonferroni corrected post hoc tests. Here we performed tests for each value of *timeout* (as explained above it defines the start and end of *usage windows* with values: 2.5, 5, 7.5 minutes). We explain how all the three smartphone usage variables *session_duration*, *number_of_sessions*, and *session_kind* significantly relate to intimacy.

The *session_duration* relates significantly to intimacy. As presented in Table 4 (LMM column) the effect is significant for all the values of the *timeout*, except for '2.5' min. All the 20 users present all the cases for the *session_duration* variable for each *timeout*: *i.e.*, there are 20 'short', 20 'medium', and 20 'long' sessions.

The post hoc tests (Figure 26 and Table 4) show that when users are in **lower intimacy** (higher z-score), they tend to have **shorter sessions than when in higher intimacy** (lower z-score). Thus, users interact for a shorter time with their smartphones when their perception of intimacy is lower and for a longer time when their intimacy is higher. For the categories 'short' and 'long' the LMM estimated means intimacy of opposite sign for all the *timeout* values. The difference is significant for '5', and '7.5' values of *timeout*. For the 'medium' category, there is a change of sign with the increase of the *timeout* value. The bigger the *timeout* value is, the higher the number of sessions in the window is (Table 3). The 'medium' category contains a mix of session durations that are not clearly belonging to 'short' or 'long' categories. For values between '5' and '7.5' the mean *intimacy* of 'medium' changes sign, indicating a different balance of (short) sessions with low *intimacy*, and (long) sessions with high *intimacy*.

timeout (min)	**LMM results**	**short vs. medium**	**short vs. long**	**medium vs. long**
2.5	F(2, 43.808) = 2.17, p = 0.126	0.04, p = 1	0.18, p = 0.138	0.14, p = 0.481
5	F(2, 40.581) = 5.26, ***p* = 0.009**	0.09, p = 0.363	0.23*, ***p* = 0.011**	0.14, p = 0.244
7.5	F(2, 42.094) = 7.79, ***p* = 0.001**	0.11, p = 0.940	0.23*, ***p* = 0.002**	0.18, p = 0.242

Table 4: LMM analysis and pairwise comparisons results for *session_duration* for each *timeout*.

*The mean difference is significant at the 0.05 level with Bonferroni correction included.

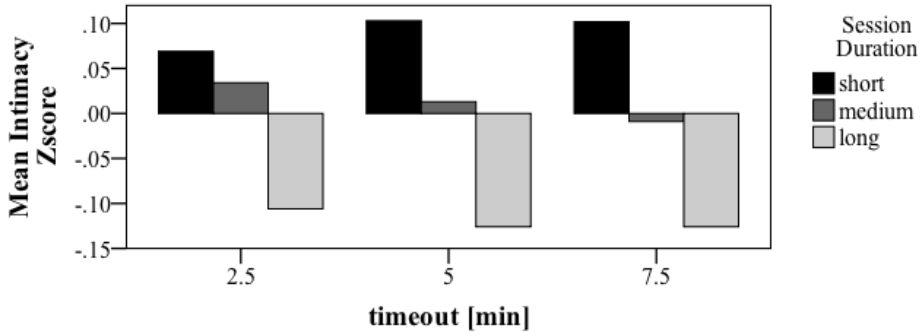

Figure 26: Estimated mean z-score of context intimacy for the short, medium, and long *session_duration* for each *timeout*.

The *number_of_sessions* relates significantly to intimacy. In all the *timeout* cases the effect is significant (Table 5, LMM column), that means that *number_of_sessions* are related to the intimacy level. Not all the users had all the three bins for each minimum distance: '2.5' min = 6 few, 20 several, 20 many; '5' min = 12 few, 20 several, 19 many; '7.5' min = 14 few, 19 several, 19 many.

The post hoc tests (Figure 27, Table 5) show that users tend to have **a higher number of interactions when in low intimacy** (higher z-score) and **a lower number of interactions when in higher intimacy** (lower z-scores). Apart from the '2.5' value, all the other cases present a significant difference between the extreme categories 'few' and 'many', indicating that the intimacy plays a role in the number of consecutive usage session users perform. We observe a similar pattern for '2.5', '5', and '7.5' *timeouts* between the categories 'several' and 'many'.

timeout (min)	**LMM results**	**few vs. several**	**few vs. many**	**several vs. many**
2.5	F(2, 9.110) = 15.06, ***p* = 0.001**	0.14, *p* = 1	-0.41, *p* = 0.442	-0.55*, ***p* < 0.001**
5	F(2, 23.137) = 9.27, ***p* = 0.002**	-0.25, *p* = 0.238	-0.46*, ***p* = 0.008**	-0.21*, ***p* = 0.036**
7.5	F(2, 28.738) = 14.3, ***p* < 0.001**	-0.25, *p* = 0.106	-0.50*, ***p* = 0.001**	-0.25*, ***p* = 0.001**

Table 5: LMM analysis and pairwise comparisons results for *number_of_sessions* for each *timeout*.

*The mean difference is significant at the 0.05 level with Bonferroni correction included.

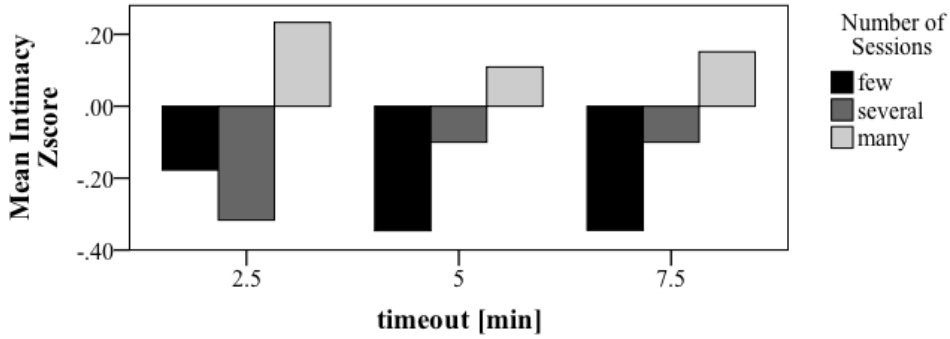

Figure 27: Estimated mean z-score of context intimacy for the few, several, and many *number_of_sessions* for each *timeout*.

Finally, also, the *session_kind* relates significantly to intimacy. The effect of *session_kind* is significant for all the *timeout* except for '2.5' (Table 6, LMM column). All the users present all the cases for this variable for each class: 20 glance, 20 review, and 20 engage.

With post hoc tests (Figure 28, Table 6) we reveal that users tend to **glance quickly at the phone when in lower intimacy** (high z-scores) while they tend to be **engaged when in higher intimacy** (lower z-scores). In lower intimacy, users tend to perform phone activities that require a minimum time of interaction; while they engage more with their devices and perform activities that require longer interactions in higher context intimacy. This finding indicates that the activity carried out on the phone correlates with changes in the users' intimacy perception. Also, in this case, there is a significant intimacy mean difference between the extreme categories 'glance' and 'engage' for '5' and '7.5' *timeouts*. The 'review' category is only significantly different from 'glance' when the mean intimacy value becomes negative (Figure 28, at '7.5').

timeout (min)	**LMM results**	**glance vs. review**	**glance vs. engage**	**review vs. engage**
2.5	F(2, 29.166) = 3.19, *p* = 0.056	-0.28, p = 1	0.16, p = 0.076	0.19, p = 0.290
5	F(2, 31.300) = 8.45, ***p* = 0.001**	0.09, p = 0.560	0.21*, **p = 0.001**	0.12, p = 0.276
7.5	F(2, 32.900) = 8.84, ***p* = 0.001**	0.15*, **p = 0.045**	0.20*, **p = 0.001**	0.05, p = 1

Table 6: LMM analysis and pairwise comparisons results for *session_kind* for each *timeout*.

*The mean difference is significant at the 0.05 level with Bonferroni correction included.

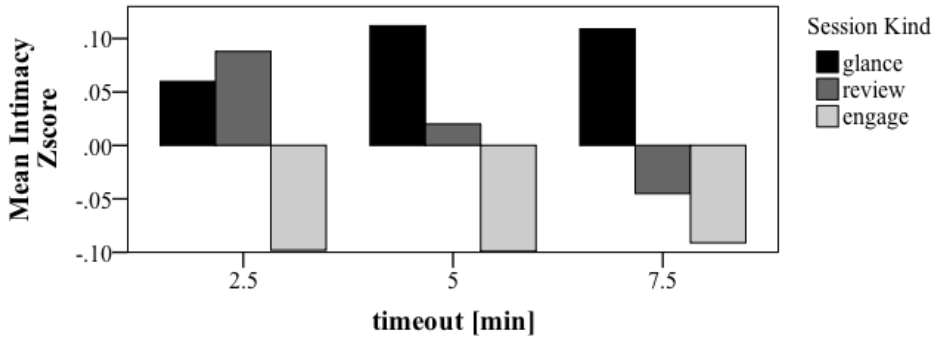

Figure 28: Estimated mean z-score of context intimacy for the glance, review, and engage *session_kind* for each *timeout*.

6.3 Experiment 1: Discussion

Quick smartphone glances can indicate low intimacy and engaging sessions can indicate high intimacy.

From the literature [24], [26], [27], we have hints that *the intimacy perception can influence the behavior of people*. Our results indicate that users have different interaction patterns with their smartphone when their perception of intimacy changes. Users perform shorter, more interrupted (high frequency of sessions), and less engaging tasks in lower intimacy. Vice versa, they perform longer, more continuous (less frequent sessions), and more engaging tasks when in higher intimacy.

These changes in behavior correlate with the subjective perception of the environment (in our case described by three objective contextual elements). When the perception of intimacy is high, users are usually in a place like home, with fewer and usually only close people. This environment facilitates the users (1) to feel more secure [24], and perform certain smartphone interactions that they would not perform elsewhere (that may require more attention, precision, concentration, and trust). (2) To be less distracted [26], and complete in “a single shot” the intended smartphone interaction and tasks. (3) To be willing to share their experience directly and indirectly with other people present [27], and engage more with their devices.

6.4 Experiment 2: Data Preparation

In this second experiment, we evaluated the intimacy to deeper smartphone usage variables. Also, in this case, we used the data we collected in US1, but we conducted our analysis more in a hierarchical way. For this experiment, we used the ESM intimacy answers, the smartphone screen events *ON*, *PRESENT*, *OFF*, the apps the participants were using, and the screen touches they were performing.

We started with the analysis of screen (*ON* and *OFF*) and presence events (*PRESENT*). Then, we analyzed applications at their category level (we categorized application used using the same categories of Google Play store: Lifestyle, Social, Productivity and others). Finally, we picked the category, *Communication*, with the highest number of different and most used applications. Screen touches were included transversally in all the steps.

We created three main categories in which we analyzed these variables separately:

1. PRESENT-OFF transition: *hour*, *day*, *interval*, *app_switched*, *touches*, and *m_touches*;
2. Top 20 applications used: *hour*, *day*, *interval*, *application*, *category*, *touches*, and *m_touches*;
3. Communication apps category: *hour*, *day*, *interval*, *sub-category*, *touches*, and *m_touches*.

To generate the variables for the analysis we performed six main steps. (1) We identified valid screen events transitions. (2) From each transition, we extracted the variables just cited above. (3) We focused on the top 20 applications and also extracted the needed variables. (4) We assigned the intimacy to variables records from ESM answers. (5) We identified which was the category of applications that users used the most and had intimacy attached. (6) We created the intimacy models using the extracted variables for each of the three variables categories.

6.4.1 Valid Transitions

In the first analysis step, we defined which were the valid transitions in which the smartphone can be following screen and presence events. Therefore, we defined the state machine depicted in Figure 29 where we present the transitions that we considered for our analysis: ON-OFF, ON-PRESENT, PRESENT-OFF, and OFF-ON.

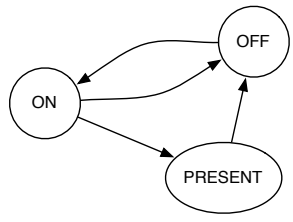

Figure 29: State machine that represents valid transitions between the *screen* and *presence* events.

6.4.2 Applications Switches and Screen Touches

In a second step, for each single valid transition identified in the entire dataset (*i.e.*, all participants together) we derived (as we present in Figure 30): (a) its *day* of the week and *hour* of the day, (b) the transition duration (state-to-state time *interval*), (c) how many applications were switched in between *PRESENT* and *OFF* states (*app_switched*), (d) the number of touches (*touches*), (e) the median of the intervals between touches (*m_touches*).

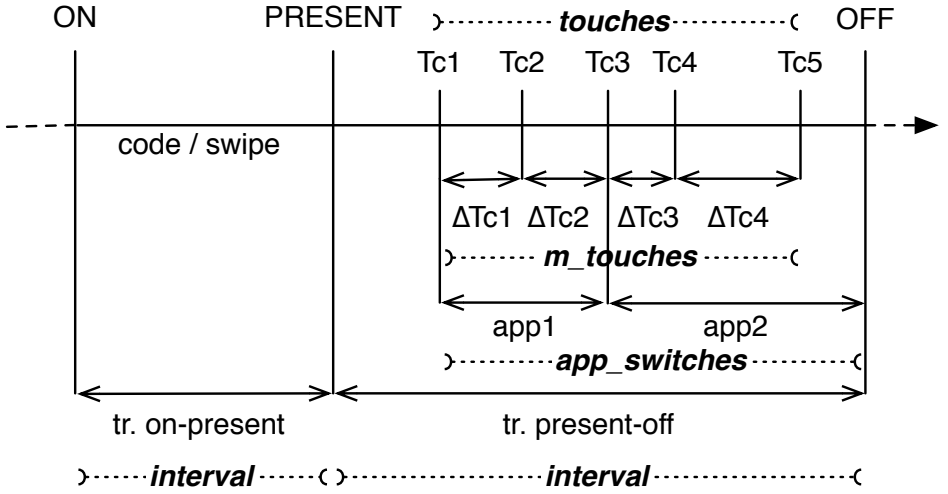

Figure 30: The transition model presents the variables considered in our research, especially in the PRESENT-OFF transition.

As a third step we analyzed each single applications used and for each of them we noted: (a) *day* of the week and *hour* of the day, (b) the usage time (*interval*), (c) the *application* id, (d) the application *category* (derived from Google Play store), (e) the number of *touches*, (f) the median of the intervals between touches (*m_touches*).

6.4.3 Intimacy ESM and Intimacy Models

The fourth step consisted of the assignment of the intimacy level (as the participants answering the ESM survey declared) to each transition and applications used. We considered an intimacy state valid for a window of 15 minutes. 7.5 minutes before and 7.5 minutes after the ESM answer time. As we established in the previous experiment, a reasonable time in which the intimacy state would remain constant, and a time interval in which we found most of the significant results. We assigned the current intimacy to all the measurements of transitions and applications used falling inside these 15 minutes window. When there was no intimacy state specified, we marked the data with an “NA” intimacy level.

In the fifth step, we performed a quantitative analysis of the applications used. We removed all the records without an intimacy level (the ‘NA’ ones), and we identified which were the most used categories of applications. We selected the *Communication* category having the highest number of different applications *used*. For this particular category we extracted the following variables: (a) *day* of

the week and *hour* of the day, (b) the usage time (*interval*), (c) the *sub-category* (we split further the *Communication* category into *Browser*, *Email*, *Messaging*, and *Phone*), (d) the number of *touches*, (e) the median of the intervals between touches (*m_touches*)

Finally, with all this information we created several models to observe the evolution of the intimacy under different combinations of the variables we depicted in the previous steps. We are going to present the most significant model results together with basic statistics for the full data set of the results (Section 6.5). However, before the results we need to explain the data cleaning procedure that we performed before the modeling phase.

6.4.4 Data Cleaning

Before cleaning the data, for screen and presence events transitions, we had a total of 18048 ON-OFF, 19625 ON-PRESENT, 19357 PRESENT-OFF and, and 36923 OFF-ON for a total of 93953 transitions. We removed the invalid transitions (*i.e.*, not following the transition model Figure 29) resulting from some data loss, as follows: 271 OFF-OFF, 491 OFF-PRESENT, 153 ON-ON, 758 PRESENT-ON, and 1290 PRESENT-PRESENT, accounting for a total of 2963 (3.15%) discarded records. As expected ON-PRESENT and PRESENT-OFF transitions are almost symmetric and OFF-ON cover nearly all other transitions starting with the *ON* event.

The number of transitions that we assigned to an intimacy state is 12181 (14.6% of the total), and the probability distribution is: 36.9% *completely* [intimate], 28.8% *yes*, 7.3% *more yes than no*, 6.6% *more no than yes*, 14.3% *no*, 6.1% *not at all* [intimate]. In Table 7 we present the statistics for the variables we extracted for each transition, before data cleaning, (as explained in Section 6.4) taking into account all the participants and transitions we tagged with their intimacy.

	Max				Mean				Std			
Var. / Tran.	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Interval [min]	5.6	7.7	481.2	656	0.2	0.06	1.9	3.3	0.3	0.2	10.3	18.1
app_switched	0	0	19	0	0	0	1.96	0	0	0	1.9	0
Touches	9	7	1335	0	0.1	0.04	33.5	0	0.7	0.3	70.3	0
m_touches [s]	33.1	55.1	54.3	0	4.24	4.44	1.45	0	6.5	11.6	1.7	0

Table 7: Basic statistics for transitions subset, A=ON-OFF, B=ON-PRESENT, C=PRESENT-OFF, D=OFF-ON.

Statistics for the applications used are as follows. In total, we have identified 326 different applications (over a total of 35 categories), but for the further analysis, we retained only the applications that users used at least 50 times (in the study) and only when intimacy state ground truth was available. Therefore, we are left with a total of 24 (7%) applications, with an intimacy states probability distribution for these, as follows: 41.2% *completely* [intimate], 27.9% *yes*, 7.2% *more yes than no*, 6.9% *more no than yes*, 12.4% *no*, 4.4% *not at all* [intimate]. The applications selection process leads to 7 (20%) different categories over a

total of 35 categories. In Table 8 we present a summary of the variables selected for the application used before the data cleaning.

Variable	Min	Max	Mean	Std
interval [min]	0	3006.7	7.6	62.8
Touches	0	1251	14.9	55.1
m_touches [min]	0	226.3	0.08	2.7

Table 8: Basic statistics for the application used.

Furthermore, the *Communication* category has the highest number of applications used. In this category we have a total of 11 applications that we further separated in 4 sub-categories: *Browser* (3 apps), *Email* (3 apps), *Phone* (1 app), and *Messaging* (4 apps). Most of the cleaning performed in these discrete variables relates to the frequency of the various categories. To avoid to deal with the many very rare events and categories we decided to focus on those that were very representative.

To avoid the influence of outliers in temporal variables (*i.e.*, resulting in long transition times), we cleaned the dataset from extreme values for the subsets' variables. We further defined bounds of variables in which we want to model our data. Data was removed either by considering the variable's distribution (*i.e.*, cutting the "long tails" by fixing a maximum value at the third quartile of the data) or by common sense (*i.e.*, removing the unreasonably long time intervals for some variables).

We considered only the records that we recorded from 5h in the morning to midnight (removed 5.2% of data). As we show in Figure 31, the intimacy level ground truth are infrequent overnight.

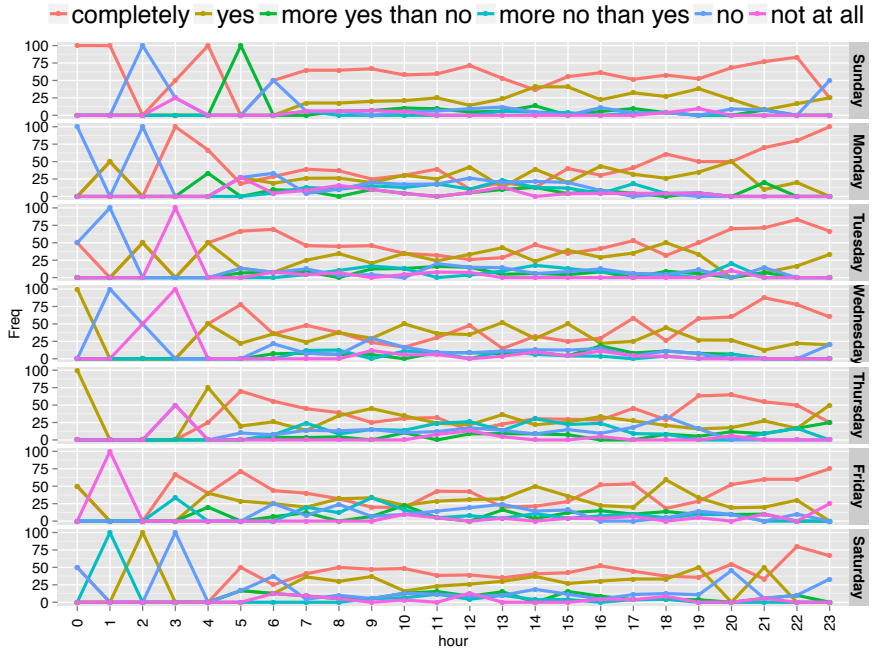

Figure 31: Frequency for each intimacy state for each hour and day.

