KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Contextual Privacy: The Interplay of Sensitivity and Context

Rula Sayaf

Supervisor:
Prof. Dr. Dave Clarke
Prof. Dr. Ir. Frank Piessens,
co-supervisor

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering Science: Computer Science

June 2016

Contextual Privacy:
The Interplay of Sensitivity and Context

Rula SAYAF

Examination committee:
Prof. Dr. Ir. Joos Vandewalle, chair
Prof. Dr. Dave Clarke, supervisor
Prof. Dr. Ir. Frank Piessens, co-supervisor
Prof. Dr. Bettina Berendt
Prof. Dr. Claudia Diaz
Prof. Dr. Ir. Wouter Joosen
Prof. Dr. Jo Pierson
(Vrije Universiteit Brussel)

Dissertation presented in partial
fulfillment of the requirements for
the degree of Doctor in Engineering
Science: Computer Science

June 2016

Acknowledgements

> A teacher affects eternity. He can never tell where his influence stops.
>
> *Henry Brooks Adams*

The research and writing of this dissertation has been a joyful process. I have received brilliant mentorship from many people. Nevertheless, I alone am responsible for any possible mistakes. I am deeply grateful to the following mentors and friends.

Professor Dr. Dave Clarke is the Promotor *Par Excellence*. He is a genuine mentor and goes the *Extra Mile* in helping his students. I am eternally indebted to him for giving me the opportunity to become his student. He is like a parent nurturing my intellectual growth and development.

I am thankful to Professor Dr. Frank Piessens, Co-Promotor. He provided me with valuable advice about my research. He made it possible for me to be simultaneously a PhD student at KULeuven and as a visiting researcher in Sweden. Professor Dr. Wouter Joosen has provided invaluable support and encouragement over and above the call of duty for which I thank him here and now. Special thanks to Professor Bettina Berendt for being a mentor and a role model. Her creative and inspirational ideas had a significant role in shaping the path to my PhD. Her ethics, dedication, and benevolence are invaluable. I would also like to express my appreciation to Seda Güses for steering me in my research, and for being so supportive. *Tacksåmycket* for the kindness and support of the Uppsala university staff, the head of the department of the Information Technology Professor Michael Thuné, Professor Dr. Tobias Wrigstad, Ulrika Andersson, and Anna-Lena Forsberg.

I am also grateful to the members of the PhD jury for their feedback. Their critique made me reflect and generate new ideas.

I am particularly appreciative to many iconic researchers at the Microsoft Research in Cambridge, England. John Guiver has generously provided valuable guidance and feedback. Richard Harper has mentored my work, helped, and encouraged me to think outside *The Box*. Natasa Milic-Frayling has shown incredible guidance and support in developing and critiquing my research. I was honoured to meet, discuss, and learn many things from Sir Tony Hoare. I am grateful for his interest in my work and all the inspiration he has given me. The unique opportunity to work at Microsoft has been made possible by my mentor Sören Preibusch. I am grateful for his guidance which made my research remarkably exciting and fun.

A healthy and essential part of pursuing a PhD is having the love and support of good friends. I feel blessed to have had many homes, families, and friends in Sweden, Belgium, the UK, the USA, and Syria. Moving to Sweden could not have been a delight without Adriaan Larmuseau. He is insightful, caring, witty, and wonderful. Adriaan is like a “flashlight" in darkness. “Never change, Rula", he counselled. Likewise, Adriaan! Thanks to my colleagues at Uppsala University Albert Mingkun Yang, Francisco Fernandez Reyes, Stephan Brandauer, Kim-Anh Tran, Huu-Phuc Vo, and Elias Castegren; thanks for the many lovely *fikas*.

Heart-felt thanks to my 16-year old friends Iyad Zikra and Samer Almoubayed, and 2-year old friend Susanne Gylesjö for being my family in Stockholm. You are a source of warmth in the northern winters. Waseem Aldahan, thanks for adding joy and hope to the Stockholm family.

The immense contribution to this dissertation comes from my big family and friend in Belgium. *Ik will graag al mijn KULeuven vrienden en vriendinnen bedanken omdat ...* you provided required support, humour, stability, and love. *Dank u* Philippe De Ryck for being such a wonderful, and shining-star friend. *Dank u* Raoul Strackx, your candour, and warm spirit makes you a precious friend. *Dank u* Katrien Janssens for your “lily-white" you. You welcomed me warmly in your loving family. *Hartelijk bedankt* to your marvellous daughters Sara and Lisa.

I would also like to thank my office mates; thank you Koosha Paridel for bringing smiles and positive energy; thank you Dimiter Milushev for your brotherly support and encouragement; thank you Ilya Sergey for sharing your experience to help and make my PhD life easier, thank you Pieter Agten for adding up to the perfectly enjoyable work days. Special thanks to Arun Kishore Ramakrishnan, you are a precious friend.

Endless thanks to Professor Dr. Ansar Yasar for your unlimited support, and for being my family together with your loving wife Sehrish Karamat. Wholehearted thanks to Dimitar Shterionov for being an ultimate generous friend. Thank you Syeda Nayyab Zia Naqvi for being the kind friend with whom the PhD years became delightful. Thank you Milica Milutinovic for being a gentle and caring friend. Thank you Caren Crowley for your friendship, and thanks to you and Professor Dr. Danny Hughes for your inspiration. Thank you Ero Balsa, you are a dear friend. Thank you Mario Henrique Cruz Torres for your kindliness. *Dank u* Frédéric Vogels for being an meticulous friend providing joy, and chocolate. *Dank u* Davy Preuveneers for your valuable advice. *Dank u* Jesper Cockx for your gentleness, and for particularly transforming the abstract into a proper Dutch text.

Thanks to many motivating, fun, interesting stories, and coffee breaks shared with José Proenca, Nessim Kisserli, Bo Gao, Wang Yuyi, Steven Van Acker, Mathy Vanhoef, Jan Tobias Mühlberg, Neline van Ginkel, Ping Chen, Kim Wuyts, Marco Patrignani, Radu Muschevici, Nick Nikiforakis, Dominique Devriese, Bert Lagaisse, Lieven Desmet, Sven Akkermans, Willem De Groef, Andreas Nuyts, Tom Van Goethem, Aram Hovsepyan, Koen Yskout, Thomas Heyman. Also, thanks are due to Ghita Saevels, Annick Vandijck, Karen Spruyt, and Marleen Somers for arranging the practical aspects required especially for my frequent travels.

I would like to thank numerous persons have provided infinite support from different places in the world. Thanks to Wacek Kusnierczyk for your precious friendship. Thanks to Valerie Elliot, she is like my second mom, a loving spirit in Cambridge. Thanks to Bashar Awwad, Noura Almoubayed and your loveable little daughter Julie for being my family in the UK. Thanks to 3-year old Roo Dave Clarke for the many smiles she brought to my life. Thanks to John Knott, and Jolien De Decker for your support. Thanks to Billy Lee, our benevolent “Godfather", and his family. Thanks to my friend Maridel Andres in Colorado.

Thanks to many people in Syria. Thanks to my spiritual father Igantios Darouj for encouraging my creativity. Thanks to my cousin Rita Farah for your beautiful heart. Thanks to my beloved cousin Tuleen Farah. Thanks to the family members: Susan, Amal, Ellen, Firas, George, Afifa, Afif, Rana, Fadi, Fahd, Jan, Mary, Souad, Entoinet, and Raif in Canada. Thank you to my best friend Nivine Tama, my twin soul mate in Washington DC.

Finally, I would like to lovingly dedicate this dissertation to my promotor Professor Dr. Dave Clarke, and my family: my constant companion in the daily grind my brother Habib, Yamen, Daleen, Nahi and Wadi my parents, to my dearest grandpa Habib.

Abstract

Privacy management is becoming a fundamental task of everyday use of the Web. People often disclose sheer amounts of data on the Web. In different web services, and in particular social software, people communicate through data disclosure. By selecting what data to disclose and to whom, people build and manage their online identities. The disclosed data can vary in its sensitivity. Handling sensitive data inappropriately or disseminating it in inappropriate contexts can drastically affect users' identities, privacy and lives. To avoid data misappropriation, a high degree of control over data and context is required. Through contextual privacy, users could have such control.

Contextual privacy management can be a complicated process. It requires reasoning about data and context changes. It also requires assessing the sensitivity of a data item and how it might change when context changes. Due to its complexity, most technological approaches offer a simplistic and limited degree of contextual privacy management. A fundamental step towards addressing this complexity without limiting the degree of contextual privacy is investigating the relationship between data, context and privacy.

The approach of this thesis is a multidimensional investigation of contextual privacy. Firstly, the investigation is performed from an empirical point of view. Through big data analyses and machine learning, we investigate the effect of context on data sensitivity and users' behaviour. The analyses show that sensitivity cannot be defined by what is commonly considered as sensitive topics, e.g., sex and health. The modelling of sensitivity management behaviour demonstrates that sensitivity is affected by context as well. The modelling demonstrates also the effect of time and subjectivity on data sensitivity. Moreover, our analysis demonstrates the effect of context on data disclosure patterns.

Secondly, the investigation involves a conceptual examination of the role of context in communication. This investigation highlights the role of

context in facilitating the interpretation of disclosed data and estimating its sensitivity. We propose controlling data sensitivity and interpretation to manage contextual privacy. We propose facilitating the inference and management of these ingredients and context by machine learning tools. The inference would facilitate the automatic monitoring of changes of data sensitivity and interpretation to identify misappropriation attacks. Through this approach, contextual privacy management can be effective without overloading users.

Thirdly, the investigation extends to analyse contextual privacy in the legal framework. This analysis compares how privacy is tackled in the technical and legal frameworks to assess the possible degrees of control, privacy and surveillance. It puts forward criteria to assess these degrees. The analysis shows the interdependence between privacy and surveillance in both frameworks.

In summary, the thesis puts forward an extensive exploration and analysis of contextual privacy. It decomposes contextual privacy and shows through big data analysis that it is an interaction between data sensitivity and context. The information provided in this thesis could contribute to developing usable and effective contextual privacy management mechanisms.

Beknopte Samenvatting

Privacybeheer is een fundamentele taak aan het worden voor het alledaags gebruik van het web. Mensen geven vaak grote hoeveelheden gegevens vrij op het web. Op verschillende webdiensten, in het bijzonder op sociale media, communiceren mensen door middel van het vrijgeven van gegevens. Door te kiezen welke gegevens ze openbaren en aan wie, bouwen en onderhouden ze hun online identiteiten. Deze vrijgegeven gegevens kunnen variëren in hun gevoeligheid. Het ongepast omgaan met of verspreiden van gevoelige gegevens kan de identiteiten, privacy en levens van de gebruikers echter drastisch beïnvloeden. Om het ongewenst vrijgeven van gegevens te voorkomen is een hoge graad van controle over gegevens en context noodzakelijk. Contextuele privacy geeft de gebruikers zulke controle.

Contextueel privacybeheer kan een ingewikkeld proces zijn. Het vereist dat we kunnen redeneren over gegevens en contextveranderingen. Het vereist ook een beoordeling van de gevoeligheid van de gegevens en hoe deze wordt beïnvloed door de context. Door deze complexiteit bieden de meeste technologische benaderingen een beperkte en simplistische mate van contextueel privacybeheer. Een fundamentele stap om deze complexiteit aan te pakken zonder de mate van contextuele privacy te beperken, is om het verband tussen gegevens, context en privacy te onderzoeken.

Deze thesis pakt dit probleem aan door middel van een multidimensioneel onderzoek naar contextuele privacy. Ten eerste wordt het onderzoek uitgevoerd vanuit een empirisch standpunt. Door middel van big data-analyses en machine learning onderzoeken we het effect van de context op de gevoeligheid van gegevens en het gedrag van de gebruikers. Deze analyses tonen aan dat gevoeligheid niet gedefinieerd kan worden aan de hand van wat doorgaans als gevoelige onderwerpen worden beschouwd, bvb. geslacht en gezondheid. Het modelleren van het gedrag van hoe gebruikers hun gevoelige gegevens beheren, toont aan dat de gevoeligheid wordt beïnvloed door de context. Het modelleren toont ook het effect aan van tijd en subjectiviteit op de gevoeligheid.

Onze analyse toont bovendien het effect aan van de context op patronen van onthulling van gegevens.

Ten tweede houdt de studie een conceptueel onderzoek in naar de rol van context in communicatie. Dit onderzoek belicht de rol van context in het ondersteunen van de interpretatie van vrijgegeven gegevens en het inschatten van hun gevoeligheid. We stellen voor om de gevoeligheid en interpretatie van gegevens te controleren om contextuele privacy te beheren. We stellen tevens voor om werktuigen van machine learning toe te passen om het beheer en de gevolgtrekking van deze eigenschappen en hun context te ondersteunen. De gevolgtrekking zou het automatisch opvolgen van wijzigingen van gevoelige gegevens en interpretaties ondersteunen om identiteitsdiefstallen te detecteren. Door deze aanpak kan contextuele privacy effectief beheerd worden zonder gebruikers te veel te belasten.

Ten derde wordt het onderzoek uitgebreid tot de analyse van contextuele privacy in het wettelijk kader. Dit onderzoek vergelijkt hoe privacy aangepakt wordt in technische en juridische kaders om de mate van beheer, privacy en toezicht in te schatten. Het stelt ook criteria voor om deze mate in te schatten. Deze analyse toont de wederzijdse afhankelijkheid aan tussen privacy en toezicht in beide kaders.

Samengevat geeft deze thesis een uitgebreid onderzoek naar en analyse van contextuele privacy. Het breekt contextuele privacy op en toont aan door gebruik te maken van big data-analyse dat het een interactie is tussen de gevoeligheid van gegevens en context. De lessen uit deze thesis kunnen leiden tot de ontwikkeling van bruikbare en doeltreffende mechanismes voor het beheer van contextuele privacy.

Glossary and Abbreviations

AIC	Akaike Information Criterion
CAP	Context-approximation Parameters
Context	Any information that can be used to characterise the situation surrounding a data item
CPML	Contextual Privacy Management Layer
CPS^2	Contextual Privacy Framework for Social Software
Deletion Pattern	The information that characterise how data is deleted from search history. The information may describe the topic of data, the context surrounding the data or the user
Functional Surveillance	The monitoring and surveillance required for operating data control approaches to offer privacy management
Identity	The information that characterise only one individual. An online identity is the information extracted from the data this individual discloses on the web.
LTP	Long-term Pattern
Offline Context	Any information that can be used to characterise the real-world situation surrounding a user
Online Context	Any information that can be used to characterise the situation on the web surrounding a data item
OR	Odds Ratio
PaC	Privacy as Control
PCA	Principal Component Analysis
SD	Standard Deviation

Sensitivity	The degree of inappropriateness of handling a data item by a certain party or in a certain context
Sensitivity Management Pattern	The information that characterise how sensitive data is managed by users. The information may describe the topic of data, the context surrounding the data or the user
Social Software	Application software for the exchange of personal data and social interaction with a large number of users
STP	Short-term Pattern
Trust	The belief of the truster that the trustee would act in the truster's best interest [53]
UP	User-specific Pattern

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Privacy on the Web

The recent advances in technology have changed our lives and made us increasingly dependent on web services. The web incorporates a wide spectrum of services relevant to almost every single daily activity. There are many types of online activities that people perform daily, such as emailing and searching the web. At the same time, people can use the web to complement offline activities, such as finding the nearest restaurant given a particular location. Moreover, people communicate and even socialise using web-based social software. Through using social software, users can foster relationships and connect to other users. Users can create and manage one or more online identities. Such identities may mirror or differ from the real identities of users. Consequently, the web is becoming an essential part of every day life.

The dependence on web services may come at a price of privacy. The wide spectrum of web services makes performing daily activities much more easier than ever before. Users disclose different data depending on the web service. In social software, users upload data to communicate with each other. This data becomes available online to facilitate further interaction with the web. By having one's data available online, different parts of one's life becomes available online. When users are not aware of who can access the data and how it is handled, the user's life can be affected. The availability of data online is risky if the user does not have the proper means for privacy management. In some cases, privacy risks can have drastic consequences. An example is losing one's job when the employer accesses the employee's data that reveals an

improper behaviour. Such a behaviour may not necessarily be related to work, e.g., smoking marijuana at a party. Yet, the availability of such data online increases privacy risks, and hence, undesired and unexpected consequences. The problem in many web services is that users do not have sufficient technical means to protect their privacy on the web.

In this thesis, we focus on technical issues of privacy on the web. We focus mainly on social software because personal data is often disclosed via such software. Personal data can reveal private details about users. Disclosures of personal data can be coupled with privacy issues. Addressing privacy issues in social software requires exploring the different privacy-related dimensions. These dimensions involve social software communication, data sensitivity, and the causes of privacy issues.

1.1.1 Social Software

Social software is the class of web services that facilitates online communication and building online communities. With social software, users can create profiles reflecting their identities. Users can connect to each other and foster social relationships. Relationships link users who may be friends in the offline world or strangers. Users communicate through social software via disclosing their personal data. A user can disclose a data item to communicate a particular messages a particular audience. The type of disclosure and communication varies based on the design of the software. One variant of social software (Facebook-like software), such as Facebook and Google+, are designed to allow users disclose data to the public or to a particular audience. Another variant facilitates public data disclosure such as Twitter, LinkedIn, blogging services, peer-to-peer, collaborative and content sharing sites such as Youtube and Flickr, and social bookmarking services such as CiteULike.

Using social software can be coupled with high privacy concerns. Web users can disclose personal or non-personal data. *Personal data* is "any information relating to an identified or identifiable natural person [...]; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity", according to the EU Directive 95/46/EC [EU, 1995], Art. 2 (a) [38]. In social software, users mostly share personal data. Personal data can vary in its sensitivity. According to the sensitivity of a data item, the user may have various concerns regarding who can view the data, in which context, and how it is handled. In contrast, when a user shares non-personal data, privacy risks can be less due to the fact that the data does not identify this user.

The ease of communication via social software gives to rise to privacy concerns. In this communication, the user can communicate with friends, and strangers. When communicating with strangers, the user may not be able to anticipate how a data item will be handled. Consequently, the user may encounter privacy violations. Even when the user communicates with friends who are trusted, they could behave inappropriately towards the user's data and privacy. To facilitate communication and avoid privacy issues, users are offered technical means to control their data and manage their privacy.

1.1.2 Privacy as Control

"Privacy as control" (PaC) is a one of three research paradigms proposed by Gürses. This class focuses on approaches that offer privacy management through control [48]. PaC focuses on approaches that offer the possibility to control aspects related to privacy. Such control may involve controlling data, the audience who can access it or are prohibited from accessing it, and various parameters concerning the contexts in which data is put. Privacy as confidentiality focuses on approaches that offer the possibility to hide and keep data confidential. Privacy as practice focuses on approaches facilitate the assessment of the constructed identity.

PaC is a broad paradigm that subsumes other privacy management approaches. Besides PaC, Gürses defines another two paradigms, namely, *privacy as confidentiality* and *privacy as practice*. These paradigms refer to approaches that offer privacy management through keeping data confidential and practicing data and identity management. However, these two paradigms are also based on control. In privacy as confidentiality, users have control over hiding or revealing data. In privacy as practice, users have control to observe and assess how their identities are constructed, and can intervene when the construction is inappropriate. In our view, these two paradigms are sub classes of PaC. For this reason, we focus on PaC approaches in this thesis.

In particular, we focus on contextual privacy within the PaC paradigm. PaC approaches facilitate controlling data or the situations or contexts in which data is disclosed. The term *contextual privacy* refers to privacy management through context control. Different approaches offer different methods to control context. Context is the informational construct that describes the online or offline situation surrounding the user and the data. The complexity and offerings of each context-based approach vary. In this thesis, we investigate contextual privacy, the related problems, and possible solutions.

1.2 Problem Statement

Privacy management can be a complicated task for social software users. A user often posts a high number of data to a variant number of users. A data item can be disclosed in a private space to a particular set of recipients; alternatively, it may be posted publicly for a large, and a priori unrestricted, audience. Given that humans reason based on context [41], privacy management requires reasoning about context. Thus, decisions related to what data to disclose, where to disclose and with whom are context-dependent. To manage privacy, the data owner needs to control every single data item, the audience who can access it, the context in which the data is handled. However, it can be complicated for users to control these various aspects all at once.

Controlling context to manage privacy is in particular, a challenging task. Many accounts have confirmed the role of context in how users control their privacy [66, 75, 79, 117]. Due to the high dimensionality of context and its complex nature [109], controlling it can be complicated. Moreover, in social software, context includes data from the real (offline) and online worlds, possibly from different users, making it challenging to separate one context from another. The resulting uncertainty and ambiguity of online context makes controlling context more challenging. Additionally, the data owner has to control contexts in which the data could be disseminated. Such type of control requires knowing the possible contexts in advance in order to list the ones that are appropriate and inappropriate, depending on closed- or open-world assumption of possible contexts. Given the theoretically infinite complexity of the social situation and its context that can be one of an unbounded number of possible contexts [108, 99], it is not feasible for users to exhaustedly provide the list of all contexts.

Most technological approaches are not sufficient to offer effective contextual privacy management. In a previous study of privacy management approaches, we reviewed many approaches that offer context control [90]. However, these approaches have various problems. Some approaches lack simplicity for average users to use. Other approaches simplify context to offer simple control. Such a simplification results in approaches that do not offer a high degree of control and protection.

A particular class of privacy violations that are related to context control and is challenging to mitigate is data misappropriation. After the data subject discloses data in a particular context, this data can be disseminated further by others. When data is disseminated into an inappropriate context, the data is said to be *misappropriated*. Data misappropriation can result in privacy violations. An example is when Alice shares her breastfeeding photo with

public and then Bob disseminates her photo from the context she put it in into *pornographic* context. Such dissemination is not appropriate to Alice and can affect her identity, and therefore her privacy. However, not every act of data dissemination has to affect privacy, e.g., when Alice's photo is disseminated into a context of *mothers and babies*.

Protecting against data misappropriation is rather challenging for users. To avoid data misappropriation, a user has to handle the complexity of controlling contexts in which the data could be disseminated. Also, the user may need to monitor the actions of others on their data, which can be viewed as surveillance of others. To understand the complexity of monitoring the actions of others and reasoning about them, the task can be viewed as a Bayesian game of incomplete information. In such a game, the players, or users, have incomplete information about the actions of others. Players can be said to be acting rationally when an equilibrium is achieved in a Bayesian game [44]. Acting rationally means that each player has the best strategy in response to other players' strategies. In other words, each player has the best strategy towards an attacker's strategy. However, it has been showed that reaching an equilibrium is NP-complete [44]. Thus, such reasoning is highly complicated for users. To protect privacy effectively, it should be possible to detect data misappropriation without burdening users with context control and the monitoring of the actions of others. A first step towards achieving that is understanding the effect of context on data is. Such an understanding would show when misappropriation occurs, how to depict it, and mitigate it.

We focus on the following research questions towards offering better context-based or *contextual privacy* management:

- How does context affect data and privacy?
- How do users manage their privacy given different contexts?
- How to detect and protect from data misappropriation?
- How to assist users in the burden of managing contextual privacy given the large amount of data and size of audience in social software?
- Is it possible to offer a high degree of control to achieve a high degree of privacy without limitations, whether technically or legally?

Towards investigating these questions and enhancing contextual privacy management, we adopt the following approach.

1.3 Approach

The approach in this thesis is based on a multidimensional investigation of context and privacy. To better identify the role of context role in privacy management, we perform empirical and conceptual analyses of context-based privacy management.

The empirical analyses facilitate inferring information about how users behave and manage their privacy in different contexts. To perform such analyses, we need to have access to a large amount of data, or *big data*. Through big data analysis, it is possible to analyse various aspects of privacy management. Firstly, we investigate privacy management patterns and user subjectivity. Secondly, we investigate the sensitivity of data. Thirdly, we investigate the effect of context on sensitivity management.

We perform conceptual analyses to understand the role of context in communication and privacy. We mainly discuss the dominant issue of context unclarity and ambiguity in social software [17]. This analysis aims at understanding data misappropriation to identify the relevant attacker model. Based on the analyses we perform, we propose a conceptual framework for contextual privacy management that addresses the problems we discussed above.

In the following, we summarise the research approach towards proposing a contextual privacy management approach.

1.3.1 Privacy and Big Data

To develop privacy management approaches, it is required to observe how users manage their data using technological tools offered by web services. Privacy is a domain of an interdisciplinary nature. Privacy is studied in social sciences, psychology, law, and computer sciences. There exists a relatively large number of studies related to humans and privacy perception and issues, in social and psychological disciplines. Comparatively, there is relatively a small number of empirical research that investigates humans and privacy. In many cases, privacy management approaches, are security management approaches that are adapted to social software. Examples are access control and encryption approaches that are originally for security management, but they are adapted for privacy management in social software [2]. Given the difference between privacy and security, it is required to collect empirical data about how users handle their data and privacy in order to propose the appropriate approach.

The main challenge of analysing privacy behaviours is the difficulty of provisioning relevant data. Analysing privacy behaviours requires accessing data about how users handle their sensitive and non-sensitive data in order to understand how privacy is managed. Such a dataset is difficult to access due to avoid exploiting users. Mostly, access to a dataset is limited to big service providers. Such providers have big data sets that describe a large number of users, and their behaviour in different situations. Fortunately, we got access to Microsoft's internal data from Bing the search engine. The data set represents 226,000,000 deleted and kept searches disclosed by 413,000 users.[1] The dataset encompasses a wide spectrum of topics and various contextual information that describe web searches issued to Bing. Searches of a user are kept in the search history and are used for serving new searches. Past searches can be displayed while entering a search to help the user in case the search has been issued before. The display of some searches may cause a privacy issues for the user. By studying the patterns of disclosing, deleting and keeping search items, we indirectly analyse privacy management behaviour in web search. The analysis explores the effect of context and content on disclosure and deletions of data, and hence, on privacy management.

To avoid overgernalisation, we refine our analysis of privacy to an analysis of data sensitivity. Because the Bing dataset does not include information about users' motivations, we avoid interpreting deletions as privacy management behaviour. Deletions can be random or motivated by a particular concern of the user due to the sensitivity of data. These concerns may not necessarily be privacy concerns. Rather, they can be related to the appropriateness of using the data, e.g., the appropriateness of using data for ads. Our hypothesis is that deletions are motivated by sensitivity of data. Commonly, topics related to sex, health, and finances are considered sensitive. Our analysis evaluates whether topics that are commonly considered sensitive are sufficiently informative and can explain deletions of searches. By investigating deletions of searches, it is possible to infer what makes data sensitive, and what affects it. Through this investigation, we discuss whether it is valid to link sensitivity to privacy in Chapter 2. Next, we elaborate on our analysis of sensitivity.

1.3.2 Sensitivity

Studying sensitivity is a rather challenging task. Since the Joint Computer Conference 1967, the difficulty of identifying sensitive data had been realised [50]. The early solutions were to classify military-related data as sensitive and non-military-related data as insensitive. Since then, there has been

[1] This work has been partially done during my internship at Microsoft–Cambridge.

more focus on identifying sensitive data in non-military-related contexts [50]. In reality, sensitivity may vary based on data type and content, as well as on the surrounding context. Investigating sensitivity patterns by interviewing people is ineffective and provides limited information. Alternatively, it is possible to investigate sensitivity patterns by analysing a large dataset that describes a wide spectrum of data in varying contexts. Ultimately, investigating sensitivity could be performed through quantitive and qualitative analyses.

We analyse the effect of context on data by assuming a latent variable of data that represents its sensitivity. Sensitivity indicates the degree of concern a user has when her data is accessed by others, whether privacy-related or otherwise. Sensitivity can be also viewed as the degree of appropriateness of putting an item in certain situation and giving access to certain audience. The latent nature of sensitivity means, by definition, that observing this variable is not possible in most cases. Rather, it is possible to estimate this variable through other variables that can be observed—following a latent structure analysis approach [62]. In our case, sensitivity is a variable that affects actions on data, whether disclosure or post-disclosure related. Post-disclosure actions refer to the set of actions performed on data after disclosing it on the web. For instance, sensitivity affects deletions and privacy management actions on data. By observing the actions on data, it is possible to infer information about sensitivity. However, the possible actions on data in our dataset are rather limited. These actions are data disclosure, or submitting data to the search engine, deleting or keeping the data. The observation benefits from large amounts of data to understand sensitivity in different situations.

We utilise machine learning techniques to investigate whether it is possible to infer an accurate mathematical model of sensitivity. The aim of our analysis is to investigate whether, firstly, content and contextual features affect sensitivity. If there is such an effect, then we investigate the underlaying model. We choose to learn the sensitivity pattern from our dataset by modelling how data is deleted or kept. The pattern indicates when data is deleted, and thus, when it is sensitive. We also investigate the effect of time and subjectivity on sensitivity. The investigation provides evidence that context, content, time and subjectivity affect sensitivity, and hence privacy management.

1.3.3 Context

A fundamental aspect of understanding the effect of context on sensitivity is analysing the effect of each context separately, as well as the effects of all contexts together. We model context in terms of categories that represent situations. Each category can have multiple values. An example is the category

Home that represents the *Home* context. This category has two alternative values representing whether the user is *at home* or *at work*. We are interested in examining whether the context effect differs based across its possible values. Additionally, we are interested in observing the effects of all contexts at once. Our dataset describes actions of data disclosure—submitting searches—and post-disclosure—managing searches by deletions. We analyse the effect of individual context categories on data disclosure. We analyse the effect of multiple context categories at once on post-disclosure actions.

Firstly, we quantify the effect of context on disclosure patterns. We consider a disclosure pattern to be represented by the counts of items across the content features per a context category, i.e., the counts of items across the content features in the context category *Home*. We quantify the effect of context by analysing the patterns of each value of one context at a time. We test whether the effect of the values of one context varies significantly, i.e., whether the context value *at home* affects the disclosure pattern significantly different from how the value *at work* affects the data. We apply the test of homogeneity (χ^2) to investigate whether the tested patterns are drawn from the same distribution, which is the null hypothesis. Rejecting the null hypothesis provides evidence that the context values affect the disclosure of data differently.

We also analyse the effect of context on post-disclosure patterns. By modelling the deletion pattern, discussed above, we also investigate the effect of all context categories on sensitivity. The model shows what contexts—contextual features—affect deletions, and sensitivity. To quantify the effect of context, we investigate whether adding more contextual information would affect the inferred pattern. In particular, we compare the two patterns by adding extra contextual parameters that describe the user offline context, e.g., occupation, proxy type, etc.

A relevant research question is whether it is possible using the information about data content to predict information about context. The analysis discussed above investigates the dependency of content on context. It is equally important to investigate whether the context depends on content. The dependency means that by knowing the data content that is disclosed, it is possible to predict the context in which the data was disclosed. We use the disclosure patterns to model the probabilities of being in a particular context category using content features. For this purpose, we utilise the multinomial logit modelling [5].

We also review the literature on the role of context in communication. We mainly focus on two extreme ends of the communication spectrum, namely, cooperative and adversarial communication. These two ends incorporate varying roles of context, trust and privacy degrees [53]. In general, when the data is put in a context, it is possible to infer the relevant interpretation

of data, and thus the communicative message. Similarly, it is possible to infer the sensitivity of data. We also discuss how ambiguous contexts affect communication and privacy by affecting the inference of the interpretation, and sensitivity of data.

1.3.4 Subjectivity

We investigate whether subjectivity plays a role in sensitivity and privacy management. In our investigation of the possibility to model the data, the focus is on finding a general model that describes the behaviour of users. However, if subjectivity has an effect on how users behave, the individual patterns that describe the behaviour of individual users may vary. By taking a sample of users and inferring their individual sensitivity management patterns, we investigate the effect of subjectivity.

1.3.5 The Attacker Model

An essential aspect of developing privacy management for online communication is modelling the attacker the approach should protect against. Traditionally, privacy management approaches are not based on the explicit modelling of attackers. Attackers are generally any user who is not given permission to access data. Privacy approaches usually facilitate delivering data to the appropriate audience. The appropriate audience are trusted to handle the data in a manner that is appropriate to the data owner. The attacker model in most approaches implicitly refers to any user the data owner does not grant access to. There are cases, however, where even the authorised and appropriate audience could perform an inappropriate action, or an attack, after accessing the data. For this reason, it is required to define attackers, and have means to detect them, whether they are authorised or not. A first step towards achieving that is modelling the attacker to protect against and detect.

To identify the attacker model in online communication, it is required to identify the role of data disclosure in communication. By reviewing the literature of communication, we infer that data disclosure aims at delivering a particular communication objective. The misappropriation of data is an act that affects the communicative message when the context changes. The attack is, hence, any act that affects the communicative message. In the following is a high level model of the data misappropriation attacker:

- A (trusted) system: a social software system that facilities social communication functions. The system enforces users' privacy policies and allows actions that are not prohibited otherwise by the data owner.
- A data owner: a user who discloses a data item to communicate a message in a private or a public space.
- An attacker: a user who can access the data item, and by performing a particular action the context is changed into an inappropriate one. The change can be achieved by putting the data in a new context, or by causing the current context to evolve by an action that adds or removes data from the context. An example is when Alice posts her *breastfeeding* photo in a *breastfeeding* context, and Bob (the attacker) changes the context to a *pornography* context by adding a comment that changes the conversation topic.

Alternatively, misappropriation attacks can result by users' actions that unintentionally affect the communicative message. We refer to such misappropriations as unintentional attacks.

1.3.6 CPS^2: a Contextual Privacy Framework for Social Software

We present a conceptual framework to facilitate communication to protect against misappropriation without burdening users with the management of context. Based on our attacker model and the role of context, we propose to decrease the complexity of controlling context by managing the possible interpretation of data. The framework provides means to ensure that the interpretation of data is appropriate in any context. The framework proposes to lift the burden of reasoning about context to the level of the social platform. In theory, misappropriation attacks can be prevented by monitoring actions on data. In practice, monitoring and detecting attacks requires complete information about all users' actions. The framework assumes the utilisation of artificial intelligence approaches to monitor users and detect misappropriation attacks.

We propose a contextual privacy management framework to maintain the appropriate interpretation and sensitivity of data. An attack results when the data is put in a context in which the interpretation changes. To counter attacks, the framework offers can allow users to specify the appropriate interpretation upon disclosure. When the interpretation of an item is different to what the user specified, an attack is detected. Another form of managing contextual privacy is without the need to specify the interpretation. In such a case, when

the interpretation changes from the interpretation in the original context the data is put in, an attack is detected. The framework can notify the user to judge whether the change is appropriate or not. Alternatively, contextual privacy can be managed by maintaining the sensitivity of data. The framework can monitor the changes of the sensitivity in comparison to the value of the data in the original context the data is put in. The sensitivity can be monitored instead of the interpretation to detect misappropriation.

We propose a three-layer architecture for the framework. The architecture design includes two layers to infer context, and the interpretation and sensitivity of the data. These layers can be embedded in the social software platform, and have access to all users' data. The third layer is a contextual privacy management layer. It allows the user to specify the appropriate interpretation or sensitivity, and interacts with the user upon changes. The design is assessed for usability and is compared with the well-known theory of Contextual Integrity [75]. The assessment shows that the usability of our platform is potentially higher than the usability of the Contextual Integrity. We discuss the possible implementation of the framework, and argue how the framework could facilitate privacy in private and public spaces.

1.3.7 Privacy and Surveillance

Another relevant aspect of understanding privacy is to examine the various privacy management approaches in the technical and legal frameworks. We investigate the approaches of contextual privacy and the consequences of the control offered.

Privacy management approaches can have varying aspects of surveillance. Different approaches offer varying means to protect privacy. The degree of privacy they offer can also vary, according to the adopted approach. However, in many cases, many parties can be involving in the functioning of an approach. Such parties can monitor the data and the behaviour of users. With such a capability, these parties can apply surveillance on users.

To understand the possible degree of privacy in PaC approaches, we perform a conceptual analysis of privacy management approaches in the technical and legal frameworks. To perform the analysis, we put forward criteria to evaluate the degree of control and privacy and the degree of surveillance entailed by each approach. The criteria are based on requirements to achieve a high degree of contextual privacy. The analysis shows how certain aspects of surveillance are deeply rooted in the realisations of PaC. We argue that data control approaches

should offer transparency, reciprocity and a balanced degree of control as a first step towards addressing the interdependency of privacy and surveillance.

1.4 Contribution

The main contribution of this thesis is an extensive exploration and analysis of contextual privacy. The first main contribution is an extensive information of the interaction of context and content in relation to privacy management. The second main contribution is deploying various data analysis methods to test various hypotheses about context and privacy management. The third main contribution is a proposal for usable contextual privacy management using artificial intelligence. In the following, we detail the contributions of this thesis.

- An exploration and modelling of sensitivity and privacy management patterns. These patterns describe the following:
    1. Modelling data disclosure patterns: by modelling how data of different content topics is disclosed in different contexts.
    2. Modelling post-disclosure patterns: by modelling how context and content topics affect how users manage the sensitivity of their data.
- An analysis of the effect of context on sensitivity and privacy management, including:
    1. Analysis of the effect of individual contexts on data disclosure patterns: the analysis quantifies the effects of context on the intensity of disclosures, and identifies the significance of the effects.
    2. Analysis of the effect of multiple contexts on post-disclosure patterns: the analysis quantifies the effect of context on two post-disclosure patterns with variant contextual parameters, and identifies the significance of the effect.
    3. Modelling the context based on content of data: the modelling shows that by knowing the content of data it is possible to predict the context in which the data is disclosed.
- A demonstration of the effect of users' subjectivity and time on sensitivity and privacy management.
- An exploration of different analyses methods to test different hypotheses and explore various aspects of data. The analysis in this thesis is performed on the same dataset. However, by selecting different

sub-datasets and different analysis methods, we demonstrate how to investigate various hypotheses and aspects of contextual privacy and users' behaviour.

- A design of a conceptual framework for contextual privacy using artificial intelligence to protect from data misappropriation and overcome usability issues of privacy management.
- Criteria to evaluate the possible degree of privacy and surveillance entailed by a privacy management approach, whether technical or legal.

1.5 Outline

Chapter 2: Exploring Sensitivity and Privacy Management on the Web Explores privacy management patterns through exploring data sensitivity. This chapter investigates sensitivity patterns in a big data of 226 million searches from the search engine Bing. The searches are deleted or kept, indicating sensitivity or insensitivity. The identified patterns provide insights into sensitivity and privacy management that go beyond the common world knowledge of sensitivity. The chapter demonstrates evidence for contextual effects on sensitivity such as temporal evolution, and subjective diversity.

Chapter 3: Quantifying the Effect of Context on Sensitivity Based on the effect of context that we report in the previous chapter, this chapter explores the effect of context in more detail. It explores the effect of context on sensitivity in disclosure and post-disclosure patterns. In this chapter, we perform a large-scale analysis to extract sensitivity patterns from 226 million searches. We assume sensitivity is a latent variable in search patterns. Each pattern is a subset of searches characterised with content-related features in a particular context. We observe the sensitivity variation across contexts and quantify the significance of the effect of context. Through these tests, we identify contexts in which sensitivity patterns significantly vary, additionally, such identification is achieved on even the content-features levels. Thus, it is possible to predict searches of certain features that may or may not vary across contexts.