For screen and presence events (*ON*, *OFF*, and *PRESENT*) transitions, we removed all the data with any of the followings characteristics. (1) *Intervals* longer than 180 seconds (8.8% of data meeting this condition). (2) More than ten *app_switches* in the same session (0.1% of data meeting this condition). (3) The *m_touches* out of the third quartile, values less or equal 1.72 seconds (6.5% of data meeting this condition). (4) The *m_touches* with NA value due to 0 or 1 touch in the interval considered (26.2% of data meeting this condition). In total, we have removed 28.4% of the data (meeting one or several of the above conditions). Also, we subset the data to deal mainly with the transition PRESENT-OFF (in the transitions ON-OFF, ON-PRESENT, and OFF-ON there was no particular interaction of the users with the smartphone).

Then, for the applications used, we removed all the records with any of the following characteristics. (1) We removed usage *intervals* longer than 180 seconds (16.8% of data meeting this condition). (2) We removed all the applications not belonging to the categories of *Communication*, *Social*, and *System* as minor contributors in the data (25.4% of data meeting this condition, with the extra fact that we also removed our logger application records). (3) The *m_touches* out of the third quartile, values higher or equal 1.87 seconds (16.3% of data meeting this condition). (4) The *m_touches* with NA value due to 0 or 1 touch for the app entry considered (65.2% of data meeting this condition), for a total of 75.4% removed data (meeting one or several of the above conditions). Finally, for the last data subset, given that the *Communication* category is a subset of the cleaned applications used, we did not need any further manipulation.

6.5 Experiment 2: Results

The analysis consisted of the brute force generation of all the possible models defined by all the possible combinations without repetitions of the subset of variables. We combined the model variables using the additive method only (*e.g.*, with three variables: intimacy=var1+var2+var3). We processed these models' definition with the Ordinal Regression Model (ORM) approach [55], because of our ordinal intimacy scale, from 1="most intimate" to 6="least intimate". Then, for each subset, we selected the most significant model (smallest χ^2 *test* p-value) using the *ANOVA's* χ^2 *test* [55] between each generated model and the baseline model. The baseline model is represented only by the intimacy threshold coefficients (*i.e.*, 1|2, 2|3, 3|4, 4|5, 5|6) without any model variables coefficient. This kind of model, it is a model without description parameters, just based on the "raw" distribution of intimacy states. Also, we analyzed how all the models composed by a single variable related to intimacy (*i.e.*, evaluating their χ^2 *test* p-value and confidence intervals).

We created the most significant models with 60% of randomly sampled data from the full data set, and we tested them against the remaining 40%. We generated the model and test data set 10 times (denoted in machine learning as 10x Cross Validation). When possible, we plotted the predictions probabilities for each intimacy level for the different model variables, and we obtained similar results from all the ten different trials.

We present the intimacy models in the following sections divided into three broad categories: (1) PRESENT-OFF transaction models, (2) applications used models, and (3) *Communication* category models, but first we start with some considerations about intimacy over time.

6.5.1 Intimacy in Time

Two variables in common for screen and presence events transitions and applications used that relate to the time of interaction with the smartphone are the day of the week and the hour of the day. As showed above, in Figure 31 we plot the probabilities for each intimacy state (from 'completely' to 'not at all', in different colors) for each day of a week (separate graphs, Sun-Sat) and hour (X axis). From the chart, we can note how the intimacy state ground truth does not cover the period from midnight to 5h (not many ESM responses from the participants). This fact is partially due to our random ESM events. We were issuing them only in waking hours (from 8h to 22h), and usually, people sleep at this time of the night. The surveys in the intervals from 5h to 8h and from 22h to midnight originated whenever the phone was un/plugged from the charger, and some are late answers to earlier-triggered notifications in the interval 8-22h. From Figure 31 we conclude that the general trend for study participants is to be more intimate in the early morning and the evening and, additionally Sunday seems to be the most intimate day of a week. The least intimate hour appears to be the ones around noon on a weekdays and Saturday. These results confirm the output of the feasibility study we presented in Chapter 3 and particularly the results we presented in Figure 7.

6.5.2 PRESENT-OFF Transition Intimacy Models

For the data concerning the transitions event PRESENT-OFF we generated models with the combination of the variables: *hour*, *day*, *interval*, *app_switched*, *touches*, and *m_touches*. We obtained a total of 63 models (combinations without repetition of the six variables from this subset, taking 1, 2, 3, 4, 5, 6 variables at time). Out of 63 models, 4 were not significant (p-value > 0.05), 59 were significant (among them 41 had p-values < 0.001). The single variable models contributed in the following way (p-value ordered from most to least contributing): $p=2.4*10^{-5}$ variable *hour* (rank 28), $p=2.3*10^{-3}$ for *day* (rank 44), $p=2.4*10^{-3}$ for *app_switched* (rank 45), $p=2.3*10^{-2}$ for *interval* (rank 57), $p=1*10^{-1}$ for *touches* (rank 61, not significant), and $p=3.7*10^{-1}$ *m_touches* (rank 63, not significant).

The model *hour+day+app_switched* is the most significant (denoted as *Most_Sign_Tr*). It has a p-value < $2.3*10^{-8}$, and a condition number of the Hessian (cond.H) = $1.7*10^{4}$ (measures if the model is ill defined; cond.H > 10^{6} [55], indicates that the model can be simplified), and maximum model gradient (max.grad) = $6.38*10^{-13}$ (a value indicates if the model converges: usually for value max.grad < 10^{-6} [55])).

In Table 9 we present the summary statistics of the variables for the whole data set based on which we defined the model and tested it to obtain the predictions (we divided the data per intimacy state).

		Intimacy					
Variable	**Statistic**	**1**	**2**	**3**	**4**	**5**	**6**
hour [5h-23h]	min	5	5	6	6	5	5
	max	23	23	23	21	22	23
	mean	13.7	13.9	14.2	12.4	12.2	12.9
	std	4.8	4.2	4.2	3.5	4.3	4.2
day [0-6]	min	0	0	0	0	0	0
	max	6	6	6	6	6	6
	mean	2.8	3	3.7	3	3.1	2.8
	std	2	1.9	1.9	1.7	1.8	1.9
app_switched	min	0	0	0	0	0	0
	max	9	9	6	6	7	7
	mean	1.7	1.8	1.5	1.8	1.6	1.5
	std	1.2	1.2	1	1.2	1.3	1.5

Table 9: Basic statistic of *Most_Sign_Tr* model data after cleaning.

In Table 10 we present how the single variables contribute to the *Most_Sign_Tr* model and their confidence intervals. From the table, we can observe the confidence intervals, and we can conclude that the most likely values of the variables are in between a small range.

Variable	**P-value (*chisq*)**	**CI 2.5%**	**CI 97.5%**
hour	$8.9*10^{-6}$ ***	-0.058	-0.023
day	0.001 **	0.025	0.107
app_switched	0.001 **	-0.180	-0.044

Table 10: Most_Sign_Tr model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '), confidence intervals (CI).

In Figure 32 we present one of the plots of the probabilities for each intimacy level as resulted from the *hour+day+app_switched* model prediction (total subset size 1957 records, data to define the model 1174 (59.9%) and testing size 783 (40.1%)). In Figure 32, we can note how the probability to be *completely intimate* increases with the hour of the day. It changes depending on the day of the week (in particular *Sunday* is more intimate than the rest of the week, as in Figure 32) and has different behaviors depending on how many applications users switched in the PRESENT-OFF transition. The more applications users switches, the higher is the probability to be *completely intimate*.

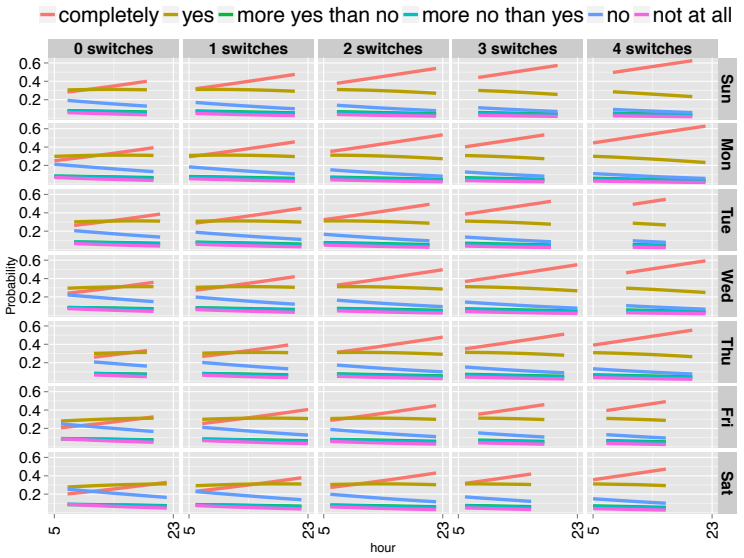

Figure 32: Probability (left Y axis) for each intimacy state depending on the *hour* (bottom X axis), *day* (right Y axis), and *app_switches* (top Y axis).

6.5.3 Applications Used Intimacy Models

For the applications used we generated models with the combination of the variables: *hour*, *day*, *interval*, *application*, *category*, *touches*, and *m_touches*, we obtained a total of 127 models. Out of them one was not significant (p-value > 0.05), 126 were significant (among them 118 p-values < 0.001). The models composed by a single variable contributed in the following way (p-value ordered from the most to least contributing): $p=1.4*10^{-14}$ for *application* (rank 63), $p=8.5*10^{-5}$ *touches* (rank 107), $p=1.5*10^{-4}$ *hour* (rank 110), $p=2.1*10^{-4}$ *m_touches* (rank 111), $p=1.4*10^{-3}$ *day* (rank 122), $p=7.6*10^{-3}$ *interval* (rank 125), $p=2.2*10^{-1}$ *category* (rank 127, not significant). As before, we present details of the most significant model *hour+day+app+touches* (denoted as *Most_Sign_App*) with its p-value < $2.9*10^{-19}$, cond.H = $5.4*10^{6}$, max.grad = $1.95*10^{-7}$.

In Table 11 we present how the single variables contribute to the *Most_Sign_App* model and their confidence intervals. Also, in this case, we have variables that are significant for the models, in particular, *application* and *hour*. We omitted *application* CI; we would need to list results for 20 applications. Differently, from the other models, we are not going to plot the results of predictions for this case. Due to the model size (4 variables) is difficult to plot these results, needing 4 + 1 (probability) dimensions.

Variable	P-value (*chisq*)	CI 2.5%	CI 97.5%
hour	0.005 **	-0.041	-0.007
day	0.011 *	0.012	0.090
application [20 apps]	$1*10^{-9}$ ***	omitted	omitted
touches	0.034 *	-0.004	-0.000

Table 11: *Most_Sign_App* model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ') and confidence intervals (CI).

6.5.4 Communication Category Intimacy Models

For the *Communication* category, we generated models with the combination of the variables: *hour*, *day*, *interval*, *sub-category*, *touches*, and *m_touches*. We obtained a total of 63 models. Out of them 2 were not significant (p-value > 0.05), 61 were significant (among them 51 p-values < 0.001). The models composed by a single variable contributed in this way (p-value ordered from the most to least contributing): $p=3.3*10^{-4}$ for *sub-category* variable (rank 42), $p=3.3*10^{-4}$ *m_touches* (rank 43), $p=4.5*10^{-4}$ *touches* (rank 45), $p=2.4*10^{-2}$ *interval* (rank 59), $p=2.4*10^{-2}$ *day* (rank 60), $p=3.2*10^{-1}$ *hour* (rank 63, not significant). As before, we present details of the most significant model *day +sub_category+touches* (denoted as *Most_Sign_Com*) with its p-value < $8.1*10^{-6}$, cond.H = $4.8*10^{5}$, and max.grad = $1.91*10^{-12}$.

In Table 12 we provide a summary of statistics for the two variables of the model for the data from which we sampled the model definition and testing data.

		Intimacy State					
Variable	**Statistics**	**1**	**2**	**3**	**4**	**5**	**6**
day	min	0	0	0	0	0	0
	max	6	6	6	6	6	6
	mean	2.9	3.2	3.9	2.8	3.2	2.9
	std	2	1.8	1.9	1.7	1.9	2.6
sub_cat.	frequency(Browser)	54	45	7	9	36	18
	frequency(Email)	43	25	10	5	17	4
	frequency(Mess.)	372	285	57	57	107	29
	frequency(Phone)	19	9	5	5	4	0
touches	min	2	2	2	2	2	2
	max	865	584	223	313	336	112
	mean	53.7	48.9	39.9	52.2	36	33.5
	std	76.1	68.3	46.9	60.2	44.4	25.6

Table 12: Basic statistics of *Most_Sign_Com* model data after cleaning.

In Table 13 we present how the single variables contribute to the *Most_Sign_Com* model and their confidence intervals. In this case for the *Most_Sign_Com* model we have the *day* variable, presenting not significant contribution to the model. Instead, *sub_category* and *touches* are significantly contributing to the model.

Variable	**P-value (*chisq*)**	**CI 2.5%**	**CI 97.5%**
day	0.06 .	-0.003	0.103
sub_cat [Email]	0.002 **	-0.959	-0.050
sub_cat [Messaging]	0.002 **	-0.851	-0.226
sub_cat [Phone]	0.002 **	-1.667	-0.342
touches	0.003 **	0.430	1.174

Table 13: *Most_Sign_Com* model variables significance (0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ') and confidence intervals (CI).

We present the predictions of the *Most_Sign_Com* model in Figure 33 (total subset size is 1213 records, model definition size 728 (60%) and testing size 485 (40%)).

The *Messaging sub_category* is the one related to the most of the touches per its usage session, and particularly on *Sunday*. *Email* and *Phone* have little interaction, *i.e.*, very few touches per session. *Browser* has a regular short number of *touches* across the week. The intimacy level change with the number of *touches* (particularly in *Messaging*, but also in *Browser*). In *Messaging*, the *completely* intimacy level probability increases by increasing the number of touches. Instead, in *Browser* with very few *touches,* we have a higher probability for the *no* intimate level that slightly decreases when the *touches* increase (*Saturday* may indicate a general trend for the week).

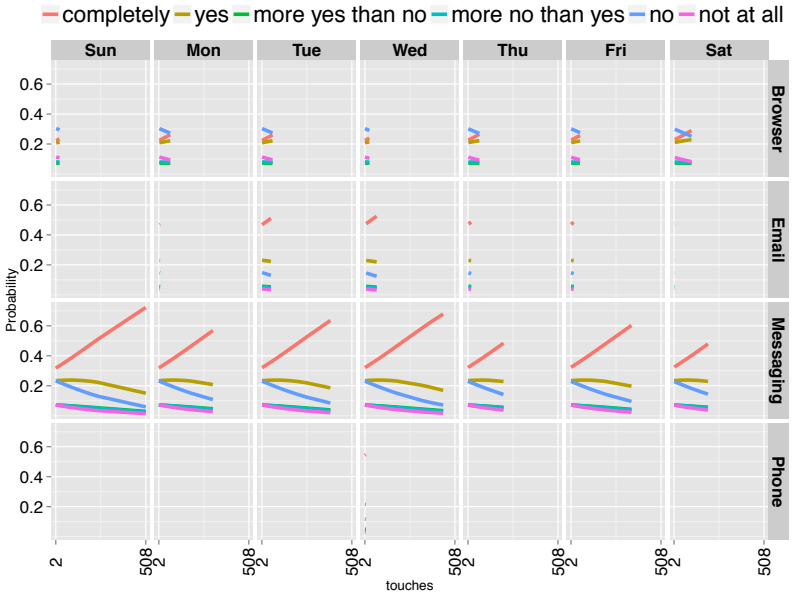

Figure 33: Probability (left Y axis) for each intimacy state (colored lines) depending on *touches* (bottom X axis), *sub_category* (right Y-axis), and a *day* (top Y axis).

6.6 Experiment 2: Discussion

Based on the results, we conclude that there are some differences on how users interact with their smartphones depending on their intimacy. Effects of these interaction changes are mostly visible at the extreme intimacy levels *completely* and *no*. Users are switching more applications when in a high intimacy (more engagement), writing shorter messages when in lower intimacy (quick operations and glances), and so on.

Time variables are correlated to intimacy, but how we look at the smartphone data activity can influence this correlation.

The *hour* and *day* variables are contributing to all the most significant models, from which we conclude that this time variable is relevant to identify different user's intimacy patterns. *Hour* and *day* are very significant in the PRESENT-OFF *transition* as a single variable in models of intimacy (high rank for single models variable). However, if we look at the *application used* set, time variables

lose the significance, and they become even less relevant for the *Communication category* where *hour* alone is not even significant. We may conclude that probably time variables also relate to smartphone usage variables, as the *hour* of the day and the *day* of the week also influence the usage of the smartphone. They become usage variables themselves. An extra note on these variables is that they are related and most probably they could be treated as an input to the models as dependent variables, instead of just addictive model terms, as we have done. These could be verified with further ANOVA tests on such model definition compared to the one we have already performed.

A high number of applications switches in the same interaction session can indicate high intimacy.

For the **PRESENT-OFF transition** model the only variables that alone are not significant to derive a single variable model for this set of data are *touches* and *m_touches* that relate to each other. Namely, there are no particular changes on the number of touches or the interval between them for different intimacy levels along the PRESENT-OFF interaction with the smartphone. Also, as the *interval* variable alone is not so powerful, we conclude that the interaction time of a single PRESENT-OFF interaction is not an excellent indicator of the user's intimacy state (from Experiment 1, Section 6.2, we know that we need several PRESENT-OFF interactions to understand the user engagement).

Furthermore, the *app_switched* variable changes depending on the intimacy. In particular, as reported in Figure 32, the higher the number of switches, the higher is the probability of being in the *completely* intimate state and less in the *no* intimate state. Interaction with the phone without switching applications (*i.e.*, when the user lands straight at screen application after the OS fires the PRESENT event) can indicate that the user may not be in an intimate situation (from Experiment 1, Section 6.2, the user may be doing quick repeated interaction with the same app). Instead, after four applications switches the probability to be intimate is higher (from Experiment 1, Section 6.2, the users may be more engaged with their smartphone in this case).

Additionally, the *day* variable is mostly contributing to the differences between weekday and weekends, as on Sunday users tend to be more intimate (see Figure 32 and Figure 33). Finally, the *hour* is the single variable most significant for the intimacy level given our set of data. When one uses the phone in the morning, he/she tends to be less intimate than when using it in the evening (Figure 31).

Some applications are mostly used in high intimacy and the number of touches when using an application increases when in high intimacy.

For **applications used** model the application's *category* and *interval* variables are not significant for the intimacy state. Also, *day* and *hour* variables, although they are present in the most significant models do not seem to contribute so much, as well as *m_touches*.

The application is the most significant variable. Depending on which *application* the user used we have a different probability of being on the *completely* intimate level. We can divide the 20 applications we identified in two groups: (a) in which the likelihood of being *completely* intimate state is high accounting 14 of them (2 Email apps, 4 Messaging apps, 3 App Launcher, 1 Browser app (but with small amount of data available), 1 Contacts app, 1 Phone app, 1 Settings app, and 1 Social Contact app (app to meet people)); (b) in which the intimacy level probabilities are equivalent (*i.e.*, users use such applications in any state) accounting 6 apps (2 Browser apps, 1 Email app, 1 System app, 1 App Market (official play store), 1 Social Network (Facebook)). For the moment, we did not investigate further these differences, and we limit ourselves to acknowledge that some applications are used mostly when in high intimacy and some in any conditions. Finally, *touches* are contributing as follows: increasing number of *touches* increases the *completely* intimate probability.

Long interactions with communication applications and more touches can indicate high intimacy.

For the last subset of data defining the **category communication** model, the temporal variables *hour*, *day* and *interval* are the least significant ones, probably because the more equally distributed use of *Communication* applications across time reduces the temporal effect. *Touches* and *m_touches* relate to how the user interacts with the phone. In particular, in this data subset, the number of *touches* depends on the *sub-category* variable. This *sub-category* variable combined with the *touches* and the *day* creates the most significant model for this subset. From Figure 33 we can notice that for the *Messaging* (4 apps) sub-category, the longer are the messages or the conversation time (more *touches* to write longer messages or longer threads), the higher is the probability that the user is intimate. We can rationalize this finding as follows. We interact or text longer messages when we feel more comfortable and secure in the environment we are at that moment (*i.e.* at home). From *Browser* (3 apps) we see that we tend to navigate with the mobile browsers when we are not intimate. If we do that with more *touches,* it probably means, that we are more intimate (*e.g.*, we interact more by clicking on several links, or we do not use bookmarks, but we type the website fully). This assumption cannot be fully confirmed from our data, because the *Browser* interaction, in term of *touches*, is quite short.

6.7 Conclusions

From our experiments, we can conclude that there is a relation between how users use their smartphone and their perceived intimacy. Both experiments showed that intimacy has a transversal influence from high-level variables of interaction down to the analysis of single communication applications usage. Understand intimacy can help to provide to users mobile applications and services that are even more intelligent.

Given the potential of intimacy, we can define the following (but not limited to) design implications for pervasive systems. (1) Service providers can notify the users (*e.g.*, via a notification on the smartphone) about promotions, particular

events or other non-real time notifications when the users are more likely to spend more time on their device once it captures their attention (*i.e.*, in a *high intimacy*). (2) Particular applications can adapt their interface to offer easy tasks or shortcuts when users are in *low intimacy*. (3) Some services may reduce their intrusiveness when the users are in *high intimacy* (*i.e.*, context elements are more familiar and valuable to the users), for example avoiding the collection of some personal context information that may be more privacy sensitive in this context. (4) The smartphone itself may adapt its functionalities to never lock the device or change the way notifications are delivered (*e.g.*, sound, vibration and light) in *high intimacy*. (5) Application designers may use intimacy to analyze different unexplored users' behaviors.

Proven that intimacy has a role in the whole users' context and we can exploit it for mobile applications and services provisioning, we now focus on two main tasks that we present in the next chapters:

1. Actionable intimacy: in Chapter 7 we study a way to predict the intimacy of the users from information that we can collect from users smartphones, data available from US1 (Section 4.2.1);
2. Leveraging intimacy: in Chapter 8 we apply intimacy in the context of mobile data collection campaigns and see if this context can be a factor that influences the willingness of participants to the campaigns to share their data anonymously or not. It is a study defined following the point 3 above where we presented possible design implications of intimacy).

7. Predictive Modeling of Intimacy

We proved that intimacy is a contextual variable predictive for smartphone usage in different intimacy conditions. In this chapter, we exploit the knowledge acquired by the previous data analysis to present a machine learning approachable to predict the perceived intimacy of users. The objective is to predict intimacy to make it available as further contextual information when users are using their smartphones. Although tempting we do not base our intimacy prediction directly on smartphone usage variables. It would be convenient, but not practical. We cannot depend on smartphone usage if we want to predict automatically intimacy because we would need to wait that users perform their usage sessions to detect then their intimacy. Discovering intimacy at such a late stage would not be beneficial.

We proved that the perception of intimacy depends on context elements such as the place, the number and the kind of people around the user. Researchers are investigating techniques to define the number of people around a user, but as of today, there are no accurate systems able to provide a reliable estimation. At this moment, most of these techniques require an infrastructure (*i.e.*, instrumenting user's context) that we are not able to fulfill or are resource-demanding [5]–[7], [13]. Identifying, who are the people around an individual, is even more difficult from a ubiquitous computation perspective. Therefore, even if we have the data for 'number of people' and 'kind of people' for each participant in our studies, we have decided not to use these in the machine learning approach, but instead rely only on users' location data.

We divided the procedure into 5 phases: (1) data preparation, (2) features extraction, (3) transformation of intimacy variable from Classes to Ranks, (4) generation of data sets for classification, (5) testing and validating the intimacy model via three experiments: (i) per subject model via 10 cross validation (10CV) train-test, (ii) create a model per each user and test them against the data of the other users, and (iii) evaluation of accuracy of the prediction model with US2.

7.1 Data Preparation

To exploit the location data, during our user study we collected smartphone connected mobile network's cell IDs. This cell ID is the 'cheapest' location-related variable we can obtain on a user's smartphone. We recorded these cell IDs with a sampling rate of one minute. The cell ID variable is the aggregation of the cell ID of the connected tower with the Location Area Code (LAC the identifier of the geographical area in which the tower stands in the country), the Mobile Network Code (MNC the identifier of the network provider in the country), and the Mobile Country Code (MCC). There are unique cell IDs across the two users samples of US and EU (all 42 users). To further define the users' locations, we grouped cell IDs in *neighborhoods.*

7.1.1 Creating Neighborhoods of Cell IDs

To cluster the cell IDs into meaningful neighborhoods, we employed a two-step process based on Fanourakis and Wac algorithm [56]. The first step determined

if two consecutive cell ID measurements are physically near each other by analyzing the cell oscillations that often occur in cellular communications due to smartphone/cell power change, cell load balancing or some other operator-specific policies. We define a cell oscillation event between two cells, A and B, as any event where a smartphone is in a fixed location, and it connects to cell A then to cell B and eventually back to cell A in a relatively short time. However, due to the relatively low sampling rate of our logger data, we must make an assumption: the connection cannot 'oscillate' between two neighborhoods in the span of four minutes (*i.e.* the user does not oscillate between two significant neighborhoods in the span of four minutes). Thus, we guarantee that any cell oscillation event of time span less than four minutes occurs while the user was in the same significant neighborhood. We do not take into account any cell oscillation occurring over a longer time span, which could mean that the user left the neighborhood, was mobile and got back to the same neighborhood. Thus, we form a graph where the cells are the nodes, and a cell oscillation event (edge) links the two cells involved in the oscillation. The second step in determining the cell ID neighborhoods is to find the *cliques* of the specific cell ID graph that we created in the first step. So, we define each neighborhood of cell IDs as each maximal clique of the graph where a cell oscillation event connects all the possible pairs of cell IDs.

It is important to note that a cell ID can belong to multiple neighborhoods. For this reason, to assign a neighborhood to each cell ID in a sequence, we must choose the neighborhood that best matches the recent cell IDs in the sequence. To do this, we assigned weights to each cell ID in each neighborhood based on a factor related to the number of occurrences of that cell ID globally and specifically in that neighborhood. We also assigned weights to the cell IDs in the sequence based on how recently they appeared. To assign a neighborhood to a cell ID in the sequence, we then compared each weighted neighborhood cell ID with the weighted sequence cell ID and derived the most likely neighborhood for that measurement. Based on these neighborhood assignments we calculated statistics such as the mean continuous time spent, minimum time spent, maximum time spent, and variance of the time expended in each neighborhood. Due to this procedure and technical problems on the retrieval of CellIDs from users smartphone, we could not include six users out of the 42 in the machine learning approach.

7.1.2 The Neighborhoods Cell IDs Data

Each user has his list of neighborhoods, and we defined each neighborhood with its ID, the mean, variance, min, max time spent in that neighborhood, its frequency, and its size. Each cell ID is part of a neighborhood. By mapping the cell ID to its neighborhood, we can know which was the neighborhood at that given time for each user. Therefore instead of cell ID per minute we assign each user his/her neighborhood ID per minute. Finally, we normalized all the statistics listed above across all the users by scaling them between 0 and 1 (we divided each variable by its own maximum). This operation allowed us to extract location related features from the statistics of neighbors to estimate the perception of the intimacy of each user.

7.2 Features Extraction

From the users' neighborhood data, we extracted the features for each user independently. We have assigned a neighborhood to each ESM answer. Once we identified the neighborhood we took the current scaled statistics of the neighborhood as features (mean, variance, max time spent in that neighborhood, its frequency, and its size). We also selected three more features from the statistics of the previous neighborhood (the neighborhood in which the user was before the current one), namely: mean, variance of time spent in that neighborhood, and its frequency. To summarize, now for each user there are in total eight features for each ESM intimacy answer (left column of Table 14).

Original Continuous Feature	**Binary Binned Features**
mean_time	mean_time_bin1 {(-inf - 0.27] or (0.27 - inf)} mean_time_bin2 {(-inf - 0.733] or (0.733 - inf)} mean_time_bin3 {(-inf - 0.976] or (0.976 - inf)}
var_time	var_time_bin1 {(-inf - 0.2585] or (0.2585 - inf)} var_time_bin2 {(-inf - 0.9375] or (0.9375 - inf)}
max_time	max_time_bin1 {(-inf - 0.4205] or (0.4205 - inf)} max_time_bin2 {(-inf - 0.9535] or (0.9535 - inf)}
frequency	frequency_bin1 {(-inf - 0.23] or (0.23 - inf)} frequency_bin2 {(-inf - 0.969] or (0.969 - inf)}
size	size_bin1 {(-inf - 0.2795] or (0.2795 - inf)} size_bin2 {(-inf - 0.801] or (0.801 - inf)} size_bin3 {(-inf - 0.9615] or (0.9615 - inf)}
prev_mean_time	prev_mean_time_bin1 {(-inf - 0.2465] or (0.2465 - inf)} prev_mean_time_bin2 {(-inf - 0.709] or (0.709 - inf)} prev_mean_time_bin3 {(-inf - 0.976] or (0.976 - inf)}
prev_var_time	prev_var_time_bin1 {(-inf - 0.155] or (0.155 - inf)} prev_var_time_bin2 {(-inf - 0.9375] or (0.9375 - inf)}
prev_frequency	prev_frequency_bin1 {(-inf - 0.1725] or (0.1725 - inf)} prev_frequency_bin2 {(-inf - 0.969] or (0.969-inf)}

Table 14: Original continuous features and their corresponding binary binned features.

Since we envisioned the model for the intimacy prediction to work on a tree approach, we discretized the features in bins. Some tree algorithms can deal with continuous variables, but an adequate discretization can help to obtain more accurate results. Our goal is to produce models that are representative of the possible different group of users in our dataset. Therefore, we derived the bins for each feature by considering all the users together. To create the bins we employed the unsupervised discretization filter (in Weka), automatically computing an adequate number of bins for each feature. Furthermore, since we have planned to use a tree-based approach and we know that most of the tree algorithms are not able to preserve the bin ordering [57] (p. 315). Therefore, to maintain the meaning of the order of bins, we set up the filter to transform each discretized attribute into a set of binary attributes. As described in [57] (p. 315) that we quote: "if the discretized attribute has k values, it is transformed into k − 1 binary attributes. If the original attribute's value is i for a particular instance, the first i − 1 of these new attributes are set to false and the remainders are set to

true. In other words, the (i − 1)th binary attribute represents whether the discretized attribute is less than i." Therefore we transformed the initial 8 continuous features into 11 discrete features: mean_time_bin{1,2,3}, var_time_bin{1,2}, max_time_bin{1,2}, frequency_bin{1,2}, size_bin{1,2,3}, prev_mean_time_bin{1,2,3}, prev_var_time_bin{1,2}, and prev_frequency_bin{1,2}. In Table 14, we summarize the original continuous features and their binned counterparts. Each continuous variable corresponds to different intervals and number of bins for a final total of 19 features.

7.3 Transforming Intimacy ESM Data: From Classes to Ranks

In [58] Martinez *et al.* showed that a ranking approach is more efficient when dealing with ordered and subjective data (i.e., on a scale, as intimacy from 1 to 6). Therefore, we focused our work on the Ranking by Pairwise Comparison (RPC) ([59] Hüllermeier *et al.*) classification approach. The RPC classification approach allows to classify ranks, and it fits our needs accordingly to [59]. These ranks can be preferences (e.g., A is preferred over B) or elements of an ordinal scale. The intimacy values are in fact part of an ordinal scale. This procedure allows mapping users answers to an ordered preference scale, with at the first place the choice made by the user when answering the intimacy question. For RPC, the ordered scales or preferences are expressed as A>B, which means that A is preferred over B for a given instance.

We created the following mapping between the single class value of intimacy to the complete ordered preference of the scale: '1' = '1>2>5>6', '2' = '2>1>5>6', '5' = '5>6>2>1', '6' = '6>5>2>1'. We omitted the ratings '3' and '4'. We did that for two reasons. (1) It is hard to define a clear ranking in the case of the ratings '3' and '4' (e.g., which is the rank preferred for '3'? '3>4>5>6>2>1' or '3>2>1>4>5>6', etc.). (2) In [58] Martinez *et al.* suggested to remove the neutral values and '3' and '4' are representing those in our data. We then replaced the classes with these ranks in all the instances. For the other rankings the rationale is to preserve as much as possible the order and the meaning of the intimacy states. For instance, we preferred to map the level '6' of intimacy (not intimate at all) to '6>5>2>1' rather than to '6>5>1>2' to add weight to the natural order. Prefering to be 'completely intimate' over 'more intimate than not' when we have 'not intimate at all' as base, would go against the whole reasoning.

We present here in Figure 34 and Figure 35, the distribution of the *intimacy* class (*i.e.,* ratings of ESM answers) and the ones for the derived rankings. The rankings represent the same distribution of the respective rating classes.

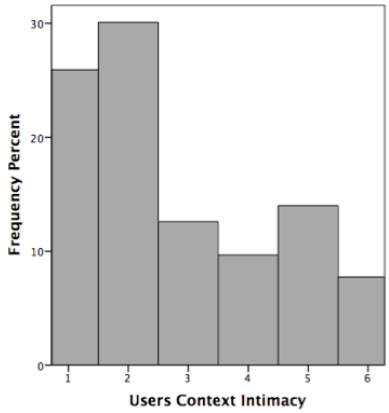

Figure 34: Distribution of the intimacy states over all the users.

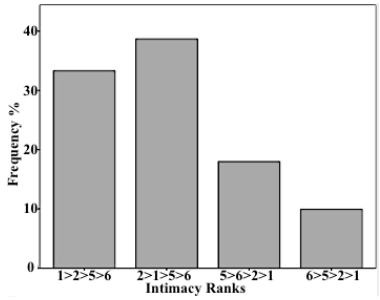

Figure 35: Distribution of intimacy ranks after transformation from intimacy scores.

7.4 Generating Data Sets for RPC Classification

Martinez *et al.* in [58] also explain some techniques to reduce the subjective noise due to user errors on answering ESM survey questions or changes in users habits and attitude during the experiment. In particular, they propose to establish a window of instances that we compare with each other inside the data of the same user. This approach allows to find cases in which the instances are similar (features values comparison), but finally, they present different or even opposite ranks. Therefore, we chose to create 6 different datasets: full, 'population_clean', 'user_clean_5', 'user_clean_10', 'user_clean_20', and 'user_clean_50'. The first dataset, denoted as full, is the full dataset as it is. For the second dataset, denoted as population_clean, we cleaned the data at population level considering all the user instances together. In this case, we analyzed the frequency of rankings represented by similar instances. We grouped the instances depending on the similarities of their features and kept counts of the frequency of the four rankings ('1>2>5>6', '2>1>5>6', '5>6>2>1', '6>5>2>1'). For each group of instances we summed up the counts of the two most similar ranks ('1>2>5>6' + '2>1>5>6' and '5>6>2>1' + '6>5>2>1'). We deleted from that group (and therefore from the full dataset) all the instances having the minor sum.

For the last four datasets, we cleaned instances at the user level with different window *Y* sizes for future instance look up, respectively *Y* = {5, 10, 20, 50} (we

denote these datasets as *user_clean_Y*). For these datasets, we checked for noise in each user dataset independently using a moving window of *Y* instances. We cleaned out instances using the same principle adopted for the *population_clean* dataset we presented above.

In Table 15, we show the percentage of the removed instances for each resulting dataset and their distribution per rank. In total, from 21474 instances from all datasets we removed 1539 (7.2%) of instances (removed distribution per rank: '1>2>5>6' = 12.2%, '2>1>5>6' = 27.2%, '5>6>2>1' = 45.2%, '6>5>2>1' = 15.4%). We removed a small part of the data, with the maximum excluded rate in the population_clean dataset (14.5%).

Dataset	Kept	Removed	Removed distribution for ranks '1>2>5>6'	'2>1>5>6'	'5>6>2>1'	'6>5>2>1'
full	3579 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
population_clean	3059 (85.5%)	520 (14.5%)	7 (1.3%)	61 (11.7%)	299 (57.5%)	153 (29.5%)
user_clean_5	3376 (94.3%)	203 (5.7%)	46 (22.7%)	78 (38.4%)	67 (33%)	12 (5.9%)
user_clean_10	3348 (93.5%)	231 (6.5%)	47 (20.3%)	78 (33.8%)	89 (38.5%)	17 (7.4%)
user_clean_20	3309 (92.5%)	270 (7.5%)	45 (16.6%)	93 (34.5%)	109 (40.4%)	23 (8.5%)
user_clean_50	3264 (91.2%)	315 (8.8%)	42 (13.4%)	108 (34.3%)	132 (41.9%)	33 (10.4%)

Table 15: The datasets generated from the six different cleaning techniques with their percentages of kept and removed instances.

7.5 Testing and Validating the Intimacy Model

To test the potential of the RPC-based intimacy prediction model we performed three experiments: (1) per subject modeling, (2) subject model vs. all the other subjects instances, and (3) the model in practice with a user study in the wild. Each experiment contributes to the model refinement. We explain their goals, detailed configuration, results, and we discuss what is their contribution to the overall research objective. In each experiment, the accuracy of the model is measured by Kendall's tau rank correlation ranging from -1 and 1 [59]. If the Kendall's tau is closer to -1 the ranks predicted by the model are inverted with respect to the expected ones, instead if the Kendall's tau is closer to 1 the rank outputted are the ones expected (high-quality result). A value of Kendall's tau around 0 means a bad classification (no correlation between ground truth rankings and predicted ones).

7.5.1 Experiment 1: Per Subject Modeling

The goals of this first experiment are: (a) evaluate the validity of our assumptions about the relation of time and place with the perception of intimacy for a significant number of users of our study; (b) understand the effect of the different datasets used for training on the accuracy of the prediction per each subject.

The configuration of the experiments is as follows. For each user independently and each dataset, we performed ten runs (ten different random seeds) of 10CV of RPC with ZeroR algorithm as an internal classifier (always predicting the majority class) with binary voting for the ranking. We use this setup to establish the baseline accuracy of our model for each user. We compare the first set up with ten runs (ten different random seeds) of 10 cross validation (10CV) of the RPC algorithm using the J48 tree algorithm as an internal classifier with binary voting for the ranking.

The RPC J48 classifier performs significantly better than RPC ZeroR (baseline). The best dataset is the population_clean set. In Figure 36 we present the Kendall's tau means across all the users for the datasets and the two RPC classifiers, ZeroR, and J48. We performed two-way repeated measurements ANOVA to verify if the reported results are significant. We report: (1) the difference between the baseline classifier RPC ZeroR and RPC J48 is significant $F(1, 33) = 14.22$, $p = 0.001$, $r = 0.28$; (2) the dataset effect on accuracy of the model (mean Kendall's tau) is significant $F(1.4, 46.4) = 8.32$, $p = 0.003$, $r = 0.45$ (Greenhouse-Geisser correction); (3) The population_clean is the most predictive dataset, but is significantly better only over the full dataset ($p = 0.006$).

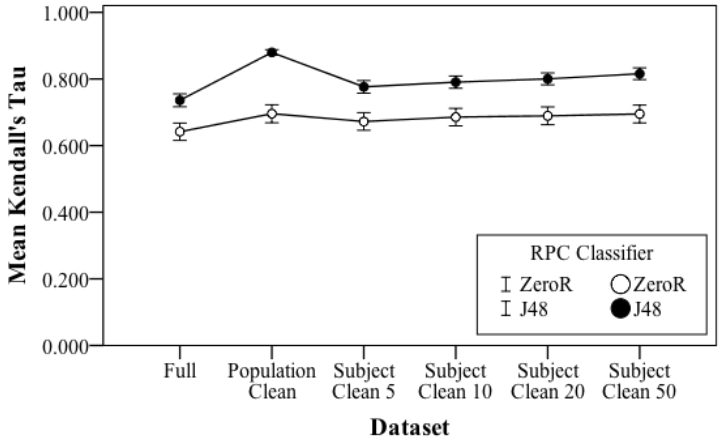

Figure 36: Mean Kendall's Tau across all the users for the datasets and the two RPC classifiers, ZeroR (baseline) and J48 for the 10CV test of experiment 1 [error bars 95% CI].

Focusing on the population_clean dataset, in Figure 37 we show the mean Kendall's tau of the ten 10CV runs of RPC ZeroR and RPC J48 for each user for the population_clean dataset. 25 users (76%, out of 33) using that datasets show an improvement over the baseline, 5 users (15%) have no improvement over the baseline, and 3 users have worse results (9%). The overall mean Kendall's tau for RPC ZeroR is 0.7032 and for RPC J48 is 0.8817 (an increase in "accuracy" of ~20%).