Chapter 4: Conceptual Analysis of Context Presents a conceptual analysis of context and privacy in social software communication. This chapter elaborates on the issues of context control and context ambiguity to manage privacy. It reviews the relationship between context, privacy, communication and identity to understand contextual privacy. Issues

of context ambiguity are analysed further to elaborate its effect on data misappropriation. Based on this analysis, the chapter defines contextual privacy and defines the data misappropriation attacker model.

Chapter 5: A Contextual Privacy Framework for Social Software (CPS^2) Presents a conceptual framework for contextual privacy management. It conceptualises a contextual privacy management framework based on maintaining the interpretation and the sensitivity of data. It presents an architecture based on the utilisation of artificial intelligence mechanisms. The design of the framework is analysed for usability aspects. The chapter provides a discussion about how the framework enhances communication and privacy in private and public spaces.

Chapter 6: The Other Side of Privacy: Surveillance in Data Control Explores privacy as control approaches in relation to surveillance issues. The chapter analyses the counter-privacy consequences of the various privacy management approaches. The analysis focuses on the technical and legal approaches. The analysis is based on criteria to evaluate the degree of control and privacy and the degree of surveillance entailed by a data control approach. The analysis shows how certain aspects of surveillance are deeply rooted in the realisations of "privacy as control". In this chapter, we argue that data control approaches should offer transparency, reciprocity and a balanced degree of control as a first step towards addressing the interdependency of privacy and surveillance.

Chapter 7: Conclusion Summarises the findings of this thesis and discusses the implications of these findings and future research.

Chapter 2

Exploring Sensitivity and Privacy Management on the Web

2.1 Introduction

Vast amounts of data are disclosed and handled on the web on a daily basis.[1] Each data item has an implicit degree of sensitivity depending on its type, content, and owner.[2] We define sensitivity as the degree of inappropriateness of handling a data item by certain parties or in certain contexts. Data ranges from being not sensitive to highly sensitive. Highly sensitive data are commonly associated with concerns of misuse and require a high degree of privacy management. For data owners, reasoning about the appropriate audience and disclosure context of a data item depends on its sensitivity. An item can be protected by limiting its availability by deleting it [17, 113, 106], or by using privacy management to define constraints on who, how, when and for which purposes data can be handled [48].

A first step towards better privacy management is understanding the nature of data sensitivity. Currently, privacy management approaches offer users the possibility to select the data they want to control. Based on the user's

[1] This work has been partially done during Rula's internship at Microsoft–Cambridge.

[2] We do not refer to the legal ownership of data. Rather, we user ownership to refer to the individual that uploads a data item to the web.

assessment of data sensitivity, the user can decide who can access the data, in which context, etc. However, users can easily disclose large amounts of data. Managing privacy of each item can be challenging, therefore. Moreover, data that can be not very sensitive in one context, can become sensitive in another context, e.g., health-related data is sensitive to disclose in work-related contexts. Managing privacy effectively requires an understanding of what makes data sensitive, and whether sensitivity changes. This understanding could also be utilised to indicate sensitive data that users should particularly manage. It can also be utilised to detect when sensitivity changes, for instance.

While there is a substantial body of work on privacy for protecting sensitive data, the knowledge about what counts as sensitive is limited. Sensitivity of data guides how, when and with whom data is shared. In the Joint Computer Conference of 1967, the difficulty of identifying sensitive data was realised [50]. The early solutions were to classify military-related data as sensitive and non-military-related data as non-sensitive. Since then, there has been more focus on identifying sensitive data in non-military-related contexts [50]. The main challenge is that sensitivity of data may be affected by the type of data and its content, as well as on the surrounding context.

To protect sensitive data, users can hide or delete it, or disclose it carefully. Typically, there is some consensus on few topics that people consider sensitive. We refer to this consensus as the *common world knowledge about sensitivity*. This knowledge comprises sex-, health-, religion-, finance-related topics, as well as other legally codified topics. However, there is a discrepancy between how the common world knowledge identifies sensitive data, and what users' actions reveal about data sensitivity. An example is the deleted items found in the top 10000 deleted searches during one day from Bing—the Microsoft search engine. In Bing, users can delete their searches from the search history. The top three most deleted search items in Bing (table 2.1) are item that are not commonly viewed as sensitive. However, it is reasonable to assume that "Syria" and "Facebook" are sensitive because they are at the top of the deleted searches. This discrepancy suggests that the assumptions of the common world knowledge may not always hold. A better-understanding of sensitivity requires an in-depth investigation of big data sets to analyse and establish descriptive knowledge about sensitivity.

Understanding sensitivity of data is challenging due to its latent nature. When uploading a data item on the web, the user can specify various attributes of this item such as the title, date of creation, etc. However, it is uncommon to specify a value that indicates the sensitivity of a data item. Rather, sensitivity is often latent. It can be possible to estimate sensitivity by observing other actions and data about a particular data item. For instance, the sensitivity can

Search	*Search Count*
Facebook	16,357,880
Syria	115,353
Lindsay Lohan	123,501

Table 2.1: The three most deleted searches in Bing during one randomly-selected day.

be estimated by estimating the item's availability and accessibility online, e.g., if the item is accessible to the public, it is assumed to have a low sensitivity. Alternatively, information about sensitivity can be extracted by interviewing individuals. However, such an option is, firstly, complicated to perform on a large scale. Secondly, the extracted information can be of limited reliability because the interviewed subjects may be too embarrassed to share what they consider sensitive [55].

As a first step to understanding sensitivity and privacy, we apply data analysis techniques on a big data set from the Microsoft search engine Bing. We access 226,000,000 data items, requested by 413,000 users. Bing allows users to manage their sensitive data by deletion. To help users formulate their searches, the auto-suggest features displays text completion suggestions as the user types, e.g., "xbox one" and "xbox live" for "xb". Bing also suggests matching searches previously issued by the user, though not necessarily popular across the entire user population. The per-user search history keeps a record of searches issued in the past, for improving and personalising the search results. Bing's interface encourages search history management. On the homepage alone, two links point to search history. Past searches are also displayed on the homepage, for ease of access. Deleting searches, however, removes them from the search history of the user, without removing them from the server. We analyse this dataset to infer sensitivity patterns of deleted and kept items. We investigate particular aspects that are captured in the following questions:

Q1) Is sensitivity defined by the content of data?

Q2) Is sensitivity temporal?

Q3) Is sensitivity contextual?

Q4) Is sensitivity subjective?

Towards answering these questions and investigating sensitivity patterns, this chapter contributes the following:

1. A discussion of challenges of investigating sensitivity, and the need for big data to perform large scale analyses of sensitivity (Section 2.2)
2. A description of a data set that includes sensitive data items, and descriptive statistics providing evidence for effect of context on sensitivity (Section 2.3)
3. An inference of sensitivity patterns and contextual effects using machine learning. Three patterns are inferred that show the effects of context, time and subjectivity on sensitivity (Section 2.5)
4. A discussion of the results and the possible motivations for deletions (Section 2.7).

2.2 Investigating Sensitivity and Privacy

This section investigates sensitivity and contextual dependencies, and argues about the relation between sensitivity and privacy.

2.2.1 The Challenge of Investigating Sensitivity

Identifying sensitive data is a challenging task. The most straight-forward method to investigate sensitivity is to interview people about what sensitive data is, and whether sensitivity varies. This method is relatively challenging to perform due to the issues of discussing sensitive data with subjects [63]. In this setting, the more detailed and sensitive the data is, the more embarrassed and reluctant subjects are to provide information about sensitivity. Moreover, it might be difficult for subjects to provide accurate information regarding whether sensitivity changes depending on location, time or any other contextual parameter. Thus, a more comprehensive investigation of sensitivity would require interviewing a large number of subjects in various contexts to ensure the reliability of the extracted information.

An alternative to interviewing people is analysing a large dataset that includes sensitive data plus sensitivity information. To infer accurate information about sensitivity, the dataset should satisfy the following requirements. Firstly, the dataset should include sensitive and insensitive data for comparison. Secondly, the dataset should represent a significantly large number of people to avoid subjectivity. Thirdly, the data should span a considerable period of time. Fourthly, to investigate the effect of context, the dataset should include as much contextual information as possible about each data item. These requirements

can be satisfied in a dataset that includes data of varying sensitivity disclosed by different users in different contexts. In the following, we discuss the big data set we use to analyse sensitivity.

2.2.2 Big Data to Understand Sensitivity and Privacy

We analyse a big dataset from Bing that contains deleted and kept searches. The dataset encompasses a wide spectrum of topics and include various contextual information that describe searches. Using these details,we investigate deletions of searches to infer what makes data sensitive and what affects it.

The prominent display of past queries in Bing gives rise to two privacy issues: incidental access and intentional access. Incidental access can happen when the list of proposed queries includes sensitive items. If the user is not in a private space, others could incidentally have access to such sensitive data. For instance, the user is at work, and she accesses Bing in the presence of a colleague. When she types in a search, some sensitive searches are displayed in the proposed searches list (Figure 2.1). Intentional access happens when the search history items are accessed by the others or the search engine. The search engine accesses the history data to enhance the search results of the user. Also, the engine accesses the history of a user for targeted advertisements. Besides the search engine, other users can deliberately access the user's search history by accessing the user's machine. All of these types of intentional accesses raise privacy concerns.

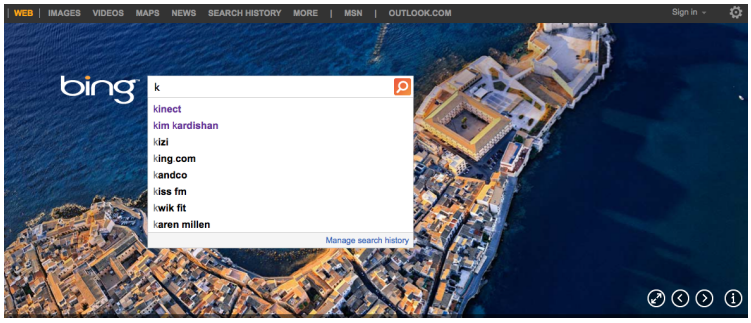


Figure 2.1: Example of privacy exposure. Auto-suggest displays stem-matching queries, sourced from generally popular queries and the user's own search history. The search 'Kim Kardashian' in this example can be considered inappropriate in certain contexts.

We avoid interpreting deletions as privacy management behaviour due to the lack of information about users' motivations. Although many studies show that people delete their data as an act of privacy management [17, 113, 106], however, it is not possible to generalise and assume that any deletion in any context is an act of privacy management without evidence. Through deletions of history items, the user can manage privacy to avoid the privacy concerns discussed above. However, deletions can be motivated by non-privacy-related concerns. A user may simply wish not to receive certain ads related to a particular search. This is similar to labelling a past purchase as for someone else in one's order history with an electronic retailer, so that it does not contaminate product suggestions. We assume that deletions are motivated by the sensitivity of items, which in turn relate to privacy or data usage concerns. Sensitivity implies that when an item is accessed by others, the user might have concerns, whether privacy-related or otherwise.

2.3 The Bing Dataset

The dataset encompasses search data issued by a sample of Bing users over the period of six months from November 2012 to May 2013. The dataset encompasses 226 million searches of 400,000 users. The dataset has both deleted and kept items. The deleted items account for 22% of the dataset. The dataset is anonymised and includes no user-identifying data. According to the Microsoft privacy policy, the users' data can be used for research [72].

In Bing, a search query is a data item. Each item is represented by a set of features. A feature corresponds to an attribute of the content or context of a search. Content features describe the submitted search and the relevant results. The context features describe the context in which the search is submitted, as well as the context the user is in upon submitting the search. Next, we classify the features in three classes, *content*, *data context* and *user context*, and describe the features of each class, as well as the deletion indicator feature. We refer to a feature A as $ft.A$, a content feature or topic B as $tp.B$, a context feature X as $cxt.X$. A context can have different values. A value Y of context X is denoted as $cxt.X.Y$.

2.3.1 Class of Content

This class includes the standard set of features corresponding to the possible topics of searches in Bing. Through these features a search is represented, and results are matched. A content feature is binary, indicating the relevance of a

search to the feature. A search can be relevant to one or more features at once. In the following is the list of features, with search examples, and the count of searches within each feature, where the value is “on":

- *tp.Adult*: “freeporn”, “adult friend finder”, (n=661,400). Bing has the feature *tp.AdultScore* a numerical value that indicates the relevance of a search to “sex" and “porn" topics.
- *tp.AppIntent*: “angry birds”, “ MSN free games”, (n=12,894,972).
- *tp.Autos*: “Cadillac”, (n=162,936).
- *tp.Book*: “game of thrones”, “Oxford dictionary”, (n=3,796,631).
- *tp.Bus*: “bus schedule”, (n=93,668).
- *tp.Celebrities*: “Kim Kardashian”, “paparazzi”, (n=14,098,552).
- *tp.ClothesAndShoes*: “Summer 2013 shoes”, (n=811,919).
- *tp.Commerce*: “buy lawnmower”, (n=14,816,246).
- *tp.ConsumerElectronics*: “xbox 360”, (n=1,085,562).
- *tp.Dictionary*: “meaning of pleasant”, (n=1,129,695).
- *tp.*tp.Download: “cnet free downloads USA”, (n=1,952,517).
- *tp.Education*: “university of Cambridge”, (n=479,337).
- *tp.Events*: “football matches”, (n=3,589,495).
- *tp.Finance*: “MSFT”, (n=1,286,520).
- *tp.Flight*: “cheap flights”, (n=792,636).
- *tp.FlightStatus*: “flight arrival schedule”, (n=3,271).
- *tp.Galleries*: “MSN games”, “Disney channel”, (n=18,501,786).
- *tp.Health*: “drugs”, “Catherine zeta-jones bipolar”, (n=3,393,061).
- *tp.Hotel*: “holiday inn”, “booking.com”, (n=1,119,532).
- *tp.HowTo*: “weight loss snacks ”, “how to tie a tie”, (n=1,779,386).
- *tp.Image*: “Kim Kardashian blonde”, (n=37,199,290).
- *tp.Jobs*: “career builder”, “job search engine”, (n=502,538).

- *tp.List*: “TV listings”, (n=964,703).
- *tp.Local*: “white pages”, “news”, (n=19,395,305).
- *tp.Maps*: “California map”, “where is Syria”, (n=512,548).
- *tp.MovieShowtime*: “Butler movie showtime” (n=1,472,319).
- *tp.MovieTheatre*: “Movie theatre in Cambridge”, (n=401,641).
- *tp.MovieTitle*: “Butler movie”, “Great Escape”, (n=1,080,991).
- *tp.Music*: “Frank Sinatra”, (n=3,798,576).
- *tp.Name*: “Einstein”, (n=5,917,571).
- *tp.NameNon-celebrity*: “Zoe Tishler”, (n=5,349,557).
- *tp.NamePlus*: “Will Smith sells mansion”, (n=16,059,194).
- *tp.Navigational*: “Hotmail.com”, “Bing”, (n=69,940,890).
- *tp.Nightlife*: “Las Vegas shows”, “prostitutes”, (n=230,957).
- *tp.Nutrition*: “calories in a banana”, “nutritional yeast”, (n=18,647).
- *tp.OnlineGames*: “addicting free games”, (n=285,657).
- *tp.QuestionAndAnswer (Q&A)*: “pregnancy test”, (n=1,615,269).
- *tp.QuestionPattern*: “how i met your mother”, (n=6,134,986).
- *tp.RadioStation*: “CNN news”, (n=1,592,502).
- *tp.RealEstate*: “real-estate news”, (n=659,574).
- *tp.Recipes*: “cabbage soup”, (n=2,081,378).
- *tp.Restaurant*: “Italian restaurants”, (n=1,430,975).
- *tp.Seasonal*: “fashionable mittens”, (n=4,066,639).
- *tp.Sports*: “NBA”, (n=2,563,947).
- *tp.Tech*: “ctrl + alt + del”, (n=2,923,418).
- *tp.ThingsToDo*: “park”, (n=1,372,730).
- *tp.Travel*: “Sweden”, (n=875,803).
- *tp.TravelGuide*: “Belgium attractions”, (n=554,501).

- *tp.TVShows*:"so you think you can dance show", (n=1,550,361).
- *tp.University*: "University of Cambridge", (n=445,014).
- *tp.Url*: "MSN.com", (n=17,402,652).
- *tp.VideoExcludesAdult*: "cartoon network", (n=32,816,037).
- *tp.VideoGames*: "gamespot.com", (n=790,141).
- *tp.Weather*: "weather channel", (n=1,468,680).
- *tp.WikipediaReference*: "Star Wars", (n=2,735,907).
- *tp.ContainsLocation*: "Microsoft in Redmond", "Syria", (n=20,645,406).

From these features, we curate S_{cwk} a set to correspond to what commonly is considered sensitive:

$$S_{cwk} = \{adult, commerce, finance, health\}$$

2.3.2 Class of Data Context

This class includes the following features characterising the online context in which the item is submitted:

- *cxt.Vertical*: the type of content to search for. Verticals can be web, images, or videos.
- *cxt.VerticalChange*: indicating that the user searched for an item within different verticals.
- *cxt.SafeSearchSetting*: is one of three possible modes to filter adult content. The modes are: moderate, strict, off.
- *cxt.AppType*: the application from which the search is submitted. It can be an app or a browser.
- *cxt.Browser*, can be Internet Explorer (IE), Chrome, Firefox, etc.
- *cxt.IsAutoSuggest*: the search is suggested to the user by Bing, based on the input of the user.
- *cxt.IsAlteration*: the search is suggested by Bing by altering the original search the user entered.

- *cxt.IsForced*: the search is submitted between quotes, by the user, to be searched for as it is.
- *cxt.IsSpellSuggestionCorrection*: the search is suggested by Bing as a correction for the search the user submitted.
- *cxt.SessionPageNumber*: is the number of the search results page the user navigates to and clicks on a link from.
- *cxt.InSearchHistory*: indicates that the search has been searched for before and is in the search history, and the number of past searches through *cxt.SearchHistoryItemCount*.
- *cxt.IsDotCom*: indicates searches for a url ending with '.com'.

2.3.3 Class of User Context

This class includes the following features characterising the user's context of submitting the item:

- *cxt.IPCount*: the number of IP addresses the user uses during the day to access Bing.
- *cxt.OperatingSystem*, the family of the operating system the user is using to access Bing, can be Windows NT 5.1, Linux, etc. This feature captures users context as it is not limited to search activities, rather, it captures other users' activities.
- *cxt.DeviceClass*: the type of device from which the search is submitted.
- *cxt.DeviceModel*: the brand of device from which the search is submitted.
- *cxt.FacebookUser*: indicates whether the user has signed into Bing using Facebook credentials, as a result searches are saved under the user's Facebook account.
- *cxt.WindowsLiveUser*: indicates whether the user has signed into Bing using windows live credentials, as a result searches are saved under the user's windows live account.
- *cxt.Hour*: the hour at which the item was deleted, and is 'Null' when it is kept.

Deletion Indicator: indicates whether a data item is deleted or kept, with values $\{1, -1\}$, respectively. A deleted item has a lifespan value that indicates the period between searching for and deleting the item.

We consider the 'deletion indicator' a dependent variable, and the rest of the features are independent variables. This means that the value of deletion depends on the values of the independent variables. In other words, the decision to delete an item is dependent on the features of this item.

We consider a deleted item to be *highly sensitive*. We consider the features of an item represent the implicit or the latent sensitivity of this item. The sensitivity, in turn, indicates whether an item should be deleted or kept. In the following, we investigate sensitivity through analysing the deletions and the other features of the data.

2.4 Preliminary Aspects of Sensitivity

This section presents preliminary aspects of sensitivity in relation to time and content. Initially, we are interested in observing whether deleted items have varying lifespans, and whether they are mainly associated with particular content features or topics.

2.4.1 Time and Sensitivity

Deletions are consistent over weekdays, except for a small rise in deletions on Mondays (1.13% above average). Deletions, however, incur a spike every 24 hours after submitting the search.

The cumulative percentage demonstrates that the lifespan varies across the dataset (Figure 2.2). More than half of all the data has a lifespan of one hour at most. After a day, 72% of all data have been deleted. This variation indicates that sensitivity is temporal. The temporality implies that not all sensitive data is deleted immediately after searching for the item. A short lifespan suggests that users are mostly aware of the sensitivity of their data that they delete them shortly after submission. A long lifespan indicates that deletions continue over time. Such a continuation means that sensitivity of items changes over time, i.e., what was not judged sensitive in the past, might be judged sensitive later. Alternatively, this continuation may be due to periodic or random deletions over time. The relationship between sensitivity and time is investigated further in Section 2.6.

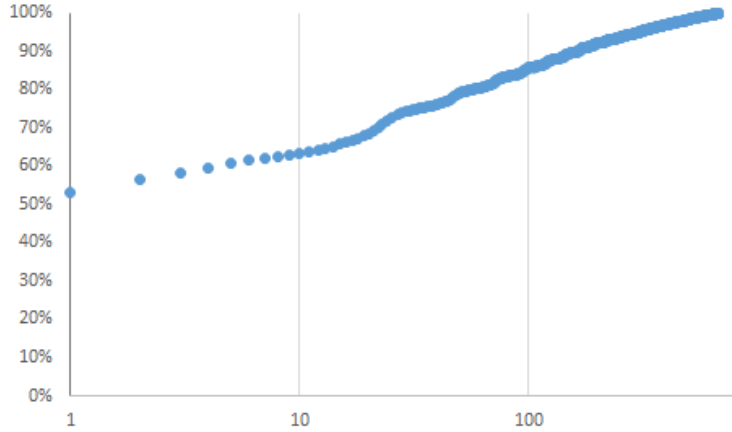


Figure 2.2: Lifespans of data: cumulative proportion of deleted data. The X axis represents the number of hours, and the Y axis represents the cumulative percentage of deletions.

2.4.2 Content and Sensitivity

Deletions and sensitivity in the dataset vary across content features. In general, adult topics are assumed to be sensitive. However, only 4% of all deleted items are related to adult. To investigate the association between sensitivity and content topics, we employ the odds ratio (OR) [16].

The OR quantifies the association between two proportions without implying causality. The OR quantifies how strongly the presence or absence of a property A is associated with the presence or absence of property B. We apply this test to analyse the association between the deletion variable (the dependent variable) and each of the independent variables in the content class. The $OR_{D,T}$ examines the prevalence of deletions, and hence sensitivity, given a particular topic, according to the following formula:

$$OR_{D,T} = \left(\frac{Count_{del,T}}{Count_{del,-T}} \middle/ \frac{Count_{kept,T}}{Count_{kept,-T}} \right)$$

where $Count_{del,T}$, $Count_{kept,T}$ represent the count of deleted or kept searches related to topic T, respectively. $Count_{del,-T}$, $Count_{kept,-T}$ represent the count of deleted or kept searches related to any topic except T, respectively. The OR computation is applied on the whole data set, as well as on sub-datasets that represent data submitted by individual users. The standard deviation (SD)

between the OR values of the users' sub-datasets is computed to quantify the subjective variation.

The OR values demonstrate the lack of strong associations with deletions, and hence sensitivity (Figure 2.3). All OR values are < 1 indicating that being related to a topic T is associated with lower odds of deletion (relative to not being related to topic T). The OR values show there is no single strong association between topics and deletions. For example, $OR_{D,tp.Navigational}$ shows that the odds of a deletion given the topic *tp.Navigational* is 0.19 the odds of a deletion given searches not related to this topic. This low association suggests the need to investigate other features that are strongly associated with deletions.

The SD values provide evidence for subjective variation of the association between topics and deletions, as well as topics and sensitivity. The computations reveal an inverse relationship between the OR values computed from the whole dataset and the SD between the OR values of individual sub-datasets. The lower the values of the OR per topic in the whole dataset, the higher the SD of individual ORs. The high SD values indicate a high disagreement between users on the association between topics and deletions. The subjectivity variation is significant particularly in the topics commonly considered as non-sensitive, e.g., *tp.Bus* and *tp.Nutrition*. The subjectivity demonstrated by the SD requires further investigation of subjective sensitivity management given contextual and content features. In the next section, we include contextual features to model the patterns of how users delete and manage sensitive data.

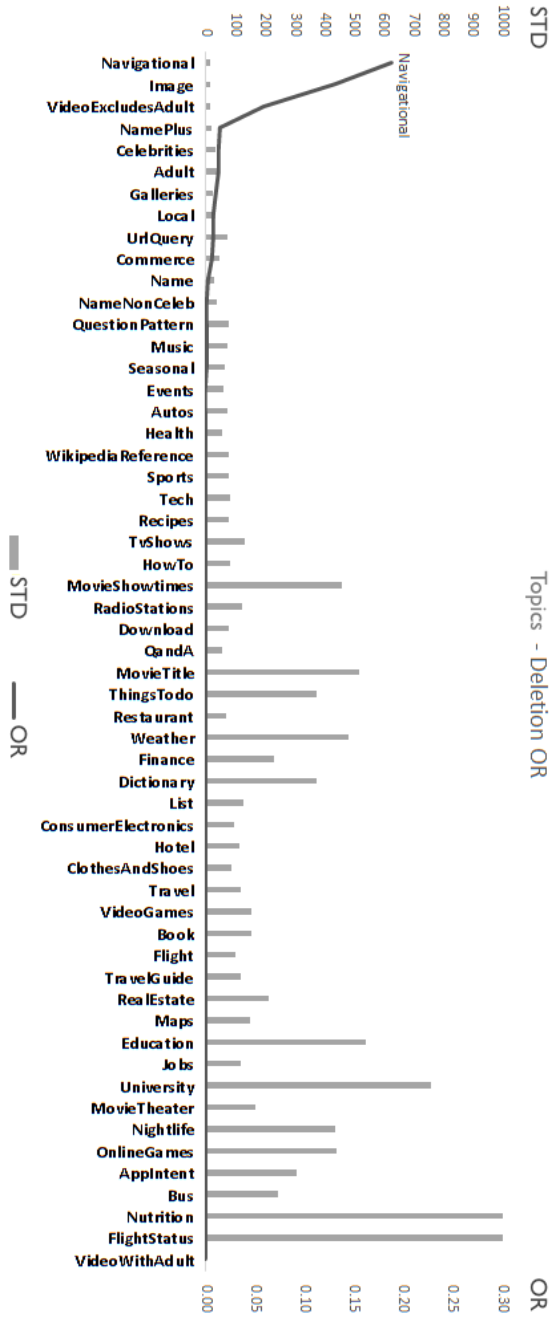


Figure 2.3: $OR_{D,T}$ per topic and the standard deviation of the OR values computed from user sub-datasets.

2.5 Modelling Sensitivity Patterns

This section presents our approach to model deletions and sensitivity.

We model sensitivity through modelling the classification of deleted and kept data. Modelling deletions is a binary classification problem. The deletion indicator is the dependent class variable. This variable can have 'deleted' or 'kept' as class values. The assumption is that when a data item is 'deleted', the deletion indicator implies that it is sensitive. Items might have different sensitivity values. When the sensitivity value is high enough, the item is deleted. By learning the underlying classification model or pattern we learn what makes an item 'deleted' or 'kept'. We also learn what makes an item sensitive to be deleted. Such a pattern explains how an item is classified. The explanation is based the independent features or *determinants* that affect the value of the dependent class variable.

2.5.1 Data Preprocessing and Analytics

Data preprocessing and analytics were performed on Cosmos—the cloud infrastructure for big data analytics developed by Microsoft Online Service Division [111]. Machine learning algorithms were executed with SCOPE language [27].

To learn statistically significant patterns, the dataset requires preprocessing. Preprocessing is required because of the *imbalanced data* problem. This problem emerges because the kept data items outnumber the deleted items. Such an imbalance is referred to as a *between-class imbalance* [28]. This problem causes bias of classification algorithms towards the majority class. The bias contributes to classification errors. To overcome this problem, we excluded users with kept or deleted items that account for less than 3% of their total items. We filtered out users who delete more than 97% and less than 3% of their items, which translates to filtering out 594,655 user and 42,724,389 data items. This filter reduced the imbalance in the overall data set and excluded users whose actions include little signal. Secondly, we applied data level random undersampling 30%. This sampling is the most conservative approach for imbalanced data [46]. It preserved the distribution of the deletion indicator values.

We also applied optimisation approaches to optimise the performance of the learning algorithms. We applied minimax normalisation to map the values to the range [0,1]. The normalisation increases the speed of the training. To learn

a proper model and avoid over-fitting given the high dimensionality of the data, we used L1 and L2 Regularisation [15].

2.5.2 Learning Deletion and Sensitivity Patterns

To learn the deletion and sensitivity pattern, we use linear regression models. The most appropriate learner algorithms for such a high-dimensional dataset are Generalised Linear Models [68]. In particular, we applied linear regression to model the data through the following functions:

$$\begin{aligned} \hat{f}(x,w) &= w_0 + \sum_{j=1}^{N} w_j x_j \\ \hat{y} &= sign(\hat{f}(x,w))) \\ \hat{y} &= \begin{cases} 0, & \text{if } \hat{f}(x,w)) \leq 0 \\ 1, & \text{if } \hat{f}(x,w)) > 0 \end{cases} \end{aligned}$$

where w_0 is bias parameter, w_j is the weight of the determinant x_j, $\hat{f}(x,w)$ estimates sensitivity, and $\hat{y}$ is the binary classification function of deletion, i.e., classifies an item as deleted or kept. A determinant is a feature with a particular value. One feature can result in one or more determinants according to its possible values, e.g., the feature *tp.Health* can result in two different determinants corresponding to the values {true, false} that we refer to as *tp.Health*, and *tp.Health=false*, respectively. Weights indicate the importance of the determinants in contributing to outcome. The higher the weight, the more significant the role of the corresponding determinant. A positive weight implies that the corresponding positive determinant indicates the sensitivity, and the deletion of the item. A negative weight implies that the corresponding determinant indicates the insensitivity, and the keeping of the item. In the following sections and due to the high number of features in our dataset, we mainly focus on the positive determinants to understand the features that directly indicate sensitivity.

To learn the weights of the function, the data was split into two sets. The dataset was split by a 70:30 ratio into a training and a test set. The function was learned from the training set. Afterwards, the function was adapted through cross validations. During the test phase the weights are adapted.[3]

[3]We also applied logistic regression and SVM algorithms for modelling the data. However, the linear regression resulted with the highest accuracy.

2.6 The Inferred Patterns

In this section, we apply the modelling approach discussed in the previous section to infer three different patterns. The aim of inferring patterns is testing the following hypotheses:

$H_{generality}$: There exists a pattern that explains the general behaviour of users in relation to managing their data through deletions at any point in time.

$H_{contextuality}$: Content and context affect sensitivity and data management.

$H_{temporality}$: Patterns learned from sub-datasets that correspond to different time frames are non-identical.

$H_{subjectivity}$: Patterns learned from sub-datasets that represent different users are non-identical.

To extract knowledge about sensitivity and test the hypotheses mentioned above, we apply linear regression to infer different patterns. The patterns are inferred from the whole dataset, and from splitting this dataset into sub-datasets. The aim of inferring patterns from sub-datasets is to infer particular information that is explained in the following:

1. *A long-term general sensitivity pattern (LTP)*, learned from the entire dataset. This pattern tests $H_{generality}$ and $H_{contextuality}0$.
2. *Short-term patterns (STPs)*, learned from 6 sub-datasets each corresponding to the period of one month. The inference of these patterns investigates the effect of time on sensitivity, and tests $H_{temporality}0$.
3. *User-based pattern (UPs)*, learned from 75 sub-datasets each representing an individual user. The inference of these patterns tests $H_{subjectivity}0$.

2.6.1 The Long-term Pattern (LTP)

The LTP validates $H_{generality}$ and provides evidence that content and context contribute to sensitivity. The LTP is inferred from the whole dataset with a 72% accuracy of the whole model. This pattern represents the general behavioural pattern of users at any point in time during the period of data collection. $H_{generality}$ is validated through this pattern that represents 226 million data items of 4 million users with a high accuracy. The pattern validates $H_{contextuality}$ by showing that the context features, in additions to content features contribute to the sensitivity of data. The pattern shows that many features of the three

classes are determinants of sensitivity. Due to the high number of determinants, in the following, we summarise the determinants of the three classes that significantly[4] contribute to the pattern, whether positively or negatively.

Class of Content This class includes various topic determinants that significantly affect sensitivity, or can be referred to as sensitive topics. These determinants are:

1. *tp.Adult* ($w = 0.2$), has the highest weight in this class. However, the *tp.AdultScore* ($w = 0.69$) has the highest weight amongst all classes. This means that an item relevant to *tp.Adult* with a high *ft.AdultScore* value is highly likely to be sensitive. Comparatively, an item with a low *tp.AdultScore* is less likely to be sensitive. At the same time, not being related to this topic, *tp.Adult=false*, has the most significant negative effect of sensitivity ($w = -0.16$).
2. *tp.Celebrities* ($w = 0.12$), this topic is not commonly considered sensitive, yet it has the second highest weight amongst topics.
3. *tp.NightLife* ($w = 0.1$), indicating night life activities may not be appropriate to keep.
4. *tp.Health* ($w = 0.074$), health is commonly considered as a sensitive topic.
5. *tp.NameNon-celebrities* ($w = 0.067$), a user may search for a certain name that she is interested in knowing more about (e.g., an ex-boyfriend).
6. *tp.ClothesAndShoes* ($w = 0.06$), this topic is not often considered sensitive.
7. *tp.Name* ($w = 0.047$), searching for names may indicate the user's interest in other people.
8. *tp.NamePlus* ($w = 0.046$), searching for names and other aspect can also indicate sensitivity.
9. *tp.VideoExcludingAdult* ($w = 0.039$), even non-adult-related videos indicate sensitive content.
10. *tp.Image* ($w = 0.034$), items related to this topic can be sensitive. At the same time, not being related to this topic, *tp.Image*=false, has a less positive effect on sensitivity ($w = 0.0059$).

[4]The aim is not to list every single determinant, rather, to mainly show that determinants of high weights that have the major effect on sensitivity. The aim is to also show that different determinants from the different classes have different weights and effects.

11. *tp.MovieTitles* ($w = 0.032$), which might be related to a movie about some sensitive or embarrassing topic, e.g., dirty dancing movie.

12. *tp.QuestionAndAnswer* ($w = 0.031$), which might expose socially awkward ignorance (e.g., how to tie a tie), reflect interest in sensitive issues (e.g., how to arrange a funeral, how to treat AIDS), or non-sensitive issues.

13. *tp.Restaurant* ($w = 0.01$), this topic indicates the user's interest in food and restaurants.

14. *tp.RadioStation* ($w = 0.0099$), this topic indicates the user's interest in radio stations and music.

15. *tp.URL* ($w = 0.0093$), this topic might expose the user's interest in certain sites.

16. *tp.Finance* ($w = 0.006$), which might expose high or low wealth, each of which can be sensitive.

17. *tp.MovieShowtime* ($w = 0.0058$), indicating the user's interest in movies and entertainment.

18. *tp.Autos=false* ($w = 0.0036$), being unrelated to this topic is indicative of sensitivity.

19. *tp.FlightStatus* ($w = 0.00035$), might expose that the user is tracking the flight of someone else (e.g., visiting or being visited by a secret affair).

20. *tp.TVShows* ($w = 0.00022$), indicating the user's interest in entertainment.

21. *tp.VideoGames* ($w = -0.14$), has a negative effect on sensitivity.

22. *tp.Jobs* ($w = -0.09$), has a negative effect on sensitivity.

These determinants are broader than S_{cwk}. Despite the overlap with the S_{cwk}, the pattern suggests that sensitivity goes beyond topics that are commonly considered sensitive. The dataset does not provide information about why certain topics are highly indicative of sensitivity.

Class of Data Context This class has various determinants that affect sensitivity. These determinants are:

1. *cxt.InSearchHistory* ($w = 0.24$), indicating that reoccurring searches are sensitive. *cxt.SearchHistoryItemCount* ($w = 0.091$) indicates the more the item is searched for the higher the sensitive it is.

2. *cxt.SessionPageNumber* ($w = 0.14$), which means that the more the user navigates the results page—to find an answer, the higher the sensitivity.

3. *cxt.IsForced* ($w = 0.1$), forced searches indicate sensitivity.

4. *cxt.AppType.SSL Bing* ($w = 0.094$), using secure SSL service indicates sensitivity.

5. *cxt.VerticalChange* ($w = 0.083$), the more the user navigates through the verticals, the higher the sensitivity.

6. *cxt.SafeSearch.Strict* ($w = 0.066$), using strict search mode can indicate sensitivity of data. *cxt.SafeSearch.Moderate* ($w = 0.049$) has a lower indication of sensitivity.

7. *cxt.IsDotCom* ($w = 0.0073$).

8. *cxt.IsSpellSuggestionCorrection* ($w = 0.0058$), an item that is corrected for spelling by the search engine is less sensitive than forced item.

9. *cxt.Browser.IE* ($w = 0.0144$), indicating that users of this browser are more likely to manage their sensitive data than users of *cxt.Browser.FirefoxAndOthers* ($w = 0.0108$). However, upon considering the version of the browser, we get more detailed information about how browsers affect sensitivity (Table 2.2). The versions 6 and 7 of IE have the highest weights although they are not recent releases (2001, 2006) as the other browsers in the table. Other browsers have relatively smaller effects. The Safari browser version 4 have a negative effect on sensitivity. The table show there is no correlation between the year of release and the effect of the browser.

The determinants in this class show that the interest of the user in searching for certain items might indicate sensitivity.

Class of User Context In contrast to the traditional view on sensitivity, these determinants show that the context surrounding the user and the data affects sensitivity. The positive effect of determinants indicates context in which data is sensitive and inappropriate to view. The negative effect, however, indicates contexts in which data is not highly sensitive and can be appropriate to view.

1. *cxt.OperatingSystem* has a varying effect on sensitivity based on its value (Table 2.3). The effect of *cxt.OperatingSystem.WindowsNT5.2* ($w = 0.22$) is the most significant effect amongst all the determinants in the three classes. At the same time, *cxt.OperatingSystem.Linux* has a significantly negative effect on sensitivity. This might be, however, due to the low number of linux users.

Browser	*Weight*	*Year of Release*
IE6	0.11	2001
IE7	0.094	2006
Chrome17	0.05	2012
IE8	0.047	2009
Chrome24	0.023	2013
Firefox5	0.0178	2011
Firefox18	0.0171	2013
Firefox13	0.008	2012
Safari5	0.006	2010
Chrome13	0.0051	2011
Chrome19	0.0044	2012
Chrome18	0.0041	2012
Firefox9	0.0041	2011
Chrome6	0.0034	2010
Chrome11	0.0019	2011
Chrome20	0.0018	2012
Chrome28	0.00039	2013
Firefox14	0.00036	2012
Safari3	0.00029	2007
Silk1	0.00023	2012
Safari4	−0.1	2009

Table 2.2: The effects of browser with versions. IE6&7 have the highest effect on sensitivity management. The negative effect of Safari4 means that Safari users are more likely not to manage their sensitive data than users of other browsers.

2. *cxt.FacebookUser.False* ($w = 0.079$), indicating that Bing users who do not sign-in with their Facebook account credentials are more likely to delete and manage their sensitive data than those who sign-in with their Facebook credentials. On the other hand, *cxt.FacebookUser* ($w = -0.11$) indicating that signing-in with Facebook credentials implies that the users are likely not to manage their sensitive data.