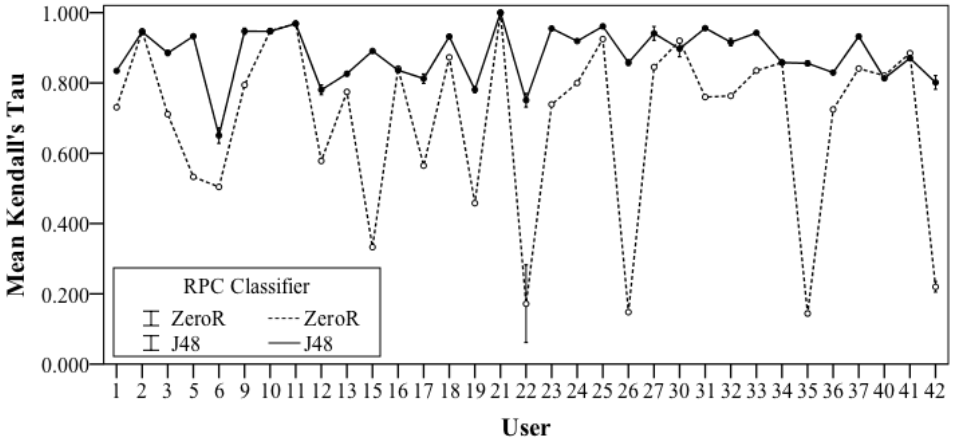

Figure 37: Mean Kendall's tau of the ten 10CV runs of RPC ZeroR (baseline) and RPC J48 for each user for the population_clean dataset [error bars 95% CI].

We can conclude that RPC J48 is significantly better than RPC ZeroR (baseline) and that the full dataset is the least suitable to generate a reliable intimacy model. Although the population_clean dataset is the best, we cannot exclude that the other cleaning method can lead to good results in practice.

7.5.2 Experiment 2: Per Subject Modeling vs. Other Subject Data

The goals of the second experiment are: (a) identify clusters of users to investigate if it is necessary to create multiple models depending on some users particularities; (b) verify if different datasets generate substantial differences on the assumption (a).

The configuration of the experiments is as follows. For each user independently and each dataset, we based the intimacy model on his/her data and tested this model on the sets of data of all the other users (one by one). We constructed all the users models using experiment one RPC J48 with binary vote configuration.

In Figure 38 we present how all the users behaved against each single models of other users across all the datasets. We can observe the presence of two clusters: a majority of users have a high average performance (higher Kendall's tau mean), and some present a low accuracy (low Kendall's tau mean).

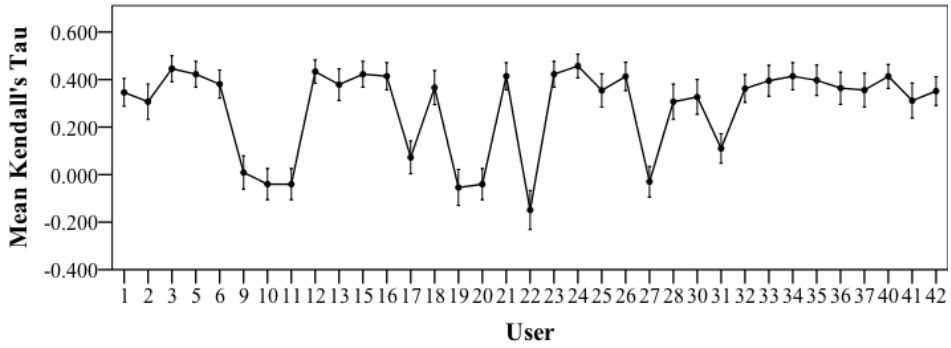

Figure 38: The mean Kendall's Tau for each user tested against the models of other users over all the datasets [error bars 95% CI].

In Figure 39 we show the performance of each user but separated between datasets. We do this to highlight that the *population_clean* dataset is still the one achieving the highest accuracy, and all the other datasets are resulting in the same accuracies. We expected this outcome since the *population_clean* dataset generation required the harmonization of all the users' data (cleaning of outliers at the population level). However, in any dataset, we can still note that we have the presence of the two clusters of users mentioned above.

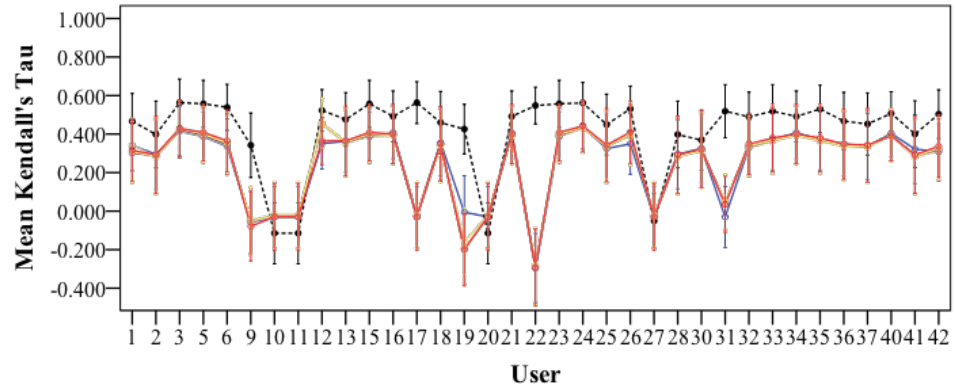

Figure 39: The mean Kendall's Tau for each user tested against the models of other users for each dataset. The dashed black is the population_clean dataset that performs better than the others [error bars 95% CI].

From Figure 40 we can notice that the *population_clean* dataset is performing significantly better than any other dataset. To confirm this fact we executed a one-way repeated measurements ANOVA and we report the following results: (1) the accuracy of the intimacy model is significantly affected by the dataset we used for the prediction $F(1.10, 37.35) = 25.98$, $p < 0.001$, $r = 0.06$ (Greenhouse-Geisser correction); (2) Bonferroni post hoc tests revealed that only the *population_clean* dataset is significantly better than all the other datasets ($p < 0.001$ for all). Focusing on the best dataset, *population_clean*, we report that the overall mean Kendall's tau for RPC ZeroR is 0.3503 and for RPC J48 is 0.4215 (an increase in "accuracy" of ~17%). This last result compared with the previous experiment (RPC ZeroR = 0.7032 and RPC J48 = 0.8817) shows how subjects have different intimacy base model and that the accuracy of the prediction

reduces greatly when models are constructed with data from different subjects instead of users' own data.

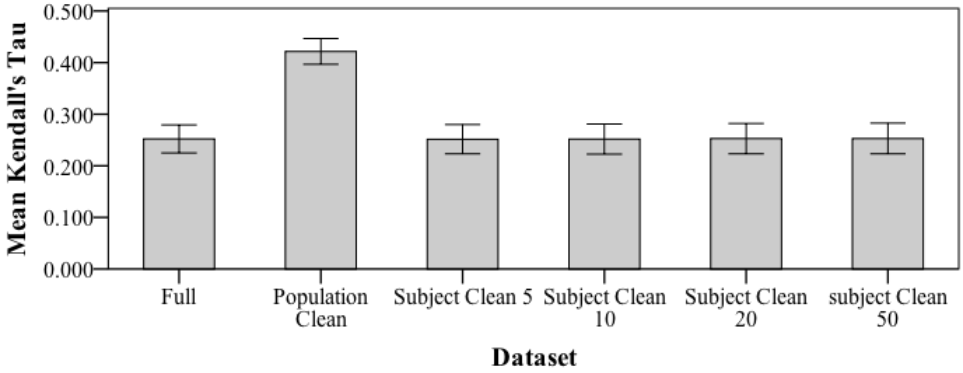

Figure 40: The mean Kendall's Tau for all the datasets over all the run of users against other users model. The dataset population_clean is significantly better than the others [error bars 95% CI].

To identify the clusters of users, for all the dataset independently we run a two-step clustering algorithm with automatic selection of the number of clusters. In Table 16 we present the clustering results. We have very similar results for all the datasets except for *population_clean* that being the best dataset presents different mean Kendall's tau centroids and less equilibrium on the number of users in the two clusters (~70% of users in the well-classified cluster, higher mean Kendall's tau). For each dataset, we have an average cluster silhouette of 0.7 (except population_clean 0.8), which means that the cluster quality is high, and the clusters of users are well separated. We can conclude that there exist, two groups of users, which indicates that we need to create at least two different models for each dataset. The dataset *population_clean* is again the significant best, but it showed as well the presence of two groups of users, and some of them are affecting its results negatively as for the other datasets.

Dataset	**# Clusters**	**Clusters Distribution** (users per cluster)	**Clusters Centroids** (mean Kendall's Tau)	**Cluster Quality** (average Silhouette)
full	2	18 (51.4%) 17 (48.6%)	0.38 / 0.12	0.7
population_clean	2	24 (68.6%) 11 (31.4%)	0.53 / 0.19	0.8
subject_clean_5	2	19 (54.3%) 16 (45.7%)	0.38 / 0.10	0.7
subject_clean_10	2	18 (51.4%) 17 (48.6%)	0.39 / 0.10	0.7
subject_clean_20	2	21 (60%) 14 (40%)	0.37 / 0.07	0.7
subject_clean_50	2	21 (60%) 14 (40%)	0.37 / 0.07	0.7

Table 16: The separation of users in clusters for each of the six datasets with cluster centroid and the average Silhouette value for cluster quality (> 0.5 represents a good cluster separation).

7.5.3 Experiment 3: User Study 'in the Wild'

The goals of the third experiment are: (a) to establish how accurate is the model in the real world (in the wild); (b) to observe if the models derived from the two clusters result in different accuracy; (c) verify that *population_clean* dataset is the most accurate intimacy model; (d) collect more ground truth to possibly refine further the intimacy model.

We configured the experiment to involve real smartphone users and run the algorithm on their own devices. First, we split all the datasets into the two clusters of users as output from the two-step clustering procedure and created the respective models using Weka. We now have 12 models in total, two per each dataset. Second, we created an Android OS application implementing the neighborhood algorithm we presented in the previous sections and the prediction logic. Third, we created a series of notifications with which the users could provide their inputs and get informed about the predictions of the intimacy model. These notifications are like ESM, and in this context they are denoted as Ecological Momentary Assessment (EMA); they are providing the ground truth data.

In Figure 41 we show the EMA notifications users were receiving on their smartphones. For the first week of the study, we displayed to users the notification A (Figure 41). They were not required to do anything. We used the first days to collect data about the users' routine and generate the statistics for the neighborhoods at the base of our approach. After the first week, we started to predict the intimacy state of users. We performed a prediction every time we detected a change in the user context. Specifically, whenever there was a change of neighborhoods we were predicting the intimacy state with all the 12 models and creating a final prediction using a majority vote. At each prediction, we notified users with notification B (Figure 41). This notification allowed them to provide their intimacy input, from *very high* to *very low*, following the intimacy ranking of our model (1. *very high* > 2. *high* > 5. *low* > 6. *very low*). Once users provided their input, we were telling them the result of our majority vote prediction using notification C (Figure 41, two example of correct predictions). All the notifications provide to users a help button to instruct them about their meaning and tasks to perform. Finally, given the availability of new users' input, we added a module to collect extra data like user activity (*e.g.,* walking, biking, staying still), top WiFi access points, and light in lumen to investigate new possible variables that could help us to refine further the intimacy model. We collected users activity with the Google Services Android API. Each time we were notified by the service of a new activity we were logging it in a file. We collected WiFi access points every minute and the light from the smartphone sensor every time the user was switching on the smartphone screen (thus with a high probability the light value corresponds to the light when users provided their intimacy ground truth).

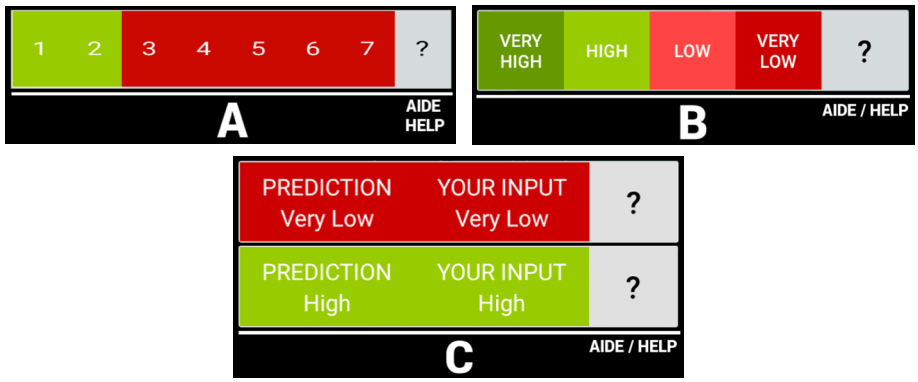

Figure 41: The notification we presented to the users during experiment 3. Notification A is the countdown of the seven days of data acquisition. With notification B users could provide their current intimacy state. We were displaying B each time the intimacy model was computing a new prediction. Notification C was presenting the users input vs. our prediction (in here two examples where our prediction was matching the user intimacy).

We involved a total of 31 users from our mQoL living lab in Geneva, Switzerland. Each of them contributed with one month of participation in the study. We collected a total of 3369 intimacy inputs and performed a total of 60307 predictions (5.6% have ground truth). In Figure 42 we present how each user contributed to the ground truth (EMA answers) and in Figure 43 we show the number of predictions we performed per user. As for the previous user study, the number of contributions is not uniform. Some users are contributing more, some less. In some cases, the contribution is bounded by the small number of predictions. This fact can mean that the user was almost always in the same context, or our neighborhood algorithm was not able to detect changes.

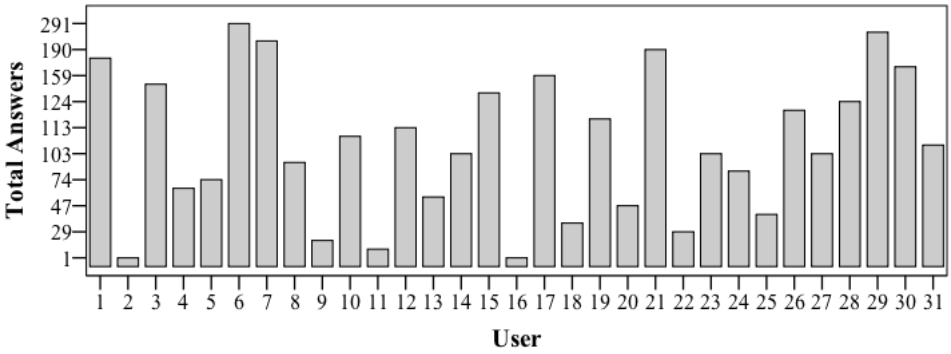

Figure 42: The number of users answers to the notifications of experiment 3. As for the first user study, some users contributed more than others.

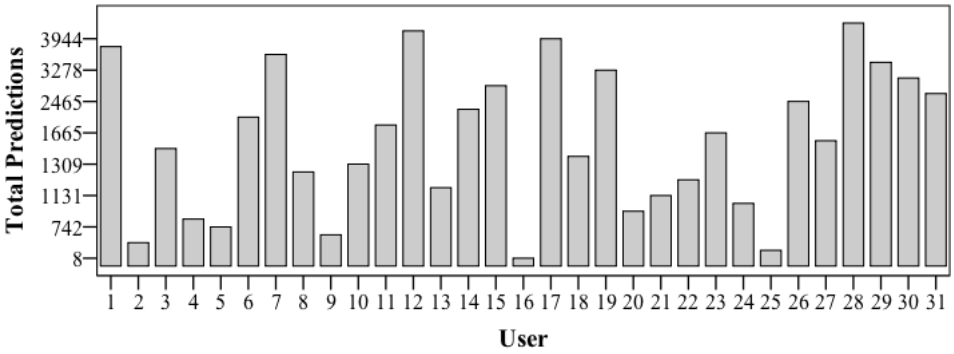

Figure 43: The number of predictions that our model did for each user involved in experiment 3. Some had a higher number of predictions probably due their higher mobility.

The intimacy ranks from participants are distributed as follow: '1>2>5>6' = 1249 (37%), '2>1>5>6' = 770 (23%), '5>6>2>1' = 628 (18.6%), '6>5>2>1' = 722 (21.4%). As the intimacy ranks from our predictions are different for each rank and model, we provide them along the analysis, as follows.

Using the Kendall's tau ranking correlation metric we compared each users' input to the corresponding 12 models outputs and the respective majority-voting outcome. We present the results of this analysis in Figure 44 with a heat map in which we display the mean Kendall's tau across all the users of each model and the outcome of the majority voting. For each model, we have the two variants we created with the data originating in the users' cluster 1 (C1) and cluster 2 (C2), as we presented in experiment 2. For some users, we can notice that we are successfully predicting the intimacy state as in the theoretical model (mean Kendall's tau from 0.4 to 0.6). While for some other users (e.g., users 7, 19, and 26) appositive rankings are observed (mean Kendall's tau from -0.4 to -0.6), and some are around 0. For users with appositive rankings, the explanation is that their intimacy rankings are reversed. Theoretically, a solution could be to inverse these rankings to obtain a better accuracy, but in practice, it is hard to detect these inverted ranks after the classification and react accordingly. The most concerning users are the ones having a mean Kendall's tau around zero (e.g., users 1, 4, and 25). For these, our model accuracy is low; there is no correlation at all between their ranking and the one our model predicts.

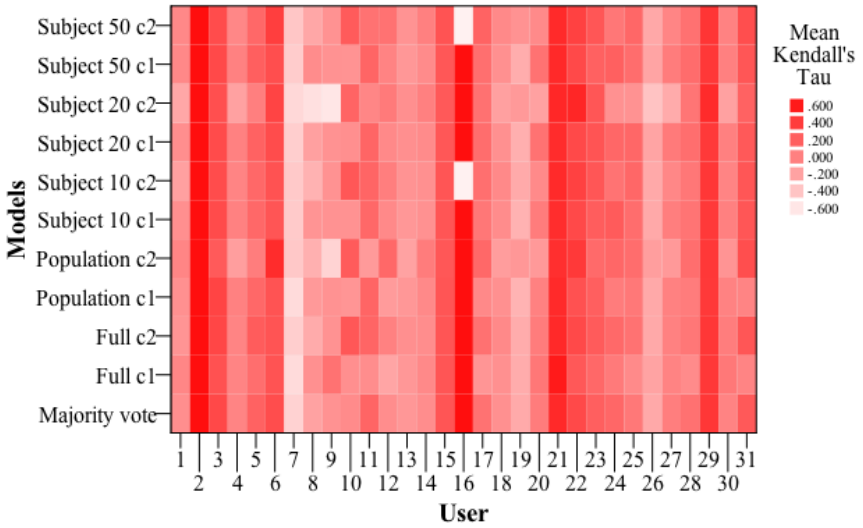

Figure 44: Mean Kendall's Tau heat map for all the 12 datasets (c1 datasets are generated by users cluster one and c2 by users cluster two) and majority vote (from the 12 datasets outcome). A positive mean Kendall's Tau means we correctly predicted the intimacy ranks; a zero means there is no correlation (wrong prediction), and a negative one represents an inverse ranking prediction.

We have investigated the raw data and results and we observe that the main reason behind the Kendall's tau around zero is that our intimacy models predicted the rank '6>5>1>2' (*e.g.*, majority-voting 2607 predictions out of 3369 marked with ground truth) instead of the rank '6>5>2>1' (recalling the expected distribution of 722 (21.4%)). The models were also unable to predict the rank '1>2>5>6' (*e.g.*, majority-voting 0 predictions out of 3369, expected distribution was 1249 (37%)). These two problems are at the origin of a low-rank correlation of users that were mostly in rank '1>2>5>6'. The users that obtained better results are the ones that were primarily in rank '2>1>5>6' (distribution of 770 (23%)). For this rank, we have a distribution for the majority-voting of 761 predictions. A deeper analysis is needed to identify the reasons behind these misclassifications. We comment further on these results in the Discussion section.

As we show in the heat map in Figure 44 there are not significant variations of each model Kendall's tau within each user. To evaluate the different models globally, in Figure 45 we provide the mean Kendall's tau of each of the models across all the users. We did not find a significant mean Kendall's tau difference between the models. We executed a one-way repeated measurements ANOVA: (1) the model variable is not significantly affecting the prediction results $F(2.10, 63.05) = 1.302$, $p = 0.280$ (Greenhouse-Geisser correction); (2) Bonferroni post hoc tests revealed that only the *subject_50_c1* model is significantly better than *population_c1* ($p = 0.035$).

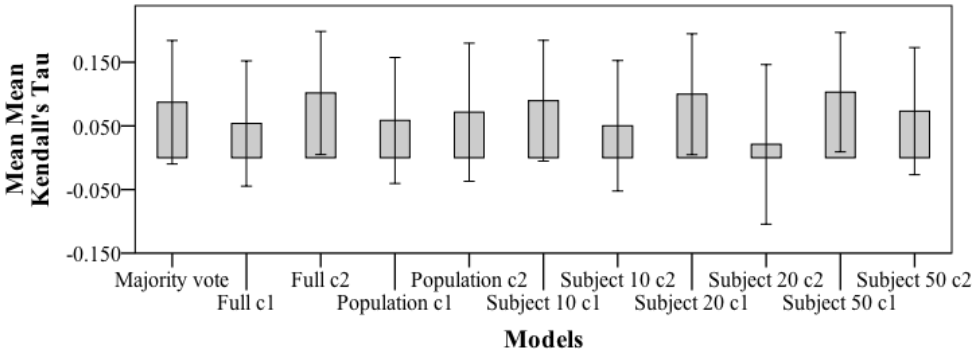

Figure 45: Mean Kendall"s Tau over all the users for all the datasets. No datasets seem to be significantly better than the others [error bars 95% CI].

Among the additional context data that we collected on users' smartphone, we found that user activity, WiFi, and light correlate with intimacy. For each of the three variables, we mapped each user intimacy vote to the closest available state of them. For example for activity, starting from the user vote timestamp we searched for the most recent previous activity state. To analyze activity and WiFi data we performed a count of the possible states of the two variables and the current intimacy state. For activity, in Figure 46 we show a heat map with the percentages of the counts normalized across each user and the four intimacy ranks, for four activities groups: 'vehicle'/'biking', 'still', 'other' (unknown + tilting), and 'on foot'. We can notice that for very high intimacy (1>2>5>6) all the users (apart 1, 2, 18, and 22) tend to be mostly in the 'still' state. The lower the intimacy, the more mobile are the users. In very low intimacy (6>5>2>1) for the majority of the users we have more activities indicating mobility. We performed a Cochran-Mantel-Haenszel Chi-Squared Test for count data and confirmed that these differences are significant: $M^2 = 227.91$, df = 9, $p < 0.001$ (for this test we removed 8 users: 1, 2, 8, 9, 16, 18, 22, and 27, because they did not have sufficient votes for each intimacy category).

Figure 46: The heat map is showing the counts in percent of the users activities in the different intimacy rankings. We normalized the counts per rank and user. Users tend to be less mobile in high intimacy than in low intimacy.

For WiFi, we created four categories as follows. "No_wifi" represents all the cases in which users were not connected to wifi. "First_wifi" is the wifi to which

the users were connected most of the time. “Second_wifi” is the second most used WiFi, and “other” is the collection of the remaining WiFi. In Figure 47, we present a heat map showing the percentage of each wifi category normalized per each user and each intimacy state. The main pattern highlighted in the graph is the high presence of “no_wifi” when users are in a low intimacy (5>6>2>1 and 6>5>2>1). Vice-versa “first_wifi” is particularly high in the very high intimacy state (1>2>5>6). Being connected to the most used WiFi can mean high intimacy and not be connected to WiFi can mean low intimacy. We performed a Cochran-Mantel-Haenszel Chi-Squared Test for count data confirmed that these differences are significant: $M^2 = 363.88$, df = 9, $p < 0.001$ (as before, for this test we removed 8 users: 1, 2, 8, 9, 16, 18, 22, and 27).

Figure 47: The heat map is showing the counts in percent of the users WiFi in the different intimacy rankings. We normalized the counts per rank and user. Users tend to be connected to a frequent WiFi when in high intimacy and not connected to WiFi when in low intimacy.

Finally, for the ‘light’ context variable, we analyzed how the mean normalized light value (lumen) changes in the four intimacy states. For each user, we transformed their absolute light values with z-scores (we set their mean lumen to 0). In Figure 48 we show a heat map with the four intimacy states, and the mean scaled lumen for each user. From the graph, we can see a possible relation between being in low intimacy and having higher mean lumen values. This relation can mean that when users are in high intimacy, they are probably inside where lumen values are lower than outside (when users are more probable to be in low intimacy). A One-Way Repeated ANOVA confirmed this assumption: (1) depending on the intimacy rank we have significantly different mean values of lumen $F(1.79, 39.46) = 4.101$, $p = 0.028$ (Greenhouse-Geisser correction); (2) Bonferroni post hoc tests revealed that when users are in intimacy rank ‘1>2>5>6’ their smartphone light sensors captured a lower lumen level than when in intimacy ranks ‘2>1>5>6’ ($p = 0.024$) and ‘5>6>2>1’ ($p < 0.001$). We found no differences between the other ranks. Also, in this case, we excluded all the users not having light values for all the four intimacy levels.

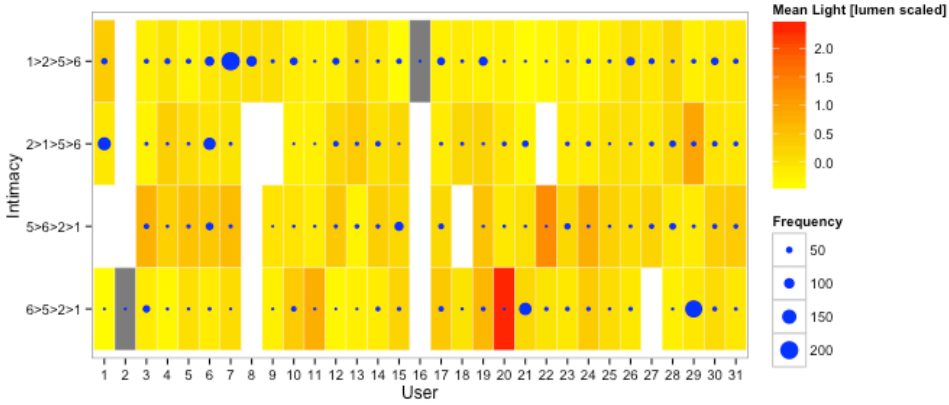

Figure 48: The heat map is showing the mean light (in Lumen) for all the users in the different intimacy rankings. We normalized the light values with z-scores for each user. Light seems to be overall more intense when users are in low intimacy (probably outdoor) than when they are in high intimacy.

7.6 Discussion

The subjective perception of intimacy can be predicted and operationalized in practice.

The results of our machine learning approach show that by leveraging features extracted from smartphone and related only to the spatial/temporal context, specifically related to locations visited by users, it is possible to predict the user's intimacy obtaining significantly better results than the baseline. This accuracy has the potential to be increased further by integrating features like the number of people around the user and possibly the kind of people. It would not be necessary to predict the exact number or the exact kind of people around the user, but it will be required to identify some features able to describe the situation the user is in as a whole. We are aware that several works were performed to study how to detect people around [5]–[7], [13], and it would be certainly useful to derive from them a combination of variables able to describe this phenomenon. Moreover, if we also consider the mood component, valence we may be able to integrate further new features. Additionally, we identified a correlation between intimacy and activity, WiFi, and light that we need to investigate and probably integrate into the prediction model. We hypothesize that adding these features to the location ones will further increase the accuracy of our prediction method.

The practical deployment of the model 'in the wild' showed that it is feasible to implement what we envision theoretically, but many variables need to be refined to make the system fully operational. Particularly, in practice, there is probably no need to create multiple models since no one proved to be significantly more robust than others in practice. Most probably the most accurate solution is to re-generate a single model by repeating the same procedure presented in this paper with the new data acquired in the latter user study. Another aspect to verify is why the models were not predicting the higher intimacy rank accurately ('1>2>5>6'). Our first assumption is that the underlining algorithm for the

neighborhood detection was failing to recognize some places or we had challenges in the computation of statistics for the different locations, the users visited. This last assumption requires further research on the neighborhood algorithm.

Our second assumption is that when we generated the different models we biased them towards predicting low levels of intimacy. The solution can be to balance the instances used to train the future models and produce a more balanced predictor.

7.7 Conclusions

We proved that it is possible to predict intimacy, and the focus of our future work is to refine the intimacy model with a cycle of prototypes and experiments. The final goal is to integrate the intimacy metric in a mobile public transport application offering real time bus departures (currently operational and available on Google Play for the city of Geneva[4]) [60]. The study will consist of 'A/B' tests where half of the user will experience app interface changes when their intimacy changes and the other half will not. We will compare the results gathered in the two groups to identify the effects of the intervention. For example, it will be done by estimating the user experience with the app derived from the app engagement data such as average usage time sessions and length of task execution. We may define other measures to verify the accuracy of the *intimacy* model while users perform usual tasks in the transport application. For example, we may introduce a new feature consisting in the identification of the 2 to 3 most intimate user's places. We will identify the intimate places with our intimacy model. We will then ask the users to label these places with *home*, *work* or a personalized place or to delete them if they are not relevant. This procedure can enable to verify the accuracy of our algorithm in identify intimate places and, therefore, being able to assess the users' high intimacy correctly for those situations. We may then use these places to provide to users several statistics about their movements between these places (e.g., the average duration of the trip depending on the crowd, time of the day and line used). We will provide the design of such study in detail in Chapter 9 where we define our future work. In the next Chapter 8, we leverage intimacy in a more practical case in which we study how users participating in MSC campaigns perceive anonymous data sharing in different intimacy context.

[4] https://play.google.com/store/apps/details?id=ch.unige.tpgcrowd

8. The Role of Intimacy in Anonymous Smartphone Data Collection

Users of mobile devices participate on Mobile Crowd Sensing (MCS) to collaboratively sense different phenomena. In such systems, the data may be collected from them anonymously or in an identifiable form. Users may be or not aware of other participants. The ubiquitous availability of smartphones propels MCS, offering an array of miniaturized sensors and the ability to sense users' environmental context (e.g., location), user's behavior (e.g., physical activity) and interactions with the environment (e.g., social interaction engagement) and the phone itself (e.g., applications used). Such smartphones can collect diverse data continuously, unobtrusively, and in real time.

Scientific researchers, marketing or advertising specialists, and end-user themselves are trying to benefit from the potential of MCS. For example, all these actors of an MCS system may be interested in the end-user location. Scientific researchers may use it to develop the next generation mobility algorithm. Marketing or advertising specialists can employ it to do geo-located advertising. End-users may simply get updates about their nearby bus departures depending on their location. In all the three cases, the problem is that a third party manipulates their location.

It is a well-known fact that the majority of users have difficulties to infer which data is collected, for which purpose, and how. For example, Lin *et al.* [48] and Kelley *et* al. [49] show how is difficult for end-users to understand the data access permissions when they download an application from the Google Play store. This difficulty does not mean that users are not becoming more conscious of the fact that other entities collect their personal data. In fact, they are demanding more control over their data. Failing to empower them with personal data control can jeopardize the MCS approach.

Recalling the definition of content disclosure in participatory sensing (at large) by Christin *et al*. [61], it is *"the guarantee that participants maintain control over the release of their sensitive information. This includes the protection of information that can be inferred from both the sensor readings themselves as well as from the interaction of the users with the participatory sensing system".* The aim of this study is to show how users perceive anonymous data collection (anonymity as defined by Pfitzmann and Köhntopp [62]) as a trusted method to maintain the complete control over their sensitive information and which is the role of intimacy in this ecosystem.

8.1 People-Centric Sensing and Mobile Crowd Sensing

To contextualize our focused research question, aiming to leverage intimacy in a particular study, we briefly analyze and explore the world of *People-Centric Sensing* (PCS) [63], [64].

As introduced above, smartphones leverage their increasing sensing and content generation capabilities (*e.g.*, user's location, physical activity and pictures, videos). This trend has given the rise PCS, where the smartphone owners become the *sensor custodians,* and smartphones opportunistically

collect and share the sensor data or content in a ubiquitous and pervasive manner.

There are different sensing scales in the PCS, as defined by Lane *et al.* [65]. (1) *Personal sensing:* the data producer consumes the data, and does not share any data. (2) *Group sensing:* individuals are producing data and sharing it with a group of people having the same interests or goals, *e.g.*, friends, family, or other closed social networks. (3) *Community sensing* (*participatory sensing* or *collaborative data collection*)*:* involves large-scale data collection. Many individuals are sharing their collected data with the vast, unknown to them group of people. The person's data is used to extrapolate knowledge for the benefit of the community.

Community sensing has been already proven to be effective in industry, not only for advertising. Google [66] has leveraged "community sensing" via machine learning techniques to learn languages and to provide accurate translation services to individuals. Additionally, community sensing enables researchers to discover and analyze the community-wide phenomena. To achieve it, one needs to involve a large group of people, who are strangers to each other; hence, we must consider ethics and privacy.

The challenge arises from the fact that smartphone application developers keep their users under the impression that their implementation leverages the "personal sensing" phenomena, while in reality, it leverages the "community sensing" assuming phone users being sensors custodians collecting specific data at large-scale, in turn enhancing the application quality provided to them. These are the same needs we have to study intimacy and its implications.

The topic of people-centric sensing builds on the same principles as distributed (wireless) ad-hoc sensor networks, with the difference that people, (smart)phone owners become the *sensor custodians*. They carry around their phones equipped with the sensors, and they are at the center of sensing allowing a new experimental scaling in space and time. Sensor networks, assuming a fixed set of sensors deployed in given locations for the purpose of sensing some phenomenon has been widely studied in the past [67], [68]. Ad-hoc sensor networks assume ad-hoc creation and operation of a network of sensors; *i.e.*, no pre-existing dedicated infrastructure is assumed. Aspects of research for the (ad-hoc) sensor networks include, but are not limited to answering a question on how to efficiently sample the environment, including how to distribute the sensors to capture the phenomena. Furthermore, the research has been conducted in this area on how to ensure the reliability of the data gathered in this network, *e.g.*, what are fault tolerance mechanisms for the malfunctioning or only temporarily available sensors. Finally, the research has been conducted on how to share the information with other components of the sensor network.

We can reapply many of these aspects to people-centric sensing. In fact since the availability of a sensor-enabled smartphone, researchers investigate how to use this new device as a new 'sensor network node' [63], [64], [69]–[71]. Research shows the potential of using the smartphone in substitution of ad-hoc sensors devices (*e.g.*, for behavioral and social networking studies) employed in the past [72]–[75]. Inspired by these, new research opportunities, developers,

and researchers deployed several application frameworks, and small/large data campaigns as follows.

MetroSense [76] is one of the attempts to create a framework to define the architectural aspects of a people-centric sensor network, including the smartphone as a sensor and mobility aspects of people carrying these devices. Mobiscopes [77] attempts to apply the knowledge of sensor networks to the mobility of people using their phones and specific sensors on vehicles to collect data useful for example for pollution estimation given the air quality analysis. CenceMe [78] explores issues in involving personal devices in the sensing loop. It proposes an implementation of a real application that their creators evaluated on the intrusiveness to its user (*i.e.*, the influence of the normal usage of the device) and performance (*i.e.*, the performance of the classifier involved in the process, power consumption of the system, CPU and memory benchmarks). The results show that is necessary to take care of several aspects ranging from pure system performance to involvement of people and their social networks when building the sensing ecosystem. Wang *et al.* [79] focus on a software framework to enable efficient sensing from a smartphone based on a sensors' management scheme. They evaluate energy consumption of the most diffused sensors available for mobile devices, Bluetooth, GPS, accelerometer, and more.

Examples of successful 'community sensing' *campaigns*, where many mobile users consent to become data providers, include work of Kiukkonen *et al.* [80] and Ferreira *et al.* [40]. The former deployed a data collection campaign in the area of Lausanne (Switzerland), with the yearlong participation of 170 people, collecting different smartphone sensors data such as GPS, WiFi, Bluetooth, and more. The latter focused on battery charging patterns, and it is one of the first attempts to conduct a research study leveraging a widget published at the Android OS applications market, *i.e.*, without meeting the participants, who reside anywhere in the world. In this case, 4000 people participated for four weeks. Other examples of data collection campaigns involve work of Eagle and Pentland [5], [13] where 100 people took part in performing studies on social networks inference from smartphone data. As we are going to see later in this work, we also deployed two small data collection campaigns to understand the intimacy phenomena (as well to investigate this particular intimacy use case) involving 70 users for a total of 2 months.

There exist five main concepts in which we group the works we present in Table 17. (1) *Sensor networks*, if the works related to the developments in sensor networks. (2) *Bridge*, if the work attempts to introduce human nodes in the sensor networks. (3) PCS *Foundation*, where researchers explicitly stated concepts and challenges of PCS, possibly using personal smartphones. (4) *Frameworks*, examples of proposal solutions for PCS addressing some problems posed by the foundation papers. (5) *PCS campaigns*, representative sensing campaigns examples (the implementation of MCS).

Additionally, we define the following challenges for the PCS based on the personal smartphones enumerated in the PCS "foundation" papers, over which we compare all the contributions. (1) *Ad-hoc sensors* represent the category of (usually) "home made" sensors deployed particularly for the research question at hand. (2) (Sensor) *Data flow* indicates that the authors research aspects of the

data flow in PCS. (3) *Efficiency*, how the extra load of operation influences the device (*e.g.*, extra battery consumption, CPU and memory usage, intrusiveness for the user), and techniques to reduce the impact of the PCS data collection. (4) *People consideration* feature applies when humans (smartphone owners) are considered in the reasoning of computational algorithms of PCS when gathering the data from the smartphone sensors. (5) *Privacy consideration* relates to papers that *mention* the privacy challenges in PCS. (6) *Privacy solutions* contributions indicate some concrete solutions to these privacy issues. (7) *Real applications* relate to the applied research – PCS used in practice, showing examples of real-world applications of the PCS-enabled systems. (8) *Smartphone as a sensor* regroups all the contributions that tackle the challenge of using the smartphone to perform sensing operations in PCS. An "X" in the table indicates whenever a research contribution exists for the given challenge in the given PCS domain concept.

As we can notice from Table 17, all the challenges we extrapolated are being researched, however, the *privacy solutions* category, despite its relation to the *privacy* group, has only a few contributions. In fact, most of the papers mention and present privacy challenges and issues in PCS but very few of them provide solutions on how to address these, and even if, these solutions are unique to the PCS system at hand and do not offer a scientific contribution to the field.

Challenge	Sensor Net.		Bridge				PCS Foundations				PCS Frameworks					PCS Campaigns			
	[67]	[68]	[72]	[73]	[74]	[75]	[69]	[63]	[64]	[70]	[71]	[76]	[77]	[78]	[79]	[80]	[40]	[13]	[5]
ad-hoc sensors	X	X	X	X	X	X		X			X	X	X					X	
data flow	X	X	X	X		X		X	X	X	X	X	X	X	X	X	X	X	X
sensing efficiency	X	X				X		X	X	X			X	X	X	X	X	X	
human aspect consider.	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
privacy consideration		X		X	X			X	X	X	X		X	X	X	X		X	X
privacy solution								X					X	X					
real-world applications	X		X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X
smartphone sensing								X	X	X	X		X	X	X	X	X	X	X

Table 17: From sensor networks to people-centric sensing campaigns – research challenges

The lack of practical solutions is stressed even more by Christin *et al.* in their survey on privacy in mobile participatory sensing [61]. They give a full picture of the privacy components essential in the PCS data flow and provide PCS challenges relating to "*including the participants in the privacy equation*". In particular: (1) *tailored privacy interfaces*, (2) *ease of use*, (3) *transparency of*

privacy protection levels, and (4) *incorporation of user feedback*. Christin *et al.* indicate that much work must be done on the PCS privacy solutions before we can deploy the PCS concept widely.

With the help of intimacy, we want to investigate how anonymity can impact the users choice of sharing data in MCS campaign and be PCS contributors. Anonymity offers the possibility to create tailored privacy interfaces, can be easy to understand for users, can avoid implementing various privacy protection levels, and can be paired with context elements such as intimacy to explore the feelings of users.

8.2 Investigating Users' Perception of Anonymity

The first step is to understand the difference in user's attitudes when we collect their data in an anonymous manner versus not. We want to verify the existence of a significant difference between the willingness to share data anonymously and the desire to share data with identifiable information. In our study, the anonymous data is defined as content shared on an open public server without any information about the producer. Instead, the identifiable shared data is any data posted publicly on Facebook (the user is directly highly identifiable).

The second step relates to the identification of the factors influencing this difference. We analyze three main factors that may influence the users' perception of anonymity in MCS: (1) people, (2) content, and (3) context. (1) *People factor*: investigates the natural propensity of people to share or not own facts (based on Westin [30] privacy clusters: unconcerned, pragmatist, fundamentalist). We investigate if being in a given category, influences the user's decision to share more data anonymously or not. (2) *Data kind factor*: we assume that the different type of data the user share may influence their decision to share it or not (*e.g.*, the location may be different from the song). We want to verify, if the kind of content shared leads to different users' decisions when shared either anonymously or not. Finally, (3) *context factor*: we investigate if the current context of the user influences the difference between sharing data anonymously or not. We treat this third factor while employing the intimacy metric.

The third step is to gather the users' opinions and understand attitudes towards their personal perception of anonymity. Amongst the others, we want to understand the role of the three factors mentioned above and particularly intimacy and its main elements: place, number and kind of people.

8.3 Public Sharing of Anonymous vs. Identifiable Data

Our work focuses on the relative difference between public sharing of anonymous vs. identifiable data. Our objective is to understand this difference from the users. In Table 18 we enumerate the studies investigating privacy that present similar characteristics than the one we propose. For each study, we consider four characteristics about the privacy concerns investigated (i) content shared, (ii) with whom users share their content, (iii) content shared is anonymous or not, and (iv) users' context is considered when sharing decision is taken (in our case the user intimacy perception). Additionally, we noted two

characteristics about the study methods: 1. data about privacy decisions emulated/narrated in lab environments? Alternatively, collected in the real user's daily life context, and 2. if researchers were interviewing the users for a deeper understanding of their sharing habits.