3. *cxt.WindowsLiveUser* ($w = 0.019$), indicating that Bing users who sign-in with their Windows live credentials are likely to delete and manage their sensitive data—more likely than those who sign-in with their Facebook credentials.

4. *cxt.DeviceModel.GameConsole* ($w = 0.014$), indicating the users accessing the internet through game consoles are likely to manage their sensitive data. Other device classes indicate sensitivity as well, such as *cxt.DeviceModel.Wii* ($w = 0.013$), *cxt.DeviceModel.BlackBerry* ($w =$

0.0006), *cxt.DeviceModel.LG* ($w = 0.0002$), *cxt.DeviceModel.Windows tablet* ($w = 0.00012$), and *cxt.DeviceModel.OperaMobileAndroid* ($w = 0.0001$). Also, *cxt.DeviceModel.Smartphone* ($w = -0.888$) indicates that users of smart phone are likely not to manage their sensitive data.

5. *cxt.IPCount* ($w = -0.12$), indicating that the more IP addresses the users uses during the day, the less likely the user is to manage sensitive data.

6. *cxt.Hour* ($w = -0.11$), indicating that the higher the hour, the less likely the user is to delete and manage sensitive data. This means that during the night hours, users are less likely to manage sensitive data.

Operating System Family	*Operating Systems*	*Weight*
Windows NT 5.2	Windows XP (64-bit), Windows Server 2003	0.22
Windows NT 5.1	Windows XP	0.043
Windows NT 5.0	Windows 2000	0.031
Windows98	Windows98	0.029
Unknown	Unknown	0.0049
Linux	Linux	−0.097

Table 2.3: The effects of different operating systems on sensitivity.

The LTP shows how various content and context features are associated with sensitivity. Some of these features are not often considered to be associated to sensitivity. For instance, the *cxt.DeviceModel=Game console* and *Wii* affect sensitivity management although they are not commonly considered within the devices that are associated with sensitive data. Usually, computers and mobile devices are the focus of privacy management patterns [20, 105]. The pattern also shows that the measured effect of a feature is affected by how specific the feature is. For instance, adding or removing the version of the *cxt.Browser* feature results with different weights. Adding the version results with a more granular information about how the browser is associated with sensitivity.

In summary, the LTP provides evidence that sensitivity is contextual. The contextual nature of sensitivity is two dimensional, based on the context of data and the user. In contrast to the traditional view, sensitivity may not be defined only by content, but rather, by many aspects of context, and any aspect that reveals the users' interests. Users' interests can be considered sensitive, even without necessarily being inappropriate. However, it should be noted that it would be beneficial to validate our results by users' feedback. In chapter 3, we investigate inferring the LTP with more contextual parameters to check whether the contribution of contextual parameter would vary accordingly.

2.6.2 The Short-term Patterns (STPs)

In this section we analyse the inferred short-term patterns. The analysis supports $H_{temporality}$ and shows divergence across the individual STPs, as well as divergence between the LTP and the STP.

The STPs vary in the number of determinants. The average number of selected determinants is 81 determinants, and the standard deviation is 8 determinants. Such a variation suggests that sensitivity determinants vary over time. By examining the persistent determinants across STPs, we find similarity with the LTP determinants. Such a similarity is expected since the transitory determinants are excluded. The main difference between the LTP and the STPs is found in the content class. This suggests that the sensitivity determination of content varies over time. The variance is evident in determinants that reflect seasonal topics, such as flight status related to holidays (Figure 2.4). The variance between the STPs and the LTP provides evidence for the temporal contextual nature of sensitivity.

2.6.3 User-specific Patterns (UPs)

This section examines subjectivity of the user-specific patterns. We inferred sensitivity patterns of 75 randomly-sampled users. According to the central limit theorem, this small dataset of 75 users is normally distributed and is representative of the whole dataset.

The analysis of the inferred UPs supports $H_{subjectivity}$ and provides evidence for the effect of subjectivity on sensitivity. Subjectivity is demonstrated by the variance between different UPs. Subjectivity implies individual differences between users with regards to what they delete and keep. The UPs vary in terms of the number of determinants and their weights. To demonstrate the difference between UPs, we present examples of mainly the commonality of determinants, without necessarily having to discuss the weights. For instance, within the set of positive determinants, only one determinant (*tp.Navigational=False*) is common in 48 UPs. The *tp.AdultScore* is common in 42 UPs. This means that adult data may not necessarily be considered sensitive to all users. The third most common determinant is *tp.Commerce=False*, common in 40 UPs.

The UPs show prevalence of both *True* and *False* values of the same feature. An example is the *tp.VideoExcludesAdult=False* that is common in 28 UPs and *tp.VideoExcludesAdult* that is common in 28 UPs. These two determinants are common at once in 6 UPs—with different weight for each determinant—and each of them is mutually exclusive in 22 UPs. The 6 UPs show that

Content Determinants	Nov	Dec	Jan	Feb	Mar	Apr
Adult	1	1	1	1	1	1
ClothesAndShoes	1	1	1	1	1	1
Health	1	1	1	1	1	1
MovieTitle	1	1	1	1	1	1
Name	1	1	1	1	1	1
NameNonCeleb	1	1	1	1	1	1
NamePlus	1	1	1	1	1	1
Nightlife	1	1	1	1	1	1
VideoExcludesAdult	1	1	1	1	1	1
QandA	1	1	1	1	1	1
FlightStatus	0	1	0	1	0	0
UrlQuery	0	0	1	1	0	1
MovieShowtimes	0	0	0	1	1	1
RadioStations	0	0	0	1	1	0
Restaurant	0	0	0	1	1	0
Dictionary	0	0	0	0	0	1
Nutrition	0	0	0	0	0	1
TvShows	1	0	0	0	0	0
Bus	0	1	1	0	0	0
Image	1	1	1	1	0	1

Figure 2.4: Persistence of content determinants across STPs. Columns represent the month of an STP, and rows represent determinants. A $cell[x,y]$ indicates '1' if the determinant x is selected in the STP of month y, '0' otherwise. Flight status only shows up in December (Christmas!) and February.

some items that are relevant to *tp.VideoExcludesAdult* may or may not be sensitive, depending on the other features of these items. The 22 UPs in which *tp.VideoExcludesAdult=False* is exclusively common indicate that what is not relevant to *tp.VideoExcludesAdult* is sensitive. The same prevalence of both *True* and *False* values of the same feature is observed in the data and user context classes. An example is the *cxt.IsAlteration* that is common in 39 UPs, and *cxt.IsAlteration.False* that is common in 37 UPs.

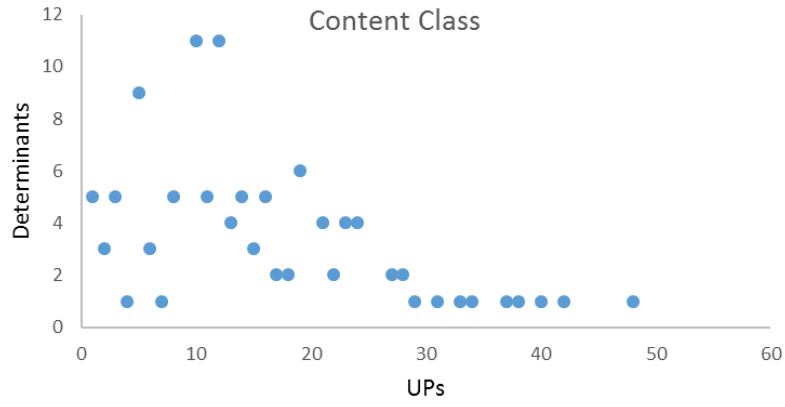
The scale of agreement in the content class is associated with a smaller number of positive determinants and a higher number of negative determinants. For each determinant, the scale of agreement reflects the number of UPs in which this determinant is common. Within the set of positive determinants, the number of common determinants decreases with the scale of agreement (Figure 2.5(a), 2.6(a)). In contrast, within the set of negative agreement, the number of

common determinants does not drop with the increase of the scale of agreement. Rather, the number of common negative determinants slightly increases with the increase of scale of agreement. In contrast to the content class, the data and user context classes do not comprise the same pattern of association (Figure 2.5(b), 2.5(c), 2.6(b), 2.6(c)).

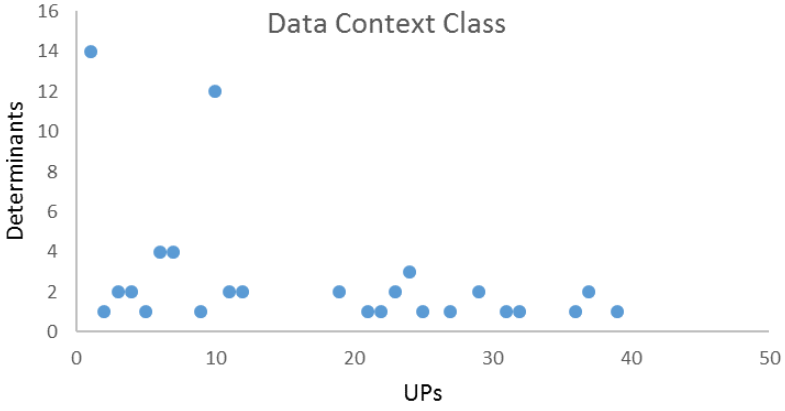
To provide more details about the positive determinants and the scale of agreement, we list the determinants in the corresponding classes.

Class of Content This class shows a varying scale of agreement across determinants with *True* and *False* values. In contrast to the LTP, the *tp.Celebrities=False* is more common than *tp.Celebrities*, for instance. Although, *tp.Celebrities* has a high effect in the LTP on the majority of users, however, this does not mean *tp.Celebrities=False* is not a common indicative of sensitivity for some users. The determinants are:

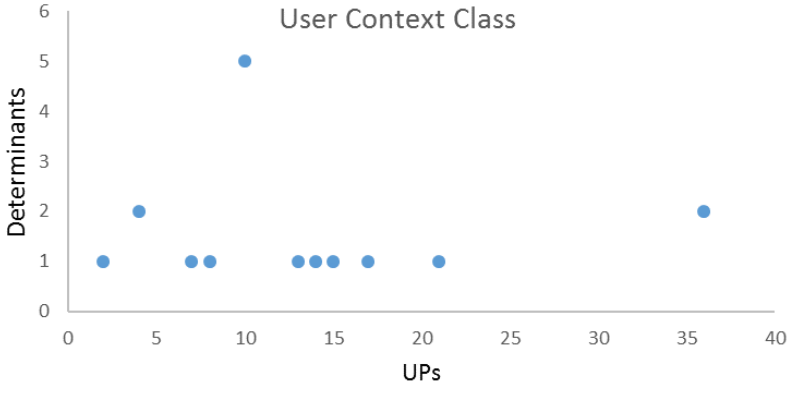
- *tp.Navigational=False* in 48 UPs, *tp.Navigational* in 23 UPs.
- *tp.AdultScore* in 42 UPs, *tp.Adult=False* in 14 UPs.
- *tp.Commerce=False* in 40 UPs, *tp.Commerce* in 21 UPs.
- *tp.ContainsLocation=False* in 38 UPs, *tp.ContainsLocation* in 22 UPs.
- *tp.UrlQuery=False* in 37 UPs, *tp.UrlQuery* in 21 UPs.
- *tp.Image=False* in 34 UPs, *tp.Image* in 33 UPs.
- *tp.Galleries=False* in 31 UPs, *tp.Galleries* in 24 UPs.
- *tp.Local=False* in 29 UPs, *tp.Local* in 24 UPs.
- *tp.VideoExcludesAdult=False* in 28 UPs, *tp.VideoExcludesAdult* in 28 UPs.
- *tp.Seasonal=False* in 27 UPs, *tp.Seasonal* in 16 UPs.
- *tp.Autos=False* in 24 UPs, *tp.Autos* in 14 UPs.
- *tp.Celebrities=False* in 23 UPs, *tp.Celebrities* in 12 UPs.
- *tp.QuestionPattern=False* in 23 UPs, *tp.QuestionPattern* in 12 UPs.
- *tp.Q & A=False* in 22 UPs, *tp.Q&A* in 5 UPs.
- *tp.ConsumerElectronics=False* in 21 UPs, *tp.ConsumerElectronics* in 10 UPs.



(a)

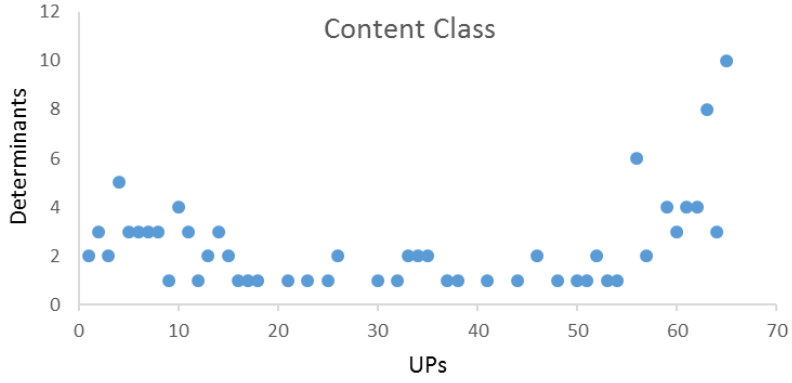


(b)

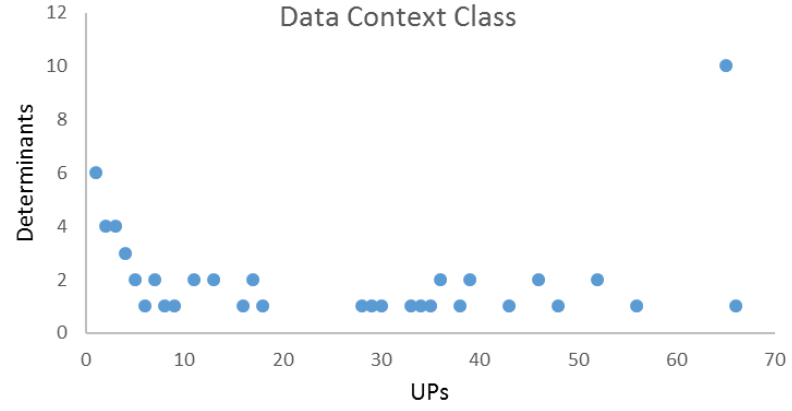


(c)

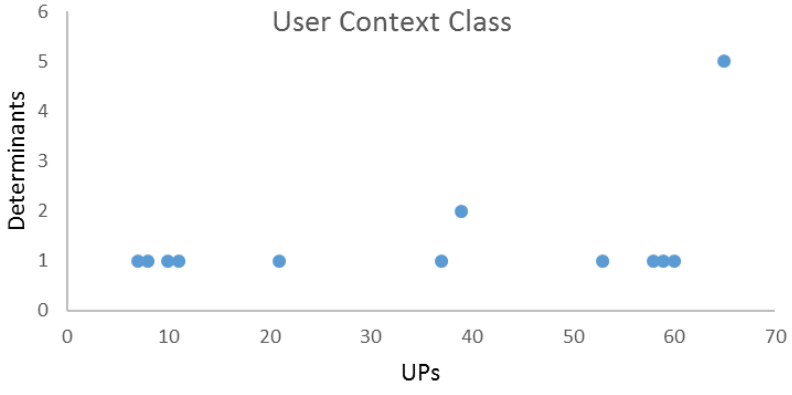
Figure 2.5: Common positive determinants across UPs per feature class.



(a)



(b)



(c)

Figure 2.6: Common negative determinants across UPs per feature class.

- *tp.NamePlus* in 21 UPs, *tp.NamePlus=False* in 19 UPs.
- *tp.Events=False* in 19 UPs, *tp.Events* in 12 UPs.
- *tp.Health=False* in 19 UPs, *tp.Health* in 11 UP.
- *tp.HowTo=False* in 19 UPs, *tp.HowTo* in 5 UPs.
- *tp.Name=False* in 19 UPs, *tp.Name* in 17 UPs.
- *tp.RadioStations=False* in 19 UPs, *tp.RadioStations* in 8 UPs.
- *tp.Book=False* in 18 UPs, *tp.Book* in 3 UPs.
- *tp.WikipediaReference=False* in 18 UPs, *tp.WikipediaReference* in 8 UPs.
- *tp.NameNonCeleb* in 17 UPs, *tp.NameNonCeleb=False* in 15 UPs.
- *tp.Music=False* in 16 UPs, *tp.Music* in 6 UPs.
- *tp.Restaurant=False* in 16 UPs, *tp.Restaurant* in 7 UPs.
- *tp.Sports=False* in 16 UPs, *tp.Sports* in 8 UPs.
- *tp.Tech=False* in 16 UPs, *tp.Tech* in 11 UPs.
- *tp.MovieShowtimes=False* in 15 UPs, *tp.MovieShowtimes* in 5 UPs.
- *tp.Recipes=False* in 15 UPs, *tp.Recipes* in 2 UPs.
- *tp.List=False* in 14 UPs, *tp.List* in 6 UPs.
- *tp.TravelGuide=False* in 14 UPs, *tp.TravelGuide* in 3 UPs.
- *tp.Travel=False* in 14 UPs, *tp.Travel* in 5 UPs.
- *tp.Hotel=False* in 13 UPs,*tp.Hotel* in 8 UPs.
- *tp.MovieTitle=False* in 13 UPs.
- *tp.RealEstate=False* in 13 UPs.
- *tp.Weather=False* in 13 UPs.
- *tp.Dictionary=False* in 12 UPs.
- *tp.Download=False* in 12 UPs.
- *tp.Finance=False* in 12 UPs.
- *tp.Flight=False* in 12 UPs.

- *tp.Jobs=False* in 12 UPs.
- *tp.MovieTheater=False* in 12 UPs, *tp.MovieTitle* in 5 UPs.
- *tp.ThingsTodo=False* in 12 UPs, *tp.ThingsTodo* in 5 UPs.
- *tp.University=False* in 12 UPs, *tp.University* in 3 UPs.
- *tp.Education=False* in 11 UPs, *tp.Education* in 2 UPs.
- *tp.Nightlife=False* in 11 UPs, *tp.Nightlife* in 3 UPs.
- *tp.AppIntent=False* in 10 UPs.
- *tp.Bus=False* in 10 UPs, *tp.Bus* in 1 UPs.
- *tp.ClothesAndShoes=False* in 10 UPs, *tp.ClothesAndShoes* in 5 UPs.
- *tp.FlightStatus=False* in 10 UPs.
- *tp.Maps=False* in 10 UPs, *tp.Maps* in 5 UPs.
- *tp.Nutrition=False* in 10 UPs.
- *tp.OnlineGames=False* in 10 UPs.
- *tp.TvShows=False* in 10 UPs, *tp.TvShows* in 6 UPs.
- *tp.VideoGames=False* in 10 UPs, *tp.VideoGames* in 1 UPs.
- *tp.VideoWithAdult=False* in 10 UPs.
- *tp.Download* in 8 UPs.
- *tp.Dictionary* in 5 UPs.
- *tp.Flight* in 4 UPs.
- *tp.Weather* in 3 UPs.
- *tp.Finance* in 2 UPs.
- *tp.Jobs* in 1 UPs.
- *tp.MovieTheater* in 1 UPs.
- *tp.RealEstate* in 1 UPs.

Class of Data Context This class also shows the variance in the agreement in the determinants. For instance, instead of one prevalent value, the different values of SafeSearchSetting are common in different UPs. The determinants are:

- *cxt.IsAlteration* in 39 UPs, *cxt.IsAlteration.False* in 37 UPs.
- *cxt.IsDotCom.False* in 36 UPs, *cxt.IsDotCom* in 23 UP.
- *cxt.IsSpellSuggestionCorrection.False* in 29 UPs.
- *cxt.SearchHistoryItemCount* in 25 UPs.
- *cxt.Vertical.Web* in 23 UPs, *cxt.Vertical.Images* in 12 UPs, *cxt.Vertical.Video* in 7 UPs.
- *cxt.SessionPageNumber* in 19 UPs.
- *cxt.IsSpellSuggestionCorrection* in 19 UPs.
- *cxt.SafeSearchSetting.Moderate* in 11 UPs, *cxt.SafeSearchSetting.Strict* in 7 UPs, *cxt.SafeSearchSetting.Off* in 6 UPs.
- *cxt.AppName.Bing* in 10 UPs.
- *cxt.AppType.Browser* in 10 UPs.
- *cxt.Browser.IE* in 10 UPs. Upon considering the version of the browser, there is a prevalence for *IE9* in 7 UPs, *IE8* in 4 UPs, *IE10* in 1 UPs.
- *cxt.IsAutoSuggest.False* in 10 UPs.
- *cxt.IsForced.False* in 9 UPs, *cxt.IsForced* in 3 UPs.

Class of User Context The variance in the agreement in the determinants is observed in this class. In contrast to the LTP, the *True* and *False* values of *cxt.WindowsLiveUser*, for instance, are common. Moreover, there is a considerable agreement on the positive effect of *cxt.Hour*, while in the LTP, this determinant as a negative effect. The determinants are:

- *cxt.Hour* in 36 UPs.
- *cxt.FacebookUser.False* in 21 UPs.
- *cxt.IPCount* in 17 UPs.
- *cxt.WindowsLiveUser.False* in 15 UPs, *cxt.WindowsLiveUser* in 13 UPs.

- *cxt.DeviceClass.PC* in 10 UPs.
- *cxt.FacebookUser* in 7 UPs.
- *cxt.OS.Windows NT 5.1* in 4 UPs, *cxt.OS.Windows NT 6.0* in 4 UPs, *cxt.OS Windows NT 6.1* in 2 UPs.

The variance of the UPs in shows the importance of learning such patterns in addition to the LTP. The information that a UP provides is not possible to capture in the LTP due to the generality of the LTP. A UP shows the particular details of the subjective preferences of the user it represents.

2.7 Discussion

The three inferred types of patterns provide complementary insight that could not be learned from any of these patterns alone. The LTP, STPs, and UPs provide information about what determines sensitivity, and how sensitivity changes over time and from one user to another. The findings show that users are not exclusively concerned with topics traditionally considered as sensitive. A second important finding is the contextual nature of sensitivity. A third finding is that the aspects of the information inferred by a pattern depends on the selection of the dataset. Generality, for instance, is the main aspect of the information inferred by the LTP. This aspect is inferred by selecting a dataset that represents a considerably high number of users. Subjectivity is another aspect inferred by selecting smaller datasets of individual users. In general, our work demonstrates the importance of specifying the aspects and the hypotheses of interest to facilitate the proper selection and modelling of data. Our approach of pattern inference can be adopted to investigate similar hypotheses and aspects of information.

Next, we discuss the relation between context, sensitivity, and privacy.

2.7.1 Context and Sensitivity

Our work provides evidence for the role of context in indicating sensitivity of data. The patterns show that a particular set of contextual parameters and situations, e.g., in *cxt.Operating system.WindowsNT* affect sensitivity. These contextual parameters capture both short-term and long-term contexts. The sensitivity of data cannot be judged only based on the content of data. Rather, sensitivity of an item may vary based on the context in which the data is

put. This means that data can be judged as sensitive in particular contexts and as not sensitive in others, e.g., breastfeeding photos can be not sensitive and appropriate to put in contexts related to health and baby care, and can be sensitive and inappropriate in other contexts.

The understanding of the effect of context on sensitivity is fundamental to adopt in privacy management approaches. Current privacy management approaches aim at incorporating context to enhance privacy. In data control approaches, most privacy management approaches are developed to offer control over context. Such control is not adaptive, it does not adapt to the changes of context [91]. The lack of adaptiveness may result in privacy risks. The risks emerge in situations where the change of context results in increasing the sensitivity of data. The challenge is that controlling every change of context can be complicated. A possible solution to address such a challenge is to incorporate mechanisms for context and sensitivity inference that assist users in detecting the changes of the sensitivity of their data, we discuss such an approach for privacy management in chapter 5.

2.7.2 Data Deletion and Privacy

In this section, we discuss the possible reasons for data deletion in relation to privacy management.

Deleting data can be a means for users to fight the predictiveness of intelligent algorithms. In general, web services utilise users' data and behavioural patterns to adapt their services to fit users' expected needs [53]. The service may use users' data to serve targeted ads, as well. Such a usage means that the web service models the user. Intelligent algorithms use the user model to predict what is relevant to the user. Users may not want some data to be included in how the service models them. Some data can be inappropriate and the user may not wish to be associated with that such data, e.g., the user may not want to be associated to his search for "lap dance in Newcastle". By being able to delete their data, users can have control over the predictive algorithms. Such an approach is even facilitated by the EU Data Protection Directive (95/46/EC) [38] The directive empowers users over predictive algorithms by allowing them to control the input to the internal algorithms with the right to object, access, and erase data, according to article 12 of the Directive. Although users do not have the right to alter the algorithm itself, controlling the input to the algorithm is sufficient to a certain extent to have an effect on the internal algorithm and its output.

Deletions can be viewed as a means to manage privacy. Fighting predictive

algorithms is an act of data management. Privacy management is also based on data management. Any data that users have in a web service reflect certain aspects of the user's interests and behavioural patterns. By deleting data, users curate their online identity [78]. From the point of view of privacy as self-determination and identity management [78], deletions are means for privacy management. Whether users delete their data to avoid receiving certain ads, or to fight the predictiveness of algorithms, they manage their identity through deletions, and hence privacy. Such privacy management addresses social privacy concerns. These concerns emerge due to potential technology harms to users social lives.

2.8 Related Work

Our investigation extends two existing streams of work: analysing sensitive topics and deletion behaviours. Lee *et al.* [63] provide evidence on the difficulty of inferring information about sensitive topics. These authors present various approaches to conduct research on sensitive topics, such as surveys. They also presents the ethical and legal issues related to conducting sensitivity research. Our work focuses on sensitivity without facing the challenges identified by Lee *et al.* because our work is quantitive.

Similar to our approach of learning patterns in different settings, Kaplowit [55] discusses two qualitative methods of extracting information about sensitivity. He states that studies based on focus groups provide less information than studies based on interviews. At the same time, the two study types extract complementary information about sensitivity. Such results are comparative to our quantitive analysis approach. In our work, the different patterns extract different information. The LTP provides general information about sensitivity, when the UPs provide more specific information that describe individuals. In total, the patterns provide complementary information that could not be extracted otherwise. In relying on opportunistically observed behaviour rather than self-professed attitudes, our work provides richer patterns about sensitivity and covers a variety of contexts.

Many works have provided evidence as to the association between deletions and privacy management. According to a study of Wang et al., users manage their regrettable Facebook posts through deletions [113]. Deletions are used for privacy management due to the lack of usability of current privacy management approaches in Facebook. According to Boyd, youngsters delete their sensitive data, rather than using access control mechanisms provided by Facebook [17]. Similarly, Tufekci states that a majority of Facebook users delete data from

their profile for privacy management purposes [106]. A previous experiment of Preibusch [82] that examined the value of privacy in Web search, delivered similar results. Users delete items from their search history to manage their privacy. This study provides information about sensitive data that users delete. Although that study was confined to a laboratory experiment, celebrity searches ("Justin Bieber"), for instance, were found to be excluded from the search history more often than tax fraud topics. All these studies have contributed to an increasing interest in studying deletion behaviour to extract insight about sensitivity and privacy management behaviour. However, to the best of our knowledge, our work is the first empirical study on such a large dataset that identifies deletions and sensitivity patterns to serve in better understanding contextual sensitivity and privacy.

2.9 Conclusion

In this chapter, we analyse a big data set of 226 million search items from Bing. To uncover sensitivity patterns, we learn three patterns: a general long-term pattern of the whole dataset over the entire period, multiple short-term patterns of subsets that cover one month each, and user patterns of a single user subset at once. The patterns provide varying information about sensitivity that is, yet, complementary. The patterns provide evidence that content and context affect the sensitivity of data. On the content level, *tp.Adult* topics are highly indicative of sensitivity. However, other topics that are not generally considered as sensitive seem to indicate sensitivity, e.g., *tp.Celebrities*. The user patterns show that other topics indicate sensitivity as well. On the context level, data and user context parameters seem to affect sensitivity. The items that have been searched for before and the type of operating system seem to indicate of sensitivity, as well. Time of the day only emerges as a significant determinant once deletions are considered on a user-per-user basis.

In the next chapter, we investigate further the effect of context on how users disclose and manage their data. The investigation aims at quantifying the effect of context on sensitivity. The quantification shows which contexts affect data disclosure patterns significantly. The investigation also analyses the effect of adding more contextual parameters on learning data management patterns.

Chapter 3

Quantifying the Effect of Context on Sensitivity

3.1 Introduction

Context is a construct that has various effects on how people behave on the web.[1] Context usually identifies what is relevant to the user in a certain situation [101]. Based on the current context, the user performs relevant tasks or handles relevant data. In the context of data management and privacy, context identifies the relevant and proper manner of handling data. In the previous chapter, we demonstrate that context affects sensitivity of data. Our work shows that context affects how and when sensitive data is managed, by means of data deletions. However, this effect does not show whether context affects data disclosure as well. If context affects data disclosure, this might suggest that sensitivity of data varies based on context. In this chapter, we explore the effect of context on data disclosure. By sensitivity we refer to the appropriateness of disclosing certain data item in a particular context. For instance, disclosing photos of *family party* or *porn* is not appropriate in *work* contexts. Sensitivity indicates the inappropriateness of certain data in certain contexts. In this example, the particular context *work* is what makes the *family party* or *porn* photos inappropriate. However, these photos are appropriate to disclose in another context.

Analysing disclosure patterns and how context affects sensitivity is fundamental

[1] This work has been done partially during Rula's internship at Microsoft–Cambridge.

to understand sensitivity. Given that sensitivity of data may be judged based on the context in which it is put, without information about what contexts may affect or not affect data, understanding sensitivity can be limited. Contexts may describe the situation wherein the data is disclosed, or where the user is upon disclosing the data. We define a disclosure pattern as the information about the intensity of data disclosure in a certain context. If disclosure patterns vary across contexts, this implies that context affects the sensitivity of data upon disclosure. To the best of our knowledge, disclosure patterns have not been studied on a large scale. Performing analysis of disclosure patterns requires access to a big dataset where users disclose data of different content types in various contexts. Investigating the effect of context on data sensitivity, requires investigating the relationship between context and content. Analysing the role of context on sensitivity involves investigating how context affects actions on data based on its content type, e.g., whether data about *celebrities* is not disclosed intensely in *work* context. Such an effect of context on content implies a dependency of content on context. Similarly, understanding sensitivity requires investigating whether context is also dependent on content, e.g., whether by knowing the content of data, it is possible to know the context it can be disclosed in. Moreover, investigating the relationship of content and context requires investigating their role on indicating sensitivity. Such an investigation may involve analysing the contribution of content and context features to the modelling of sensitivity management patterns. It may also involve, for instance, analysing whether adding more contextual features would affect the modelling.

In this chapter, we focus on analysing the effect of context on data sensitivity upon disclosure and post-disclosure. We conduct our analysis on the same dataset we used in the previous chapter. The dataset has 226,000,000 data items from the Microsoft search engine Bing. These items are disclosed and managed by 413,000 users. We incorporate extra contextual features that describe the online and offline contexts of users. The online context describes the online situation in which a data item is disclosed. The offline context describes the offline situation surrounding the user upon disclosing a data item. Towards investigating the effect of context on data disclosure and post-disclosure patterns, this chapter contributes the following:

1. A description of the extra contextual data we add to the dataset (Section 3.2)

2. A description of the analysis method. The method involves three type of analyses. Firstly, we analyse and quantify the effect of context on disclosure patterns through the utilisation of test of homogeneity through the goodness of fit for multinomial distributions test. Secondly, we model

context using content via multinomial logit modelling to investigate the dependency of context on content. Thirdly, we analyse the role of the extra contextual features and content in modelling sensitivity management patterns post to disclosure. (Section 3.3)

3. A presentation of the results of the three analyses that show that context significantly affects data disclosure, the dependency between context and context, the role of the extra contextual features in enhancing the modelling of sensitivity management patterns (Section 3.4)

3.2 Dataset

This section describes the dataset, and the disclosure and post-disclosure patterns derived from the dataset.

We use the dataset described in the previous chapter 2 in addition to extra contextual information. The extra contextual information represents the offline and online context of users and their searches. The offline context is represented by geolocation data that describes the situation surrounding the user when accessing the internet, e.g., connecting from home. The online context represents the online situation from within which the user submitted a search, in addition to data about how the user is accessing the internet, e.g., using a proxy. The anonymity of users' identities was preserved during the extraction of the extra contextual information.

In the following, we describe the extra contextual features that is added to the dataset within the *data context* and *user context* classes.

3.2.1 Class of Data Context

The following extra contextual features are added to this class:

- *cxt.SearchService*: the type of the web service from which the user submitted the search to Bing. This context has the following values:
  - Celebrities
  - Domains
  - Entertainment
  - Explore
  - Games
  - History
  - Local
  - Movies

  - Music
  - News
  - OnlineGames
  - Rewards
  - Social
  - Tags
  - Travel
  - TV

- *cxt.MSNService*: the search service offered by the MSN web portal. The MSN search services are the following:
  - Food channel
  - Autos
  - Careers and Jobs
  - Clients
  - Entertainment
  - Health
  - Healthy living
  - Homepage
  - Lifestyle
  - Living
  - Local edition
  - Money
  - Movies
  - Music
  - News
  - MSN Now
  - Photos
  - Realestate
  - Sports
  - Tech and gadgets
  - Travel
  - TV
  - Video services
  - Weather

3.2.2 Class of User Context

This class includes features that characterise the online and offline context surrounding the user upon disclosing the data.

- Online context: encompasses features of the machine or the network the user is accessing when submitting the search. The following features are added to this class:
  - *cxt.TouchDevice*: the type of device the user is using. Its possible values are *Touch* and *NotTouch*.
  - *cxt.LineSpeed*: the speed of the internet connection of the user. Its possible values are:
    * Low
    * Medium

    - High
  - *cxt.ConnectionType*: this feature has the following possible values:
    - Cable
    - Consumer satellite
    - Dialup
    - DSL
    - Fixed wireless
    - Frame relay
    - ISDN
    - Mobile wireless
    - OCx
    - Tx
    - Unknown high
    - Unknown medium

    Each value is characterised by different speeds (Table 3.1).
  - *cxt.AnonymiserStatus*: the type of an anonymisation service the user is using. The possible values are the following:
    - Active
    - Inactive
    - Private
    - Suspect
  - *cxt.ProxyLevel*: the degree of obfuscation of the used proxy. The possible values are:
    - Transparent: A proxy that forwards the user's IP address to the target service. It offers no anonymity.
    - Anonymous: A proxy that does not reveal the user's IP address to the target service.
    - Distorting: A proxy that hides the user's IP address to the target service.
    - Elite: a proxy that offers the highest degree of anonymity that the target service does not know that the user is using a proxy.
  - *cxt.ProxyType*: indicates the proxy technology used. The relevant variations are:
    - Http: A type that offers a range of anonymity levels.
    - Tor: An onion routing proxy type.
    - Web: A web-based type that conceals the real IP of the user.
    - SOCKS: A type of proxy that offers a complete anonymity. This type do not add any identifying information to the communicated information.

- Offline context: encompasses features related to the offline situation the user is in when disclosing the search. The features are related to the real world environment, as follows:

Type	*Description*	*Speed*
OCx	Fiber optic connections, which are used primarily by large backbone carriers.	High
Tx	Leased line, which is circuits used by many small- and medium-sized companies.	High
Frame relay	Frame relay circuits, which can range from low- to high-speed and are used as a backup or alternative to Tx.	High
Unknown high	Unknown connection type with an estimated connection speed as high.	High
Consumer satellite	High-speed or broadband links between a consumer and a geosynchronous or low-earth orbiting satellite.	Medium
DSL	Digital Subscriber Line broadband circuits.	Medium
Cable	Cable Modem broadband circuits, offered by cable TV companies.	Medium
ISDN	Integrated Services Digital Network high-speed technology. Offered by some major telephony companies.	Medium
Fixed wireless	Fixed wireless connections, where the location of the receiver is fixed.	Medium
Unknown medium	Unknown connection type with an estimated connection speed as medium.	Medium
Mobile wireless	Cellular network providers who employ CDMA, EDGE, EV-DO, GPRS, 3G, and 4G technologies.	Low
Dialup	Consumer dial-up modem technology.	Low

Table 3.1: Description of connection types and their speed.

- *cxt.Home*: indicating whether the user is at home or not.
- *cxt.OrganisationType*: the type of organisation the user is in when disclosing the data item. The possible values are:
  - Accounting and Auditing
  - Advertising
  - Agriculture
  - Banking
  - Business Conglomerate
  - Construction
  - Data Services
  - Dining
  - Education
  - Finance
  - Gaming
  - Government
  - Government (County)
  - Government (Federal)
  - Government (General)
  - Government (Municipal)
  - Government (State)
  - Health
  - Hospital
  - Insurance
  - Internet Cafes
  - Internet Colocation Services
  - Internet Hosting Services
  - Internet Service Provider
  - Legal Services
  - Library
  - Lodging
  - Manufacturing
  - Medical and Dental Services
  - Member Organisation
  - Motor Vehicles
  - Pharmacy
  - Private Service
  - Professional Service
  - Publishing
  - Real Estate
  - Religious Organisations
  - Research and Development
  - Retail
  - Telecommunications
  - Transportation
  - Travel Services
  - Utilities
  - Wholesale
- *cxt.Weekday*: the day on which the data is disclosed.

In the following sections, we refer to a value *Y* of context *cxt.X* as *cxt.X.Y*.

3.3 Method

This section presents the statistical approach we follow to test the effect of context on disclosure and post-disclosure patterns. It also describes our investigation of the dependency of context on content.

Our analysis investigates the interaction of content and context that affects the sensitivity of data. We consider submitting a search item an act of disclosure to the search engine that may or may not be related to privacy concerns. The user assesses the latent sensitivity of a data item upon disclosure. The sensitivity of the item affects when and where it is disclosed. The user can further manage the sensitive data post disclosure. The user can hide, delete, or limit the access to sensitive data items. We investigate the effect of context on disclosure patterns. The effect of context on disclosure patterns is captured by the change of disclosures in different content features based on context. We also investigate whether the content affect context. Lastly, we present our analysis method to investigate the effect of context on post-disclosure patterns.

3.3.1 Analysis of Disclosure Patterns

Our approach to analyse disclosure patterns is based on examining the effect of one context on the intensity of disclosing data within the different values on this context. The effect of context on disclosures implies that the sensitivity of disclosed data varies based on this context.

Initially, we compute disclosure patterns from the dataset. A *disclosure pattern* is the information that represents the intensity of data disclosures in a particular context. To compute a disclosure pattern, we group the data based on one contextual feature, or one context, at a time. We then compute the count of items within each content feature. The result is a pattern that shows the intensity of search items per content features and one context feature. The number of possible values of the context corresponds to the number of disclosure patterns of this context, e.g., two disclosure patterns are computed based on the context *cxt.Home* that has two possible values.