The column of Table 18 in gray relates to studies about anonymous sharing of data, which is the focus of our paper. Only Leon *et al*. [81] investigated some effects of anonymization on data collected for advertising purposes on the web explicitly. None of the other studies consider anonymity as a factor.

For the factor: "with whom users share their data", we focused on *public sharing* only, because past research identified it as an important and most disruptive sharing condition. We decided to limit this factor not to increase further the complexity.

In our studies, we aimed at understanding user's attitude in the current, real user's context, i.e., "in situ". To have a better picture of users situations and understand better the users choices and attitudes, we also employed face-to-face interviews with the users. In the current studies, as we illustrate in Table 18, the last two presented study methods are rare.

Study	Privacy concerns depending on data kind/content shared	Privacy concerns depending with whom users share their content	Privacy concerns depending if users share their content anonymously or not	Privacy concerns considering the users' context when they decide to share or not	Situation when answering surveys is the actual user context (i.e., if "in situ" real random, dynamic context)	User personal interviews based on surveys' answers
[82]	X	X			Narrated/Emulated	
[46]	X	X			Narrated/Emulated	
[83]	X	X			Narrated/Emulated	
[84]	X				Narrated/Emulated	
[81]	X	X	X	X	Narrated/Emulated	
[85]	X	X		X	Narrated/Emulated	
[86]	X	X		X	Narrated/Emulated	
[87]	X	X		X	Real	
[88]	X	X		X	Real	X
Ours	X	X*	X	X	Real	X

Table 18: Comparison of different studies performed to understand differences in the users' privacy concerns depending on diverse sharing aspects (*only public sharing).

8.4 Data Anonymisation

The recent survey by Christin *et al.* [61] on privacy in participatory sensing and the work of Ganti *et al.* [89] about challenges in MCS indicate anonymity as a possible partial solution to the MCS privacy challenges. From these works, the ones we present in Table 18, and Oswald [90] suggestions (she argues about the importance of anonymity and possible risks), we conclude that there is very limited work on understanding how MCS users perceive the anonymisation of their data. Most of the works relate to processing of the gathered sensor data towards their anonymisation (Shilton [91]), spatial obfuscation by combining users' data (Hara *et al*. [92]), data aggregation (Shi *et al*. [93]), selective hiding

of data (Mun *et al*. [94]). The most significant difference between the existing studies and ours is the fact that we focus solely on the users, to understand the factors influencing the acceptance of anonymisation.

8.5 Analysis of Factors Influencing Anonymity

For this analysis, we used again the dataset we collected during the US1 (Section 4.2.1). The data analysis procedure, data preparation and the method applied are the same one we employed and motivated in Section 5.2 when we verified the validity of intimacy. The general results of ESM answers and users contributions of Section 5.3 also apply here.

The first objective of this chapter is to verify that there is a significant difference between sharing data anonymously or not. The second goal is to investigate if the difference in deletion choices is related to one or multiple of these factors: (1) *people factor*, (2) *data kind factor* and (3) *context factor*.

8.5.1 User's Deletion Choices: Public on Facebook vs. Anonymously on a Public Server

We analyzed if the data sharing option is statistically significantly different between Facebook (FB) and Server (SE) across all the participants and for the whole duration of the study. The participants have answered with the same frequency the surveys relating to sharing content either on FB or SE (3295 beeps, 50.6% / 3214 beeps, 49.4%).

We prepared the data by selecting only the variable representing how the system shared the data (not anonymously FB, or anonymously SE) and the dependent variable delete choice. We collected all the *beeps* of users in the *beep file* that we transformed in a *subject file* (Larson and Delespaul [50], Section 5.2). In this case, the *subject file* contains three columns: user ID, sharing modality, and the mean delete value for both sharing modality for each user. In this way, we obtained two rows of data per user for a total of 84 rows.

Users are more likely to delete Facebook data than the one shared anonymously on the server. We validated our basic hypothesis: the sharing modality, either anonymous or not, influences the willingness to share data. Users are substantially more prone to share data when no information about them is disclosed with the collected content (as for this particular case with anonymous data shared on a publicly accessible server).

In Figure 49 we indicate the mean users' mean delete choice (to recall from Section 4.2.1, denoted Del_m,'1' being very likely to remove the content, i.e., not to share, while '5' being the opposite). For FB we have $Del_m = 2.53$ and for SE a $Del_m = 2.92$. The matched-pairs *t*-test, $t(41) = -4.17$, $p < 0.001$, $r = 0.55$ shows that the difference in deletion choices between FB and SE is high, and therefore substantial.

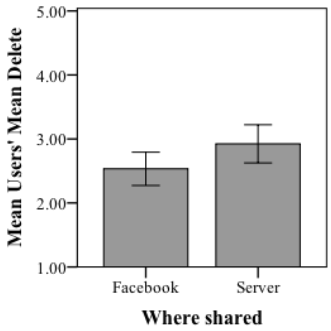

Figure 49: Mean of users' mean delete for sharing on Facebook (not anonymously) and on Server (anonymously) ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI).

8.5.2 People Factor: Users Privacy Clusters

The data preparation to analyze this factor consisted of two steps: (1) cluster the study participants in 3 privacy clusters and (2) the transformation from *beep file* to *subject file*. We clustered the participants on their privacy attitude by Westin's, i.e., being privacy fundamentalist (never share), pragmatist (sharing depending on personal decisions and situation) or unconcerned (always share) [30]. We have done so considering the mean 'delete' choice value *over the whole study for a given participant*. We have employed Jenks natural breaks optimization [95]. We assumed 3 breaks, and we have acquired mean 'delete' choice breaks being Del_m = [1.19, 2.18] for 14 participants being *fundamentalists* (Figure 50, in dark grey), Del_m = (2.18, 3.26] for 16 people being *pragmatists* (Figure 50, in medium grey), and Del_m = (3.26, 4.31] for 12 people being *unconcerned* (Figure 50, in light grey).

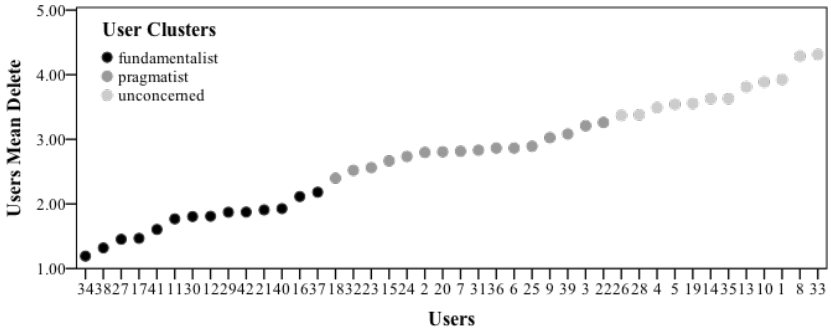

Figure 50: The users divided into three privacy clusters using Jenks Natural breaks, 14 fundamentalists Del_m = [1.19, 2.18], 16 pragmatists Del_m = (2.18, 3.26], and 12 unconcerned Del_m = (3.26, 4.31] ('1' very likely to delete the content, '5' being the opposite).

We added this privacy clusters to the *subject file* as a new column labeling the respective users' entries. In this case, the *subject file* contains four columns: user ID, sharing modality, the mean delete value for both sharing modality for each user, and the user privacy cluster (same file as previous Section 8.5.1 with an additional column).

The three privacy clusters are all significantly different from each other, but they are not significantly influencing the difference between FB and SE. We conclude

that cluster privacy memberships are not influencing how users, being in a particular privacy cluster, perceive anonymisation. Only 'unconcerned' are even less "concerned" when the system shared data anonymously, but we cannot prove this difference to be significant. Also, the other two groups do not present significant differences between the two sharing modalities. Therefore, the *people factor* (privacy attitude) *does not explain alone* the difference between FB (not anonymous) and SE (anonymous) sharing preferences.

In Figure 51 we present the mean of the user's mean delete choice for where the system shared the data for the three privacy clusters. We performed a Two-Way Mixed ANOVA, and we can conclude, again, that the deletion choice is significantly different between FB (more deleting) and SE (less deleting), $F(1, 39) = 19.207$, $p < 0.001$, $r = 0.57$. The deletion choice is significantly affected by the three clusters: $F(2, 39) = 173.128$, $p < 0.001$, $r = 0.95$ (expected very high effect, because we derived clusters from the data). With post-hoc test with Bonferroni correction, we confirm that the three clusters are significantly different from each other. However, the interaction between where data is shared (FB or SE) and users privacy clusters is not significant, $F(2, 39) = 1.682$, $p = 0.199$, $r = 0.28$.

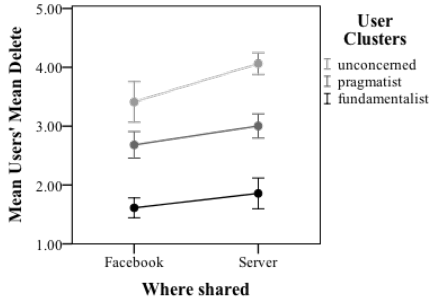

Figure 51: Mean users' mean delete for sharing on Facebook (not anonymously) and on Server (anonymously) separated in the three privacy clusters ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI).

8.5.3 Data Kind Factor: User's Deletion Choices and Content Type

We analyze where data is shared (FB, SE) versus kind of data the system shared (i.e., content factor). In Table 19 we show the distribution of the seven possible content types (~14% of beeps each, uniform distribution).

Data type	Number of 'beeps'	Percent
activity	947	14.5
air	872	13.4
audio	933	14.3
location	915	14.1
photo	929	14.3
song	972	14.9
video	941	14.5

Table 19: Distribution of content types for the random sharing survey question.

Once again we transformed the *beep file* into the *subject file*. In this case, the *subject file* contains four columns: user ID, sharing modality, the kind of data

shared, and the mean delete value for both sharing modality subset for each data kind for each user.

In the results, we show that the change in deletion due to the type of data shared is significantly different when the system shared data on FB compared to the SE one. We discovered that the *audio*, *activity*, *air* and *song* contents are responsible for the changes between FB and SE (i.e., this content gets more deleted on FB) while *location*, *photo,* and *video* are not (i.e., very frequently users deleted these data for both FB and SE). In Figure 52 we present a graph of the ANOVA estimation of means of delete choices for FB vs. SE for the seven different content types. As we present in Figure 53, post hoc tests with Bonferroni correction revealed that we can group content types into four groups. (A) *Air* and *song* are equally deleted and less than everything else. (B) *Activity*, *location,* and *audio* are equally deleted and less deleted than *video* and *photo* (exception *audio*), but more than *air* and *song*. (C) *photo* and *audio* are equally deleted, and *photo* is less deleted than *video* but more than everything else. (D) *video* is more deleted than everything else.

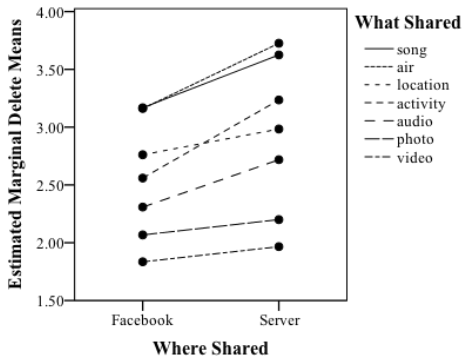

Figure 52: Estimated mean users' mean delete for sharing a different kind of data, to highlight how different content is treated depending on the sharing format Facebook (not anonymously) and on Server (anonymously) ('1' very likely to delete the content, '5' being the opposite).

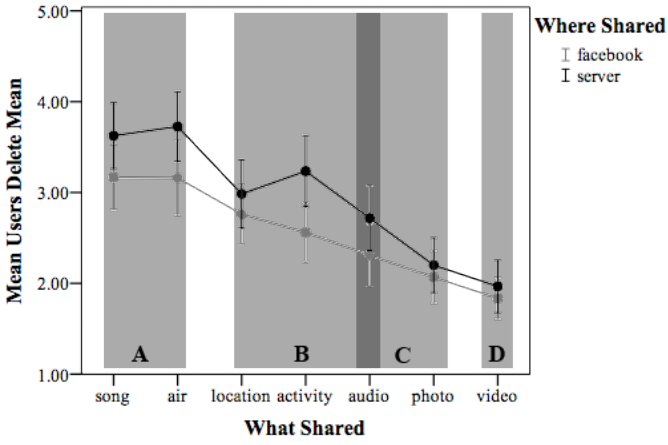

Figure 53: Mean users' mean delete for sharing different kind of data in Facebook (not anonymously) and on Server (anonymously) with the groups of similarly shared contents: A easily shared, B more protected depending on information that is conveyed, C and D highly protected contents ('1' very likely to delete the content, '5' being the opposite, error bars: 95% CI).

The details of the Two-Way Repeated ANOVA test results for delete means (i.e., within-subject effects) shows, once again that the choice of Facebook and Server have significantly different delete ratings, $F(1, 41) = 15.906$, $p < 0.001$, $r = 0.53$.

The data type factor alone is also significantly influencing the deletion choice, $F(4.46, 182.85) = 38.01$, $p < 0.001$, $r = 0.69$ (sphericity Greenhouse-Geisser correction). To detail the results, we show the difference between all the content types: (1) *activity* is not significantly different from *audio* and *location* (Bonferroni corrected $p > 0.05$), and *air* ($p = 0.008$) and *song* ($p = 0.049$) are deleted significantly less than *activity* and *photo* ($p < 0.001$), and *video* significantly more ($p < 0.001$); (2) *air* is not significantly different from *song* ($p > 0.05$), and everything else is deleted significantly more than *air*; (3) *audio* is not significantly different from *activity*, *location,* and *photo* ($p > 0.05$), and significantly more deleted than *air* and *song* ($p < 0.001$) and significantly less deleted than *video* ($p < 0.001$); (4) *location* is not significantly different from *activity* and *audio* ($p > 0.05$), and significantly more deleted than *song* ($p = 0.035$) and *air* ($p = 0.010$) and significantly less deleted than *photo* and video ($p < 0.001$); (5) *photo* is not significantly different from *audio* ($p > 0.05$), and significantly less deleted than *video* ($p = 0.016$) and significantly more deleted than everything else (all $p < 0.001$); (6) *song* is not significantly different than *air* ($p > 0.05$), and significantly less deleted than everything else; (7) *video* is significantly more deleted than everything else (all $p < 0.001$ except for *photo* $p = 0.016$).

Finally, the interaction between where data is shared (FB or SE) and kind of data is significant, $F(4.451, 182.501) = 5.64$, $p < 0.001$, $r = 0.35$ (Greenhouse-Geisser correction). To discover which kind of data is involved in this difference we conduct multiple paired samples test with post-hoc tests using Bonferroni correction of significance levels (in this case significant when $p < 0.007$). These tests results reveal that *activity*: $t(41) = -4.82$, $p < 0.001$, *air*: $t(41) = -3.49$, $p = 0.001$, *audio*: $t(41) = -1.89$, $p = 0.001$, and *song*: $t(41) = -3.81$, $p < 0.001$ are deleted differently depending if the data is shared on FB (more deletion) or on SE (less deletion). Instead, there is not significant difference for *location*, *photo*, and *video* (all $p > 0.007$).

8.5.4 Context Factor: User's Deletion Choices and Intimacy

Until now we analyzed how individual's attitude and content are influencing how data is shared either on FB or SE. In this last part, we analyze the role of context in the users' deletion choices by applying intimacy (proven in Section 5.3 being able to enclose the perception of users about the place, number and kind of people around them). As before, we transformed the *beep file* into the *subject file*. In this case, the *subject file* contains four columns: user ID, sharing modality, the intimacy level (1 - completely intimate to 6 - not intimate at all), and the mean delete value for both sharing modality subset for each intimacy level for each user.

We found that the intimacy factor influences sharing, but it does not interact significantly with where the system shared the data (FB vs. SE). It is only relevant to decide *if* to share data at all when in extreme intimacy states 1 or 6. The context of the user does not seem to imply significant differences in the

choice of sharing data anonymously or not. In Figure 54 we provide the means of delete choices for where the system shared the content (FB, SE) for different levels of context intimacy across all the participants.

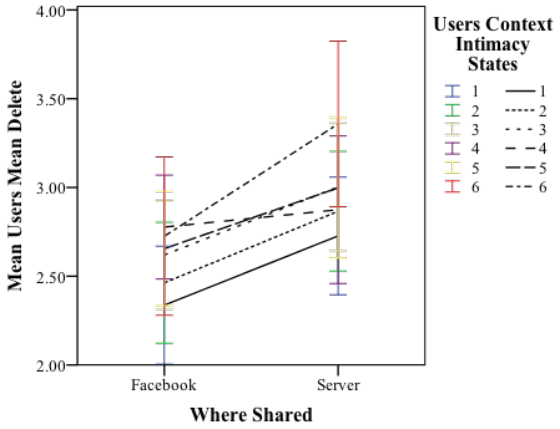

Figure 54: Mean users' mean delete for sharing different kind of data in Facebook (not anonymously) and on Server (anonymously) and intimacy states (delete: '1' very likely to delete the content, '5' being the opposite, intimacy: '1' completely intimate – '6' not intimate, error bars: 95% CI).

The data is unbalanced across the intimacy levels, presenting the following frequencies of participants: *Facebook:* '1' = 35, '2' = 40, '3' = 39, '4' = 37, '5' = 35, '6' = 28 and *Server:* '1' = 33, '2' = 41, '3' = 38, '4' = 36, '5' = 34, '6' = 33. Given the missing data, we employ the Linear Mixed Model Repeated Measurements for significance tests.

We show that the fact that data is shared on FB or SE is once again significant, $F(1, 376.653) = 30.376$, $p < 0.001$. Furthermore, the intimacy is also proven to be significant in respect of the data shared, $F(5, 378.448) = 2.352$, $p = 0.04$. The pairwise comparisons post-hoc tests with Bonferroni correction reveal that the difference in deletion mean of our study is only significant between intimacy states 1 and 6 ($p = 0.019$). Finally, the interaction between where data is shared (FB or SE) and users context intimacy is not significant, $F(5, 376.522) = 0.860$, $p = 0.508$.

8.5.5 Summary of Quantitative Results

To summarize the quantitative results we obtained in our analysis; we proven that only the shared 'data type' has a significant effect on the individual's choice to share data anonymously or not anonymously. We found significant differences between the data kinds: *audio*, *activity*, *air* and *song*. Instead, the kinds: *location*, *photo*, and *video* are not shared differently if anonymous; for both cases they would be rather deleted than shared. All the other factors, like cluster privacy membership and users' context (intimacy), do not significantly influence the individual's sharing choice between FB and SE.

In the following section, we present the results of the interviews we conducted with the participants of our US1 to investigate deeper the personal point of views of users on data anonymity. We further elaborate on the intimacy and sharing options given subjective user opinions.

8.6 DRM: Participants Interviews

In this section, we report the themes and the main related findings on the users' perception of anonymity in MCS, from the analysis of qualitative data originated from the DRM interviews. With the information we collected from our participants during the interviews, we want to understand how our participants perceived the difference between sharing data publicly on FB (i.e., not anonymously) and anonymously on an open-data SE (considered as a case of MCS). We divide the analysis into two parts: one regarding FB and one regarding the SE. For each of them, we particularly investigate the kind of data shared (a relevant factor when sharing data anonymously or not). We first describe the users feelings in general for each data type and then derive conclusions. For both cases, we sort the data types from the most to the less likely to be deleted as depicted in Figure 53.