Disclosure patterns of a context are represented in a two-dimensional contingency table (Table 3.2). In our analysis, a context is a factor that can affect data disclosure. The possible values of a context represent the levels of the factor. The frequencies (sample sizes) of each content feature are hence independent realisations of a Poisson distribution. The frequencies of the content features within one level are a population. Thus, given the sizes of the respective content populations, a level is distributed according to a multinomial distribution. The contingency table represents the frequency of disclosures within the levels of a context. Each row is a content feature, and each column is a level.

To investigate the effect of a particular context, we apply the χ^2 test of homogeneity on the disclosure patterns of this context. The test determines

Context Levels	***ft.Adult***	***ft.Celebrities***
$Level_1$	$Count_{1,1}$	$Count_{1,2}$
$Level_2$	$Count_{1,2}$	$Count_{2,2}$

Table 3.2: A contingency table of a context factor with two values and two content features.

whether multinomial distributions are homogeneous or equal in reference to one factor [14, 96]. The test is applied on a contingency table. It summarises the discrepancy in the populations, and tests whether the population samples are drawn from identical distributions. When there is a significant discrepancy, we say that the factor affects the populations, and hence, affects the corresponding disclosure patterns. χ^2 is computed as follows:

$$X^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed frequency count of content feature i for level j, and E_{ij} is the expected frequency count of content feature i for level j. E_{ij} is computed as follows:

$$E_{ij} = \frac{(N_i * N_j)}{N}$$

where N_i is the total number of observations from content feature i, N_j is the total number of observations at level j, and N is the total sample size.

X^2 tests the null hypothesis that all the considered the multinomial distributions of a context are equal—or, in other words, drawn from the same distribution. The null hypothesis can be written for a particular factor, and a set of j levels and i content features, as:

$$H_0 : P_{i1} = P_{i2} = \cdots = P_{iJ} : \forall i$$

where P_{ij} is the probability that an observation from the ith content feature belongs to the jth level.

Rejecting the null hypothesis indicates a significant effect of the factor on the distributions, for P-value < 0.05.

We perform post-hoc analysis of χ^2 to further analyse and quantify the effect of the factor levels on content feature. The χ^2 shows whether there is a significant

difference in at least one of the populations. Given that our contingency tables are larger than (2X2), rejecting the H_0 does not provide information about which level of the contingency table is the source of a statically significant result. To identify such levels, we perform post-hoc analysis based on residual analysis. A residual is the difference between the observed and expected values for a cell. This analysis quantifies the degree to which an individual cell diverges from an overall hypothetical homogeneous distribution, and contributes to the chi-square test result [96]. In other words, this analysis quantifies the effect of each level on each content feature. We use the *Pearson* residual defined as:

$$PR_{ij} = \frac{(O_{ij} - E_{ij})}{\sqrt{E_{ij}}}$$

where O_{ij} is the observed frequency count of cell ij of the contingency table, and E_{ij} is the expected frequency of cell ij. To identify the contribution of each cell, we follow Agresti's recommendation [5] and choose a threshold value $T = 3$. The threshold indicates that the absolute residual values greater than T contribute to the statistically significant result. For a cell $C_i j$, if $PR_{ij} \geq T$, we say that $C_i j$ positively contributes to the result, or that $C_i j$ has more values than expected. If $PR_{ij} \leq -T$, we say that $C_i j$ negatively contributes to the result, or that $C_i j$ has less values than expected. If $T > PR_{ij} > -T$, we say that $C_i j$ has no contribution, or that $C_i j$ has the expected number of values. The interpretation of these cases is that the context level j affects the content feature i positively, negatively, or has no effect, respectively.

Let us remark that two non-standard traits of the dataset pose potential problems towards the interpretation of the results of the statistical analysis. Firstly, the contingency tables are typically highly unbalanced, meaning that in an individual table the marginal sums of observed values over the rows (and columns) vary substantially (see also Section 2). This may result in an inaccurate post-hoc analysis, since the rows (columns) with high number of observations are typically more likely to indicate significance. Nevertheless, this does not affect the overall results of the χ^2 tests, and the main results remain in order. Secondly, in the used dataset it is possible that one search item relates to more than one content feature at once. However, it might be noted that this phenomenon does not significantly affect the results, as the ratio of the total number of contributions to the total number of items is still rather low (1.53).

3.3.2 Modelling Context based on Content

In this section, we present our approach towards investigating the effect of content on context. We investigate the effect of content on context by investigating the dependency of context on content. The possibility to model the probabilities of being at a particular context using content features implies the dependency of context on content. Context modelling means that through knowing the content we can predict the context appropriate for the sensitivity of this item. This also means that through knowing the content, the sensitivity of an item is implicitly assessed to predict the relevant context. In contrast to the previous analysis, this analysis demonstrates that sensitivity is not only affected by context, rather by content as well.

Such a modelling is possible given the high dimensionality of the content features in contrast to the low dimensionality of each context—except for the *cxt.Organisation* that has a relatively equal number of dimensions to the content features. For this context, the modelling can be performed in both directions, modelling the context based on content or modelling the content based on context. However, we only focus on modelling the context using the content in this section.

Content Feature	$cxt.Context.Category_1$	$cxt.Context.Category_2$
ft.Adult	$Count_{1,1}$	$Count_{1,2}$
ft.Celebrities	$Count_{1,2}$	$Count_{2,2}$

Table 3.3: A contingency table of two content features and a context with two categories.

The modelling is based on a contingency table of the modelled context (Table 3.3). The determinant variables are the content features and the dependent variable is the context. The category values of the context are mutually exclusive. Given that the probability distribution of the context is multinomial distribution, we use multinomial logit modelling [5]. In this modelling, a category of the dependent variable is selected as the baseline category. The model computes the log-odds for other categories relative to this category. The log-odds of each category of the dependent variable follow a linear model:

$$\eta_{ij} = \log \frac{\pi_{ij}}{\pi_{iJ}} = \alpha_{ij} + \beta_{ij} X_{ij}$$

where $\pi_{ij} = Pr\{Y_i = j\}$ indicates the probability of an item of the ith content falls in the jth context category, β_{ij} is a vector of regression coefficients, for $j = 1, ..., J-1$ context categories. The model describes the effects of X on the $J-1$ logits.

Through modelling context, we also investigate the similarity of content features. Given the relatively high order of content features, we are interested in dimensionality reduction of these features. The investigation aims at discovering whether using a smaller number of features, we can model context. If certain content features are similar, by clustering these features together, we can reduce the dimensionality and thus model the context. We investigate similarity through applying clustering of content features before modelling context.

We use Ward hierarchical clustering analysis method [114]. This method clusters data in a multivariate Euclidean space, similar to the approach of the principal component analysis (PCA). Ward's method is used for correspondence analysis to represent categorical data in a low-dimensional Euclidean space [74]. This method follows a bottom-up approach and is low cost, compared to k-means [74]. The method starts with an individual cluster for each data item, and then clusters are merged together to minimises the within-cluster sum of squares over all partitions.

For each context, we apply clustering with different numbers of clusters. We assess the fit of each clustering in terms of the accuracy of predicting context.

3.3.3 Analysis of Post-Disclosure Patterns

Our approach to investigate the effect of context, and content as well, on post-disclosure patterns is through modelling the sensitivity pattern from deleted and kept data. It is not possible to apply χ^2 test on post-disclosure patterns because we aim at analysing how context, as well as content, affect the deletion indicator variable at once. Rather, we adopt the approach of modelling the sensitivity pattern. The modelling approach is the same as the approach of the previous chapter 2.5. The current analysis focuses on whether adding more contextual features would affect the inferred sensitivity pattern.

The preprocessing and modelling of the sensitivity pattern are performed according to (Section 2.5). We refer to the inferred pattern with the extra contextual features as $\mathbf{LTP_{new}}$, and to the pattern with fewer features inferred in the previous chapter as $\mathbf{LTP_{old}}$. We compare these two patterns to each other and compare the contribution of the three feature classes to each pattern.

3.4 Results

This section presents the results of the analysis as described in the previous section. The analyses were performed on the whole dataset with the external contextual features.

3.4.1 The Effect of Context on Disclosure Patterns

This section presents the results of applying our analysis method on disclosure patterns of the contextual features, including the extra features. Each feature represents one context, as previously mentioned. The χ^2 is applied on each context. When the test provides evidence for a significant difference between the disclosure patterns of a context, we perform the post-hoc analysis to quantify the difference between the disclosure patterns.

To represent the results of the post-hoc analysis, we use heat map plots. The heat map represents the residual analysis table. Each cell of this table shows the degree of contribution of each corresponding cell in the context contingency table to the difference between disclosure patterns. A negative contribution to the disclosure pattern means that the context level, of a cell, affects sensitivity of the content feature, of the same cell, positively to the degree that disclosures of this feature decrease. In contrast, a positive contribution means that the context level, of a cell, affects sensitivity of the content feature, of the same cell, negatively to the degree that disclosures of this feature increase. When the cell has no contribution to the difference, this means that the counts of the content feature do not differ from the expected counts.

The application of χ^2 on each context resulted with a significant difference. In other words, for each context, the intensity of searches across at least one of the values of this context vary significantly. The significant variance of disclosure patterns may be due to users being more interested in a particular set of topics in a particular context variable than in other context variables. Nevertheless, the variation of users' interest affects what people search for, and hence their disclosure patterns. In the following, given the significance effect of all contexts separately, we mainly present the results of the post-hoc analysis of each context.

Vertical

The *cxt.Vertical* affects disclosure patterns significantly. The patterns vary significantly across the values of this context (Figure 3.1). *cxt.Vertical.Web*, *cxt.Vertical.Video*, and *cxt.Vertical.Image* context have a varying effect on the content features. For instance, *cxt.Vertical.Video* negatively affects—decreases—disclosures of *tp.Adult*, while *cxt.Vertical.Image* positively affects—increases—disclosures of the same feature.

Safe Search Setting

The *cxt.SafeSearch* affects disclosure patterns significantly. Across the three possible values, the *cxt.SafeSearch.Moderate* has mostly a positive effect on the content features, while *cxt.SafeSearch.Off* has a mostly negative effect, and *cxt.SafeSearch.Strict* has a varying effect on content features (Figure 3.1). For instance, *cxt.SafeSearch.Strict* negatively affects *tp.Adult* disclosures, while *cxt.Safe Search.Off* and *cxt.Safe Search.Moderate* affect disclosure positively.

App Type

The *cxt.AppType* affects disclosure patterns significantly. The *cxt.AppType.Browser* has no significant effect on disclosures across all content features, while *cxt.AppType.App* has a varying effect on content features (Figure 3.2). The significant effect of *cxt.AppType* means the two disclosure patterns vary significantly, although *cxt.AppType.Browser* has no effect on disclosures. However, *cxt.AppType.App* and *cxt.AppType.Browser* have no effect on the same content features, e.g., on *tp.Adult* disclosures. Such an effect on *tp.Adult* means that disclosures of this topic do not vary based on the value of *cxt.AppType*.

Forced Searches

The *cxt.ForcedSearch* context affects disclosure patterns significantly. The *cxt.ForcedSearch.True* has a varying effect on content features although while *cxt.ForcedSearch.False* has no effect on disclosures (Figure 3.2). For instance, *cxt.ForcedSearch.True* has a positive effect on *tp.NameNonCeleb* and *cxt.ForcedSearch.False* has no effect on *tp.NameNonCeleb* disclosures.

Search Service

The *cxt.SearchService* affects the disclosure patterns significantly. Certain values of this context have no effect on most of content features, e.g., *cxt.SearchService.Tags* and *cxt.SearchService.Travel*. Most of the context values, however, have a relatively more negative effect on content features (Figure in Appendix A.1). Positive effects can result from the relevance of the content to the context, for instance, *cxt.SearchService.Social* has a positive effect on *tp.Sports* disclosures and *cxt.SearchService.Local* has a positive effect on *tp.Maps*. Such a correlation suggests that disclosure increase when the content of the data is relevant to the context in which the data is disclosed.

MSN Service

The *cxt.MSNService* affects the disclosure patterns significantly. The majority of the values of this context have a negative effects on content features, e.g., *cxt.MSNService.Photos* and *cxt.MSNService.Travel* (Figure in Appendix A.2). The effect of *cxt.MSNService.Homepage* has a positive effect on most of the content features. Similar to the *cxt.SearchService*, the positive effect of context values on content features is correlated to the relatedness between the content and context, for instance, *cxt.MSNService.Jobs* has a positive effect on *tp.Jobs*, like *cxt.MSNService.Health* has on *tp.Health*.

Browser

The *cxt.Browser* affects the disclosure patterns significantly. All the values of this context has either a negative or a positive effect on content features (Figure in Appendix A.3). Despite the variance of the effect of the different browsers, *cxt.Browser.Safari* and *cxt.Browser.Opera*, as well as *cxt.Browser.Chrome* and *cxt.Browser.Firefox* are more similar in their effects than *cxt.Browser.IE* and *cxt.Browser.Silk*. An example is *cxt.Browser.Safari* and *cxt.Browser.Opera* have a positive effect on *tp.Dictionary*.

Device Class

The *cxt.DeviceClass* affects the disclosure patterns significantly. The *cxt.DeviceClass.True* has a varying effect on content features, but mostly positive. The *cxt.DeviceClass.False* has no effect on the majority of content features, and a positive effect on few of the content features, e.g.,

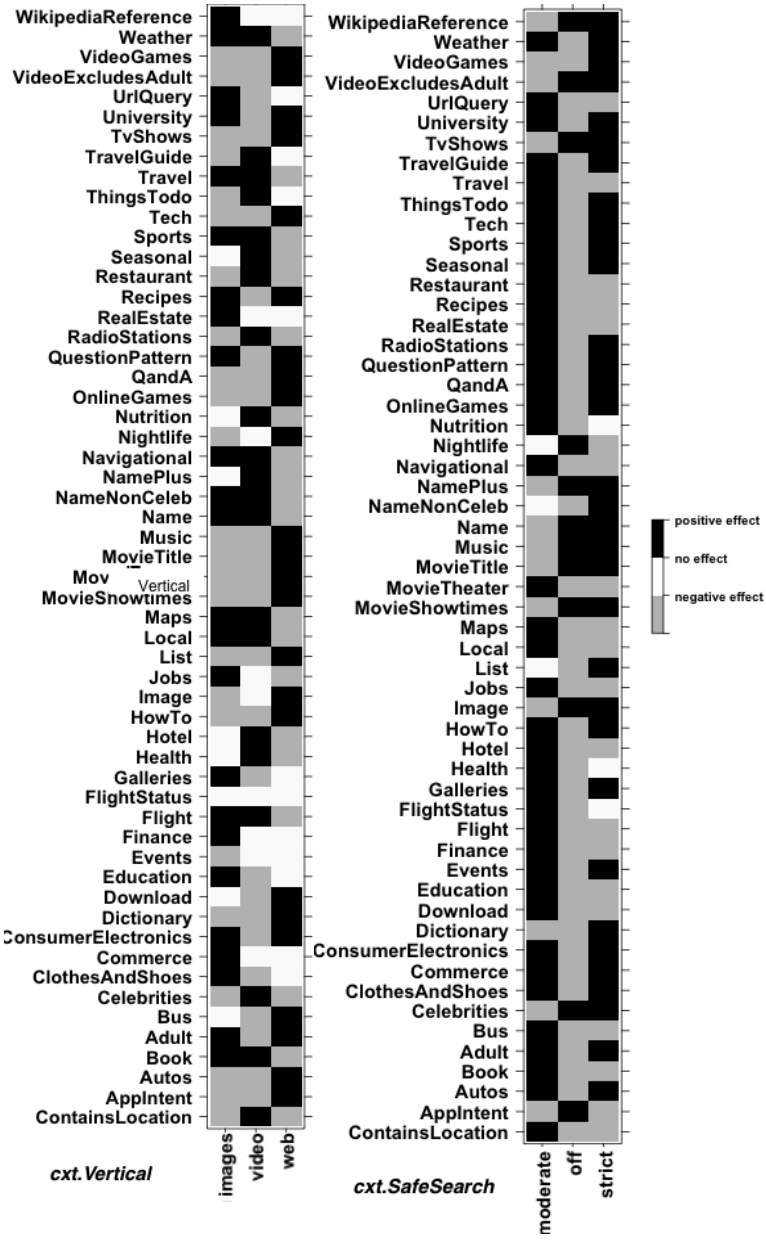


Figure 3.1: The heat maps of *cxt.Vertical* and *cxt.SafeSearch*. *cxt.Vertical.Video* negatively affects—decreases—disclosures of *Adult*, while *cxt.Vertical.Image* positively affects—increases—disclosures of *Adult*.

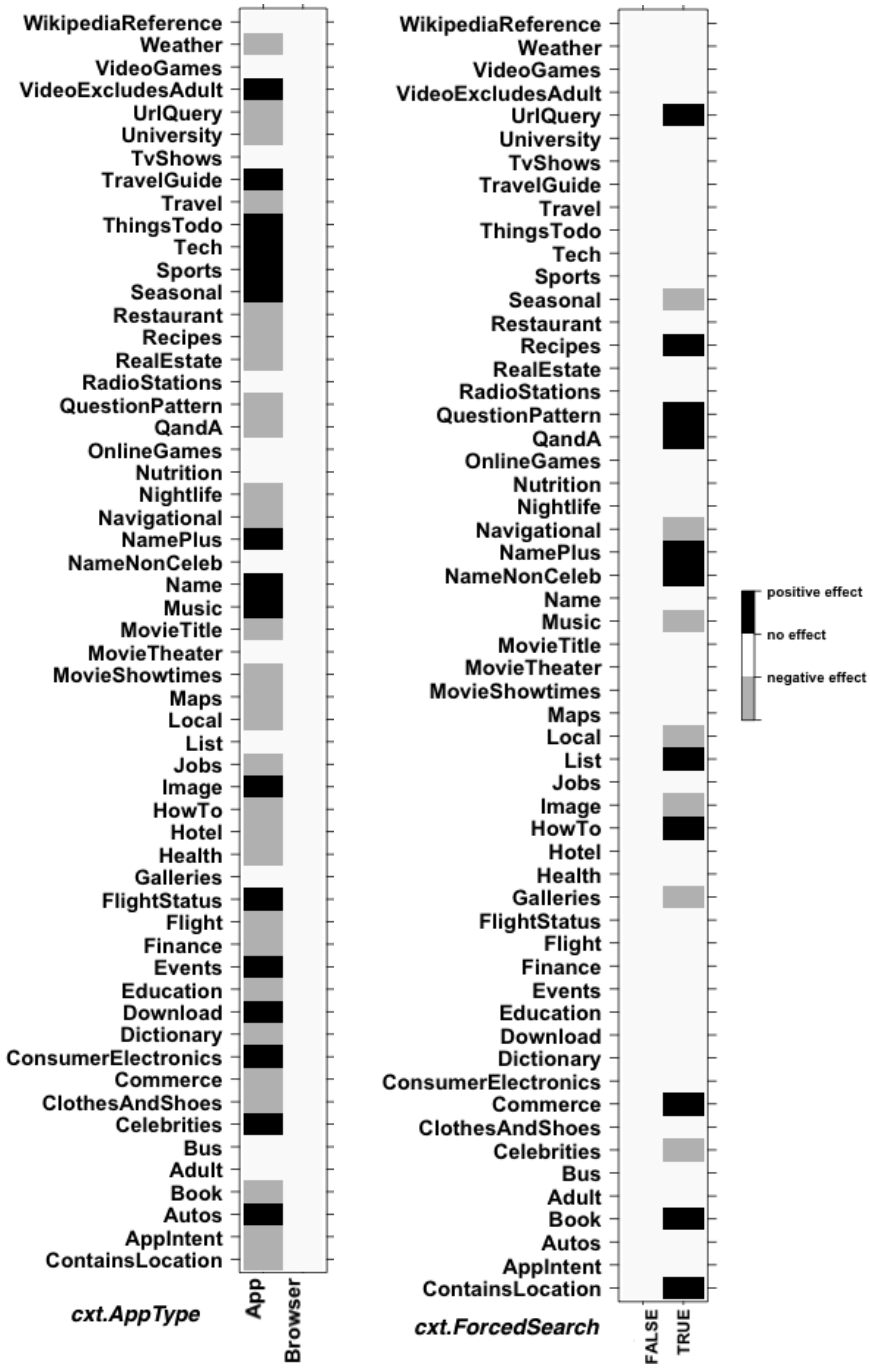


Figure 3.2: The heat maps of *cxt.AppType* and *cxt.ForcedSearch*.

tp.Navigational, *tp.Local*, and *tp.ContainsLocation* (Figure in Appendix A.4). The effect of *cxt.DeviceClass.False* means that the intensity of items submitted through a computer—not a mobile—is not affected, except for a very few content features.

Touch Device

The *cxt.TouchDevice* affects the disclosure patterns significantly. The two values of this context *cxt.TouchDevice.False* and *cxt.TouchDevice.True* have a varying effect on different content features (Figure in Appendix A.4). For instance, *cxt.TouchDevice.False* has a positive effect on *tp.RealEstat* and a negative effect on *tp.Celebrities*.

Facebook

The *cxt.Facebook* affects the disclosure patterns of its possible values significantly. The two possible values have a varying effect on the content features (Figure in Appendix A.5). For instance, *cxt.Facebook.True* has a negative effect *tp.Education* disclosures and a positive effect on *tp.Adult* disclosures.

Windows Live

The *cxt.WindowsLive* affects the disclosure patterns significantly. The patterns vary significantly across the context values (Figure in Appendix A.5). In contrast to *cxt.Facebook*, *cxt.WindowsLive.True* has a negative effect on most content features that *cxt.Facebook.True* has a positive effect on. An example is the negative effect of *cxt.WindowsLive.True* and the positive effect of *cxt.Facebook.True* on *tp.Adult*.

Connection Type

The *cxt.ConnectionType* affects the disclosure patterns significantly. The values of this context have a varying effect on data disclosures. For instance, *cxt.ConnectionType.UnknownMedium* and *cxt.ConnectionType.UnknownHigh* have mostly no effect on content features, while *cxt.ConnectionType.Consumer-Satellite* has mainly a negative effect on content features (Figure in Appendix A.7).

Line Speed

The *cxt.LineSpeed* affects the disclosure patterns significantly. The *cxt.LineSpeed.High*, *cxt.LineSpeed.Low*, and *cxt.LineSpeed.Medium* have varying effects (Figure in Appendix A.6). The *cxt.LineSpeed.High* have more positive effect than the other two values. This effect can be due to the ease of disclosing more items when the line speed of the internet is high, and not necessarily related to sensitivity of data. For instance, *cxt.LineSpeed.High* has a positive effect on *tp.Maps* and a negative effect on *tp.Navigational* disclosures.

Anonymiser Status

The *cxt.AnonymiserStatus* affects the disclosure patterns significantly. The values of this context have mainly no effect on most of the content features (Figure in Appendix A.6). The *cxt.AnonymiserStatus.Suspect* and *cxt.AnonymiserStatus.Active* have the most positive effect on content features.

Proxy Level

The *cxt.ProxyLevel* context affects the disclosure patterns significantly. The values of this context have mainly no effect on content features (Figure in Appendix A.8). While the *cxt.ProxyLevel.Elite* has no effect on all content features, *cxt.ProxyLevel.Distorting* and *cxt.ProxyLevel.Transparent* have the most positive effect on content, and *cxt.ProxyLevel.Anonymous* do not have negative effect on any feature.

Proxy Type

The *cxt.Proxy Type* affects the disclosure patterns significantly. The values of this context have a varying effect on disclosure. The *cxt.ProxyType.Web* has mainly negative effect on most of the content features, and *cxt.ProxyType.Tor* has the most of positive effect on content features (Figure in Appendix A.8).

Home

The *cxt.Home* affects the disclosure patterns significantly. The values of this context exhibit various contribution of the content features to the difference between disclosure patterns (Figure in Appendix A.9). The contribution is

mainly negative or positive. Only *tp.ThingsTodo* and *tp.FligthStatus* are not affected by any of the context values.

Organisation Type

The *cxt.OrganisationType* affects the disclosure patterns significantly. The possible values of this context very in their effect on content features. The context values have mainly positive effect on features such as *tp.ContainsLocation*, *tp.Name*, *tp.NamePlus*, *tp.Celebrities*, *tp.Restaurant* and *tp.RadioStation* (Figure in Appendix A.10). The positive effect on content features represent the topic that users disclose mostly in work places. The contexts *cxt.OrganisationType.Advertising*, *cxt.OrganisationType.Agriculture*, *cxt.OrganisationType*DataService, and *cxt.OrganisationType.InternetCafes* are examples of context values that have mainly no effect on most of the content features. Such content features represent topics that are not affected by the type of work place of a user.

Weekday

The *cxt.Weekday* affects the disclosure patterns significantly. The context values have varying effect on content features. The *cxt.Weekday.Saturday* and *cxt.Weekday.Sunday* have similar effects on content features. Context values that represent week days have more negative effects on content in comparison to weekend days (Figure in Appendix A.11).

3.4.2 Modelling Context based on Content

In this section, we present the results of context modelling and cluster analysis of content features. We show that context is dependent on content. We also show that it is possible to reduce the dimensionality of the content features and cluster them to capture their similarity in affecting disclosure patterns. The clustering exhibits semantic similarity between content features.

The modelling shows that all the contexts in our dataset are dependent on content. The multinomial logit regression was performed with maximum 2000 iterations to guarantee the conversion of the model. The regression model converges for all contexts. The convergence means that the regression algorithm infers a model of the data. The algorithm requires different number of iterations for each context depending on the number of values of the modelled context (Table 3.4).

Context	*Number of Iterations*	*Number of Values*
cxt.Organisation	1780	45
cxt.MSNService	960	24
cxt.Connection	760	12
cxt.SearchService	480	16
cxt.Browser	380	6
cxt.ProxyLevel	240	4
cxt.ProxyType	210	5
cxt.AnonymiserStatus	210	4
cxt.Line	160	3
cxt.Vertical	140	3
cxt.SafeSearch	130	3
cxt.Weekday	100	7
cxt.TouchDevice	90	2
cxt.Facebook	80	2
cxt.Mobile	70	2
cxt.Home	70	2
cxt.Windowslive	70	2
cxt.Forced	1	2
cxt.AppType	1	2

Table 3.4: The number of iterations to model a context, and the number of values of each context. The number of iterations is high for contexts with high number of values ≥ 12. This number varies for smaller number of values.

The cluster analysis of content features demonstrates that there is a similarity between different content features in relation to modelling context. The clustering groups content features based on the similarity of searches in relation to a context. We modelled the context after clustering the content features based on this context. We applied cluster analysis with a varying number of clusters $C = \{15, 30, 56\}$. The clustering shows similarities between the content features based on on their contribution to the disclosure patterns of one context. The similarity varies based on the context and the number of clusters. For instance, for the *cxt.Vertical*, the clusters differ from the clusters of the *cxt.Weekday* (Figures 3.3, 3.4). The clustering shows that semantic similarity plays a role in this clustering. For example, *tp.Autos* and *tp.Bus*, and *tp.MovieShowTime* and *tp.MovieTitle* are in the same clusters (Figure 3.3).

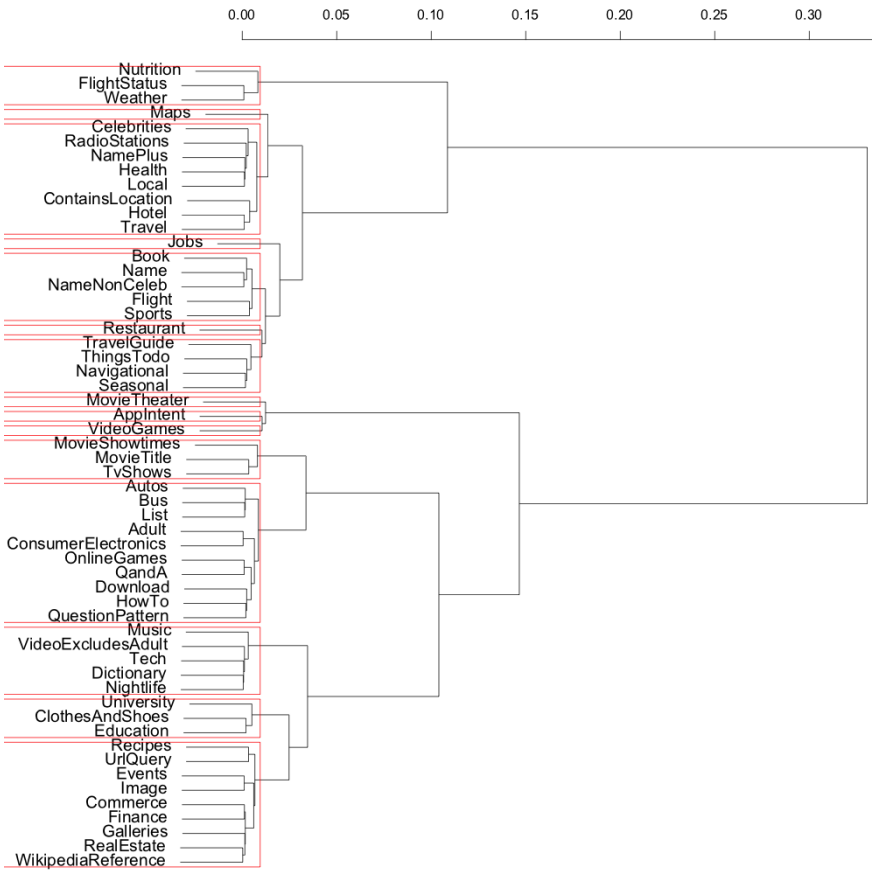


Figure 3.3: Dendrogram of the cluster analysis of *cxt.Vertical*. The horizontal axis represents the distance or dissimilarity between the content features, e.g., the distance between *tp.Nutrition* and *tp.Jobs* is 0.10. The clusters group certain content features with semantic similarities together, e.g., *tp.ContainsLocation*, *tp.Hotel* and *tp.Travel* are in the same cluster.

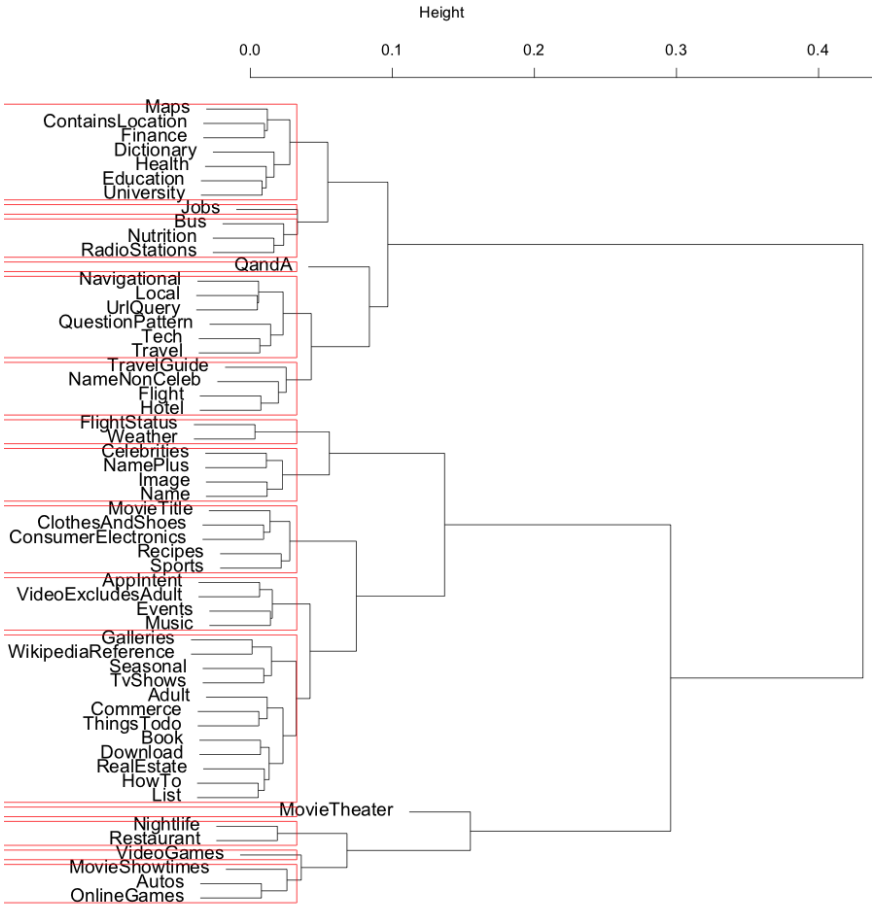


Figure 3.4: Dendrogram of the cluster analysis of *cxt.Weekday*. The clustering is different from the clustering in Figure 3.3. The horizontal axis represents the distance or dissimilarity between the content features, e.g., the distance between *tp.Nutrition* and *tp.Jobs* is < 0.1. The clusters include features with semantic similarities, e.g., *tp.Maps* and *tp.ContainsLocation* are in the same clusters, which belong to different clusters in Figure 3.3.

We measured the relative quality of clustering and modelling using the Akaike information criterion (AIC) [88]. The AIC is a measure for the relative quality of a statistical model for a given dataset. The smaller the AIC the better the model. For example, the *cxt.Weekday* and *cxt.Vertical* models, the AIC increases with the decrease of clusters numbers (Tables 3.5, 3.6).

Number of Clusters	*Number of Iterations*	*AIC*
56	100	1,356,463,884
30	120	1,356,498,105
15	70	1,356,588,757

Table 3.6: The number of clusters, iterations and AIC values per the *cxt.Weekday* model. The higher the number of clusters, the smaller the AIC, and the better the model

As expected, the 56-clusters models for these two contexts are the most accurate models (Figures 3.7(a), 3.5). However, the accuracy of the 15-clusters model is accepted, despite that some models are better than other models, e.g., the model of the *cxt.Vertical* is better than the model of the *cxt.Weekday* (Figures 3.7(b), 3.6). To summarise, based on the modelled context, the content features exhibit similarities that facilitate clustering and modelling of context with a reasonable accuracy.

Number of Clusters	*Number of Iterations*	*AIC*
56	140	570,082,929
30	80	570,084,608
15	50	570,092,223

Table 3.5: The number of clusters, iterations and AIC values per the *cxt.Vertical* model. The higher the number of clusters, the smaller the AIC, and the better the model

3.4.3 The Effect of Context on Post-Disclosure Patterns

This section presents the results of modelling the post-disclosure patterns, and the role of the different feature classes in determining sensitivity.

Modelling Sensitivity

In the following, we present the sensitivity pattern $\mathbf{LTP_{new}}$ inferred from the dataset and the extra contextual parameters. Since our focus is on sensitivity and given the high dimensionality of our dataset, we mainly discuss the positive determinants of sensitivity.

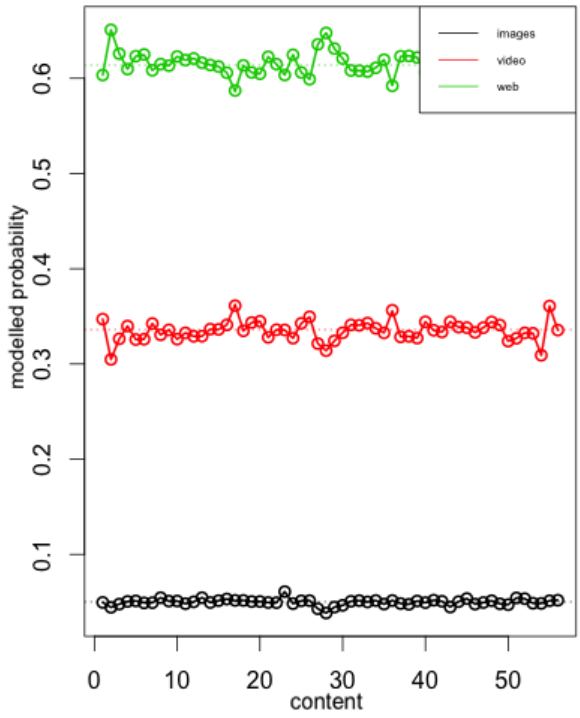


Figure 3.5: The fitted model of *cxt.Vertical* with 56 cluster. The circles are the observed values, the lines represent the predicted probabilities, and the dotted lines represent the mean probabilities. The model predicts the probabilities accurately—the line passes through the circles centres.

The $\mathbf{LTP_{new}}$ shows that sensitivity is affected by the online and offline context of the user. In this pattern, the three classes of features affect sensitivity, positively and negatively. The extra contextual parameters are amongst the positive and negative determinants of sensitivity as well. The type of organisation of the user (*cxt.OrganisationType*) is an indicator of sensitivity management. The pattern shows also that not being at home (*cxt.Home.False*) is correlated with sensitivity management, while being at home (*cxt.Home.True*) is an indicator of not managing sensitive data. The pattern shows that using proxies or anonymisation services (*cxt.ProxyType.Tor*, *cxt.AnonymiserStatus.Private*) is a negative indicator of sensitivity. Similar to the $\mathbf{LTP_{old}}$, the $\mathbf{LTP_{new}}$ shows that being not signed-in with Facebook credentials to Bing (*cxt.Facebook.False*) is an indicator of sensitivity management, while being signed-in with Windows Live credentials

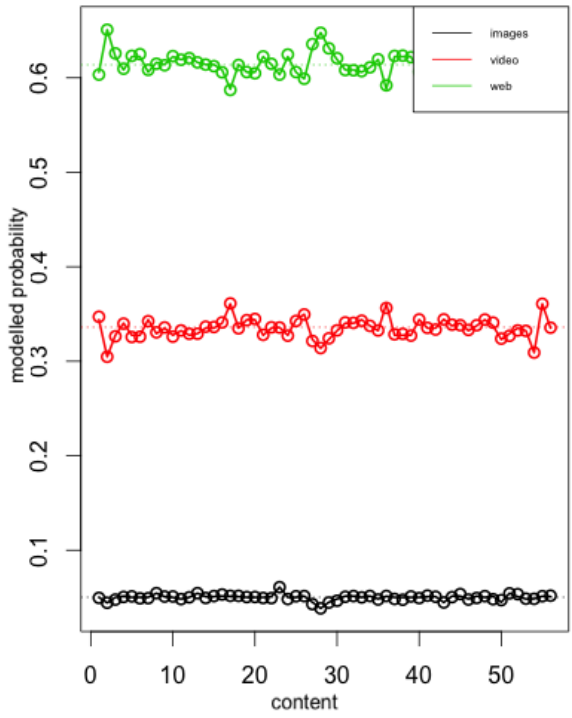
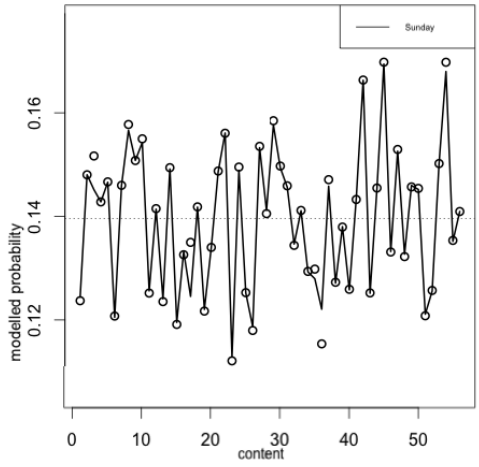


Figure 3.6: The fitted model of *cxt.Vertical* with 15 cluster. The *X* axis represents the indexed content features. The circles are the observed values, the lines represent the predicted probabilities, and the dotted lines represent the mean probabilities. The model predicts the probabilities relatively accurately—the line passes through the circles, although not always through the centre.

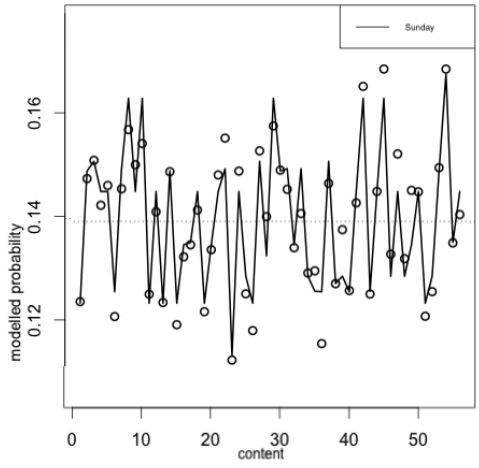
(*cxt.WindowsLive.True*) is an indicator of sensitivity management. However, the weights in the $\mathbf{LTP_{new}}$ differ from the weights in the $\mathbf{LTP_{old}}$. The difference is due to the extra contextual features. Following is the list of the most significant determinants:

Content Class

The effect of the determinants in this class is compared to the effect inferred in the $\mathbf{LTP_{old}}$. Topics like *tp.Health* and *tp.Name* have a high effect on sensitivity. The main difference between the $\mathbf{LTP_{old}}$ and the $\mathbf{LTP_{new}}$ is the effect of sex (*tp.AdultScore*) on adult data. In the $\mathbf{LTP_{old}}$ the *tp.AdultScore* has the highest weight across the three feature classes.



(a) The fitted model of *cxt.Weekday.Saturday* with 56 cluster.



(b) The fitted model of *cxt.Weekday.Saturday* with 56 cluster.

Figure 3.7: The circles are the observed values, the lines represent the predicted probabilities, and the dotted lines represent the mean probabilities. The accuracy of the model drops with the decrease of the number of the clusters. The low accuracy of the 15-cluster model is demonstrated by the line not meeting all the circles.

- *tp.Adult* (w=0.27), and *tp.AdultScore* (w=0.14).
- *tp.Health* (w=0.088)
- *tp.NameNonCeleb* (w=0.073)
- *tp.Celebrities* (w=0.071)
- *tp.Name* (w=0.046)
- *tp.VideoExcludesAdult* (w=0.038)
- *tp.NightLife* (w=0.35)
- *tp.QuestionAndAnswer* (w=0.032)
- *tp.MovieTitle* (w=0.028)
- *tp.Navigational* = False (w=0.021)
- *tp.Image* (w=0.016)
- *tp.ClothesAndShoes* (w=0.0128)
- *tp.NamePlus* (w=0.0113)
- *tp.Finance* (w=0.0082)
- *tp.RadioStations* (w=0.0059)
- *tp.Commerce*=False (w=0.005)
- *tp.Galleries* = False (w=0.0036)
- *tp.Local*=False (w=0.0033)
- *tp.UrlQuery* (w=0.002)
- *tp.TvShows* (w=0.0019)
- *tp.MovieShowtimes* (w=0.0012)
- *tp.Autos* = False (w=0.0008)
- *tp.ContainsLocation* = False (w=0.0007)
- *tp.Restaurant* (w=0.0003)
- *tp.VideoGames* (w=−0.15)
- *tp.Adult* (w=−0.14)

- *tp.Jobs* (w=−0.1)
- *tp.Recipes* (w=−0.084)
- *tp.MovieTheater* (w=−0.082)
- *tp.Book* (w=−0.081)

Data Context Class

This class exhibits more effect of the *cxt.Browser* than in the $\mathbf{LTP_{old}}$. The selected values of this context differ from those selected in the $\mathbf{LTP_{old}}$. However, the other determinants of sensitivity are comparable to those in the same class in the $\mathbf{LTP_{old}}$, for instance, reoccurring searches have a high effect on indicating sensitivity in the two patterns.

- *cxt.InSearchHistory* (w=0.24)
- *cxt.Browser*, the selected values are:
  - IE7 (w=0.089)
  - IE6 (w=0.087)
  - IE8 (w=0.042)
  - Firefox18 (w=0.014)
  - Safari (w=−0.047)
  - Chrome (w=−0.041)
- *cxt.VerticalChange* (w=0.068)
- *cxt.IsForced* (w=0.059)
- *cxt.SafeSearch*
  - Strict (w=0.026)
  - Moderate (w=0.016),
  - Off (w=−0.06)
- *cxt.AppType.SSLBing* (w=0.0131)
- *cxt.IsDotCom* (w=0009)

User Context Class

This class shows the effect of the extra contextual features on sensitivity. Similar to the $\mathbf{LTP_{old}}$, the *cxt.OperatingSystem.WindowsNT* has the highest effect. Different values of the *cxt.OrganisationType* are significant indicators of sensitivity, e.g., working for health and governmental institutions is associated with sensitivity management. The following determinants show the role of the extra contextual features in the inferred pattern.

- *cxt.OperatingSystem*
  - Windows NT5.2 (w=0.071),
  - Windows NT5.0 (w=0.0136)
  - Linux (w=−0.04)
  - Android (w=−0.038)
- *cxt.OrganisationType*: this feature indicates sensitivity to a greater degree than other contextual features. The following values of this feature have varying effects on sensitivity:
  - Government State (w=0.066)
  - Medical and Dental Services (w=0.062)
  - Government County (w=0.052)
  - Government Municipal (w=0.034)
  - Finance (w=0.023)
  - Motor Vehicles (w=0.019)
  - Manufacturing (w=0.015)
  - Government Federal (w=0.0145)
  - Lodging (w=0.0142)
  - Banking (w=0.0129)
  - Government (General) (w=0.011)
  - Insurance (w=0.008)
  - Utilities (w=0.007)
  - Private Service (w=0.006)
  - Gaming (w=0.005)
  - Transportation (w=0.004)
  - Religious Organizations (w=0.003)

  - Data Services (w=0.0023)
  - Legal Services (w=0.002)
  - Health (w=0.000002)
  - Internet Hosting Services (w=−0.064)
  - Telecommunications (w=−0.02)
  - Library (w=−0.003)
- *cxt.MSNService*
  - Entertainment (w=0.05)
  - Local edition (w=0.04)
  - Homepage (w=0.0105)
  - Sports (w=0.009)
  - Lifestyle (w=0.007)
  - Money (w=0.006)
  - Food Channel (w=0.0031)
  - Career and Jobs (w=0.0016)
- *cxt.Facebook.False* (w=0.035) and *cxt.Facebook.True* (w=−0.15)
- *cxt.WindowsLive* (w=0.028)
- *cxt.ConnectionType*
  - OCx (w=0.066)
  - Tx (w=0.062)
  - DSL (w=0.0088)
  - ISDN (w=0.0087)
  - Framerelay (w=0.0023)
  - Consumer Satellite (w=−0.07)
  - Mobile Wireless (w=−0.03)
- *cxt.LineSpeed*
  - High (w=0.0125)
  - Low (w=−0.007)
  - Medium (w=−0.008)

- *cxt.Home*False (w=0.0108), being outside home is an indicator of sensitive data management, while being at home has a negative effect on sensitivity management
- *cxt.ProxyType*
  - Http (w=0.0034)
  - Tor (w=−0.001)
  - Web (w=−0.003)
  - Socks (w=−0.000004)
- *cxt.AnonymiserStatus*
  - Suspect (w=0.0032)
  - Inactive (w=0.000037)
  - Private (w=−0.004)
- *cxt.SearchService*
  - Explore (w=0.0006)
  - Music (w=0.0003)
  - News (w=0.00007)
  - Rewards (w=−0.007)
  - Domains (w=−0.005)
  - Movies (w=−0.001)
  - Social (w=−0.00007)
- *cxt.ProxyLevel*
  - Distorting (w=0.0003)
  - Elite (w=−0.002)
  - Transparent (w=−0.0001)
- *cxt.DeviceClass* (w=−0.06), mobile devices negatively affect the management of sensitive data
- *cxt.IPCount* (w=−0.018)
- *cxt.DeviceModel* (w=−0.014)

In summary, the context affects post-disclosure patterns to a high degree. The $\mathbf{LTP_{new}}$ shows that most of the extra contextual features are selected as sensitivity determinants. Moreover, the extra contextual features enhance the accuracy of the inferred pattern. The accuracy of the inferred pattern $\mathbf{LTP_{old}}$ increased from 0.70 to 0.75 in the $\mathbf{LTP_{new}}$. The increase of the accuracy is mainly due to the extra contextual information, since the same learning algorithm was used. The extra contextual information enhances the accuracy of predicting the management of sensitive data.

3.5 Discussion

The results of our analysis demonstrate the effect of context on data sensitivity upon disclosure and post disclosure. When data is disclosed, the sensitivity can be affected by context. The user can assess the sensitivity of data post-disclosure, and limit its availability on the web. The analysis shows that data disclosure patterns vary based on the online and offline context. The effect of the offline context might indicate that the user reasons about the appropriateness of disclosing or viewing certain data in a particular offline context. The appropriateness of data may not be limited to situations where the data can be viewed by others. The user might judge the data as sensitive, inappropriate, and embarrassing based on imaginary audience [37]. Such a judgement may be the reason why the online context affect data sensitivity.

The results of our analysis can be incorporated into developing privacy management approaches. Our results provide detailed knowledge about the effects of different contexts on different content types. Privacy management approaches offer users functionality to manage data on the web [49]. Data management involves controlling what to disclose and in which context. By incorporating knowledge about the effect of context on disclosure patterns and post-disclosure patterns, it is possible to enhance how a privacy management approach assists users in managing their data. An approach can assess the sensitivity of disclosing a data item in a particular context. It can also detect the data that needs to be managed post to disclosure.

3.6 Conclusion

Context is an essential ingredient in reasoning about sensitivity. This chapter provides evidence as to the role of context in determining sensitivity in different phases of data management. The evidence shows that context affects data

disclosure. The evidence also shows that context affects the management of sensitive data post-disclosure.

The chapter demonstrates the dependency between content and context. The analysis of the effect of context on content shows that content is dependent on context. The modelling context using content shows that context depends on content as well. The analysis shows that all the contexts in our dataset affect data disclosures significantly. The analysis shows the effect of context on each content feature in our dataset. The analysis also shows that adding more contextual parameters leads to an increasing the contribution of context towards determining sensitivity. The effect of context on content of data need to be adopted when developing various data and privacy management approaches in order to enhance the usability and effectiveness of these approaches.

Chapter 4

Conceptual Analysis of Context

4.1 Introduction

In the era of social software, privacy management is essential to protect one's data. Social software user can disclose different types of multimedia (e.g., images, videos) rather than the exchange of explicit messages. Data are subject to mishandling and inappropriate dissemination. Through privacy management, users can control dissemination of data to avoid manipulation of their data. Privacy mechanisms offer various types of control over context and the audience. However, misappropriation may not be avoided by current privacy mechanisms. These mechanisms should guard against misappropriation of data such as the following scenario:

Scenario 1. Proud mother Alice posts her photo breastfeeding her new born baby and shares it with public. The intended message is how she loves and cares for her baby. However, she did not anticipate that pervert Bob could also have access. Bob disseminates the photo further in an 'pornographic' context. Alice finds out about the inappropriate dissemination, and reports the abuse to the social software provider as it is a privacy violation due to the identity damage she and her daughter have experienced.

Avoiding misappropriation of data is complicated and requires controlling dissemination contexts. Controlling dissemination contexts requires reasoning about the current context of a post and how it may change [69]. It also requires

allowing appropriate or prohibiting inappropriate contexts [76]. Such control requires the ability to control the various contextual parameters to constrain change of context. Due to the high dimensionality of context and its complex nature [109], controlling it can be complicated. The complexity of context control can be even more complicated due to ambiguity, which is a characteristic of social software contexts. Without the ability to control misappropriation, users may incur privacy violations and identity damage.

Towards offering contextual privacy approaches to counter data misappropriation, it is required to examine the role of context in communication, data misappropriation and privacy. This chapter contributes the following:

1. An elaboration on the issue of controlling context in social software communication (Section 4.2)
2. A presentation of the concepts relevant to contextual privacy and its relation to communication (Section 4.3)
3. An analyses of context ambiguity in relation to data misappropriation (Section 4.4)
4. A definition of a contextual privacy and data misappropriation attacker model (Section 4.5).

4.2 Problem Statement

In this section, we discuss the problem of privacy management in a context-based manner in online communication to avoid data misappropriation.

Managing privacy to mitigate data misappropriation requires a high degree of control over context. From a 'privacy as control' point of view [48], users should be able to control their data and the context wherein data is put. By controlling context, it is possible to limit the changes of context or actions of disseminating data into new contexts. The main challenge for context-based privacy management is the complexity of controlling context in online communication. In the following, we describe context control issues that are related to context and communication.

4.2.1 Context-related Issues

This class includes issues that are caused by the nature of context. Context can be complex in certain situations, and it can be ambiguous as well. These two

characteristics of context contribute to the complexity of controlling context, as we discuss in the following.

Context Complexity

Context is the information construct that defines a situation. Context can define a situation with a certain degree of specificity. A context can be represented by a set of parameters from within the situation, such as the location, time, the type of users, their age, and the topics discussed in a situation. The more parameters, the more complex the context is. To control context, users need to control the various parameters of this context. Thus, context complexity may hinder context control.

Simplified context representation is insufficient to mitigate data misappropriation [90]. To avoid the complexity of context control, most approaches adopt simple means for control. The simplification is achieved by capturing context by a few parameters such as roles of users [13], location or time [6]. This simplification may fail to actually capture contexts that users may want to control. In Scenario 7, limiting access to the photo to users with the role 'mother' is not enough to avoid the dissemination in the 'pornographic' context. Such a simplification results in offering limited control over the disclosure context, and a lower degree of control over dissemination context [92]. Consequently, users cannot control every change of context to avoid inappropriate changes. Controlling context changes is essential because sensitivity of data may change, as we demonstrate in Chapters 2 and 3. When context changes, the sensitivity of data may increase. When the data with increased sensitivity is put in an inappropriate context, the data owner may incur privacy violations.

Mitigating data misappropriation contributes to the complexity of context control. To avoid disseminating data in inappropriate context, users can specify the set of appropriate and inappropriate contexts. However, this specification is infeasible due to the theoretically unbounded number of contexts [91]. Moreover, users may not be willing to invest much time in managing their data [60].

Context Ambiguity

Controlling context becomes more challenging when context is unclear and ambiguous. Social software contexts can possibly be ambiguous due to the mix of audience and data [17]. Context ambiguity has been firstly identified

by Meyrowitz as the main issue of broadcast media [71]. Boyd extends this concept to social software and suggests that the dynamics that cause ambiguity in the unmediated spaces apply to the mediated networked spaces [18]. These dynamics are: the audience invisibility and the obscured viewing of others' data, the contexts collapse due to lack of boundaries in social software situations, and the blurred boundaries between private and public and how posts can be accessed. These issues hinder estimating the context of a situation. Context ambiguity can affect reasoning about data sensitivity. When sensitivity is not estimated properly, reasoning about the appropriateness of putting data in a particular context can be affected. Incorrect assessment of appropriateness can result in misappropriation.

4.2.2 Communication-related Issues

Online communication gives rise to a class of issues concerning managing privacy in both private and public spaces. Communication is based on the delivery of a communicative message. Online communication can occur via disclosing data within private or public spaces. Private spaces are situations in which a message is communicated to a limited number of audience. In public spaces, the message is accessible to anyone. In private spaces, the user can estimate in advance the audience, their actions, etc. and hence control the context to manage privacy. However, the complexity of context control makes privacy management challenging. In a public space, the user may not be able to predict who will access the data, what actions can be performed, how the context can evolve, etc. In such a case, data misappropriation can occur easily due to the lack of context control.

In the next section, and as a first step towards addressing context problems and data misappropriation, we conceptually investigate the interaction between context, communication and privacy.

4.3 The Interaction of Context, Communication and Privacy

In this section, we examine the main concepts related to contextual privacy. We discuss the interaction of context, privacy and communication. The discussion provides the conceptual platform required to understand contextual privacy.

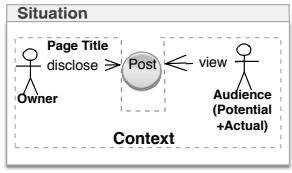
4.3.1 Context

Context is “any information that can be used to characterise the situation" [1]. An online context is any information that can be used to characterise an online situation. The context indicates the topic of communication and possibly some characteristics of the interlocutors. According to the theory of relevance and context [101], the communication discourse context can be approximated by the set of available informational parameters in the situation. We refer to those parameters as *the context-approximation parameters* (CAP). The inaccessibility of these parameters hinders the correct approximation of context, which makes the context *ambiguous*.

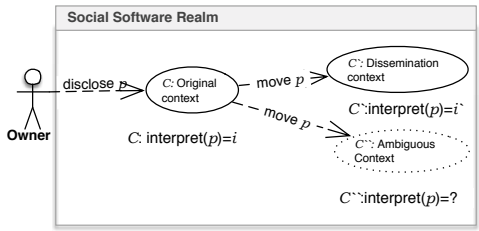
A situation in social software can be characterised by a post and the context wherein the post is put (Figure 4.1(a)). The communication context includes the data surrounding the post, the data owner and an audience. These data are the CAP of context in most social software, e.g., Facebook, Google+. The *data owner* is the user who discloses the post in the original context. The *original context* is the situation in which the post is originally disclosed via the software (Figure 4.1(b)). The audience in a context can be potential or actual audience. The *potential audience* are the users that can view the post. The *actual audience* are the members of the potential audience who have already viewed the post. When the actual audience contribute to the communication (e.g., comments or likes), they become *subordinate owners*. Another class is the *extended audience*, which is a feature of Facebook (Jan. 2016). This feature characterises the friends of a friend who become part of the audience of public posts. When a friend of the owner likes a public post, the post is displayed in the newsfeed of the friends of this subordinate owner, as if the data owner shared the post with these friends.

By controlling context, it is possible to affect the approximation of context. Adding or removing data to context can affect the CAP, and as result the context changes. Whenever the observer is unaware of this change, e.g., a member of the audience, there can be discrepancy between the *perceived* and the *actual* context resulting in ambiguity (Figure 4.1(b)). Ambiguity disrupts the interpretation of the communicated message. By controlling the CAP, approximating the context can be facilitated.

Context approximation is required to infer information relevant to the actions and data in a situation. Context provides guidance about the relevant, and appropriate behaviour in a situation [107]. Also, context provides information about how to perceive data in this context. In particular, context affects the following two latent values of data, as follows:



(a) A simplistic representation of a communication situation.



(b) The post p is disclosed in C, where it has the interpretation i. When p is in C', the *dissemination* context, the interpretation is i'. When p is in the ambiguous context C'', an observer may not be able to interpret p.

Figure 4.1: Context in social software.

- Data Interpretation: Context provides the appropriate interpretation in a situation [22]. The interpretation captures the meaning of the message [9]. A data item can have a limited set of possible interpretations. Based on the context, the relevant interpretation can be disambiguated [7, 11]. The relevant interpretation implicitly reflects that the data item is in that context (Figure 4.1(b)). By controlling the CAP to facilitate the context approximation, it is possible to facilitate a particular interpretation of data.

- Data Sensitivity: A data item has a latent value of sensitivity. Sensitivity indicates how appropriate it is to disclose the item in a particular situation. This sensitivity is affected by the context whether it is online or offline. We have demonstrated how context affects data sensitivity in Chapters 2 and 3. By controlling the CAP to facilitate the context approximation, it is possible to maintain the sensitivity of data.

  The sensitivity is related to the interpretation of data. In Scenario 7, when interpreting the photo as a *motherly action*, the sensitivity is not

as high as when interpreting the photo as *pornographic action* in the *pornography* context.

The interpretation and sensitivity are the ingredients that are affected by context when data is contextualised. *Contextualisation* is the act of putting a data item in a particular context [67]). Also, *decontextualisation* is the process of taking a data item out of the current context, to where the interpretation is unavailable [67]. We identify a third process of moving data items between two contexts as *recontextualisation*. This process decontextualises a data item and contextualises it in another context. Recontextualisation can affect the interpretation and sensitivity of data. Similarly, the change of these ingredients indicate a context change, or a recontextualisation of data. However, not every recontextualisation may necessarily change these ingredients.

Next, we discuss the role of context in facilitating communication.

4.3.2 Communication in Social Software

Social software communication is initiated by the data owner to interact with her audience. A data owner[1] posts her data through the social software to convey a particular message. Each communication is characterised by an owner who is the communication initiator, and an audience selected by the data owner. The *interlocutors*, the owner and the audience, communicate about the owner's data.

In general, communication can take various types based on how the interlocutors act and reason about privacy and context. On one hand, the interlocutors trust each other and cooperate in *cooperative communication*. On the other end, they may not fully trust each other and communicate adversarially in *adversarial communication* [53]. These two types of communication have varying privacy concerns. Additionally, context plays different roles in these two types of communication. To emphasise the roles of context and privacy, we focus in the following on the two extreme ends of the communication spectrum, namely, cooperative and adversarial communication.

Cooperative Communication

In cooperative communication, interlocutors act jointly to understand the communicated message. Any contribution to the conversation should be clear,

[1] We do not imply the legal ownership.

relevant and easily understood. According to Grice, cooperative communication can be achieved by following the following four maxims [45]:

1. Quantity: provide a sufficient amount of information
2. Quality: provide statements that are true
3. Relation: provide the relevant information that is appropriate and needed
4. Manner: avoid ambiguous, obscure and unnecessary information while being orderly.

By means of these maxims, Grice describes how interlocutors make the context explicit to infer the *conversational implicature* of what is being communicated. The conversational implicature is the meaning that can exceed the literal meaning of the utterance given the *intentions* of the speaker and the context [45]. To infer the conversational implicature, the communicative message, the interlocutors abide by the maxims and rely on the communication context. At any point, if the context is ambiguous, any of the interlocutors trust that the others will cooperate in clarifying it, based on the presumed shared knowledge [100]. Accordingly, the conversational implicature can be inferred properly, and the speaker maintains her self-presented identity [42].

In cooperative communication, the privacy concerns of the interlocutors are minimal, or at reasonable levels. When disclosing data to deliver a message, the owner trusts the audience to cooperate to perceive the message correctly [53], and that they would not violate her trust. When the message is unclear, the audience should clarify the message to avoid performing actions that might misappropriate the message. In this communication, the clarity of the context and the CAP is essential to facilitate communicating messages. Therefore, the privacy of the data owner is maintained, against misappropriation—unless there is an intentional attack.

In social software, however, *cooperative communication is challenging to achieve*. Although users aim to communicate cooperatively, they may not necessarily succeed in acting cooperatively [60]. The setting and design of social software does not fully support such trust and cooperation. A user could disclose a large amount of data and share it with a large audience. The overload of data and audiences makes abiding by Grice's maxims a highly-demanding task. Abiding by the maxims is more problematic when the audience members and the owner are not highly familiar with each other. The failure in following the Gricean maxims results in ambiguous context or ambiguous data. Consequently, audience members fail to behave in an appropriate manner that

preserves other owner's privacy [60]. Additionally, the difficulty of managing privacy can result in making users less trusting and less cooperative.

Adversarial Communication

Adversarial communication is characterised by the manipulation of the communicated message. An interlocutor—the adversary—acts maliciously and misleads others into misinterpreting the message to disrupt the communication or force others to reveal certain information [36]. Alternatively, when context is ambiguous, communication can also become adversarial [100]. In such communication, users can protect privacy by providing less information [110], with detrimental consequences to the clarity of context and the inference of the communicative message. Consequently, data can be misinterpreted resulting in a data misappropriation attack. In this form of communication, privacy concerns are high, the degree of trust is low, and data interpretation is manipulated.

In social software, it is challenging for users to identify adversarial communication and act properly. One of the main features of social software is the possibility to communicate with large audiences. Users may not always be familiar with the audience they communicate with. The unfamiliarity results in a low degree of trust [53]. The low degree of trust may prevent users from providing sufficient information to avoid ambiguity. Ambiguity, in turn, hinders users' ability to distinguish situations where others are acting adversarial, and situations where some users are simply unfamiliar with each other. In other cases, users may use adversarial communication to protect their privacy. Users can act adversarially and use steganography to hide one message and mislead the audience into perceiving another inaccurate message. Steganography is mostly teenagers [19] to disclose data that has more than one meaning for which only a subset of interlocutors is able to perceive the intended meaning or the conversational implicature. To elaborate steganography, consider the following scenario reported previously by Boyd [19]:

Scenario 2. Carmen was sad because she broke up with her boyfriend. She wanted to express that to her friends but not to her mother so that she would not worry. Carmen posted lyrics from "Always Look on the Bright Side of Life" from the film "Life of Brian", which is about the main character who was about to be crucified. She knew that some of her friends would infer her exact implicature, while her mother would infer a literal meaning of the post.

This scenario shows how users can adopt certain strategies when they are unable to control context or to avoid investing time and effort in selecting the

appropriate audience [60], for instance, Carmen could have excluded her mother from the audience. Although the intended message in the example is delivered to the right audience, such a strategy is insufficient to guide the appropriate behaviour of the audience and preserve one's privacy. A friend of Carmen can be unaware of the presence of her mother and can act inappropriately by making Carmen's implicature explicit. Moreover, using steganography can be safe in the original context, but not when the data is put in another context.

4.3.3 Identity and Privacy

Privacy management is one of the means to manage identity in online communication. Through interacting and communicating with others, people express their identity. In social software, users communicate with others by means of the data they disclose. Through this communication, users build and manage their online identity by managing who they communicate with and what data they disclose [119]. In other words, thee act of identity management is achieved through privacy management [78].

To make proper privacy management decisions, the owner needs to be aware of how others would perceive and interpret her data [78]. Based on the expected interpretation of others to the owner's data, the owner could make the privacy decision of to whom disclose her data. The owner's expectation is that the audience would most probably interpret the owner's data in a way that corresponds with the identity the owner is aiming to express. Therefore, the interpretation of the disclosed data is of a central role in the privacy management process. In other words, privacy management aims at making disclosure decisions that facilitate the proper perception of the owner's data in order to express the owner's desired identity. When context is ambiguous, the expression of identity can be affected, as we demonstrate in the following section.

4.4 Context Ambiguity and Data Misappropriation

In this section, we discuss context ambiguity and how it facilitates data misappropriation.

As discussed above, the ambiguity of context can be the main reason for data misappropriation. The analysis of context in the previous section suggests

that ambiguity results due to invisibility of some of the CAP in a situation. The issues identified by Meyrowitz as the causes of context ambiguity [71], despite being widely accepted as causes of ambiguity, are not specific enough. Since context approximation is based on the contextual parameters available in the situation, we argue that the invisibility of these parameters is the main cause of context ambiguity. We use the following scenarios to demonstrate how ambiguity affects the communicative message. These scenarios illustrate how data can be misappropriated, and how data misappropriation affect the user's privacy and identity. The scenarios are based on the CAP we identify in Section 4.3.1. They are cases where communication can be sensitive such as activism-related communication.

Invisible Owner

Scenario 3. Bill lives in San Francisco where he is an activist against gentrification. He anonymously posts a photo of a demonstration with violent protesters. He shares the photo with his wider group of friends through the social software. Some of the actual audience are unable to infer the reason for the violence and its relation to the people in the photo.

The anonymous post disrupts the audience's ability to infer that the photo is to report violence in Bill's neighbourhood. Beside the identity of the owner, the owner's attributes play a role in approximating the context. Knowing Bill is an activist, makes possible approximating that the context as social uprising. Invisibility of owners can be a privacy feature of anonymous social software. However, anonymity may come at a price of ambiguity.

Invisible Subordinate Owner

Scenario 4. Sam is a police officer and comments on Bill's photo, saying that he was attacked and injured. Dean, being unable to see that Sam posted the comment, assumes a fellow protester was injured. He comments back saying that the police are brutal and the protests should continue against them.

The invisibility of the subordinate owner Sam, affects the approximation of context by Dean. Dean's comment is inappropriate because misrepresents the police as brutal. The comment disrupts the communication. Had Dean been able to know that Sam is a police officer, he could have approximated the context more accurately.

Invisible Potential Audience

Scenario 5. Rex, a friend of Bill who works for the secret service, is a member of the potential audience. If he views the communication after Dean has commented, he might disseminate the comment in a page about people who encourage violence against the police.

Had Dean been aware that the potential audience include Rex, the intelligence employee, he could have been able to approximate the context more accurately and reason that his comment was inappropriate to be disclosed or that it may be assigned an interpretation that is not right. The possible act of taking Dean's comment and putting it in another context is a recontextualisation of his post. The recontextualised comment is interpreted different to what intended. The sensitivity of this comment is higher in the new context in comparison to the sensitivity in the old context. Recontextualising the comment is a violation of Dean's privacy, and does not contribute to the identity he is expressing. In this case, the communication becomes adversarial regardless of Rex's intentions.

Invisible Actual Audience

In the previous scenario, if the actual audience were visible, Dean would have been able to detect the context change, namely Rex becoming a member of the actual audience, assuming that he has his profession accessible in the social software. To Dean, Rex is an adversary who may (mis-)interpret his message. Being aware of this transition, Dean could have removed his comment, or taken precautions.

Invisible Extended Audience

Scenario 6. (Scenario 4 cont.) After Sam commented on Bill's post, Sam's colleagues, who are also police officers, become part of the audience and see the interaction. They think that Dean has urged the crowd to attack the police.

When Sam becomes a subordinate owner, his friends become part of the audience. This extension of the audience changes the context. The invisibility of the extended audience challenges Dean to approximate the context to reason about how the new audience may perceive his message. The extension of audience is a case of blurred boundaries between private and public. The new audience can be total strangers for Dean, and sharing his comment with them

takes it out of the private space he thought his comment would only be viewed in. Moreover, to the extended audience, the new communication context can be ambiguous, and this is a problem because the audience who get accidental access to personal data of others do not have the same ethical obligations and responsibilities towards respecting the privacy of the owners [81].

The scenarios discussed above demonstrate how ambiguity negatively affects privacy, as well as communication. Context ambiguity disrupts the interpretation of data, and hence disrupts communication and privacy. The scenarios also show how privacy is affected when data is misappropriated. In many cases, social software users wish their data to be disseminated in different contexts. The problem arises when the data is disseminated to convey a message different from the one intended by the data owner. This means that the misappropriation occurs when the interpretation of the data item is different from the interpretation in the context the owner shared the data within originally. The misappropriation occurs also when the sensitivity of data changes.

4.5 Contextual Privacy and Data Misappropriation

In this section, we present our definition of contextual privacy based on the analysis in the previous sections, and define the data misappropriation attacker model.

4.5.1 Defining Contextual Privacy

Generally, contextual privacy is the aspect of privacy concerned with controlling data in a context-dependent manner. Privacy concerns the control of data to manage one's identity or the mediated self [49]. In particular, privacy as practice concerns having the ability to construct one's identity without constraints [4]. The analysis in the previous section demonstrates how the identity can be affected by the communicative message or the interpretation of data [119]. Controlling context aims at controlling the change of context to avoid changing the communicative message. This control results in maintaining the sensitivity and interpretation of data, as well. Thus, we conclude that the interpretation and sensitivity of data essential ingredients of contextual privacy management. We define *contextual privacy* as the concept that concerns controlling data in a context-dependent manner to preserve the *interpretation* and the *sensitivity* of data, in order to preserve the user's expressed identity.

We propose using the interpretation of data and sensitivity to manage contextual privacy. A data item can have limited possibilities of interpretation and sensitivity. The possibility to manage the appropriate interpretations or sensitivity values could potentially overcome the complexity of controlling context.

4.5.2 Data Misappropriation Attacker Model

One of the main objectives of contextual privacy management is countering data misappropriation. Towards achieving this goal, we firstly discuss the need for defining data misappropriation attacker model, and put forward our representation of such a model.

Commonly, and within *privacy as control* [49], privacy attack models are often rather general. Privacy attack models, or violations, often represent how an unauthorised or inappropriate action on the data is performed [25]. Such descriptions of attacks assume that the data owner can specify all the possible authorised or unauthorised users and actions. However, data owners may not be able to perform such an extensive specification. They may not be able or willing to invest time to list the authorised actions and users. Consequently, without users' specification of what is unauthorised, it may not be possible to properly detect attacks. Data misappropriation can be viewed as an unauthorised act of disseminating data in an inappropriate context. However, in order to mitigate this attack and given the complexity of context control, we need to define what exactly makes certain dissemination acts inappropriate. By such a definition, it would be possible to detect attacks even when the user does not list all possible inappropriate actions.

We model data misappropriation based on the interaction of context, communication and privacy. Data misappropriation is commonly identified as data *decontextualisation* [43, 115, 118]. But decontextualisaiton refers only to taking data from one context to another. It does not refer to taking data from one context to an inappropriate context. Based on the analysis of the role of context in communication and privacy (Section 4.3), context affects that communicative message of the data owner. When the context changes, the communicative message (the interpretation) of data, as well as the sensitivity change. We can state that decontextualisation misappropriates the data if the new context changes the interpretation or sensitivity of data. We define the data misappropriation attacker model given the following items:

- A (trusted) system. A system that offers social communication functions. The system enforces users' privacy policies and allows actions that are

not, otherwise, prohibited by the data owner.

- A data owner. A user who discloses a data item to communicate a message—that is appropriate. This user can be targeted by the attacker.
- An attacker. A user who can access the data item, and by performing a particular action, the communicative message of the data item becomes inappropriate—in comparison to the message intended by the data owner. The attacker can misappropriate the data by causing the context to change. The change can be achieved by putting the data in a new context, or by causing the current context of the data to evolve.

The attack can be intentional or unintentional. The intentional attack is characterised by the deliberate intent of the attacker to perform the attack to misappropriate the data item. An unintentional attack occurs as a result of an action that does not aim to misappropriate the data. In both cases, the interpretation and the sensitivity of the data can change, resulting in a privacy violation of the data owner. Contextual privacy approaches should offer means to protect the interpretation and sensitivity of data from changes to counter data misappropriation.

4.6 Conclusion

This chapter provides a conceptual analysis of the role of context in privacy management. The chapter mainly focuses on the role of context in facilitating communication in social software. Given the common context ambiguity in social software, the chapter provides an analysis of the causes of context ambiguity. The analysis shows how ambiguity affect communication and the privacy of users. Based on this analysis, we conclude that contextual privacy aims at maintaining the interpretation and sensitivity of data. Lastly, we define the data misappropriation model considering our definition of contextual privacy. In the next chapter, we propose a framework for managing contextual privacy.

Chapter 5

CPS^2: a Contextual Privacy Framework for Social Software

5.1 Introduction

Privacy management is essential to facilitate communication in private and public spaces within social software. Social software communication can be achieved through data disclosure. Through each disclosure, a user expresses a particular communicative message and builds a desired identity. A data post can be disclosed in a private space to a particular set of recipients; alternatively, it may be posted publicly for a large, and a priori unrestricted audience. Data are subject to mishandling and inappropriate dissemination. Inappropriate dissemination can affect the communicative message, and the user's identity [91]. Privacy as self determination is the key to protect data and identity [78]. Through privacy management, users can control dissemination of data to avoid manipulation of their data and the communicative messages. While privacy is commonly viewed as a means for data protection, it can also facilitate communication in private and public spaces.

To facilitate communication, privacy management mechanisms should offer control of contexts to preserve the communicative message. However, misappropriation may not be avoided by current privacy mechanisms. These mechanisms should guard against misappropriation of data such as the

following scenario:

Scenario 7. Proud mother Alice posts her photo breastfeeding her new born baby and shares it with public. The intended message is how she loves and cares for her baby. However, she did not anticipate that pervert Bob could also have access. Bob disseminates the photo further in an 'pornographic' context. Alice finds out about the inappropriate dissemination, and reports the abuse to the social software provider as it is a privacy violation due to the identity damage she and her daughter have experienced.

To avoid the complexity of controlling context, security and privacy management mechanisms focus on simplifying the control offered to users. The user can select the disclosure context to disclose a post and can control various aspects of this context. Also, the user can allow or prohibit dissemination in a few contexts that form a small subset of all possible contexts. This context control may not be sufficient to limit misappropriation of data [90]. At the same time, a higher degree of control requires more complex mechanisms. Complex mechanisms can be too sophisticated for non-technical users to handle.

Offering users usable privacy management mechanisms requires a shift in the conceptualisation and design of privacy management mechanisms. Most of privacy mechanisms are security mechanisms applied in social software contexts. An example is access control mechanisms for privacy management [90]. Using security mechanisms for privacy management can address various privacy concerns. However, it may not necessarily result in mechanisms that strike a balance between ease-of-use and expressivity. Privacy management mechanisms should offer context control to facilitate identity management and communication with ease-of-use for non-technical users. Proposing such contextual privacy mechanisms requires a shift in designing such mechanisms towards utilising artificial intelligence to assist users in managing their privacy and controlling context [51]. Such assistance is required particularly to overcome context ambiguity.

We define contextual privacy as the concept that concerns controlling data in a context-dependent manner to preserve the *interpretation* and the *sensitivity* of data, in order to preserve the user's expressed identity. Based on this definition and the analysis of context and communication in the previous chapter, in this chapter, we propose a framework for contextual privacy management (CPS^2). The framework is based on the management of the interpretation and sensitivity of data. It aims at overcoming limitations of current privacy management approaches. This chapter explores possible designs of effective and easy-to-use contextual privacy mechanisms by contributing the following:

1. A conceptualisation of a contextual privacy management framework CPS^2 to address the data misappropriation model that we presented in the previous chapter (Section 5.2)

2. A proposal for a design, and a discussion of deep learning approach for the possible implementation (Section 5.3)

3. An assessment of the usability of CPS^2 and comparing it to Contextual Integrity [76], and a discussion the implications of the proposed design on users experience (Section 5.4).

4. A conceptual assessment of how CPS^2 can address context ambiguity, enhance privacy in private and public spaces; as well as a demonstration of how it facilitates cooperative communication and helps avoiding adversarial communication (Section 5.5).

5.2 Contextual Privacy Management

In this section, we propose a conceptual framework CPS^2 to facilitate communication with an increased level of privacy without burdening users with the management of context. We also discuss a possible realisation of our framework.

In order to detect misappropriation attacks, it is required to use quantifiable measures. In theory, misappropriation attacks can be prevented by monitoring actions on data. In practice, monitoring and detecting attacks requires complete information about all users' actions. However, users cannot monitor and know all actions of other users. The social software provider, on the other hand, can perform such monitoring. Given the complexity of context control and the challenge of having complete information about other users actions, it is required to integrate automatic approaches to detect such an attack. Such an approach can be utilised to monitor acts on data that lead to misappropriation. The monitoring requires quantifiable measures that can be any value related to the data that changes with context change. Given our definition of contextual privacy, the interpretation or the sensitivity of data can be used as these measures, as we discuss next.