8.6.1 Identifiable Public Sharing on Facebook

a. *Video and photo - the most likely deleted contents*: For both content types the most recurring participants' explanation for the deletion is that they "*don't share private content*", especially when at home, in the bedroom or bathroom, at a security-sensitive job or when they are relaxing, tired or sleepy. The video/photo content gets also deleted if users consider it silly (*e.g.*, regular meeting with friends) or not attractive (*e.g.*, a lecture) to others. This last factor is very subjective and depends on what the given user perceives as interesting or not. For both content types, participants are least likely to delete it, when already publicly exposed, *e.g.*, in public transportation, and if users consider the situation interesting for others to see. To summarize, users consider 'video' and 'photo' too private to be shared *publicly* with identifiable information.

b. *Audio - what gets recorded matters*: users consider audio content also very private, and they expressed, even more, interest in the audio content on its deletion choice. If its content is not interesting (*e.g.*, waking up, being alone, or at the lecture), or not relevant for "*polluting the Facebook wall*" (*e.g.*, watching a movie) it will get deleted. Some participants were worried about respecting other people's privacy when the recording would take place at school, on a bus or other public space, and they would likely to delete it. Users do not delete recorded audio when the user sharing the content perceives it as potentially interesting for others (*e.g.*, what I watch on TV, what song I listen in the background), and/or the situation does not reveal something users consider personal (*e.g.*, just a background noise). Summarizing, users may share the 'audio' publicly with identifiable information depending on its interest, but they usually prefer not to share it due to the respect for the privacy of others potentially taking part in this recording (i.e., out of the fear that someone recognize them as the audio record source).

c. *Sport/General Activity - social appearance and approval matter*: users are least likely to delete activities if they are active (*e.g.*, walking). They probably delete it, if others may conclude from it that the users are doing something that is not good for their health or well-being (*e.g.*, inactive at work (sitting), relaxing on the sofa). Additionally, based on the content, users will delete

the activity if the person thinks that others will not find it interesting (*e.g.*, going to bed, doing homework, just inactive). To summarize, the 'activity' content sharing publicly with identifiable information relates to the willingness of the user to expose himself to a social 'judgment' of an activity and its interest to others.

d. *Location - users delete routine locations:* for participants it is important to protect their frequently visited places, like home (especially) or work, and not to reveal their routines and places in which most of their personal life situations occur. For rarely visited or entirely new places, the choice of deletion is more related to how interesting the location would be if shared on Facebook (special event or a place of particular interest), than for any other particular reason. Users do not delete the location when it is temporal and mutates quickly, like when being on a bus, or in an unknown place on the route. Summarizing, users publicly share 'location' content with identifiable information when it is related to specific events and does not reveal users routines.

e. *Song and air quality - the least likely deleted contents:* users do not perceive these content types as personal or significant and deleting them would be a "waste of time". The deletion choice seems to be agnostic to the context, from which we extract the content. Users that do not want to "pollute" their Facebook wall will likely delete song and air contents. Summarizing, users share easily the 'song and air quality' contents publicly with identifiable information.

Key findings on the Identifiable Public Sharing (Facebook)

From the results we conclude that the mobile users motivation to share identifiable content or not depends on three main factors, which in certain cases may overlap:

1) *The ability of the content to indicate to third parties about the situation in which the user is, when we collect the content*. The level of description accuracy of personal facts in the content is its 'power' to infringe upon the *context* of the user; that applies to, *e.g.*, video and photo (to a very great extent) and recorded audio (lower extent). Users delete location especially when it reveals routines and exposes places where users are likely to live their most personal life moments. Instead, users do not consider song and air quality as highly descriptive to the others, they are not revealing enough information;

2) *The reputation of the users sharing this content*: users are selecting content to be deleted depending on the level of reputation they want to maintain in the face of others. Even in a public setting, they may delete content, if they feel that their reputation could change based on the content, *e.g.*, not wanting to be seen as sedentary, sitting at work, sitting in public transport, and sitting on the couch at home. Users may think that is important that society does not label them as "loafers";

3) *The personal interest of the user in the shared content*: if users do not perceive the shared content as useful or interesting, it is more likely that they delete it so as not to "pollute" their Facebook wall. It is important for users to

keep 'clean' their sharing space to publish only what (they think) matters to others to know about them. For example, an audio recording that contains only environmental sounds is likely to be deleted, being considered as "pollution". Another example is air quality, for which the users do not see utility and will delete it.

8.6.2 Anonymous Public Sharing on a Server

a. *Video and photo - the most likely deleted contents*: even if shared anonymously, users are likely to consider video and photo private and personal and delete them. Particularly, the users want to minimize as much as possible the probability of revealing who is the source, or the subjects, of the photo or video. They choose to delete this content independently of the users' situation. They do not delete video and photo when the activity or the situation the user is in is not relevant to the user (*i.e.* the content does not express something specific about the user, but is describing more general situations such as waiting for a bus or attending a concert).

b. *Audio - what gets recorded matters, but to a less extent*: users express similar motivations as for the video and photo, but in this case, they delete this content less. Users perceive that they are less identifiable, and the situation is less detailed/salient without the identity of the user (*e.g.*, studying silently or waiting at tram stop). In the case of anonymous audio, the content of any conversation, rather than the whole situation plays a fundamental role in deletion choice. When the audio track does not reveal the direct interaction of the user with other people (*e.g.*, at a lecture with a professor, speaking to the audience or when walking around in a shopping center), they do not perceive it as an infringing of their privacy and thus is not deleted.

c. *Location – users delete routine locations:* the deletion choices are very similar to Facebook choices: users who delete location do not want their routine to be revealed, even anonymously. If we look at situations when they do not delete location content, we notice that if the user's location cannot be easily linked to the individual, e.g., indicating a busy street in the city, the user is less likely to delete it. To summarize, the 'location' content is one of the least to benefice from the anonymisation of data.

d. *Sport/General Activity - social appearance and approval does not matter*: sharing decisions for this content are significantly different from Facebook, as participants tend not to delete the content. The reason behind their choice is: if the activity does not relate to an identity executing that activity, it is not seen as a private content independently from the actual situation. The few people that delete this content, even when anonymous, are more concerned about its interest to others (*e.g.*, sitting) rather than their privacy. To summarize, the 'sports/activity' content is the one that presents a change of paradigm when shared anonymously.

e. *Song and air quality - the least likely deleted contents:* users delete these contents less than for FB and the reasons are very similar to FB ones. However, in particular for air quality, the motivation that the Facebook wall could be 'polluted' with useless information is not present in the SE case.

Participants that decided to delete this content from the anonymous server are ones that cannot see its utility. Additionally in the EU, we noticed few concerns about revealing the location where the system collected the air quality and thus content becoming identifiable because users' location routines may be revealed.

Main Findings on Anonymous Public Sharing

For the general trends, we notice the following differences when the system shared the content anonymously:

1) *The ability of content to indicate to third parties about the situation in which the user is, when the system collected the content*: the impact of this factor was stronger and became even more relevant for the anonymous sharing case; even though the content is anonymous, issues of reputation are still important. From the users' perspective, the content shared should not reveal who is the source of it. For example, video and photo are still very powerful descriptor of the situation and can easily reveal who shared them. Location data is also an indicator of routines, and there are some evidence that users are aware or, at least, suspicious of the risk of simply removing their personal metadata from locations they shared. We assume that they see the possibility that a third party may recognize them anyway.
2) *The reputation of users sharing this content*: the impact of this factor was lower in the case of anonymous sharing, and users delete it, if the shared audio, video or photo content is not appropriate given the social context of the users and if revealed it can be disadvantageous to the user's reputation. This reasoning does not apply to other content types such as activity, song, and air quality and so the users are willing to share such content.

The subjective interest of users in the shared content: the impact of this factor is less emphasized in the case of anonymous sharing because users cannot identify the source of the "pollution". However, there are still a few users in disagreement about the motivations behind sharing what they consider as a 'useless' content, even anonymously.

8.7 Discussion

The results we present in this chapter show that there is a significant difference for mobile users between sharing their content data anonymously or not. We have investigated this difference on three main factors: privacy clusters (people), kind of data shared (content) and intimacy (context of users when they make sharing decisions). Additionally, from the interviews with the study participants we extracted some themes that are relevant to our analysis.

8.7.1 Anonymous vs. Identifiable Sharing

Users prefer to share data anonymously than when they are identifiable (e.g., on Facebook).

With the results, we show that there is a significant difference in the deletion choice between data shared publicly on FB (not anonymously) and publicly on

an SE (anonymously). This result agrees with the work of Leon *et al*. [81] where they showed that users are more willing to provide personal information to web advertisers anonymously. Although, our sample of 42 participants may not be representative of the whole population, this significant difference signals that anonymity can help to encourage mobile users to contribute with their data to MCS.

8.7.2 Factors Influencing Anonymous vs. Identifiable Sharing

Privacy unconcerned, pragmatists, and fundamentalists do not change their sharing habits depending on data anonymity.

In general, as Westin [30] pointed out, our participants' sample confirmed the existence of three main groups of privacy categories. We have users that tend not to delete any automatically shared data (privacy unconcerned). There are also other users that are more equilibrated, i.e., privacy pragmatists, who take different decisions, depending on context and kind of data shared. Finally, there are some that are privacy fundamentalists and tend always to delete content. This particular fact is known, and our results show that our sample has similar characteristics.

An important factor we investigated through our results is the influence of privacy clusters in the choice of sharing data identifiably on FB versus anonymously on a public SE. We show that this factor is not significantly influencing the difference in deletion choice between FB and SE. Only privacy unconcerned participants are sharing significantly more when data is anonymous. At this stage, this fact implies that anonymity cannot, in theory, change the mind of users in our sample on how they make their sharing decisions. Privacy pragmatists are still pragmatist, and fundamentalists remain fundamentalists.

Intimacy influences sharing in general, but does not seem to influence the choice between anonymous or not-anonymous sharing.

A second factor we investigated in our studies is the context in which user make sharing decisions, i.e., where the user is and with whom. On the difference in deletion choices on FB and SE, the intimacy influences significantly the deletion choice in general but has no significant influence on the fact that users delete data when shared anonymously or not.

Similarly to the work of Khalil and Connelly [87] (who investigated the social context and home/work as location), we can conclude that for our sample, the intimacy in which participants were at the moment of the choice may determine in general if users share data or not. However, as for privacy clusters, this factor does not explain the difference between deletion choices between FB and SE. Therefore, in practice, this probably indicates that we should perform anonymisation independently of the users' intimacy.

Different kinds of data are shared differently if sharing is anonymous.

We left the kind of data shared factor as last because we show that it influences the willingness to share the content on FB or SE significantly. In Figure 53 we reported the mean delete choice for each data type and the difference between sharing the information on FB and SE. As for previous works, e.g. Shih and Boortz [88], users share easily some types of content than others. For example, users consider *song* and *air* less important than *photo* and *video*. For our particular investigation, we show that *activity*, *air*, *audio*, and *song* are determinant for the differences between FB and SE. The remaining: *location*, *photo*, and *video* do not present a significant difference between the two cases. One of the most interesting results is for *location*. Always referring to Figure 53, *location* is a data type that users share on FB more often than *activity* and *audio*, but the effect of anonymisation is not significant. Even if anonymised, users share it to a less extent than other contents. We acquired the same results for *photo* and *video* that are the least shared kind of content types in both cases. This finding may imply that our participants see *location* anonymization as not effective, or they do not trust that we can guarantee anonymity. Instead, they may believe that removing personal information from *activity* and *audio* contents can protect their identity and thus they may disclose these contents via MDC. We cannot extract the motivations behind this particular behavior noted in our study solely from the quantitative data at disposal. We have investigated the interview answers the participant provided to understand if our reasoning holds, as follows.

8.7.3 Understanding Users' Perception of Anonymity

Overall, the analysis of participants' interview confirmed the results of the quantitative analysis and provided some extra information on factors influencing their choice of anonymity. It revealed three main themes regarding the motivations behind the difference in choice of deletion for a given content being shared on FB and SE. First, our participants would delete the kind of contents that are 'powerful' regarding describing their situations (spoil their intimacy) and habits. In this group of contents, we identified *location*, *photo*, and *video*. Second, anonymity should protect the reputation of the user producing the data. They delete more content on FB when their reputation is at risk. Furthermore, if the participants in the study did not trust that the anonymity was accurate enough to protect their reputation, they were choosing to delete the content. The same thematic spans over all of the data kinds shared in this study and gets accentuated by highly descriptive data. Third, users share data depending on the subjective feeling of our participants about its actual utility and interest to others. If the users cannot see the utility of the data collected they will not engage in its collection, not even anonymously; they will choose to delete it.

8.8 Conclusions

To summarize our findings we showed how anonymous or not anonymous data sharing is influenced significantly by the kind of content shared and how this data impacts on the user privacy and reputation. Particularly for location, we

discovered that users do not share it more when anonymized. Instead, data like users' activity greatly benefits from a simple anonymization like the removal of personal meta-data. Additionally, if the data does not seem useful to the users, they are not willing to share it even anonymously.

Since the results we reported in this chapter show that intimacy seems not to be a significant factor to make a choice on sharing data anonymously or not, we do an *excursus* and focus our conclusions and future work on the more significant factor, data sharing, that it is significantly influencing the behavior of users.

Based on our findings, to uncover possible future research, we formulate some hypotheses and define further research questions. We extracted both from the results acquired in our study and their discussion. We have identified a fundamental challenge on user's perception for *location* anonymization. From our study, we conclude that even if we anonymize the *location*, and despite that it has no information about its producer, from the point of view of the user, it has still the power to compromise the user's privacy. Most probably users do not trust that is possible to hide their identity and at the same time keep their location usable for MDC. As the results show, for the users, it seems that the anonymisation of their *activity* is much accurate easily, because once we remove the users' identity information, all is left is the name of the activity performed at a given time. We are aware that there have been several works about *location* anonymization (*e.g.*, Hara *et al*. [92] and Gedik and Liu [96]). We would like to note, that the results of these works are likely to be theoretically valid, but, as we conclude from our users studies, the mobile users do not understand and trust the anonymisation techniques. From the point of users, anonymisation is still limited to the point of just removing their name (and other personal information) from the data collected. This finding of ours may lead to an interesting future research in user's perception of anonymised *location*.

Another hypothesis, which has been derived based on the results we have got is: the mobile users are more confident about the possibility to identify people from their anonymous *location* traces than from the processing of an anonymous *audio* track. Recall that, as derived from the participants' interview, the *audio* track to be considered anonymous must not contain names or references to people in the scene (e.g., the user is interacting with), neither the characteristics of the content producer. Most probably the users estimate *audio* recognition as a more difficult task (than *location*). They may think that the task is difficult for a person re-listening the recorded *audio*, so they do not see how a computer may do it. This same thinking does not happen for *photo* and *video* - that from the users point of view reveal too much, and a person may be able to identify the situation, the producer of the content or the people involved easily. Therefore, these content types get deleted, even if anonymous.

Finally, some of our participants delete anonymous data even when they just believe useless and uninteresting for others. We assume that without clear goals of content collection, the mobile users are not able to see why they need to collaborate to collect data, even with the incentive to stay anonymous. As a result, if the users do not see the benefit of their effort, they do not collaborate.

What we want to highlight with these hypotheses is that most probably anonymization algorithm, particularly for *location* content, need not only to

guarantee a certain anonymity threshold but also be able to address fundamental users concerns. Any system and services in the use of anonymization algorithms should somehow assure their users and be capable of showing to them that we anonymize the collected content and cannot be back-tracked and threaten the users' privacy.

The future research questions we propose to investigate are, but not limited to (1) Why the users do not trust *location* anonymization, or anonymisation in general (following our findings as hints for proper investigation) specifically? (2) What do we need to do to explain to users that anonymization algorithms are accurate and effective (for *location* and in general for other types of content)? (3) Related to the previous question: How do we explain to users that anonymisation is not simply removing their name from the data? (4) Besides theoretical proofs, how we show to users that their data is still valuable, even if anonymous, and not 'polluting' the MCD system, and at the same time, that the data is not recognizable to threaten their privacy? Our research results indicate that more human concerns are influencing the anonymous data sharing that literature shows. As our current and future work, we follow the path of research to unveil these concerns and address them in MCD systems.

9. Future Work and Conclusions

To leverage intimacy further in a concrete way, we first present a design of an experiment focusing on the deployment of intimacy in a real mobile application. Second, we conclude by summarizing the contributions of our work and at the same time, we discuss several still open questions that we contributed to surface with the study of intimacy.

9.1 Leveraging Intimacy Further in the UnCrowd TPG Mobile Application

The scope of the experiment is to integrate intimacy in our UnCrowd TPG mobile application [60].

Uncrowd TPG is a crowdsourcing mobile application that provides users of Geneva's public transport system real-time information on tram and bus schedules as well as current or predicted passenger volumes for any selected service. The system estimates passenger volumes from live subjective assessments of public transport users. Users can check which connections are leaving from nearby stops and whenever the user reaches a stop they have shown interest in the application automatically provides them with information on the next departures. The application asks the users to provide their subjective crowd assessments at the stops and in the vehicles (i.e., large crowd, medium, small). This time, in exchange for their second estimate, the user gets real-time updates about the current service (e.g., coming stops) and the status of following connections. A user is expected to use an app at least twice a day while going to work/school and while returning from it back home. At the application installation phase, we inform all the users of UnCrowd TPG about the fact that the application may collect anonymous data that QoL lab leverage for its research. By installing the application, an individual agrees to these terms. We provide detailed contact information to the lab to the individuals, in case they wish to ask more questions about the app or the data we collect.

The goals of the study are: (A) apply the intimacy model to a real specific scoped use case and evaluate its accuracy, speed, and dependability 'in the wild'; (B) based on the intimacy model, define the interventions on the Uncrowd TPG user interface and verify that these interventions are effective and provide an improved experience for its users.

9.1.1 Users Involved

Since we conduct the study using UnCrowd TPG (a real Android application we deployed in Google Play Store[5], 100+ users as of May 2015), we will start the study with a small number of beta testers of the application. For this first phase, we will involve 14 beta users that subscribed to our community. In a second phase, we will extend the study to the whole UnCrowd TPG user base involving a total of 150+ existing users and future new users. We will split the users into three groups due to the experiment design we explain in the next subsection: (1)

[5] https://play.google.com/store/apps/details?id=ch.unige.tpgcrowd

interface changes depending on intimacy, (2) always old interface, and (3) always new interface. The separation of users in the three groups will be done randomly when the users will launch the updated application for the first time. The participation is anonymous, i.e., we do not know the names or socio-demographics of the participants, as we only monitor the unique user ID generated by the application every time the user install and use the application for the first time (each reinstallation generates a new ID).

9.1.2 Study Methods

We will instrument the operational UnCrowd TPG application to fulfill our study goals. We will configure it with the integration of the intimacy model and two of the general methods we propose in the mHUMAC methodology to perform user research in the wild: ESM and an analytic logger (the same methods we explained in Chapter 4).

The intimacy model will provide intimacy level changes events to the UnCrowd TPG application. We will use the ESM and the logger to measure its accuracy, speed and dependability (goal A) and to verify the effects of the intervention on the user interface based on intimacy changes (goal B). We will instantiate these two user study methods in UnCrowd TPG transparently to users.

First, for goal A, we will integrate ESM surveys in the ordinary flow of the application. The ESM deployment will be as follows. The users will be not aware that they will answer to actual surveys questions. We will add a new application feature and further enhance the existing questions present in the application to gather users crowd assessments.

Specifically, the new feature will consist of identifying the 2 or 3 most intimate significant user's places. We will use these places to provide to users several statistics about their movements between them (e.g., average duration of the trip depending on the crowd, time of the day and transportation line used). We will identify the intimate places with our neighborhoods detection algorithm (implementation of the algorithm presented by Fanourakis and Wac [56] and the intimacy model (Chapter 7), based on one week of data we will collect from the user. Then, we will ask the users to label these places with home, work or a personalized place or, in case they will be not interested in labeling the places, they will be able to delete them if they want. When users will decide to remove a place we will suggest, we will ask them the reason of this by providing 6 possible answers: "it is just a temporary place / not in routine", "I do not want to let you know which place is it", "This place is private", "I am not interested in this place", and "I am never there". This procedure will help us to verify the accuracy of our algorithm in identify intimate places and, therefore, being able to assess correctly users' high intimacy for those places. On the other hand, we will be able in which cases our intimacy labeling is wrong. For example, imagine that we will propose to a user a place that is not significant, but it is where she/he spends lots of time due to traffic. In future, we may be able to add such situations to our model and correctly assess intimacy also for these special cases.

Additionally, to this new feature, we will add to the crowd inputs (at a stop and in the vehicle) some extra details to better understand the users context when

answering the crowd surveys. This new information that we will collect, together with the knowledge of the crowd provided by the user at a stop and in the vehicle, helps us to verify the accuracy of the intimacy model to identify not intimate places and situations. We know from our previous studies (Section 5.3.1) that bus and street (bus stops) and/or a high number of people (crowd on the bus or at a stop) indicate low intimacy. In these situations, we expect a low intimacy prediction.

Secondly, for goal B, we will use the analytic logger to capture applications and user interface events, such as user clicking on application's screens and buttons or the time they will take to perform some specific tasks within the application. We will perform small modifications to the user interface of the original UnCrowd TPG app to obtain two different ways to present data and tasks to users differently depending on high or low intimacy. Following the results we acquired in our past research [97] (Chapter 6) we will design the new interface for faster access to information and faster task performance within the app, in low intimacy cases.

We will randomly select 1/3 of the application users (group 1), and we will apply the interface changes when we detect that they are using the application in low intimacy. Vice-versa we will present to the same users the standard (old) interface when we detect that they are using the app in a high intimacy state. Out of the remaining 2/3 of users, 1/3 will not experience any interface change (group 2, standard, old interface independently from their intimacy state) and 1/3 will always get the new 'low intimacy' interface (group 3, independently of their intimacy state). These last two groups of users will be our control groups to verify that intimacy influences the variables we are investigating. With this interface changes, we aim to understand if interface interventions based on users' intimacy are effective or not in the case of our mobility application. We explain the changes to the interface describing the variables involved in the experiment more in details.