5.2.1 CPS^2: Contextual Privacy Framework for Social Software

We propose CPS^2 to manage contextual privacy by managing the interpretation and sensitivity of data within the communication context in social software. The framework facilitates communication with an increased level of privacy without burdening users with the management of context. The framework differs from other approaches in the way it addresses context complexity. Rather than simplifying context or imposing reasoning about context on users, the framework proposes to manage contextual privacy by having means to ensure that the interpretation of data is appropriate in any context. The framework proposes to lift the burden of reasoning about context to the level of the social platform, using the technological advances in context inference, user intent inference [24, 85], automatic data interpretation mechanisms [26], and the sensitivity inference approaches we demonstrated in Chapters 2 and 3.

In CPS^2, we propose the separation of the context inference and the inference. The assumption is that the interpretation and the sensitivity in a specific context are *appropriate* if the owner allows the disclosure in this context. Thus, the user needs to only indicate the appropriate interpretation explicitly. Alternatively, these values can be inferred from the data item when disclosed by the user. The platform can be responsible of monitoring the context changes, as well as the interpretation and sensitivity of data. Monitoring only the interpretation or sensitivity is sufficient to indicate change of context, and potential data misappropriation in online contexts. This approach, however, is not intended to and cannot detect misappropriation in offline contexts. This approach is also not intended to control the interpretation the audience infer, it rather aims at facilitating the inference of the intended interpretation by the audience. The framework aims at prohibiting the dissemination of data in contexts that would facilitate inferring an inappropriate interpretation.

In the following are the steps using contextual privacy management in CPS^2:

1. The user discloses a data item in a context.
2. The user can specify a value of the data's interpretation or sensitivity. Alternatively, the framework infers the value of the the data's interpretation or sensitivity in this context.
3. The framework monitors the value specified or inferred in the previous step to avoid changes.

5.3 An Architecture Design for Contextual Privacy Management

In this section, we propose an architecture design for CPS^2.

The framework lifts the burden of reasoning about context to the level of the social platform, and proposes three layers to manage contextual privacy. The realisation of CPS^2 implies a system with three main functions: context inference, data inference, and contextual privacy management. The inference layers need not be managed by users, but by the social software provider, for instance.[1] While such layers are seemingly challenging to have in practice, such systems are available, e.g., the *Cyc* system [64], which provides a knowledge base to represent contexts and data interpretation.

In the following, we provide a high-level description of an architecture proposal to facilitate the realisation of CPS^2. We present also the interaction between these layers (Fig. 5.1), and investigate techniques to implement the inference layers.

5.3.1 Context Inference Layer

This layer is responsible for processing data to approximate the current relevant context of a situation in social software. The input is the social software data: users, their attributes, data items, and relations, ads, and the structure of its pages. When data is added to the platform, this layer adapts to the change by adapting the approximated context. For example, and inspired by the model of Buvač [23], this layer can represent two types of knowledge, namely, general knowledge and discourse context knowledge. The general knowledge reflects the propositions that hold in the world. The discourse contexts reflect the communication context labels. The realisation of this layer can be based on the realisation of the following two modules:

1. Approximating the communication or discourse context: this module is required to approximate context based on specific variables that characterise the communication.
2. Logic of Context: this module is required for reasoning about context and its transitions. In certain cases, it might be needed to reason about

[1] Such an assumption raises concerns regarding the control the social software provider will have [93] But the addressing of such concerns can be of a similar approach to that of decentralised social software, where the context layer can be embedded at the user level or at distributed trusted parties [98].

a context other than the current one. For instance, allowing an action on a data item might result in changing the current context, or allowing a recontextualisation of a data item. In both cases, it is required to transition to another context in order to test the appropriateness of the interpretation in the new context. For instance, a logic of McCarthy offers means to transition between contexts by offering two contextual operations *enter* and *exit* context [67] to move between contexts.

5.3.2 Data Inference Layer

This layer is responsible for inferring the interpretation and sensitivity values of data. This layer can infer these values based on the context inferred by the previous layer. The inference is similar to how a search engine performs page ranking to match a query to a document: the document is the context and the query is the post. The query has a specific interpretation in a document, based on the popularity of this interpretation, the engine judges the relevance of the document. Similarly, the interpretation, or the sensitivity, of a post can be inferred in an online context.

The inference layers need to be embedded in the software platform. These layers can be implemented by machine learning models, especially deep machine learning generative models. Deep learning focuses on computational models for complex information representation [10]. Generative models are useful for unsupervised learning with a high number of parameters [54]. These models are useful in social software situations because the parameters are many and vary across users; and because it may not be possible to have context and interpretation labels during the training phase. Generative models can learn a joint probability distribution over observable data and labels. This means that it is possible to estimate the conditional probability $P(O|L)$ and $P(L|O)$, where L is a label and O is a set of observable data variable. In CPS^2, the observable data is the CAP, and labels are information about context names and interpretations, whether meanings or sensitivity values.

In the following, we discuss two possible approaches to infer the interpretation or the sensitivity.

Inferring of data interpretation A relevant model for such an inferences is the Multimodal Learning with Deep Boltzman Machine proposed by Srivastava and Salakhutdinov [104]. A Deep Boltzman Machine is a network of symmetrically coupled stochastic binary units [89]. The network consists of interconnected hidden and visible layers. It can be used to model situations with millions of parameters. The model learns a

multimodal data representation to perform classification and information retrieval tasks. Multimodal data includes different representations at the same time, such as an image with text. Data disclosed in social software can be multimodal, for instance, the user can post a photo with a title or tags of friends in the photo.

The model classifies images and tags them. The tags in this case refers to the interpretation. The learning, however, is based on the availability of observable data with tags. After learning, when a data item is disclosed, given the CAP, the model infers the relevant interpretation. It is also possible to allow the user to give feedback on the inferred interpretation to enhance the inference.

Inference of data sensitivity A simpler approach is to infer the sensitivity of data and monitor its changes. A relevant model is similar to the models developed in Chapters 2 and 3. Those models infer sensitivity by observing users' behaviour. In the disclosure patterns, we assume that the intensity of data disclosures indicate the appropriateness and sensitivity of disclosing data in various contexts. In the post-disclosure patterns, we assume that the act of deletion indicates sensitivity. However, this model infers only two values of sensitivity, sensitive and not sensitive. Ultimately, the inference of sensitivity can be made more fine grained given the availability of a wide spectrum of actions that indicate varying sensitivity values. An example is the availability of information about the number of people who could access an item before it is deleted. Alternatively, users can provide a value of sensitivity for each item they post. The post-disclosure model can be adopted to infer the pattern of sensitivity in social software. When the context in which a data item changes, the sensitivity of the item can be inferred accordingly.

5.3.3 Contextual Privacy Management Layer (CPML)

This layer facilitates contextual privacy management by maintaining the appropriateness of interpretation. This layer can follow two 'privacy as control' approaches, access control or accountability and auditing approaches [91]. In access control, CPML allows users to specify the appropriate interpretation of their posts. CPML verifies any action or change of context to maintain the appropriateness of the interpretation. Alternatively, without specifying the appropriate interpretation, CPML notifies the owner when the interpretation changes, following an accountability and auditing approach. The owner judges the appropriateness of the new interpretation, and accordingly the change of context is allowed or prohibited.

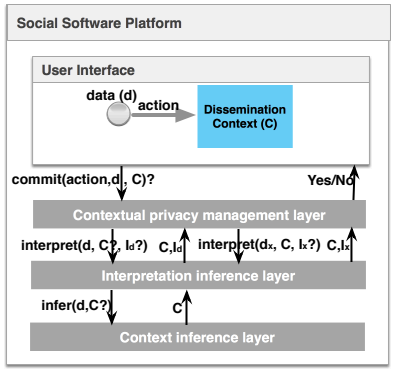


Figure 5.1: Interaction between layers. Upon adding a post d, CPML checks whether the action can be committed by consulting the inference layer. To infer the interpretation, the context inference layer is consulted to check if the current context changes by simulating the action. Based on the inferred context, the interpretation layer infers the new interpretation I_d. If I_d is appropriate and the context changes, CPML checks the appropriateness of the interpretations of other posts d_x before allowing the action.

This layer requires monitoring of data. CPML is responsible for the continuous monitoring of context and interpretation of posts through interacting with the other layers (Fig. 5.1). The interaction is triggered by by actions on data in any situation. Upon an action, e.g., adding a photo in a situation, the CPML checks whether this action can be performed. Firstly, the CPML consults the data inference layer to infer the values of the photo within the current situation, namely the interpretation or the sensitivity of the photos. This layer also consults the context inference layer to check whether the context transitions by committing this action of adding the photo. The context inference layer assesses the context by simulating committing the action. The context layer sends the inferred context to the data inference layer. The latter infers the values of the data and sends them to the CPML. The CPML checks if the values are appropriate. If the context transitions in the simulation of the action, other data items in the same situation will be checked for appropriateness post the transition. When the values of all the data items in the situation are appropriate, the action is authorised to be committed.

5.4 Conceptual Analysis of Usability

In this section, we present a comparative assessment of the design of the CPS^2 and the conceptual framework of contextual integrity (CI) proposed by Nissenbaum [76]. CI is a widely accepted framework for addressing recontextualisation of posts. CI is based on controlling four parameters: contexts, actors, attributes, and transmission principles. CI requires the specification of the norms including: terms of information flow; the prevailing contexts and possible sub- and super-contexts; subjects, senders, recipients; and transmission principles.

Usability is an important aspect in achieving the objectives of security and privacy management mechanisms [21]. If a mechanism is not easy to use, non-technical users would fail to use it to preserve their security or privacy regardless of the effectiveness of the mechanism. Assessment of usability is difficult and time consuming. It requires implementing a system and testing it with users for a particular period of time. This approach requires therefore a significant engineering effort. To avoid such time-consuming tasks, many works have focused on assessing usability at the design phase [95, 21]. Given the similarity in the objectives of CPS^2 and CI, we assess their usability using the 'Security Usability Model' proposed by Braz et al. [21] for designing usable security mechanisms. In their model, they select metrics from the Quality in Use Integrated Measurement model for usability standards [95] to assess the usability of a security mechanism. This model is considered to provide the best usability standard [21]. We use these metrics for our assessment.

The assessment is an estimate of the performance of the designed system. We only use two degrees 'high' and 'low' to indicate the estimated degree of satisfaction of each metric (Table 5.1). In principle, CI requires more effort from users and may pose challenges to usability in contrast to CPS^2. CI requires specifying parameters that may be challenging to specify in advance, for instance, users may not be aware of the terms of information flow in the system, or they not be able to predict how the terms may change over time. The most challenging aspect of CI is that it is based on the prohibitive requirement of specifying appropriate contexts. On the practical level, when CI is deployed in formal access control models or technical mechanisms [13, 59], users are still required to specify the same number of parameters stated in CI. Such models and mechanisms do not offer significant simplification to enhance the usability of CI. In contrast, CPS^2 limits the number of parameters users need to specify for the interpretation of their data. It also requires the incorporation of intelligent mechanisms to overcome the burden of handling context. These two aspects make CPS^2 satisfy the most of the metrics to a higher degree than CI.

Usability Metric	Description	CPS[2]	CI
UM1- Minimal Action	The amount of action required to achieve the task	Low	High
UM2- Minimal Memory Load	The amount of information the user should have in mind to complete the task	Low	High
UM3-Operability	The amount of effort required to operate an application	Low	High
UM4-Privacy	Whether users' personal information is protected	Yes	Yes
UM5-Security	Whether of the application protects information in The system against security threats	Depends on the hosting system	

Table 5.1: Usability metrics relevant to contextual privacy management approaches.

Even if CI incorporates intelligent mechanisms, its usability may not significantly increase. Assuming CI can be deployed with the incorporation of intelligent mechanisms (e.g., context inference). The intelligent mechanism can help to present the users with the possible values of the parameters. However, the users would still need to reason about the appropriate values of these parameters. Such reasoning is not required in CPS^2, except for the interpretation values. On the other hand, the effectiveness and the usability of CPS^2 is dependent on the accuracy of the inference layers. Accuracy may take time to achieve in the system and to adapt to different users.

5.4.1 Implications of CPS^2

With our vision and proposed architecture of CPS^2, we foresee three main design features that enhance the user interaction experience of social software, contribute to the usability, and offer better privacy management. These aims are in line with privacy requirements by various privacy management approaches [3]:

Context change alerts. Besides alerts of inappropriate interpretations, users can be alerted when CAP change. Users will also be alerted of unexpectedly expanded audiences or new co-owners in posts, as hidden audiences and invisible owners are at the source of potential recontextualisation. Users will be given the opportunity to prevent those new audiences, before the damage happens.

Awareness tools. More generally, users will be made more aware of how their communication evolves. As demonstrated in the previous chapter, communication may become adversarial when new interlocutors are brought in who cannot see the full context (e.g., invisible owners) or who have missed previous parts of the discussion. The communication owner will be notified of such potential sources of context change and be given the opportunity to exclude new participants. Alternatively, social software users may selectively disseminate other's posts without violating other's privacy.

Feedback loops. CPS^2 relies on the application of intelligent mechanisms to take the burden off the users and offer them easy identity and privacy management. Users can have the opportunity to provide feedback to the system (e.g., rate alerts or confirm blocked interlocutors) to improve system recommendations. Over time, social software can refine how the intended interpretation is concisely presented to users.

Besides the contribution of the implications discussed above on the usability, we describe in the next section how the proposed concept and the design contribute to minimising user effort in managing contextual privacy.

5.5 Applying CPS^2

In this section, we present how CPS^2 can be applied in private and public communication situations and also how it can enhance privacy management in ambiguous contexts. This section also delivers scenarios to demonstrate the usability aspects of the framework [31].

5.5.1 CPS^2 in Private Contexts

We define private contexts as the communication contexts in which owners constrain access to data to protect their privacy. In the following, we describe using CPS^2 to manage privacy and interact with the software, and demonstrate how the framework requires minimal involvement of users during the following phases.

Disclosure of a Post

The owner provides values for the various CAP, such as post attributes and the audience. The context inference layer infers the context. The interpretation layer infers a set of relevant interpretations. CPML prompts the owner with the set of possible interpretations to specify the appropriate interpretation—in case it follows an access control approach. In case it follows an accountability and auditing approach, CPML saves the inferred interpretations from the original context, or can also allow the user to specify the appropriate interpretation for accuracy.

Context Evolution

CPML checks changes in context and allows only those that continue to preserve the appropriateness of data interpretation. The change is simulated so that the context inference layer and interpretation layer infer the context and the interpretation after the change. Based on the appropriateness of the interpretations of all posts in the new context, CPML either allows the change,

or prohibits it, in case it follows an access control approach. If the change affects the interpretation or the sensitivity of any post and if CPML follows an accountability and auditing approach, it notifies the relevant owners to judge the appropriateness in the new context.

Recontextualisation

When a post is added to a situation, CPML interacts with the data inference layer to infer the post interpretation or sensitivity in the new context. If these new values have not been specified as appropriate by the owner of the post, the recontextualisation is prohibited. If an accountability approach is followed, the owner can judge the appropriateness.

Upon any misappropriation attack, the framework would be able to detect the misappropriation and prohibit it, or consult the attacked user.

5.5.2 CPS^2 in Public Contexts

Besides managing privacy in private spaces, CPS^2 offers contextual privacy management in public spaces. The sharing of posts with the public is not always safe. As an example, some Facebook users suffered from privacy violations by the misappropriation of their profile photos—that are by default public—in the incident of 'hookers of Antwerp' [34]. Profile photos of some girls were put in a 'prostitutes of city of Antwerp' context. The possible interpretation in the new context negatively affected the identity of the girls and counted as a privacy violation for the girls and was reported to the police and Facebook [34]. With CPS^2, users can have a certain degree of control when posts are public to avoid inappropriate dissemination.

5.5.3 Enhancing Communication

In the following we discuss how CPS^2 can enhance communication.

Adversarial Communication

In scenario 7, the manipulated interpretation of the photo through the recontextualisation makes the communication adversarial. This type of

adversarial communication can be mitigated by CPS^2, as we describe in this section.

The steganography Scenario 2 of the previous chapter reflects how users adopt particular strategies when they are unable to control context or want to avoid investing time and effort in restricting the audience [60]. Carmen keeps the context ambiguous and chooses to disclose a post that has two interpretations so that the correct interpretation is not inferred by all the audience member. By doing so, she misleads those in the audience who believe she is truly happy when they are unaware of the actual context. The interpretation is disambiguated based on the knowledge of the audience with the film and not only the online context. This way, the post will be potentially perceived correctly by the friends but not the mother.

This approach of Carmen is referred to as social steganography [19]. It is based on managing the interpretation to be conveyed to the appropriate audience. It is convenient for users who are faced with the complexity of privacy management approaches. It is also similar to the concept of CPS^2, yet, insufficient for contextual privacy management: any of Carmen's friends could comment in a way that reveals an interpretation that Carmen does not want to make explicit. Another problem with this approach is that it obstructs communication. The deliberate interpretation ambiguity may lead to ineffective communication It also involves the risk of inappropriate behaviour by the audience who are unable to perceive the intended interpretation.

CPS^2 offers privacy management and better communication by guarding the interpretation. The affordance of CPS^2 allows the audience to communicate without being concerned that they might reveal an inappropriate interpretation, and thus, they may not become unintentionally adversarial. With CPS^2, owners do not need to resort to adversarial communication, and the audience does not need to be concerned about violating the privacy of others, as reported in the study of Lampinen et al. [60] that shows that the audience are concerned about acting appropriately to avoid violating others' privacy.

Cooperative Communication

CPS^2 can guard cooperative communication where trust in the audience is high leading to violations. Assume the same scenario as Scenario 2 but instead Carmen does not have her mother as a member of the audience. In that case, she can post a status expressing that she feels sad for breaking up with her boyfriend. In this scenario Carmen trusts the audience in perceiving the message and protecting her privacy. However, it is possible for any of the audience to

disseminate the post into an inappropriate context, e.g., put the photo in a context about 'single loser girls'. CPS^2 could assist users in guarding their data in situations where the audience are assumed to be trusted.

5.5.4 Critique

One objection to the proposed contextual privacy approach is that it might affect freedom of speech in public contexts. In principle, it is not possible to practice total freedom of speech. In most cases, freedom of speech is a process of negotiation within the boundaries of social norms, and ethics. For instance, sexists and racist comments are socially not accepted. From an HCI point of view, technologies are required to support freedom of speech and the rights to privacy. Our work seeks to explore a more delicate approach to reach freedom of data communication while respecting boundaries of others. When a user approves the dissemination of her data, the user can still be in danger of manipulating the message of the data. In the setting of an accountability and audit approach, our work does not affect freedom of speech. Rather, it offers keeping data subjects informed about the usage of their data. Once the data subject is aware of how his/her data is used, the possible action is dependent on the software design and legal rules that specify what the subject's rights are. For instance, the EU Data protection Directive allows subjects to rectify inappropriate usage of their data [93]. In other cases, subjects may have the right to prohibit the misappropriation of their data or to negotiate the usage of their data.

5.6 Related Work

Various works realise the importance of context in privacy management and focus on context-based privacy management approaches. However, most approaches lack the dynamic adaptivity to changes in context [90]. Moreover, most works on context-based privacy management address the complexity of controlling context by simplifying the representation of context, as we discuss in the following.

The simplification of context representation can be seen in various models. In the access control model proposed by Fong [39], the context is approximated by relationships between the audience and owner, regardless the type and semantics of the posts they are communicating about. Current social software such as Facebook and Google+ adopt models similar to Fong's. In their implementations, users can pre-specify contexts by specifying groups of

friends. Users can disclose their posts to a specific group or context. The contextual privacy approaches based on Nissenbaum's work [76] also require the prohibitive complexity of specifying details in advance. To overcome such complexity, such approaches also simplify contexts and replaces them with roles [13, 59].

There are multiple shortcomings with the simplistic context representation, mentioned above. Firstly, grouping contacts is a time-consuming process and users are not willing to invest time in managing social software communication [65]. Secondly, such a process continues to be a challenge given changes in friends that cannot be reflected easily in the grouping. One model addresses the challenge of the manual grouping of friends by utilising clustering algorithms to group friends [40]. However, this model does not adapt the groupings over time and does not use disclosure patterns of users in clustering the friends. It may result in groups that feel unnatural to the user. Thirdly, this model does not offer protection against recontextualisation. Fourthly, and most importantly, empirical studies with Facebook users showed that grouping friends is not relevant to privacy management [58]. In contrast to the models discussed above, our conceptualisation of contextual privacy reduces the parameters needed to be controlled without simplifying the representation of context, and proposes the integration of intelligent mechanisms to assist users.

5.7 Conclusion

Context is an essential ingredient for communication and privacy management. This chapter demonstrates that by managing contextual privacy through managing the interpretation of posts, users could manage their privacy without being faced with the complexity of controlling context. The proposed architecture design using intelligent mechanisms is promising for addressing the complexity of controlling context, and enhancing communication. It is promising for offering a social software experience preserving privacy in private and public spaces with a relatively high degree of usability, as well as offering other functionalities related to feedback and awareness.

Chapter 6

The Other Side of Privacy: Surveillance in Data Control

6.1 Introduction

The concepts of privacy and surveillance take new forms in social software technologies. Privacy and surveillance are two important concepts in relation to data disclosure and communication. Traditionally, privacy refers to the right of an individual to be isolated or anonymous [116, 80]. Through privacy, it is possible to avoid surveillance. However, the nature of social software communication affects how these two concepts are viewed and practiced. In social software, privacy may not be achieved through being alone if one is to use the software. Rather, it is a compromise between disclosing data to a set of trustees and hiding it from others. Surveillance takes a new form of monitoring users through their data disclosure. Surveillance is achievable through the utilisation of social software to monitor users. In such cases, privacy may not necessarily counter such surveillance [33].

"Privacy as control" (*PaC*) is one of the most fundamental aspects of daily use of social software. PaC is a research paradigm of privacy management approaches through data control [49]. Social software users can disclose their personal data to various types of audiences. To manage their privacy, users can employ data control approaches. Data control approaches offer users control on where

and to whom their data is disclosed to avoid inappropriate access, tracking, and surveillance. Data control approaches include two classes, namely, access control and accountability. Access control offers means to control who can access data, how, and for what purpose. Accountability offers the verification of the correct enforcement of data control users have. Data control approaches are realised in different ways in technical and legal frameworks.

Although PaC aims at facilitating privacy, in practice, it involves a certain degree of surveillance. PaC approaches are realised differently in technical and legal frameworks. In the two frameworks, giving total control to users is challenging and may not be feasible [93]. In the technical framework, the complexity of data control approaches requires involving *functional entities*—other than users—to deploy these approaches. Such entities control users' data and monitor their actions. With such control, functional entities have surveillance powers. In the legal framework, similar entities are required to monitor users to enforce laws and detect violations. We refer to the surveillance that is required for the functioning of data control approaches as *functional surveillance*. Functional surveillance is essential to ensure privacy management. At the same time, there is no guarantee that functional surveillance may not be used for surveillance of users. In such a case, functional surveillance can turn the social software into a panopticon [53]. In a panopticon, it is not possible to tell whether people are surveilled or not. Either way, people would behave cautiously under the assumption that they are being surveilled. Similarly, social software users have no means to tell whether functional surveillance is used to surveil them or not. Thus, users may have to act under the assumption that they are continuously under surveillance.

The interdependency of privacy and surveillance hinders the assessment of the degree of control and privacy users can have. Functional surveillance in PaC implies that there is an interdependency of privacy and surveillance. While privacy aims at countering surveillance, it may utilise functional surveillance. Functional surveillance can facilitate and hinder privacy. As a result, the control required by functional entities may limit users' control. Functional surveillance affects the offerings of data control approaches and their effectiveness. The main challenge to assessing the offerings of data control approaches is that control is an abstract concept. Control cannot be quantified; however, it can be assessed by the aspects it affects. In this chapter, we investigate the degree of control, privacy and the related surveillance issues in PaC approaches. Our contribution can be summarised as follows:

- A review of PaC in theory and in practice to understand the limitations of PaC (Section 6.2)

- A presentation of the main characteristics of data control approaches based on which criteria is proposed to assess the degrees of control, privacy and surveillance (Section 6.3)
- An evaluation of the technical and legal frameworks using the proposed criteria (Section 6.4)
- A demonstration that transparency and reciprocity are the most essential requirements towards addressing the interdependency of privacy and surveillance (Section 6.5).

6.2 Privacy as Control

"Privacy as control" (*PaC*) is one of three research paradigms of privacy management approaches in the technical and legal frameworks [49]. "Privacy as confidentiality" and "privacy as practice" are the other two. Privacy management approaches vary across paradigms in their approach, assumptions and objectives. In this chapter, we only focus on the PaC paradigm and data control approaches—as opposed to anonymity, feedback, and awareness approaches that belong to the other two paradigms [49]. Our focus aims to investigate the interdependency of privacy and surveillance rooted in data control approaches in social software.

6.2.1 PaC in Theory

Theoretically, PaC concerns offering users as much control as possible to control their data and disclosure contexts in social software. Users can disclose their data in communication contexts within the software. These data can be accessed and may be used inappropriately. To manage their privacy in social software, users should be able to control their data, in terms of how it is accessed and handled in contexts, as we argue in the previous chapters.

Controlling context means that a user should be able to control the various ingredients in a context [91]. In social software, a *context* is the information that identifies a situation within which a user can disclose a data item. For example, in Facebook, a context is the information in a page within which a user posts her graduation photo. The context is defined by the data, the poster and the audience. The *poster* discloses a data item and selects the audience. The *audience* is the set of users who can view an item of a specific poster. A *data subject* is a user that the item relates to, referred to as a subject in this chapter.

Generally, the poster discloses data for which she is the subject. However, it is possible to post items about others; we refer to those subjects as *participants*. Participants can also be members of the audience who contribute to the context with their data, e.g., by posting comments. A user should be able to control the *original context*, wherein the data is first disclosed, as well as control any dissemination context [91].

6.2.2 PaC in Practice

In practice, PaC is realised by data control approaches to offer control to users over their data. Data control approaches aim at facilitating the expression, the enforcement, and the verification of users' control over their data. Expressing control over data requires verifying the correct observance of this control. Access control and accountability approaches offer users various aspects of control. Access control approaches enable users to control access to their data. Accountability approaches verify the enforcement of users' control and identify misconduct.

Data control approaches offer a varying degree of data control, based on the assumptions and objectives of each approach [93]. In many cases, these approaches involve functional entities that perform particular functionalities, e.g., an access control enforcement entity or an accountability and audit entity. Functional entities are necessary in certain approaches to perform tasks that users are unable to perform. Functional entities are required to have a certain degree of data control to perform their tasks. With such control, functional entities have access to users' data and actions. This control, in turn, limits the degree of control users can have, e.g., by not allowing users to hide their data from functional entities. Moreover, the realisations of these approaches and their offerings vary based on the underlying framework. Such variance means that a consistently high degree of control is not possible in PaC in practice, as we discuss next.

6.2.3 Limitations of PaC and Surveillance

The main issue of PaC is that it may not be practically possible to offer a high degree of control to social software users [49]. In practice, users may be faced with various limitations. Such limitations can be technical, legal, and ethical. From a technical point of view, the degree of control depends on the capabilities of the social software system, the design of the data control approach, and the data it protects. Additionally, the degree of control depends on the usability of

the approach. In general, users face difficulties in using PaC approaches [90]. However, in the context of this chapter, we do not focus on the usability issues of PaC. From a legal point of view, the degree of control varies depending on international boundaries, e.g., there is a significant difference between the control offered by the EU Directive and US privacy legislation [86].

From an ethical point of view, assigning a high degree of control to users may have consequences counter to privacy [93]. Users with a high degree of control can conceal their malicious acts of violating others' privacy, e.g., leaking data. In such cases, assigning a high degree of control to users can run counter to what is dictated by law. According to the law, when data may affect other users or concerns criminal or illegal acts, a certain degree of supervision and the limitation of users' control are needed. Supervision of users' actions can be achieved by accountability approaches, for instance. Such approaches entail a certain degree of surveillance as well.

Social software may be utilised for different forms of surveillance. Social software are seen as a realisation of surveillance in modern society [50]. Many parties can apply surveillance such as parents, marketing, recruiting companies or governments [12]. This form of surveillance is achievable through monitoring the data users disclose. Another form of surveillance is functional surveillance achievable through PaC. Functional surveillance is a fundamental part of many data control approaches (Section 6.3). Functional surveillance facilitate monitoring users disclosures, actions, trust and privacy management patterns.

The main challenge is that there are currently no means to assess the degree of control the users and other parties could have, and the involved degree of functional surveillance. Both users and researchers need to understand the offerings of data control approaches. The inability to assess the degree of control users can have affects assessing the degree of privacy management that can be achieved. To address this problem, an understanding of data control approaches is needed. Based on such understanding, it is possible to assess—at a high level—the control users can have and functional entities can have. In the following, we provide an understanding of data control approaches and propose criteria for evaluating the control offered to users, and the entailed functional surveillance. We apply the criteria on the general aspects of data control approaches at a high level. Such an application demonstrates how the criteria can be used to assess any data control approach.

6.3 Data Control Approaches

In the following, we discuss the realisations of data control approaches in the technical and legal frameworks.

6.3.1 The Technical Framework

The technical framework realises data control approaches via technical mechanisms to express, enforce and verify data control, as described in the following.

Access Control

An access control mechanism enables users to define policies regarding how their data can be accessed. The data control policies can be defined using a specific set of features of the system to regulate and authorise access to data [90]. For instance, a policy can state that only users who are of a certain age can access a specific data item. Most of the mechanisms assign control to the poster assuming that the poster is the subject. Fewer mechanisms may assign control to posters and participants, e.g., the mechanism of Squicciarini et al. [102]. For each access request of a data item, the relevant policy of the controller is enforced. When the constraints of the policy are satisfied, access is authorised. The enforcement of policies is executed by an enforcement mechanism. The enforcement mechanism is a functional entity that controls users' data and policies.

An access control mechanism may involve one or more functional entities depending on the architecture of social software. In *centralised architectures*, a central authority is responsible for providing the social software services. The central authority can be the enforcement functional entity. In *decentralised architectures*, the data is distributed on multiple servers. These entities are the functional entities. Another option is deploying the services on users' own machines, e.g., the work of Cutillo et al. [32]. In this case, the user's machine acts as a functional entity.

Accountability

Accountability mechanisms are based on auditing the system to identify misconduct and anomalous actions [97]. These mechanisms perform auditing

of system logs, policy enforcement transactions and users' actions. By such auditing, they judge the compliance with privacy rules [94]. When no violations are detected, it is an indication of observance of the control users have expressed.

The functioning of accountability mechanisms requires the involvement of functional entities. The number of functional entities depends on the social software architecture. In *centralised architectures*, the central authority is the functional entity that deploys the mechanism. This entity tracks users' actions. In *decentralised architectures*, the entities hosting the social software are required to cooperate and exchange data. These entities have to record all the information exchanged between users and then link them to data of other entities [52].

6.3.2 The Legal Framework

The legal framework adopts PaC through data protection legislation to protect individuals [2]. We focus on the legislation of the European Directive 95/46/EC (EU Directive 1995), referred to as "The Directive" in this chapter. The Directive is (in comparison to other data protection legislation) one of the most privacy-friendly legal regulations [86].

6.3.3 Access control and Accountability

The Directive offers data protection via adopting accountability [8]—and implicitly access control. The Directive states the rights and liabilities of entities that can access and process data. By specifying who can access and process data, The Directive also formulates access control regulations.

The Directive differs from the technical framework in the set of entities that can have control over data. Instead of assigning control to users, The Directive distinguishes two main entities: the data subject, hereafter referred to as subject, and the data controller. The controller "determines the purposes and means" of the processing (Article 2(d)), and is responsible for ensuring compliance with The Directive. In contrast to the technical framework, The Directive considers the social software provider or a third party to be the controller. According to the personal use exemption (Article 3(2))—when data is accessed for "purely personal or household activities"—subjects are not the data controller of their own data. Subjects may not solely determine the purposes and means of the processing of their data. However, subjects can be considered as controllers of other subjects' data, not their own [30].

The Directive distinguishes between controlling and processing data. The Directive defines a data processor role as the entity that can process and perform specific operations on data on behalf of the controller. The processor is usually not a subject. Although the data controller can solely determine how subjects' data is processed, subjects have the right to be informed and give consent to the control of the data controller (Article 7). The data controller is obliged to inform subjects about its identity and purpose of processing in order to obtain their consent (Articles 10–11). The data controller should maintain the accuracy of data or else delete or rectify them. However, the data controller is not subject to those constraints in specific exceptional cases (Articles 13 and 7(b–f)). The first exception (Article 13) applies when the processing of data is necessary for the completion of tasks of legal authorities, such as the protection of the security or economic interests of the state, criminal investigations, or the protection of subjects, their rights and freedoms. The second exceptional case (Article 7(b–f)) applies when the processing is required to comply with a contract the subject is part of, fulfil a legal obligation, or protect the interest of the subject or the controller.

The legal framework varies from the technical framework in the parties they offer control to. The legal framework assigns the highest degree of control to the service provider, who is also a functional entity. In a distributed architecture, the number of service providers increases, and thus the number of functional entities.

6.4 Evaluation Criteria for Compliance with PaC

Measuring the degree of fulfilment of PaC is not possible in general [49]. However, it is possible to assess the degree of control offered by a data control approach by investigating the aspects that can be controlled and the degree of functional surveillance entailed. According to the identified context ingredients (Section 6.2) and a previously proposed set of requirements for offering a high degree of control and privacy [90], we propose the following criteria. Each of the following criteria concerns a particular data control aspect.

Control over Data Which types of data items can a subject control? What subjects have control over their data, posters or participants? And what is the degree of control the subject has? Data items can be posted items, actions produced while using the social software, or other inferable data.

Identifiability of the Data Subject Is the subject identifiable in the original context? Is the subject identifiable in any dissemination context where her data is put?

Audience Control What is the degree of control a subject has over her audience? And what is the degree of control the audience have over the subject they are the audience of? Such control means that an audience member can select the subject to view data from.

Control over Context What is the degree of control a subject has over the original context within which her data is first disclosed? And what is the degree of control over any dissemination context? The degree of satisfaction of this criteria is dependent on the satisfaction of the above criteria. As an example, when participants cannot control their data, this criterion is not satisfied in relation to participants.

Degree of Functional Surveillance What is the degree of functional surveillance applicable by a functional entity?

To use the criteria for evaluation, we discuss the related aspects and state the approximate degree of satisfaction. The degree of control can be either, high, moderate or low. If the evaluation results in a high degree of control and a low degree of functional surveillance, we conclude that users can have a high degree of control, and in principle a high degree of privacy. The evaluation does not focus on a particular realisation, rather, it is performed at a high level on the general aspects shared amongst realisations.

6.4.1 Evaluating the Technical Framework

The criteria are applied on the main characteristics of technical data control mechanisms.

Evaluating Access Control

In the following, the satisfaction of the criteria is applied on the main characteristics shared amongst various access control mechanisms, which we surveyed earlier [90].

Control over Data Access: Most mechanisms offer posters control over their posted data items. In relation to offering control to participants, only very few mechanisms offer such control [90], e.g., voting-based access control [112].

This criterion is satisfied to a high degree in terms of controlling the data users disclose, and to a low degree in terms of controlling actions or inferable data. It is satisfied to a high degree in terms of giving control to posters, and to a low degree in terms of giving control to participants.

Identifiability of the Data Subject: Most mechanisms maintain posters' identifiability in the original context. Identifiability is maintained when the policies of the poster are enforced on the poster's data. The identifiability is not always possible in dissemination contexts, unless sticky policies are used [56]. The lack of participants' control over their data makes them not identifiable. This criterion is satisfied to a high degree in the original context, and to a low degree in any dissemination context in relation to posters.

Audience Control: Access control mechanisms offer subjects control over their audience. This control is mainly offered to posters. The control is applicable within the social software boundaries. Posters' control does not apply to functional entities and what they may access [51]. It is not alway possible for the audience to have control over subjects. An example is Facebook-like access control, once a poster specifies the audience, the audience will have the poster's item in their newsfeed. The audience cannot specify the poster and what of her data they would like to view. This criterion is satisfied to a moderate degree in terms of giving control to posters over the audience, to a low degree in terms of giving control to participants, but is not satisfied in terms of giving control to the audience.

Control over Context: Most mechanisms offer a high degree of control over the original context to posters, and a limited control over dissemination contexts, as discussed in Chapter 4. Offering control over dissemination contexts to users requires assigning a higher degree of control to functional entities over users' data and the contexts [103]. This criterion is satisfied to a high degree in terms of controlling the original context, and often a low degree in terms of controlling dissemination contexts.

Degree of Functional Surveillance: the degree of functional surveillance depends on the underlying architecture. In a centralised architecture, the enforcing entity must have access to users' posted data — unless the approach incorporates encryption to hide the content of the posted data. Upon enforcing a policy, the entity gains knowledge about who is denied or allowed to access a particular data item. The enforcement entity could infer information about users' trust patterns, amongst other information. In this architecture, the degree of functional surveillance is high due to possible accessibility of users' disclosed, actions or inferable data.

In decentralised architectures, users' data and actions are accessible by more

than one functional entity. Even when these entities have access to a subset of the information, they can still aggregate the information. The challenge in this architecture is defining “trust" and “trusted entities". With the assumption that the trusted entities act as trusted and do not aggregate users’ data, the functional surveillance is lower than that encountered in the central architecture. If the trusted entities aggregate users’ data, the degree of functional surveillance is higher than that in the centralised architecture because more than one entity are surveilling users. In the case of deploying the enforcement mechanism on the user’s own machine, no external functional entities are required and the degree of functional surveillance is low.

Access control mechanisms involve a relatively high degree of functional surveillance. This functional surveillance affects the degree of the control resulting in a moderate degree of control offered to users. The limited control offered to users is dependent on the control offered to functional entities. The more control functional entities have, e.g., over contexts inside and outside the social software, the more control users can have.

Evaluating Accountability

In the following, the satisfaction of the criteria is applied on the main characteristics shared amongst various accountability mechanisms.

Control over Data: Accountability mechanisms facilitate indirect verification of subjects’ control over only their posted data. Since subjects cannot define policies over actions, or inferable data (Section 6.4.1), accountability mechanisms cannot verify control over such data. Since participants cannot specify policies over their data, most accountability mechanisms can only verify the observance of posters’ control. This criterion is satisfied to a high degree in relation to posters and their posted data, and is not satisfied in relation to actions or inferable data. It satisfied to a low degree in relation to participants.

Identifiability of the Data Subject: Accountability mechanisms are based on the identifiability of subjects. To identify accountable entities and the affected entities, every action and data is linked to its subject. However, the identifiability is mainly possible to the poster. This criterion is satisfied to a high degree in relation to posters, and to a lower degree in relation to participants.

Audience Control: Generally, these approaches check the handling of data by other users but not the functional entities [51]. Thus, these mechanisms verify the subject’s control over the audience within the software. Accountability mechanisms do not check the audience control over their subjects when such control is missing in the access control mechanism. This criterion is satisfied to

a high degree in terms of subjects controlling the audience, and is not satisfied in terms of the audience controlling subjects.