9.1.3 Variables Involved

To summarize, the variables involved in the study to investigate goal A will originate from: (1) users detected most significant places, and (2) user crowd surveys (at a stop and in the vehicle). We want to verify that automatically detected users' places such as home/work correspond to intimate places (Section 5.3.1) and when users are at a stop or in the vehicle providing the crowd our intimacy model predict that users are in low intimacy. We refer to the former as *starting_place_vs_predicted_intimacy* and the latter as *crowd_survey_vs_predicted_intimacy*.

As far as the second goal (B), we have some ideas for the changes we will perform to the interface of the original application. We expect that these changes will influence specific variables such as user engagement time, the number of clicks, UI operations flow, and more. We want to use these variables to identify effects of the new interface and if they correlate to the predicted users' intimacy.

9.2 Conclusions and Open Research Questions About Intimacy

This work explores intimacy in several ways: give its definition, validates its definition, checks that intimacy relates to some smartphone usage patterns, analyzes if we can apply intimacy to other fields like anonymity in MSC, and finally it proposes a computational model to estimate intimacy. We can investigate each of these steps deeply and can open new research questions that we could not answer in this work. Therefore, to guide researchers that may be interested in the topic in the future, we present an open research question for each step.

We start with the intimacy definition. We defined intimacy as new subjective context information that encloses three primary users' context elements: the current place, number and kind of people around (Chapter 2). We included in the definition only three critical context information, but most probably there is room for more. With the analysis we performed in Chapter 5, we proved that definition (as it is) matches the perception of users. The place, number and kind of people around the users are adamant representatives of the users' intimacy. By extending the analysis to users' mood (Section 5.4), we prove that most probably valence (is the situation pleasant or not?) has a role in the intimacy perception. Most probably, as we see from the analysis of the intimacy model also what the person is doing (i.e., user activity) is an important intimacy indicator. Investigate how to integrate these extra variables in the basic intimacy definition can be surely a new research path. Clearly more variables are considered more complex the modeling of intimacy may become. This complexity is another challenge that can force to understand until where to go. Which are the actual boundaries of intimacy? Which are the risks of including more variables in the definition? Shall we stop at the point we arrived and limit our self to the initial definition?

The second point is the study of the effectiveness of the changes to application interfaces with the help of intimacy. In this work, we focus only on determining that some smartphone usage patterns relate to the perception of intimacy (Chapter 6). We also made some assumptions on how users can benefit from actual changes in the user interface with the help of intimacy (Section 6.7). However, we did not investigate if user interface changes dictated by intimacy are beneficial for smartphone users. In our future work (Section 9.1) we already designed and planned a user study with the real mobile application deployed and available to users, but we have not yet the necessary results to derive meaningful conclusions on this subject. Additionally, the future study investigates only a reduced set of possible interface changes. We should consider more variables, and we need to design new evaluation measures. Which are the most effective interface changes based on intimacy and in which conditions? How can we measure and evaluate the influence of intimacy and the efficacy of its intervention on users interfaces? Finally beside user interfaces, more domains of intimacy applicability may be explored. As we did with MSC and anonymity in Chapter 8, are there more domains in which we can apply intimacy? Can we explore more in the MSC domain? We are conscious that to facilitate the task to researchers a more reliable automated intimacy prediction

system should be in place. This system will surely help a wider community to test intimacy in “the wild” and applied to real cases.

With the considerations of the intimacy prediction system, we arrive at our final discussion point. We showed the importance of having a reliable system able to predict intimacy. We think that this should be the first focus of our future work, and we included this aspect in the new user study we proposed in Section 9.1. Our goal is to refine the system to predict intimacy that we presented in Chapter 7 and be sure that it is sufficiently reliable to explore intimacy in practice. The first open question on this subject is to determine how to measure the practical reliability of such system? When our predictions are right, when wrong, when does being wrong is bad or when it does not? Just to answer to these questions it is probably necessary to design new user studies. Once we defined the targets, the main issue is to evaluate the intimacy prediction under real conditions (as we designed in Section 9.1). Furthermore, the scenario of Section 9.1 may not be sufficient to cover all the aspects of the prediction verification. For these reasons, our plan is to deliver finally a software package (open source and in a form of service in Google Play store) that can be used by researchers and developers in their studies and applications. With the feedback from a larger community, many steps can be shortened, and the validation procedure can be faster.

Bibliography

[1] V. Woollaston, “How often do YOU look at your phone? The average user now picks up their device more than 1,500 times a week,” *MailOnline*. [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-2783677/How-YOU-look-phone-The-average-user-picks-device-1-500-times-day.html. [Accessed: 03-Feb-2015].

[2] N. Banovic, C. Brant, J. Mankoff, and A. K. Dey, “ProactiveTasks: the Short of Mobile Device Use Sessions,” in *MobileHCI*, 2014.

[3] R. Likamwa, Y. Liu, N. D. Lane, and L. Zhong, “Can Your Smartphone Infer Your Mood ?,” in *PhoneSense*, 2011, pp. 1–5.

[4] R. Likamwa, Y. Liu, N. D. Lane, and L. Zhong, “MoodScope: Building a Mood Sensor from Smartphone Usage Patterns,” in *MobiSYS*, 2013, pp. 389–401.

[5] N. Eagle, A. S. Pentland, and D. Lazer, “Inferring friendship network structure by using mobile phone data.,” *National Academy of Sciences of USA*, vol. 106, no. 36, pp. 15274–8, Sep. 2009.

[6] B. Adams, D. Phung, and S. Venkatesh, “Sensing and using social context,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 5, no. 2, pp. 1–27, Nov. 2008.

[7] J. Zheng and L. M. Ni, “An unsupervised learning approach to social circles detection in ego bluetooth proximity network,” *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp ’13*, p. 721, 2013.

[8] G. Chen and D. Kotz, “A survey of context-aware mobile computing research,” *Technical report*, pp. 1–16, 2000.

[9] C. Bolchini, C. Curino, and E. Quintarelli, “A data-oriented survey of context models,” *ACM Sigmod*, vol. 36, no. 4, pp. 19–26, 2007.

[10] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, “A survey of context modelling and reasoning techniques,” *Pervasive and Mobile Computing*, vol. 6, no. 2, pp. 161–180, Apr. 2010.

[11] J. Ye, S. Dobson, and S. McKeever, “Situation identification techniques in pervasive computing: A review,” *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 36–66, Feb. 2012.

[12] A. K. Dey, “Understanding and using context,” *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, Feb. 2001.

[13] N. Eagle and A. (Sandy) Pentland, “Reality mining: sensing complex social systems,” *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, Nov. 2005.

[14] C. Ryan and A. Gonsalves, “The effect of context and application type on mobile usability: an empirical study,” in *Research and Practice in Information Technology*, 2005, vol. 38, pp. 115–124.

[15] C. Shin, J.-H. Hong, and A. K. Dey, “Understanding and prediction of mobile application usage for smart phones,” in *UbiComp*, 2012, p. 173.

[16] M. Böhmer, B. Hecht, J. Schöning, A. Kruger, and G. Bauer, “Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage,” in *MobileHCI*, 2011.

[17] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J.-H. Hong, and A. K. Dey, “Factors influencing quality of experience of commonly used mobile applications,” *IEEE Communications Magazine*, no. April, pp. 48–56, 2012.

[18] J. Floch, C. Frà, and R. Fricke, “Playing MUSIC—building context-aware and self-adaptive mobile applications,” *Software: Practice and Experience*, no. 7465, pp. 359–388, 2012.

[19] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, “Diversity in smartphone usage,” in *MobiSYS*, 2010, p. 179.

[20] T. Soikkeli, J. Karikoski, and H. Hammainen, “Diversity and end user context in smartphone usage sessions,” *International Conference on Next Generation Mobile Applications, Services, and Technologies*, pp. 7–12, 2011.

[21] K. Prager, *The psychology of intimacy*. The Guilford Press, 1997.

[22] O. Dictionary, “Intimacy definition,” 2014. [Online]. Available: http://www.oxforddictionaries.com/definition/english/intimacy.

[23] O. Dictionary, “Intimate definition,” 2014. [Online]. Available: http://www.oxforddictionaries.com/definition/english/intimate.

[24] L. C. Manzo, “Beyond house and haven: toward a revisioning of emotional relationships with places,” *Journal of Environmental Psychology*, vol. 23, no. 1, pp. 47–61, Mar. 2003.

[25] D. Seamon and J. Sowers, “Place and Placelessness, Edward Relph,” *Key Texts in Human Geography*, no. Seamon 2000, pp. 43–51, 2008.

[26] S. Saegert, “Crowding: Cognitive Overload and Behavioral Constraint,” *Environmental design research*, pp. 254–260, 1973.

[27] E. Diener and M. E. P. Seligman, “Very Happy People,” *Psychological Science*, vol. 13, no. 1, pp. 81–84, Jan. 2002.

[28] R. S. Miller and H. M. Lefcourt, “The assessment of social intimacy.,” *Journal of personality assessment*, vol. 46, no. 5, pp. 514–8, Oct. 1982.

[29] I. Altman, *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* ERIC, 1975.

[30] A. F. Westin, *Privacy and Freedom*. Bodley Head, 1970.

[31] R. S. Gerstein, “Intimacy and privacy,” *Ethics*, vol. 89, no. 1, pp. 76–81, 1978.

[32] C. Tossell, P. Kortum, C. Shepardb, A. Rahmati, and L. Zhong, “An empirical analysis of smartphone personalisation: measurement and user variability,” *Behaviour & Information Technology*, vol. 31, no. 10, pp. 995–1010, 2012.

[33] K. Geihs and M. Wagner, “Context-Awareness for Self-adaptive Applications in Ubiquitous Computing Environments,” *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 109, pp. 108–120, 2013.

[34] C. Efstratiou, K. Cheverst, N. Davies, and A. Friday, “An Architecture for the Effective Support of Adaptive Context-Aware Applications,” *Lecture Notes in Computer Science*, pp. 15–26, 2001.

[35] K. Geihs, “Self-Adaptivity from Different Application Perspectives,” *Lecture Notes in Computer Science*, pp. 376–392, 2013.

[36] C. Evers, R. Kniewel, K. Geihs, and L. Schmidt, “Achieving User Participation for Adaptive Applications,” *Lecture Notes in Computer Science*, pp. 200–207, 2012.

[37] L. Pei, R. Guinness, R. Chen, J. Liu, H. Kuusniemi, Y. Chen, L. Chen, and J. Kaistinen, “Human behavior cognition using smartphone sensors.,” *Sensors*, vol. 13, no. 2, pp. 1402–24, Jan. 2013.

[38] G. Bauer and P. Lukowicz, “Can smartphones detect stress-related changes in the behaviour of individuals?,” in *Percom*, 2012, no. March, pp. 423–426.

[39] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen, “The Mobile Data Challenge: Big Data for Mobile Computing Research,” in *Mobile Data Challenge by Nokia Workshop*, 2012.

[40] D. Ferreira and A. Dey, “Understanding human-smartphone concerns: a study of battery life,” *Pervasive Computing*, pp. 19–33, 2011.

[41] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, *Experience Sampling Method: Measuring the Quality of Everyday Life*. SAGE Publications, Inc, 2007.

[42] D. Kahneman, A. B. Krueger, D. a Schkade, N. Schwarz, and A. a Stone, “A survey method for characterizing daily life experience: the day reconstruction method.,” *Science (New York, N.Y.)*, vol. 306, no. 5702, pp. 1776–80, Dec. 2004.

[43] K. Wac, “mQoL Living Lab: Experience of Running Longitudinal Studies for Mobile Internet Measurements,” in *ACM Internet Measurement Conference*, 2013.

[44] M. Bradley and P. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *Behav. Ther. Psychiat*, vol. 25, no. 1, pp. 49–59, 1994.

[45] C. Jensen, C. Potts, and C. Jensen, “Privacy practices of Internet users: Self-reports versus observed behavior,” *International Journal of Human-Computer Studies*, vol. 63, no. 1–2, pp. 203–227, Jul. 2005.

[46] A. Braunstein, L. Granka, and J. Staddon, “Indirect content privacy surveys: measuring privacy without asking about it,” *Symposium on Usable Privacy and Security*, 2011.

[47] N. Lathia, K. K. Rachuri, C. Mascolo, P. J. Rentfrow, and U. Kingdom, “Contextual Dissonance : Design Bias in Sensor-Based Experience Sampling Methods,” in *UbiComp*, 2013, pp. 183–192.

[48] J. Lin, S. Amini, and J. Hong, “Expectation and purpose: Understanding users’ mental models of mobile app privacy through crowdsourcing,” in *Ubicomp*, 2012, pp. 501–510.

[49] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall, “A conundrum of permissions: installing applications on an android smartphone,” in *Financial Cryptography and Data Security*, 2012, pp. 1–12.

[50] R. Larson and P. Delespaul, “Analyzing Experience Sampling data: a guidebook for the perplexed,” *The Experience of Psychopathology Investigating Mental Disorders in their Natural Settings*, 1992.

[51] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*. Springer, 2009.

[52] D. Ferreira, J. Gonçalves, V. Kostakos, L. Barkhuus, and A. K. Dey, “Contextual Experience Sampling of Mobile Application Micro-Usage,” in *MobileHCI*, 2014.

[53] A. Oulasvirta, T. Rattenbury, L. Ma, and E. Raita, “Habits make smartphone use more pervasive,” *Personal and Ubiquitous Computing*, vol. 16, no. 1, pp. 105–114, Jun. 2011.

[54] T. Yan, D. Chu, D. Ganesan, A. Kansal, and J. Liu, “Fast app launching for mobile devices using predictive user context,” in *MobiSYS*, 2012, p. 113.

[55] R. Haubo and B. Christensen, “Analysis of ordinal data with cumulative link models — estimation with the R-package ordinal,” 2012. [Online]. Available: http://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf.

[56] M. Fanourakis and K. Wac, “Lightweight Clustering of Cell IDs into Meaningful Neighbourhoods,” in *HET-NETs*, 2013, pp. 698–704.

[57] Ian H. Witten, Eibe Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. 2011.

[58] H. Martinez, G. Yannakakis, and J. Hallam, “Don’t Classify Ratings of Affect; Rank them!,” *Affective Computing*, vol. 5, no. 3, pp. 1–14, 2014.

[59] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker, “Label ranking by learning pairwise preferences,” *Artificial Intelligence*, vol. 172, no. 16–17, pp. 1897–1916, Nov. 2008.

[60] M. Gustarini, J. Marchanoff, M. Fanourakis, C. Tsiourti, and K. Wac, “UnCrowdTPG : Assuring the Experience of Public Transportation Users,” in *The 2nd IEEE WiMob 2014 International Workshop on Smart City and Ubiquitous Computing Applications*, 2014, pp. 1–7.

[61] D. Christin and A. Reinhardt, “A survey on privacy in mobile participatory sensing applications,” *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, Nov. 2011.

[62] A. Pfitzmann and M. Köhntopp, “Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology,” *Designing privacy enhancing technologies*, 2001.

[63] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. a. Peterson, “People-centric urban sensing,” *Proceedings of the 2nd annual international workshop on Wireless internet - WICON ’06*, p. 18–es, 2006.

[64] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. a. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn, “The Rise of People-Centric Sensing,” *IEEE Internet Computing*, vol. 12, no. 4, pp. 12–21, Jul. 2008.

[65] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, and T. Choudhury, “A Survey of Mobile Phone Sensing,” *IEEE Communications Magazine*, pp. 140–150, 2010.

[66] A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data,” *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 8–12, Mar. 2009.

[67] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer Networks*, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[68] C. David, D. Estrin, and M. Srivastava, “Overview of Sensor Networks,” no. August, pp. 41–49, 2004.

[69] A. Oulasvirta, “Finding meaningful uses for context-aware technologies: the humanistic research strategy,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 247–254.

[70] M. Raento, a. Oulasvirta, and N. Eagle, “Smartphones: An Emerging Tool for Social Scientists,” *Sociological Methods & Research*, vol. 37, no. 3, pp. 426–454, Feb. 2009.

[71] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, “Participatory sensing,” in *World Sensor Web Workshop*, 2006, pp. 1–5.

[72] M. Laibowitz, J. Gips, R. Aylward, A. Pentland, and J. a. Paradiso, “A sensor network for social dynamics,” *Proceedings of the fifth international conference on Information processing in sensor networks - IPSN ’06*, p. 483, 2006.

[73] M. Lamming and W. Newman, *Activity-based information retrieval technology in support of personal memory*. 1991.

[74] L. Holmquist, J. Falk, and J. Wigström, “Supporting group collaboration with interpersonal awareness devices,” *Personal and Ubiquitous Computing*, pp. 13–21, 1999.

[75] M. Addlesee, R. Curwen, and S. Hodges, “Implementing a Sentient Computing System,” *Computer*, 2001.

[76] S. Eisenman, N. Lane, and E. Miluzzo, “MetroSense project: People-centric sensing at scale,” in *WSW at SenSys*, 2006.

[77] T. Abdelzaher, Y. Anokwa, P. Boda, J. Burke, D. Estrin, L. Guibas, A. Kansal, S. Madden, and J. Reich, “Mobiscopes for Human Spaces,” *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 20–29, Apr. 2007.

[78] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, “Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application,” in *Proceedings of the 6th ACM conference on Embedded network sensor systems*, 2008, pp. 337–350.

[79] Y. Wang, J. Lin, M. Annavaram, Q. a. Jacobson, J. Hong, B. Krishnamachari, and N. Sadeh, “A framework of energy efficient mobile sensing for automatic user state recognition,” *Proceedings of the 7th international conference on Mobile systems, applications, and services - Mobisys ’09*, p. 179, 2009.

[80] N. Kiukkonen, J. Blom, O. Dousse, and D. Gatica-Perez, “Towards rich mobile phone datasets: Lausanne data collection campaign,” *Proc. ICPS, Berlin*, 2010.

[81] P. Leon, B. Ur, and Y. Wang, “What matters to users?: factors that affect users’ willingness to share information with online advertisers,” in *SOUPS*, 2013.

[82] A. Felt, S. Egelman, and D. Wagner, “I’ve got 99 problems, but vibration ain't one: A survey of smartphone users' concerns,” in *SPSM’12*, 2012, pp. 33–43.

[83] J. S. Olson, J. Grudin, and E. Horvitz, “A Study of Preferences for Sharing and Privacy,” in *CHI*, 2005, pp. 1985–1988.

[84] B. Knijnenburg and A. Kobsa, “Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems,” *ACM Transactions on Interactive Intelligent Systems*, 2013.

[85] K. Hawkey and K. Inkpen, “Keeping up appearances: understanding the dimensions of incidental information privacy,” in *CHI*, 2006.

[86] S. Lederer, J. Mankoff, and A. K. Dey, “Who wants to know what when? privacy preference determinants in ubiquitous computing,” in *CHI ’03*, 2003.

[87] A. Khalil and K. Connelly, “Context-aware telephony: privacy preferences and sharing patterns,” in *CSCW*, 2006, pp. 469–478.

[88] F. Shih and J. Boortz, “Understanding People’s Preferences for Disclosing Contextual Information to Smartphone Apps,” *HAS/HCII*, pp. 186–196, 2013.

[89] R. Ganti, F. Ye, and H. Lei, “Mobile crowdsensing: Current state and future challenges,” *Communications Magazine, IEEE*, no. November, pp. 32–39, 2011.

[90] M. Oswald, “Something bad might happen: Lawyers, anonymization, and risk,” *XRDS: Crossroads, The ACM Magazine for Students*, vol. 20, no. 1, p. 22, Sep. 2013.

[91] K. Shilton, “four Billion Little Brothers ? Privacy, mobile phones, and ubiquitous data collection,” *Communications of the ACM*, 2009.

[92] T. Hara, Y. Arase, A. Yamamoto, X. Xie, M. Iwata, and S. Nishio, “Location anonymization using real car trace data for location based services,” *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication - ICUIMC ’14*, pp. 1–8, 2014.

[93] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, “PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems,” in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.

[94] M. Mun, S. Hao, N. Mishra, and K. Shilton, “Personal data vaults: a locus of control for personal data streams,” in *CoNEXT*, 2010.

[95] G. F. Jenks, “The Data Model Concept in Statistical Mapping,” *International Yearbook of Cartography 7*, pp. 186–190, 1967.

[96] B. Gedik and L. Liu, “Location privacy in mobile systems: A personalized anonymization model,” *Computing Systems, ICDCS*, pp. 620–629, 2005.

[97] M. Gustarini and K. Wac, “Smartphone Interactions Change for Different Intimacy Contexts,” in *Mobicase*, 2013, pp. 1–18.