Control over Context: Accountability mechanisms can facilitate indirect control over data even in contexts that subjects do not have control over via access control mechanisms. When a data item is leaked into a new context, accountability mechanisms can detect the leakage in this context and report it to the subject concerned. The subject can take the appropriate action and thus have a certain degree of control. This criterion is satisfied to a higher degree by accountability mechanisms than by access control mechanisms.

Degree of Functional Surveillance: Accountability mechanisms are strongly coupled with surveillance. The surveillance in these approaches takes the form of monitoring users' data and actions in order to identify misconduct. The degree of functional surveillance is dependent on the architecture. In a centralised architecture, the auditing entity has access to all users' data. In decentralised architectures, the functional entities have access to a subset of the data. Also, functional entities might need to aggregate data to identify misconduct, e.g., when a data item is leaked from one user to others, the auditing entities should aggregate their data to trace how the data item has moved from one user to another. In such cases, this architecture may entail a higher degree of functional surveillance than that of the centralised architecture, which is counter to objective of decentralised architecture.

Accountability mechanisms offer a moderate degree of control. At the same time, they comprise a higher degree of functional surveillance. In access control, users do not have control over who can access their actions and relational data. Yet, such data is utilised by accountability mechanisms to verify the observance of users' control over their posted data. Accountability mechanisms involve a high degree of functional surveillance because they utilise data users post as well as behavioural data.

In summary the technical framework offers a high degree of control on internal audiences and original contexts. At the same time, the framework involves a high degree of functional surveillance. Thus, the degree of total privacy achieved in this framework is moderate.

Evaluating CPS^2

In this section, we will evaluate our conceptual framework for contextual privacy management presented in the previous chapter 5 to understand its offerings. In principle, CPS^2 is aimed at giving a high degree of control without burdening users. The framework requires the utilisation of artificial intelligence

mechanisms to achieve ease-of-use and control. This approach offers control over data interpretation and sensitivity. It indirectly offers control over other ingredients, as discussed in the following. The framework can adopt an access control or accountability model. In both cases, the satisfaction of the criteria is the same.

Control over Data: Users have control over their posted data. The control is offered for posters, and can be offered to the participants as well. This criterion is satisfied to a relatively high degree in terms of posted data only.

Identifiability of the Data Subject: This criterion is satisfied to a high degree through maintaining the appropriateness of data of the user across contexts.

Audience Control: This criterion is satisfied to a moderate degree in terms of controlling the audience who might misappropriate data. It is not satisfied in terms of the audience controlling their subjects.

Control of Context: This criterion is satisfied to a high degree in terms of controlling context changes and data dissemination.

Degree of Functional Surveillance: The framework requires a high degree of functional surveillance. The functional surveillance is required to monitor context changes and users' actions.

In total, the degree of control and privacy is moderate given the high degree of functional surveillance. This degree demonstrates how aiming at achieving ease-of-use of privacy management approaches can come at a price in terms of possible surveillance.

6.4.2 Evaluating the Legal Framework

In the following, the criteria are applied on aspects of access control and accountability regulations in The Directive.

Evaluating Access Control

The criteria are satisfied to variant degrees as discussed in the following.

Control over Data: Subjects have control over any identifying data whether posted or processed by automatic means (Article 3(1)). The control is possible for posters, and participants as long as the data identifies them. However, the data controller solely specifies and enforces the terms of how the data can be used, and, as a result, has a higher degree of control than the subject. Subjects'

control is limited to the right to consent. This right allows the subject to either accept or reject the processing terms of the data controller, as long as the exceptions are not applicable (Article 7(b–f)). This criterion is satisfied to a relatively high degree in terms of controlling different data types of different subject types as long as exceptions do not apply. However the control users have in this framework is not as granular as the control users have in the technical framework, since users can specify their own detailed data control rules in the technical framework.

Identifiability of the Data Subject: Identifiability in any context is required by The Directive to allow subjects to access and receive information about their data when it is to be processed (Articles 10–12). This criterion is satisfied to a high degree.

Audience Control: The Directive states that subjects can specify their audience within the social software. Also, subjects have a limited degree of control over the data controller — by the right to consent — who controls external audiences, and processors. The Directive does not offer control to audience members over subjects. This criterion is satisfied to a high degree in terms of controlling audience within the social software users, and to a moderate degree in relation to controlling functional entities and external audiences. It is not satisfied in terms of the audience controlling their subjects.

Control of Context: The Directive offers control to users over contexts by being informed about the processing data (Articles 10–11). The control is limited to whether the user would give consent or not. Subjects, however, cannot limit the processing of data in a specific context if they have given their consent or if the processing is necessary (Article 7). Such control applies to contexts within the social software as well as any context beyond the boundaries of social software. This criterion is satisfied to a moderate degree.

Degree of Functional Surveillance: The access control regulations involve a high degree of functional surveillance. The data controller is the functional entity responsible for the enforcement of the control offered to subjects. The controller specifies the terms of data processing, and has access to all the data to enforce the terms. Additionally, external functional entities can access the data to conduct certain investigations. In such cases, the monitoring or surveillance of subjects facilitates performing the investigation tasks. Such surveillance is explicitly exempted from being reported to subjects (Article 13(f)). The degree of functional surveillance is higher than the degree of functional surveillance in the technical framework, as long as exceptions do not apply (Section 6.3.2).

The satisfaction of the criteria varies as discussed in the following.

Control over Data: The regulations oblige the data controller to inform data subjects of how their data are used. If the processing does not comply with what the subject has consented to, the subject can complain. If the exception of Article 13 applies, subjects may not be entitled to this right. This criterion is satisfied to a high degree in terms verifying the observance of control over all data types as long as exceptions do not apply.

Identifiability of the Data Subject: Identifiability is required by the accountability regulations to facilitate identifying subjects who are accountable for misconduct. This criterion is satisfied to a higher degree than the degree of satisfaction in the technical framework.

Audience Control: The Directive allows subjects to verify how their data is being processed by the controller or the processor, as long as exceptions are not applicable. Thus, subjects can verify the observance of the control they have over their data by functional entities and external audiences (Article 3(2)). This criterion is satisfied to a high degree in terms of verifying the control over functional entities and external audiences, but it is not satisfied in terms of the control of the audience on subjects.

Control of Context: The Directive can verify the control over context by verifying the processing terms users consented to, whether context is within or outside the social software boundaries. This criterion is satisfied to a moderate degree.

Degree of Functional Surveillance: to verify the observance of terms of processing by the data controller and the processor, a high degree of functional surveillance is required. Such functional surveillance is performed by external functional entities and legal authorities that access all data available about subjects, the data controller, and the data processors. As a result, the accountability regulations entail a high degree of functional surveillance.

In summary, the legal framework involves a high degree of surveillance to enforce rules and facilitate the control to users on external entities and contexts[1]. Thus, the degree of privacy achieved in this framework is moderate.

The main difference between the technical and legal framework is the scope of the offered control (Table 6.1). The technical framework offers control of

[1]The new Directive may involve a different degree of functional surveillance as it mainly focus on data processing for criminal activities [61]

varying granularity. Users can express fine- or coarse-grained control, according to the technical mechanism used. In contrast, the legal framework offers static and large-scale control. The regulations that offer control cannot be changed by users, but they apply on a large scale—within and beyond the social software boundaries. Another difference is that it is possible in the technical framework to detect violations sooner than in the legal framework. In the legal framework, detecting violations requires checking the compliance with the regulations by external authorities. Such checking may not occur periodically, as is the case in the technical framework. With regards to functional surveillance, the main difference is the entities who perform the surveillance. In the technical framework a the surveillance is mainly performed by technical mechanisms, while in the legal framework, humans involvement is required for functional surveillance.

In the Technical Framework	*In the Legal Framework*
- Specified by the user	- Specified by the data controller
- Applied on data disclosed by the user	- Applied on data that identifies a user regardless of who discloses it
- Applicable within the system the data is disclosed in	- Applicable within and beyond the boundaries of the system
- Affects the audience within the system	- Affects the audience within and outside the system
- Functional surveillance performed by technical mechanisms	- Humans perform functional surveillance
- Violations are detected in a timely manner	- Detection time depends on how often the regulations are checked

Table 6.1: A comparison between the control offered by the technical and legal framework. The technical framework offers more fine-grained control that is mostly limited to the system within which the data is disclosed, while the legal framework offers control on a larger scale.

6.5 The Interdependency of Privacy and Surveillance

The interdependency of privacy and surveillance is inherent in the design of data control approaches. In the previous sections, we discussed how functional surveillance is part of data control approaches. Giving users a high degree of control over dissemination contexts, for instance, requires increasing functional surveillance to detect disseminations of data in any context. The

interdependency of privacy and surveillance is manifested in how functional surveillance facilitates privacy, and how privacy aims at mitigating surveillance. In the following, we propose recommendations that are essential when the interdependency is present in an approach.

6.5.1 Recommendations

The variation in the degree of control, privacy and surveillance between PaC approaches in the two frameworks suggest the need for a holistic PaC approach. The variation emerges from the differences in the perspectives and the aspects focused on by an approach. Addressing privacy issues and legal concerns at the same time requires an approach that merges the offerings of the technical and legal frameworks. Developing such an approach requires taking into consideration the interdependency of privacy and surveillance, rather than focusing on only privacy issues [33]. The first step towards developing a better data control approach is adopting transparency and reciprocity.

Transparency is essential to address the dependence of PaC on surveillance. Transparency means that the parties who are having their data processed should know who is processing their data and, possibly, why. The functional surveillance in data control approaches may turn the social software platforms into a panopticon. In a panopticon setting, individuals should be aware that they are being surveilled. If users are aware of surveillance they can choose what data to disclose in such a platform [57]. Similarly, social software users should be aware of the degree of surveillance in data control approaches [33]. In this case, users can experience being in surveillance spaces and develop appropriate strategies [70] — assuming they are not faced with usability issues of data control approaches.

Reciprocity must be adopted to support transparency. Reciprocity means that if a surveillant entity can monitor users, then users should be able to monitor this entity [29]. Once surveillance is transparent for users, the users should able to observe the conduct of the surveilling entity. Reciprocity can be replaced by feedback to users about how their data is handled. Feedback is a “privacy as practice" approach [49]. With reciprocity, users may achieve privacy as practice as well.

6.5.2 Transparency and Reciprocity in Practice

As an example of the benefit of transparency and reciprocity, consider the case of Facebook use in Syria. Facebook and Youtube were blocked in Syria until

after the uprising in Egypt. The uprising in Egypt coincided with calls on Facebook for demonstrations in Syria. At that point, the ban was lifted as a reward for the people who did not respond to the calls [77, 73]. This meant that individuals did not need to use Tor anonymous communication networks [35]. Without Tor proxies, the identities and communication of individuals were to be known to the ISP [51]. Such unintentional transparency and reciprocity about the potential behaviour of the authorities made it clear for activists that the motivation for the lift of the ban was probably to prevent anonymous communication and to surveil individuals—given the history of the country. Later reports suggested that surveillance was the reason for the ban being lifted [84]. In this scenario, it is the knowledge about potential surveillance that empowered individuals to carefully use social software and select what to disclose, as McGrath predicted people to behave in such a transparency about surveillance [70]. Such a change of behaviour is also observed in the spike of web searches about surveillance after Snowden's revelations [83]. The spike indicates that users were interested in gaining more knowledge about how surveillance can be applied, avoided, etc.

6.6 Related Work

Similar analyses of privacy management approaches have been conducted earlier. In the technical framework, Danezis and Gürses provide a review of privacy technologies and highlight the entanglement of privacy and surveillance in technologies [33]. Their review covers a wide selection of technologies developed between 2000 and 2010, but mainly focuses on anonymous communication and identity management technologies. Their review argues that total control of data is an illusion, and that privacy technologies can be turned into surveillance tools. While their review focuses on the three privacy paradigms—privacy as control, privacy as confidentiality and privacy as practice—our work differs in focusing just on privacy as control (PaC). Our work extends the analysis of PaC to the legal framework. Our work also differs in conceptualising functional surveillance as one factor that facilitates the use of technologies for surveillance. Another difference is our proposed criteria that can be applied to any approach to assess the degree of control and surveillance. The criteria can be applied on anonymous communication, identity management approaches, or any other approach within PaC.

In another work, Gürses and Diaz focus on surveillance and social privacy issues in social software [51]. These issues relate to the aspects discussed in our work. Data control approaches facilitate social privacy management to avoid violations and surveillance. The authors argue that surveillance and

social privacy issues are entangled, and that privacy management approaches should not address one of these issues and ignore the other. We also examine data control approaches comprehensively and show that this entanglement is a functional requirement for data control approaches. We argue that aiming to give as much control as possible to users may not address this entanglement, since functional surveillance is a fundamental aspect of data control approaches. Our work also differs in that we consider data control approaches in the legal framework, while Gürses and Diaz do not focus on data protection regulations. The questions proposed by Gürses and Diaz for eliciting information useful to developing a holistic privacy management approach can be integrated with our proposed criteria. Such an integration provides detailed information towards developing holistic approaches.

The concept of functional surveillance has been examined by other authors. The work of Gurevich *et al.* conceptualises *inverse privacy* [47] that relates to our conceptualisation of functional surveillance. Inverse privacy refers to the concept of collecting information about users without their knowledge. *Inversely private data* is data that the user is unaware of, yet, it is accessed by entities unknown to the user in a way that can be inappropriate. In our work, functional surveillance facilitates inverse privacy by facilitating the collection of information about users' actions and usage of data control approaches. Our proposed criteria can be applied to assess the degree of inverse privacy.

6.7 Conclusion

In PaC, users cannot control their data without relying on the control of functional entities. In the comparative analysis presented in this chapter, we show the complementarity between access control and accountability. We also show the variation in the realisations of PaC in the technical and the legal frameworks. The realisations of data control approaches offer varying degrees of control and functional surveillance. These offerings result in an interdependency of privacy and surveillance. The analysis explicates the reasons for this interdependency.

The application of the proposed criteria is promising for the assessment of the degree of privacy and the degree of surveillance of specific approaches. Such an assessment is fundamental and should not be skipped by researchers. The criteria should be adopted to decrease the ambiguity about the degree of control and privacy an approach can offer. It should be also adopted to make clear the possible surveillance that may be caused by a particular data control approach.

Chapter 7

Conclusion

> I think there is a tendency in science to measure what is measurable and to decide that what you cannot measure must be uninteresting.
>
> *Donald Norman*

7.1 Introduction

This thesis provides a framework for analysing and managing contextual privacy through measuring the interaction of data sensitivity and context. Commonly, data sensitivity and the role of context in privacy management are not studied in the field of privacy research. Different technical privacy management approaches incorporate context differently based on assumptions related to how data is disclosed and managed. This thesis analyses such assumptions. The approach of the thesis is a multidimensional investigation of the relation of privacy to context and data. The investigation is performed from an empirical point of view, through large-scale data analyses (Chapters 2 & 3). It also involves a conceptual examination of the role of context in communication (Chapter 4). Based on our investigation, we highlight the effect of context on data sensitivity. Accordingly, we propose a conceptual approach to manage contextual privacy by managing the sensitivity or the interpretation of data (Chapter 5). Additionally, the investigation of this thesis extends to how privacy

is tackled in the technical and legal frameworks to explore the limitations of privacy management approaches (Chapter 6).

Throughout the thesis, we explore contextual privacy as the interplay of context and data sensitivity. We conduct big data analyses to investigate the effect of context on users' behaviour. The investigation involves a wide spectrum of contexts and how they affect data of different topics. The analyses show how users disclose their data and manage privacy in a context-based manner. Using machine learning algorithms, we model users behaviour in relation to context. The modelling demonstrates that context has a significant effect on data sensitivity. The variation of sensitivity, based on context, affects data disclosure and privacy management behaviour. The results of the analyses are important to understand how contextual privacy can be managed effectively.

The investigative approach of the thesis is motivated by problems observed in technical privacy management approaches. The focus of the thesis is on the *privacy as control* paradigm through which users can control over their data to manage their privacy. Most privacy problems emerge due to the lack of control users have over their data. Other problems emerge from complexity and usability issues. Moreover, the lack of context control may facilitate data misappropriation privacy attacks, even by users who are authorised to access data. To address these problems, we adopt an investigative approach to understand and simplify contextual privacy management. We also model data misappropriation attacks and analyse their relation to context and sensitivity. We propose an approach to mitigate these attacks through the management of data sensitivity and interpretation. We discuss various realisation approaches to simplify context control using automatic inference mechanisms.

In total, the thesis provides answers to the research questions stated in Chapter 1, as follows:

- How does context affect data and privacy?

  By affecting the sensitivity of data, as well as its interpretation. The empirical analyses (Chapters 2 & 3) demonstrates the effect of context on data disclosure and management patterns.

- How to detect and protect data from misappropriation?

  By maintaining the appropriate sensitivity and interpretation of data (Chapter 4).

- How to achieve usability without limiting context-based control?

  By offering sensitivity or interpretation management—instead of context management (Chapter 5).

- How to assist users in the burden of managing contextual privacy given the large amount of data and audience in social software?

  By utilising automatic inference mechanisms to assist users in context management (Chapter 5).

- Is it possible to offer a high degree of control to achieve a high degree of privacy, whether technically or legally?

  It is not possible to offer a high degree of control to achieve a high degree of privacy. In the technical and legal frameworks, the degree of control and privacy users can have is often limited, and even coupled with a certain degree of surveillance Chapter 6).

The thesis provides more detailed answers to these research questions, as we summarise in the next section.

7.2 Summary of Findings

The thesis focuses on analysing parameters related to contextual privacy. Context, data sensitivity, users' subjectivity are essential parameters of contextual privacy management. These parameters, however, are usually challenging to measure. Challenges emerge from the lack of data that describe such parameters. Challenges also emerge from the nature of certain parameters that makes them not directly observable. For instance, subjectivity is not a parameter that can have a value in datasets. Rather, measuring subjectivity, requires having access to data from different users to measure the differences emerging from subjectivity in their behaviour. We overcome such challenges through analysing Bing dataset. The dataset describes how data is disclosed and managed in different contexts, by different users over a period of six months. The data is also characterised by the type of topic it relates to. Our empirical investigation contributes a set of findings that are described in the following.

7.2.1 Sensitivity

Our investigation focuses on exploring data sensitivity and what affects it. Sensitivity is a latent variable in our dataset. However, by observing how data is disclosed and managed in different situations, we inferred information about data sensitivity. This information relates to sensitivity and the following aspects:

Content

The findings show that there is an association between sensitivity and data content. Through knowing the data content, it is possible to estimate the sensitivity of data. Certain content topics are associated with high sensitivity, even those that are not commonly viewed as sensitive, e.g., *Celebrities*. However, sensitivity of topics can vary across contexts or users.

Time

Sensitivity of data varies based on time. Our findings show that users tend to manage their sensitive data soon after disclosing it. They also show that data sensitivity may change over time.

Context

Sensitivity may change based on the online or offline context surrounding the data or the user. By knowing the context in which data is disclosed or managed, it is possible to predict the sensitivity of data.

Subjectivity

Sensitivity can be assessed differently by different users. Topics that can be sensitive to some users, may not be very sensitive for others. Topics that are commonly considered sensitive may be sensitive to some users, but not all. For instance, in a random sample of 75 users, *Adult* data is sensitive for only 43 users.

In summary, sensitivity is affected by all the above-mentioned aspects. It is important to state that the effect of any these aspects cannot be considered in isolation of the effect of the other aspects. The sensitivity of a data item is affected at once by the content, context, time and the user handling this item.

7.2.2 Context

Given the importance of context in contextual privacy, our investigation focuses on exploring the effect of context on sensitivity and how users handle their

data. Our findings show how context affect data disclosure and management (post-disclosure), as we describe in the following.

Effect of Context on Data Disclosure

The findings show that all contexts affect how data disclosed. Data disclosure varies across the possible values of each individual context. This means that data of the same topics is disclosed differently within different values of this context. Thus, the context affects the sensitivity of data, according to which the data is disclosed.

Effect of Context on Data Management

The context affects how data is managed after it is disclosed. We investigated the cumulative effect of all the possible contexts in our dataset on sensitivity and users' behaviour. Our findings show that different contexts have a varying effect on data sensitivity and users' behaviour. For example, we found that *organisation type* have a significant effect on data management, in particular, *Government*, and *Hospital* types. We found that *Facebook* users are significantly less keen on managing their sensitive data, in comparison to *Windows Live* users. Also, we found that users accessing the internet using *HTTP* with no proxy are more keen on managing their sensitive data, in comparison to users using proxies. This may mean that using a proxy, users may believe they are safe and there is no need to delete sensitive data.

7.2.3 Modelling Data Management Patterns

We investigated the possibility of modelling users' behaviour with regards to sensitivity and privacy management. The modelling describes how content and context affect how users behave. By this modelling, it is possible to predict how data of a particular topic can be managed in a particular context.

Our findings show that it is possible to model users' behaviour. They show that it is possible to model the sensitivity management pattern at a high level that applies to all users. This pattern describes how data can be managed, in our case, deleted. Given the content and context attributes of a data item, it is possible to predict whether the item can be kept or deleted with a certain degree of accuracy.

Our findings show that it is also possible to model the pattern of sensitivity management at the user level. This pattern is more specific than the high-level pattern. For certain users, the user-specific pattern can vary significantly from the high-level pattern. The accuracy of predicting data management actions can be higher in the user-specific pattern, compared to the high-level pattern. This means that there are common rules that define most of the individuals? behavioural, but not completely. The difference between the high-level and user-specific patterns should be taken into account in designing privacy management approaches. Fo instance, without taking into consideration users' subjectivity, automatic privacy management approaches may not be as effective as they could be.

Our findings also show that it is possible to model context using the information about data content. Using information about disclosed data, it is possible to predict in which context the user disclosed the data. This modelling can be used to detect similarities between different content topics. Using these similarities, it is possible to group topics together without affecting the accuracy of the prediction of context.

7.2.4 Contextual Privacy

Our findings show that contextual privacy concerns the data management to maintain the data owner's communicated message. The conceptual investigation of the role of context in communication shows the importance of context in facilitating the correct delivery and interpretation of the communicated message. Users communicate their messages through disclosing data. By controlling context, users can avoid data misappropriation that affect the communicated message. Given the complexity of controlling context, we investigate the possibility to control ingredients other than context. Given the role of context in affecting data sensitivity and interpretation, we propose controlling these ingredients to manage contextual privacy. At the same time, we propose facilitating the inference and management of context by the social software platform using machine learning tools. Actions on data and context changes can be allowed as long as the values of sensitivity or interpretation specified by the user are maintained. We also propose using the sensitivity as a proxy for the interpretation, and that contextual privacy can be managed by maintaining the sensitivity of data.

Our findings show the high usability of the design of the CPS^2 and its relevance to legal privacy management approaches. The design of CPS^2 is assessed for usability. The assessment shows that the framework offers a high degree of usability, in comparison to the well-known framework of *Contextual*

Integrity [75]. We demonstrate also how our framework provides technical means to enforce what is dictated by the legal system, in particular the European Directive 95/46/EC [38].

7.3 Implications of Findings and Future Work

Our findings have various implications on context and privacy management. Firstly, our findings show the importance of examining data to infer the patterns of users' behaviour. The findings show how certain assumptions about what can be sensitive and how users behave may not always hold. They show the importance of selecting the dataset to investigate a particular hypothesis and infer a particular pattern. In other words, our findings show the implications of making different modelling and inference decisions. Moreover, our empirical investigation methods can be applied in data management applications. Our analyses methods can be applied to test similar hypotheses, and investigate different aspects of data.

Secondly, CPS^2 has implications for adaptive privacy management approaches. Managing data and privacy in social software is challenging for average users to perform effectively. Various approaches focus on facilitating methods to provide feedback and awareness to users. Such methods facilitate adapting data management and privacy decisions if needed. CPS^2 can be utilised for user feedback and awareness such as notifying users of context changes, privacy attacks, etc. It can also assist users in taking actions based on a potential change in the data sensitivity or context. The framework can offer various adaptive privacy management functionality that can be addressed in the future work discussed in the following.

7.3.1 Future Research

The work in this thesis is mainly exploratory and requires further future work to impact users' privacy management experience. This work requires further research to explore how users perceive contextual privacy, and how our framework can be realised, as we discuss in the following.

Firstly, to complement our investigative analyses there is a need to conduct user studies to investigate sensitivity and privacy management patterns. It is vital to verify the relevance of the inferred patterns to users. Although the inferred patterns are based on actual data that describe users' behaviour, however, the patterns do not capture users' motivations. The patterns are

inferred with the assumption that users take actions motivated by managing their sensitive data, and their privacy. With user studies, it would be possible to verify this assumption. User studies may provide information that could affect how patterns are inferred.

Secondly, other complementary research is validating our conceptualisation of contextual privacy through user studies to test its effectiveness. Usually, privacy management approaches are developed without prior user validation. However, the novelty of our framework requires validating its relevance to users. The validation concerns investigating the role of appropriateness and interpretation in reasoning about privacy. The validation should investigate how data can be labelled as *appropriate* or *inappropriate* in different human-human communication contexts. The validation approach could be similar to membership categorisation analysis in sociology. This analysis investigates how people practice assigning members to categories to understand how people behave in social situations [87]. The validation aims to analyse the membership assignment pattern to these categories: 'appropriate dissemination', 'inappropriate dissemination', 'appropriate interpretation', 'inappropriate interpretation', 'privacy violation'. The validation could be performed as a crowd-sourced study to facilitate the collection of a large amount of data. The collected data can be further analysed to infer what context changes affect the appropriateness of data, and what changes could affect the sensitivity of data resulting in a violation. This information could be utilised further to implement the framework in social software. The implemented framework can be then tested with real users to manage their actual data.

Thirdly, the relation between privacy and surveillance need to be further investigated. Besides adopting our proposed criteria to assess the degree of privacy and surveillance, the inevitability of surveillance in privacy management approaches requires further research to reduce or minimise surveillance. It also requires investigating methods to quantify the degree of privacy and surveillance of a particular approach. The investigation should explore the appropriate way to present such information to users. Presenting this information to users is not enough to inform them about the possible risks if users are not aware how to interpret this information. The information should convey what it means for a privacy mechanism to have a particular degree of surveillance, and what the possible consequences are. Conducting such a research may require an interdisciplinary approach between computer science, HCI, media and communication studies and social sciences.

Fourthly, a relevant future research path is investigating the incorporation of the CPS^2 framework in social robots. Social robots are new technologies that handle personal and sensitive data during communication with humans. Beside

social software, CPS^2 can be adopted in social robots to control how sensitive data can be disseminated by the robot. The framework is based on utilising artificial intelligence mechanisms. Social robots platforms offer various artificial intelligence mechanisms for context and interpretation inference for communication. The social robot platform is an appropriate platform to incorporate CPS^2 and enhance robots reasoning about privacy. However, such an incorporation requires investigating the notions of *appropriate* and *sensitive* data in human-robot interaction. This investigation can be validated in a similar approach to the human-human interaction setting discussed above.

Finally, the broad future research path of this thesis is developing a holistic privacy management approach based on sensitivity, context, subjectivity, and a particular attacker model. This thesis investigated how these parameters affect contextual privacy management. Future research could investigate how to incorporate these parameters in one approach. Such an approach could, for instance, infer the sensitivity of data from a general pattern and then adapt based on the learned subjectivity of individual users. Changes of context and data misappropriation attacks could be detected by the approach, and the user could be notified accordingly.

Appendix A

The Effect of Context on Disclosure Patterns

This appendix contains heat map plots that show the results of the post-hoc analysis of the effect of context on disclosure patterns (Chapter 3).

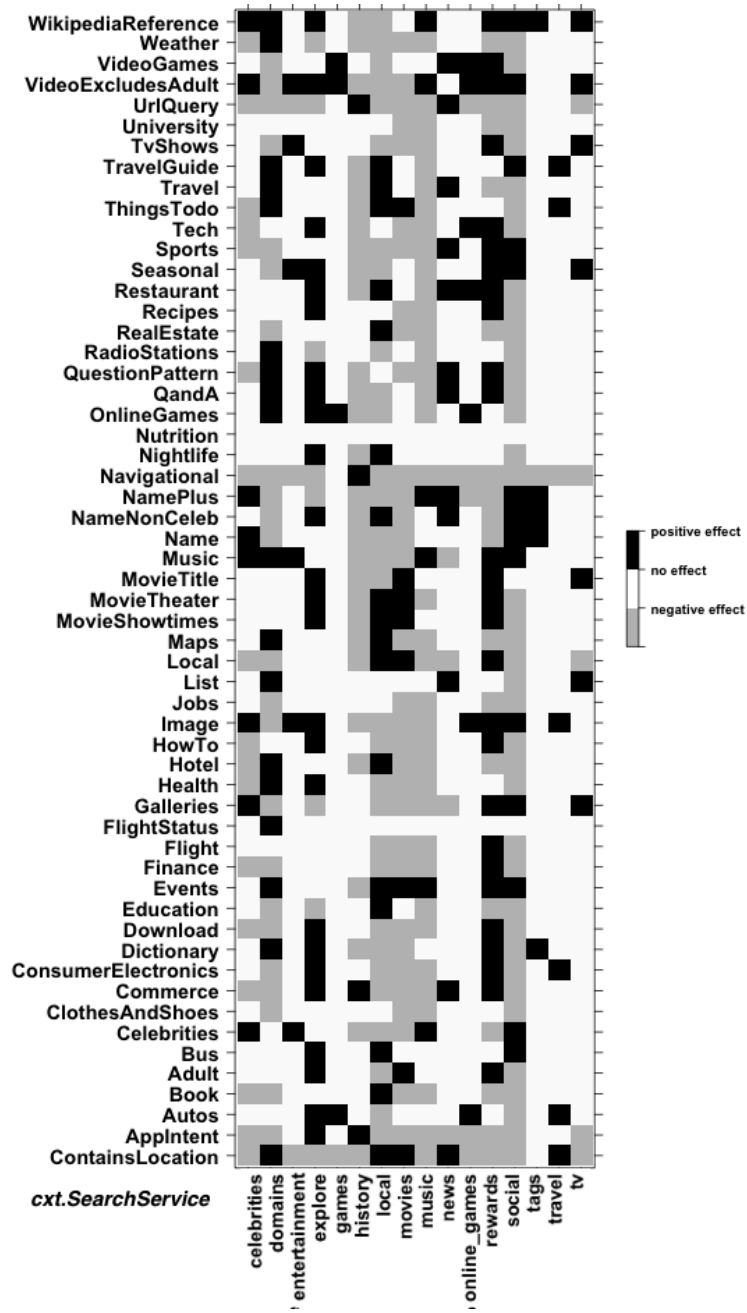


Figure A.1: The heat map of *cxt.SearchService*.

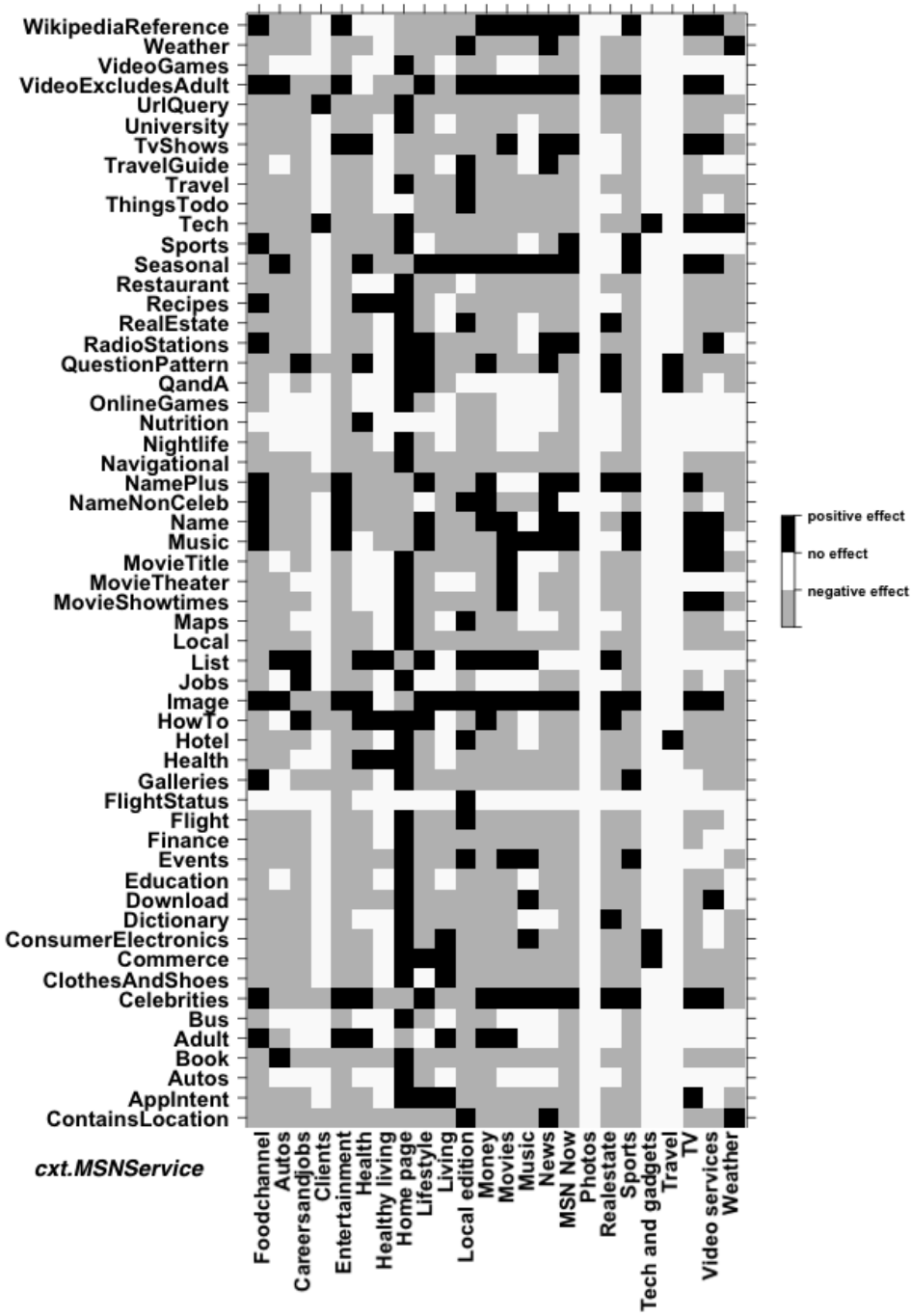


Figure A.2: The heat map of *cxt.MSNService*.

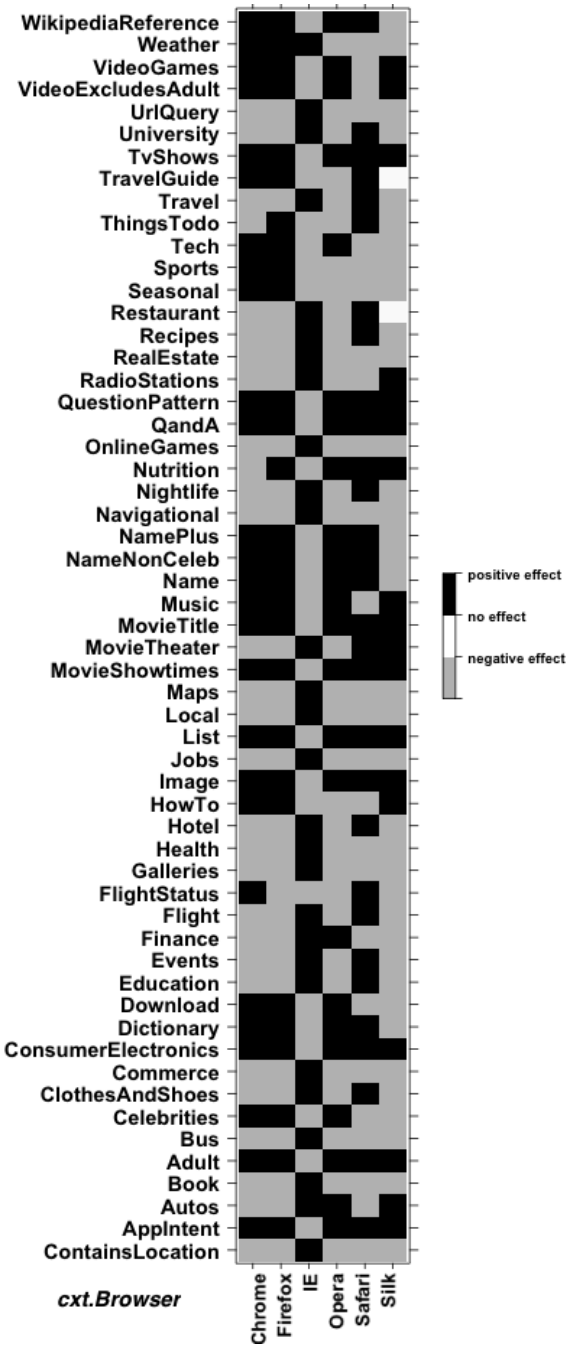


Figure A.3: The heat map of *cxt.Browser* .

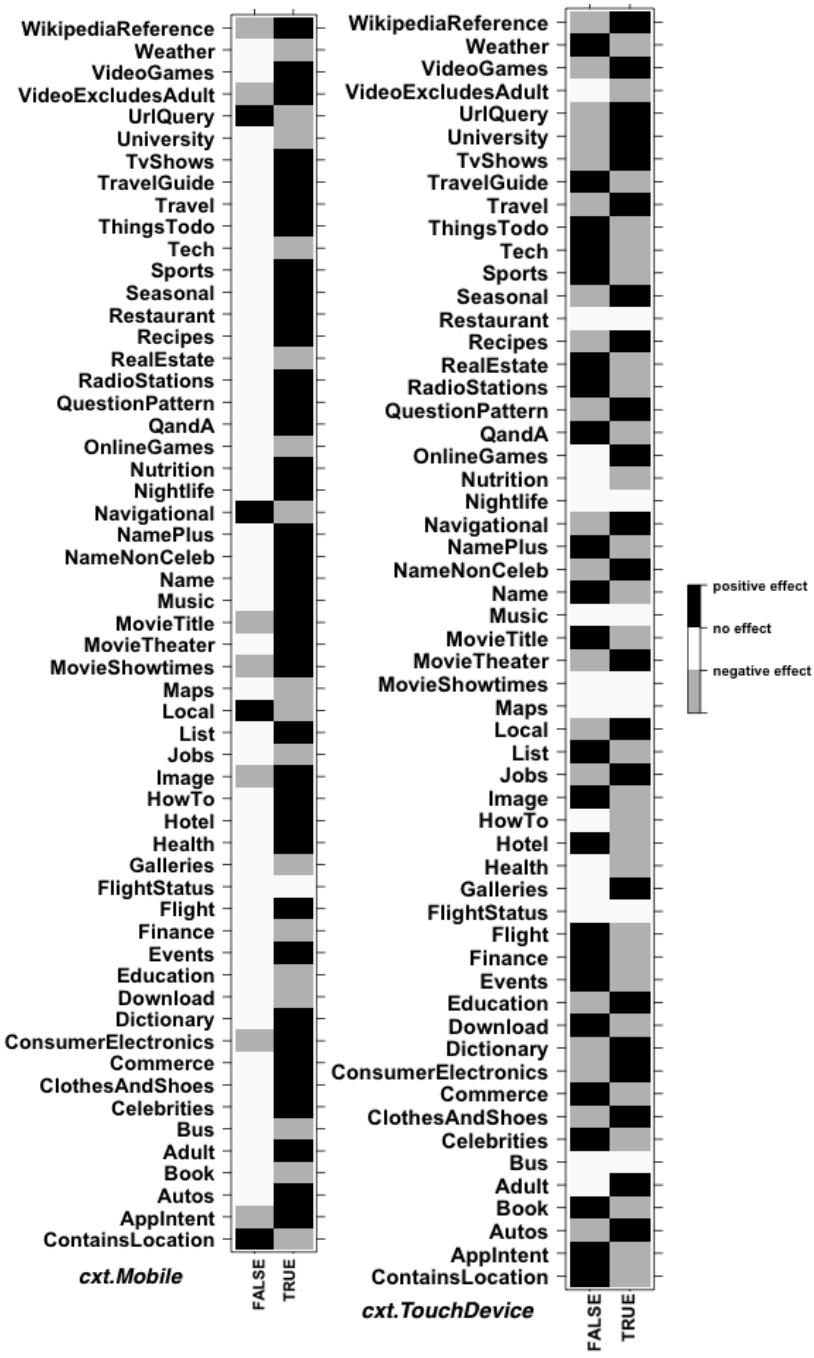


Figure A.4: The heat maps of *cxt.DeviceClass* and *cxt.TouchDevice*.

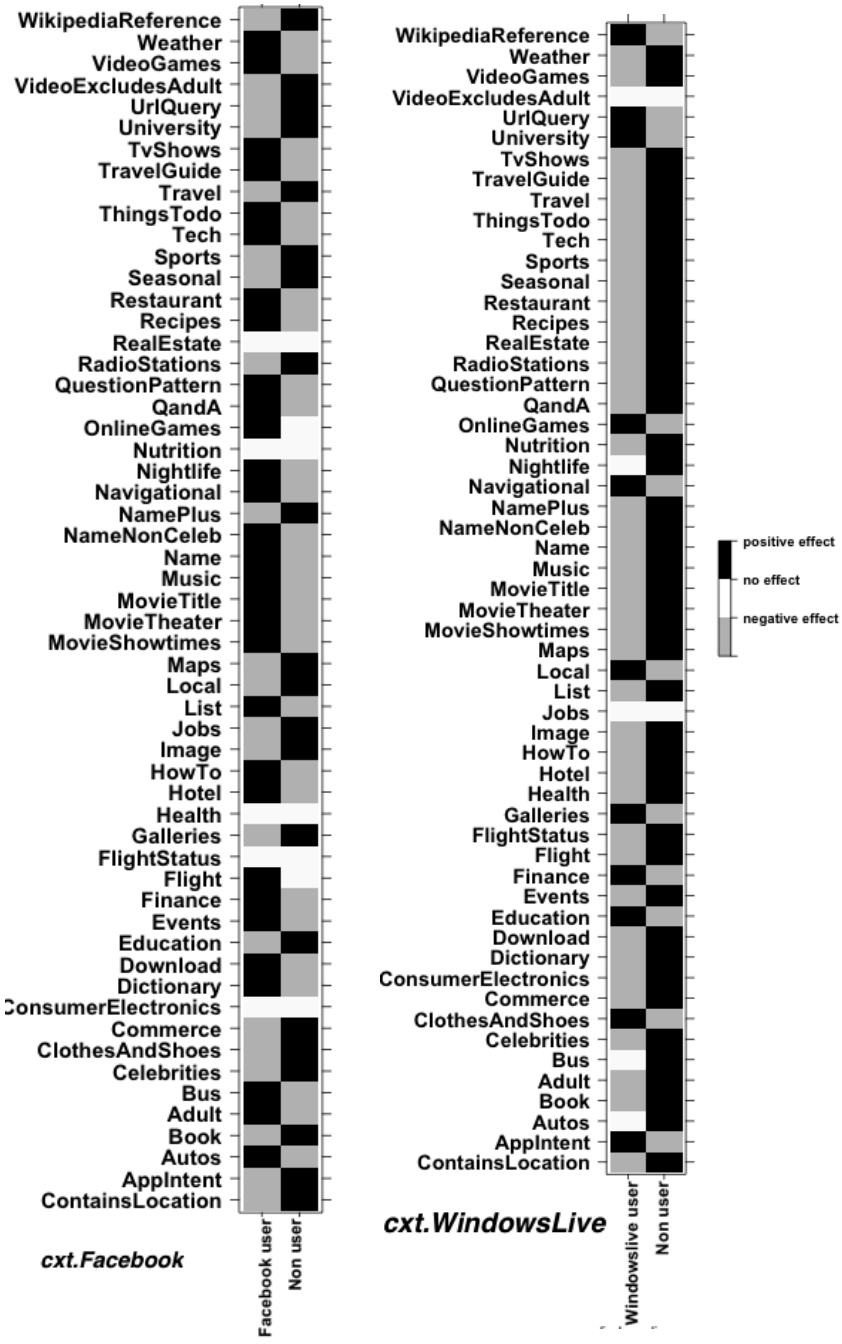


Figure A.5: The heat maps of *cxt.Facebook* and *cxt.WindowsLive*.

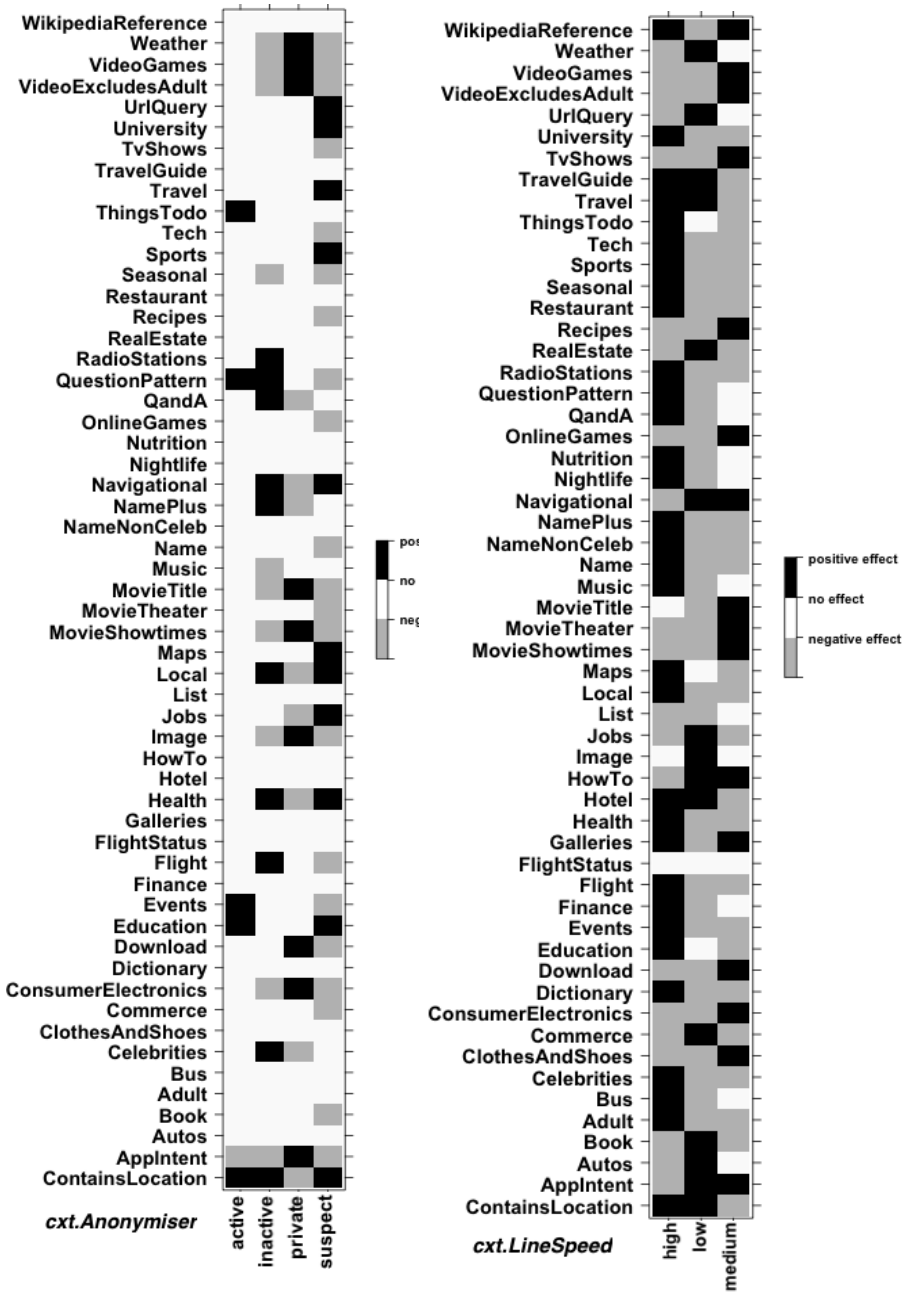


Figure A.6: The heat maps of *cxt.AnonymiserStatus* and *cxt.LineSpeed*.

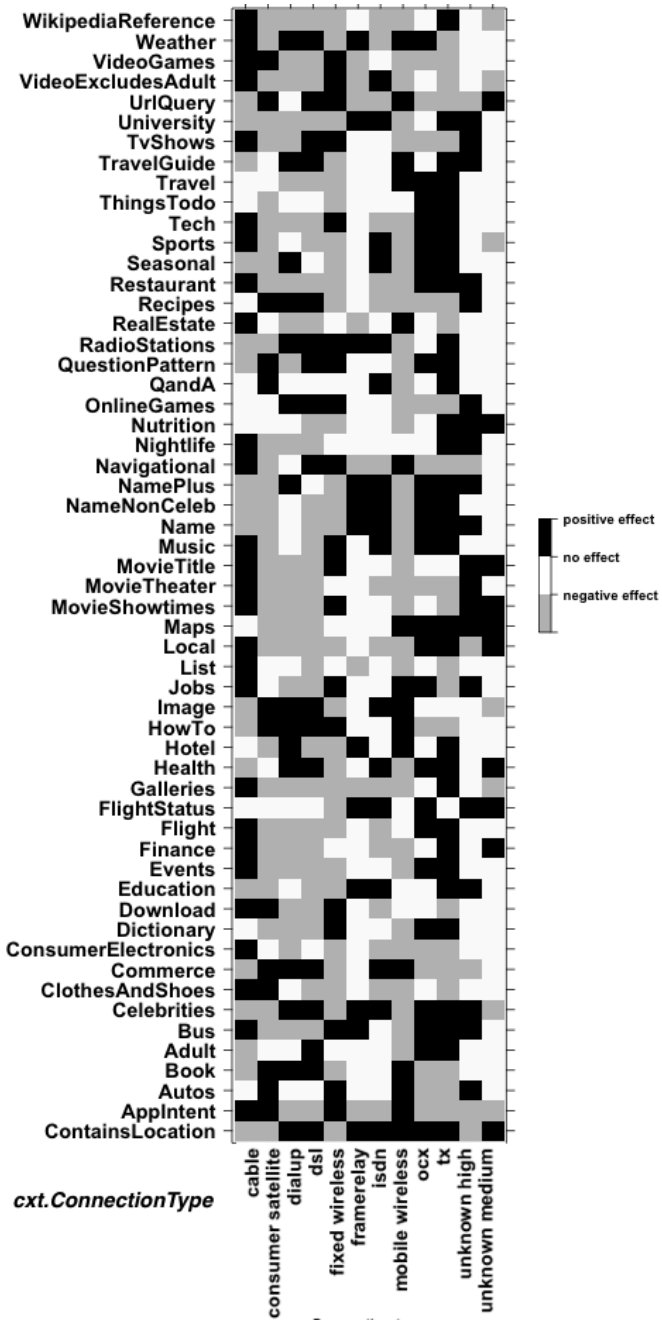


Figure A.7: The heat map of *cxt.ConnectionType*.

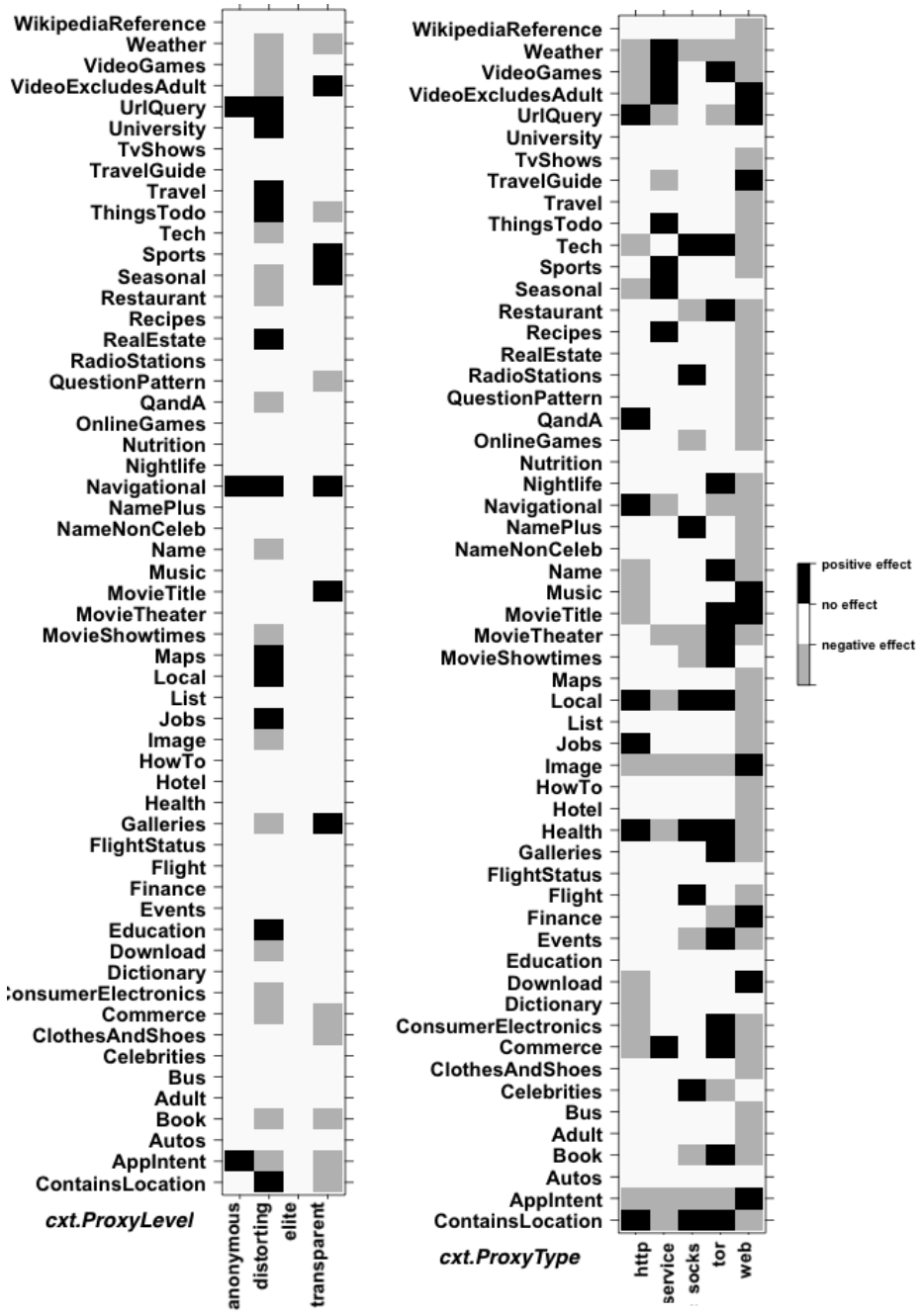


Figure A.8: The heat maps of *cxt.ProxyLevel* and *cxt.ProxyType*.

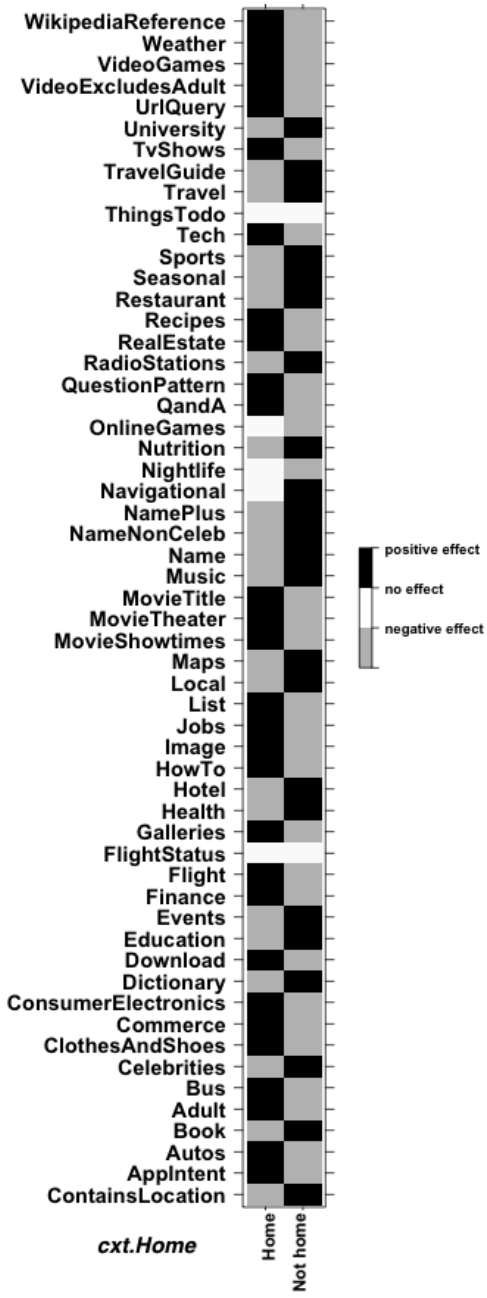


Figure A.9: The heat map of *cxt.Home*.

Figure A.10: The heat map of *cxt.OrganisationType*.

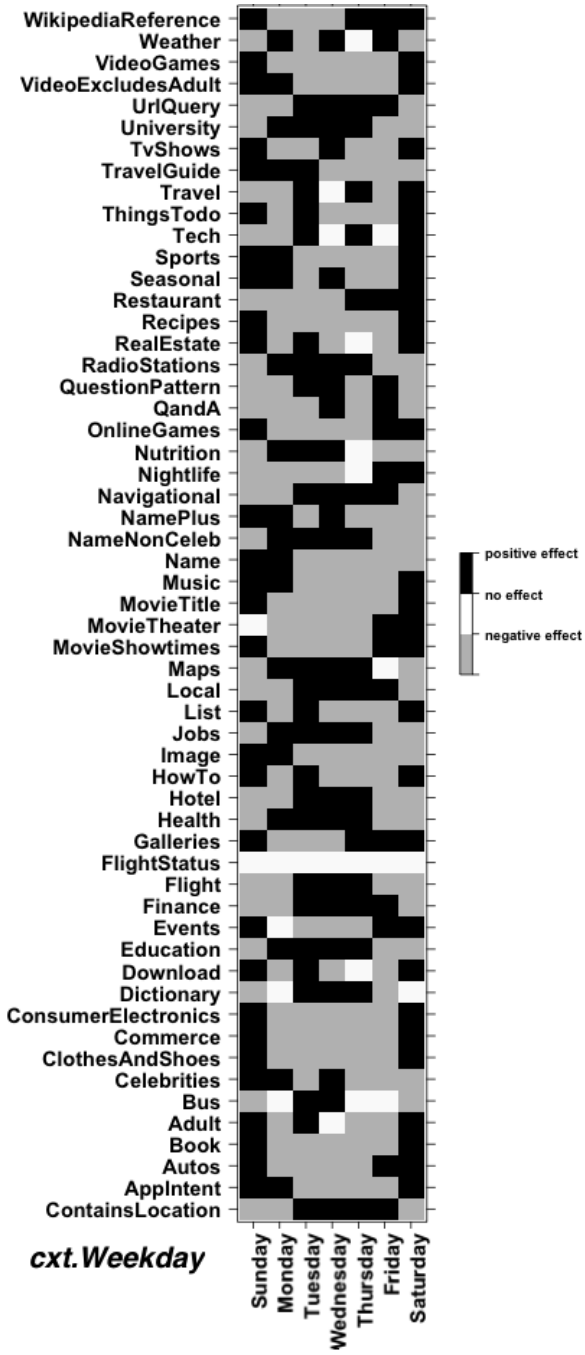


Figure A.11: The heat map of *cxt.Weekday* context variations on the disclosure patterns.

Bibliography

[1] ABOWD, G. D., DEY, A. K., BROWN, P. J., DAVIES, N., SMITH, M., AND STEGGLES, P. Towards a better understanding of context and context-awareness. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing* (1999), Springer-Verlag, pp. 304–307. pages 89

[2] ACQUISTI, A., BALSA, E., BERENDT, B., CLARKE, D., DE GROEF, W., DE WOLF, R., DIAZ, C., GAO, B., GÜRSES, S., PIERSON, J., PIESSENS, F., SAYAF, R., SCHELLENS, T., STUTZMAN, F., VAN ALSENOY, B., AND VANDERHOVEN, E. SPION project deliverable. D2.1-state of the art, 2010. pages 6, 123

[3] ACQUISTI, A., BALSA, E., BERENDT, B., CLARKE, D., DE GROEF, W., DE WOLF, R., DIAZ, C., GAO, B., GÜRSES, S., PIERSON, J., PIESSENS, F., SAYAF, R., SCHELLENS, T., STUTZMAN, F., VAN ALSENOY, B., AND VANDERHOVEN, E. SPION Project Deliverable 2.2–Requirements and Conceptual framework. Tech. rep., KULeuven, 2011. pages 111

[4] AGRE, P. E. The architecture of identity: Embedding privacy in market institutions. *Information, Communication & Society 2*, 1 (1999), 1–25. pages 97

[5] AGRESTI, A. *An introduction to categorical data analysis.* NY: Wiley, 2007. pages 9, 60, 61

[6] AJAM, N., CUPPENS-BOULAHIA, N., AND CUPPENS, F. Contextual privacy management in extended role based access control model. In *Proceedings of the 4th workshop, and 2d conference on Data Privacy Management and Autonomous Spontaneous Security* (2010), Springer, pp. 121–135. pages 87

[7] AKMAN, V. Rethinking context as a social construct. *Journal of Pragmatics 32*, 6 (2000), 743–759. pages 90

[8] ALHADEFF, J., VAN ALSENOY, B., AND DUMORTIER, J. The accountability principle in data protection regulation: Origin, development and future directions. In *Managing privacy through accountability*, D. Guagnin, L. Hempel, C. Ilten, I. Kroener, D. Neyland, and H. Postigo, Eds. Palgrave Macmillan, Basingstoke, UK, 2012. pages 123

[9] ANALYTI, A., THEODORAKIS, M., SPYRATOS, N., AND CONSTANTOPOULOS, P. Contextualization as an independent abstraction mechanism for conceptual modeling. In *Information Systems* (2007), vol. 32, Elsevier, pp. 24–60. pages 90

[10] AREL, I., ROSE, D. C., AND KARNOWSKI, T. P. Deep machine learning-a new frontier in artificial intelligence research. *Computational Intelligence Magazine, IEEE 5*, 4 (2010), 13–18. pages 106

[11] BAKER, M. Contextualization in translator-and interpreter-mediated events. *Journal of Pragmatics 38*, 3 (2006), 321–337. pages 90

[12] BARNES, S. B. A privacy paradox: Social networking in the united states. *First Monday 11*, 9 (2006). pages 121

[13] BARTH, A., DATTA, A., MITCHELL, J. C., AND NISSENBAUM, H. Privacy and contextual integrity: Framework and applications. In *IEEE S& P'6* (2006), IEEE Computer Society, pp. 184–198. pages 87, 109, 116

[14] BENTLER, P. M., AND BONETT, D. G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin 88*, 3 (1980), 588. pages 59

[15] BISHOP, C. M. *Pattern recognition and machine learning*. Springer, 2006. pages 32

[16] BLAND, J. M., AND ALTMAN, D. G. The odds ratio. *BMJ 320*, 7247 (2000), 1468. pages 28

[17] BOYD, D. *Taken out of context: American teen sociality in networked publics*. PhD thesis, PhD Dissertation. University of California-Berkeley, School of Information., 2008. pages 6, 17, 22, 49, 87

[18] BOYD, D. ocial network sites as networked publics: Affordances, dynamics, and implications. *Networked Self: Identity, Community, and Culture on Social Network Sites* (2010), 39–58. pages 88

[19] BOYD, D., AND MARWICK, A. Social steganography: Privacy in networked publics. *International Communication Association, Boston, MA* (2011). pages 93, 114

[20] BOYLES, J. L., SMITH, A., AND MADDEN, M. Privacy and data management on mobile devices. *Pew Internet & American Life Project 4* (2012). pages 38

[21] BRAZ, C., SEFFAH, A., AND M'RAIHI, D. Designing a trade-off between usability and security: a metrics based-model. In *HCI–INTERACT 2007*. Springer, 2007, pp. 114–126. pages 109

[22] BRÉZILLON, P. Context in problem solving: a survey. In *The Knowledge Engineering Review* (1999), vol. 14, Cambridge University Press, pp. 1–34. pages 90

[23] BUVAČ, S. Resolving lexical ambiguity using a formal theory of context. In *Semantic Ambiguity and Underspecification*, K. van Deemter and S. Peters, Eds. Cambridge University Press, Cambridge, England, 1996, pp. 101–124. pages 105

[24] CAO, H., HU, D. H., SHEN, D., JIANG, D., SUN, J.-T., CHEN, E., AND YANG, Q. Context-aware query classification. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (2009), ACM, pp. 3–10. pages 104

[25] CARMINATI, B., FERRARI, E., HEATHERLY, R., KANTARCIOGLU, M., AND THURAISINGHAM, B. A semantic web based framework for social network access control. In *Proceedings of the 14th ACM symposium on Access control models and technologies* (New York, NY, USA, 2009), ACM, pp. 177–186. pages 98

[26] CELIKYILMAZ, A., HAKKANI-TUR, D., AND TUR, G. Statistical semantic interpretation modeling for spoken language understanding with enriched semantic features. In *Spoken Language Technology Workshop (SLT), 2012 IEEE* (2012), IEEE, pp. 216–221. pages 104

[27] CHAIKEN, R., JENKINS, B., LARSON, P., RAMSEY, B., SHAKIB, D., WEAVER, S., AND ZHOU, J. Scope: easy and efficient parallel processing of massive data sets. *PVLDB 1*, 2 (2008), 1265–1276. pages 31

[28] CHAWLA, N. V., JAPKOWICZ, N., AND KOTCZ, A. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter 6*, 1 (2004), 1–6. pages 31

[29] CLEMENT, A. Considering privacy in the development of multi-media communications. *Computer Supported Cooperative Work 2*, 1-2 (1993), 67–88. pages 133

[30] COMMISSION, E. Article 29 data protection working party, opinion 5/2009 on online social networking. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2009/wp163_en.pdf, 2009. pages 123

[31] CRANOR, L. F. *Security and usability: designing secure systems that people can use*. " O'Reilly Media, Inc.", 2005. pages 112

[32] CUTILLO, L., MOLVA, R., AND STRUFE, T. Safebook: A privacy-preserving online social network leveraging on real-life trust. *Communications Magazine, IEEE 47*, 12 (dec. 2009), 94 –101. pages 122

[33] DANEZIS, G., AND GÜRSES, S. A critical review of 10 years of privacy technology. In *Surveillance Clutures: A Global Surveillance Society* (2010). pages 117, 133, 134

[34] DE WOLF, R. Over 'spotted', 'hoeren' en 'failed'-pagina's. Electronic article: http://www.knack.be/nieuws/belgie/dader-antwerpse-hoeren-foto-geklist/article-4000230766578.htm, Last checked Feb. 2013, 2013. pages 113

[35] DINGLEDINE, R., MATHEWSON, N., AND SYVERSON, P. Tor: The second-generation onion router. In *USENIX Security Symposium* (2004), pp. 303—320. pages 134

[36] DYNEL, M. There is method in the humorous speaker's madness: Humour and grice's model. *Lodz Papers in Pragmatics 4*, 1 (2008), 159–185. pages 93

[37] EDELMANN, R. J. Embarrassment: The state of research. *Current Psychological Reviews 1*, 2 (1981), 125–137. pages 83

[38] EU DIRECTIVE. Directive 95/46/ec of the European parliment and of the council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free. *Official Journal L 281*, 23/11 (1995), 0031–0050. pages 2, 48, 143

[39] FONG, P. W. L. Relationship-based access control: protection model and policy language. In *Proceedings of the first ACM conference on Data and application security and privacy* (New York, NY, USA, 2011), CODASPY 11, ACM, pp. 191–202. pages 115

[40] GAO, B., BERENDT, B., CLARKE, D., DE WOLF, R., PEETZ, T., PIERSON, J., AND SAYAF, R. Interactive grouping of friends in osn: Towards online context management. *International Workshop on Privacy in Social Data (PinSoDa)* (2012). pages 116

[41] GERSHENSON, C. Contextuality: A philosophical paradigm, with applications to philosophy of cognitive science. In *in Artificial Life IX* (2002), Citeseer. pages 4

[42] GOFFMAN, E. The presentation of self in everyday life. *Garden City, NY* (1959). pages 92

[43] GOLDIE, J. Virtual communities and the social dimension of privacy. *University of Ottawa Law & Technology Journal 3*, 1 (2003), 133–167. pages 98

[44] GOTTLOB, G., GRECO, G., AND MANCINI, T. Complexity of pure equilibria in bayesian games. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence* (San Francisco, CA, USA, 2007), IJCAI'07, Morgan Kaufmann Publishers Inc., pp. 1294–1299. pages 5

[45] GRICE, H. P. Logic and conversation. In *The Logic of Grammar*, D. Davidson and G. Harman, Eds. Harvard Univ., 1975, pp. 64–75. pages 92

[46] GU, Q., CAI, Z., ZHU, L., AND HUANG, B. Data mining on imbalanced data sets. In *Advanced Computer Theory and Engineering, 2008. ICACTE'08. International Conference on* (2008), IEEE, pp. 1020–1024. pages 31

[47] GUREVICH, Y., E., H., AND WING, J. Inverse privacy. Tech. rep., Microsoft Research, 2014. pages 135

[48] GÜRSES, S. *Multilateral privacy requirements analysis in online social network services*. PhD thesis, KU Leuven, 2010. pages 3, 17, 86

[49] GÜRSES, S. *Multilateral Privacy Requirements Analysis in Online Social Network Services*. PhD thesis, KU Leuven, 2010. pages 83, 97, 98, 117, 119, 120, 124, 133

[50] GÜRSES, S., AND BERENDT, B. PETS in the surveillance society: A critical review of the potentials and limitations of the privacy as confidentiality paradigm. In *Data Protection in a Profiled World*. Springer, 2010, pp. 301–321. pages 7, 8, 18, 121

[51] GÜRSES, S., AND DIAZ, C. Two tales of privacy in online social networks. *IEEE Security & Privacy 11*, 3 (2013), 29–37. pages 102, 126, 127, 134

[52] HAEBERLEN, A., AND KOUZNETSOV, P. Peerreview: practical accountability for distributed systems. In *Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles* (NY, USA, 2007), SOSP '07, ACM, pp. 175–188. pages 123

[53] HARPER, R., Ed. *Trust, Computing and Society*. CUP: New York, 2014. pages x, 9, 48, 91, 92, 93, 118

[54] HINTON, G., OSINDERO, S., AND TEH, Y.-W. A fast learning algorithm for deep belief nets. *Neural computation 18*, 7 (2006), 1527–1554. pages 106

[55] KAPLOWITZ, M. D. Statistical analysis of sensitive topics in group and individual interviews. *Quality and Quantity 34*, 4 (2000), 419–431. pages 19, 49

[56] KARJOTH, G., SCHUNTER, M., AND WAIDNER, M. Platform for enterprise privacy practices: Privacy-enabled management of customer data. In *Privacy Enhancing Technologies* (2003), Springer, pp. 69–84. pages 126

[57] KATZ, J. E., AND RICE, R. E. *Social consequences of Internet use: Access, involvement, and interaction*. MIT press Cambridge, MA, 2002. pages 133

[58] KELLEY, P. G., BREWER, R., MAYER, Y., CRANOR, L. F., AND SADEH, N. An investigation into facebook friend grouping. In *HCI–INTERACT 2011*. Springer, 2011, pp. 216–233. pages 116

[59] KRUPA, Y., AND VERCOUTER, L. Handling privacy as contextual integrity in decentralized virtual communities: The privacias framework. *Web Intelligence and Agent Systems* (2012). pages 109, 116

[60] LAMPINEN, A., LEHTINEN, V., LEHMUSKALLIO, A., AND TAMMINEN, S. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 3217–3226. pages 87, 92, 93, 94, 114

[61] LAW, E. U. Eu directive 2016/680. Electronic article: http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf, May 2016. pages 131

[62] LAZARSFELD, P. F., HENRY, N. W., AND ANDERSON, T. W. *Latent structure analysis*. Houghton Mifflin Boston, 1968. pages 8

[63] LEE, R. M., AND RENZETTI, C. M. The problems of researching sensitive topics: An overview and introduction. *American behavioral scientist 33*, 5 (1990), 510–28. pages 20, 49

[64] LENAT, D. B., GUHA, R. V., PITTMAN, K., PRATT, D., AND SHEPHERD, M. Cyc: toward programs with common sense. *Commun. ACM 33*, 8 (Aug. 1990), 30–49. pages 105

[65] LIPFORD, H. R., BESMER, A., AND WATSON, J. Understanding privacy settings in facebook with an audience view. In *Proceedings of the 1st Conference on Usability, Psychology, and Security* (Berkeley, CA, USA, 2008), USENIX Association, pp. 2:1–2:8. pages 116

[66] MAJESKI, M., JOHNSON, M., AND BELLOVIN, S. M. The Failure of Online Social Network Privacy Settings. Tech. Rep. CUCS-010-11, CS, Columbia University, 2011. pages 4

[67] MCCARTHY, J. Formalizing context (expanded notes). In *Computing Natural Language*, A. Aliseda, R. J. van Glabbeek, and D. Westerstahl, Eds. CSLI Publications, 1993, pp. 13–50. pages 91, 106

[68] MCCULLAGH, P., AND NELDER, J. A. *Generalized linear models*, 2nd ed. Chapman and Hall, London, 1989. pages 32

[69] MCCULLOH, I. *Detecting changes in a dynamic social network*. PhD thesis, Carnegie Mellon University, 2009. pages 85

[70] MCGRATH, J. *Loving Big Brother: Performance, privacy and surveillance space*. Psychology Press, 2004. pages 133, 134

[71] MEYROWITZ, J. *No sense of place: The impact of electronic media on social behavior*. Oxford University Press New York, 1985. pages 88, 95

[72] MICROSOFT. Microsoft privacy statement. Electronic article, 2015. pages 22

[73] MROUE, B. Syria Facebook, YouTube Ban Lifted: Reports. The World Post, Feb. 2011. pages 134

[74] MURTAGH, F., AND LEGENDRE, P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of Classification 31*, 3 (2014), 274–295. pages 62

[75] NISSENBAUM, H. Privacy as contextual integrity. *Wash. L. Rev. 79* (2004), 119. pages 4, 12, 143

[76] NISSENBAUM, H. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law & Politics, 2010. pages 86, 103, 109, 116

[77] ON CENSORSHIP, I. Syria unblocks Facebook and Youtube. Electronic article: http://www.indexoncensorship.org/2011/02/syria-unblocks-facebook-and-youtube/, Feb. 2011. pages 134

[78] PALEN, L., AND DOURISH, P. Unpacking "privacy" for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2003), ACM, pp. 129–136. pages 49, 94, 101

[79] PARK, R., BURGESS, E., AND MCKENZIE, R. The city: Suggestions for the study of human nature in the urban environment. Chicago: University of Chicago Press, 1925. pages 4

[80] PEDERSEN, D. M. Model for types of privacy by privacy functions. *Journal of Environmental Psychology 19*, 4 (1999), 397 – 405. pages 117

[81] PETRONIO, S. *Boundaries of privacy: Dialectics of disclosure*. SUNY Press, 2002. pages 97

[82] PREIBUSCH, S. The value of privacy in web search. In *The Twelfth Workshop on the Economics of Information Security (WEIS)*. 2013. pages 50

[83] PREIBUSCH, S. Privacy behaviors after snowden. *Commun. ACM 58*, 5 (Apr. 2015), 48–55. pages 134

[84] PRESTON, J. Seeking to disrupt protesters, Syria cracks down on social media. Electronic article: `http://www.nytimes.com/2011/05/23/world/middleeast/23facebook.html`, May 2011. pages 134

[85] REN, X., WANG, Y., YU, X., YAN, J., CHEN, Z., AND HAN, J. Heterogeneous graph-based intent learning with queries, web pages and wikipedia concepts. In *WSDM* (2014), pp. 23–32. pages 104

[86] RULE, J. B. When it comes to protecting its citizens' data, Europe is way ahead of the U.S. http://www.latimes.com/opinion/op-ed/la-oe-rule-nsa-privacy-european-union-20140513-story.html, May 2014. pages 121, 123

[87] SACKS, H. *Lectures on conversation*, vol. 1. Blackwell Publishing, 1995. pages 144

[88] SAKAMOTO, Y., ISHIGURO, M., AND KITAGAWA, G. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* (1986). pages 73

[89] SALAKHUTDINOV, R., AND HINTON, G. Deep boltzmann machines. In *Proceedings of the International Conference on Artificial Intelligence and Statistics* (2009), vol. 12. pages 106

[90] SAYAF, R., AND CLARKE, D. Access control models for online social networks. In *Social Network Engineering for Secure Web Data and Services*, L. Caviglione, M. Coccoli, and A. Merlo, Eds. IGI, 2012, pp. 32–65. pages 4, 87, 102, 115, 121, 122, 124, 125

[91] SAYAF, R., CLARKE, D., AND HARPER, R. CPS^2: a contextual privacy framework for social software. In *Proceedings of SECURECOMM'14* (2014), Springer, pp. 25–32. pages 48, 87, 101, 107, 119, 120

[92] SAYAF, R., CLARKE, D., AND RULE, J. B. The other side of privacy: Surveillance in data control. In *Proceedings of the 2015 British HCI Conference* (New York, NY, USA, 2015), British HCI '15, ACM, pp. 184–192. pages 87

[93] SAYAF, R., RULE, J. B., AND CLARKE, D. Can users control their data in social software? an ethical analysis of data control approaches. In *IEEE S& P Workshops (SPW)* (2013), pp. 1–4. pages 105, 115, 118, 120, 121

[94] SCHNEIER, B., AND KELSEY, J. Secure audit logs to support computer forensics. *ACM Trans. Inf. Syst. Secur. 2* (May 1999), 159–176. pages 123

[95] SEFFAH, A., DONYAEE, M., KLINE, R. B., AND PADDA, H. K. Usability measurement and metrics: A consolidated model. *Software Quality Journal 14*, 2 (2006), 159–178. pages 109

[96] SHARPE, D. Your chi-square test is statistically significant: Now what? *Practical Assessment, Research & Evaluation 20* (2015), 1–10. pages 59, 60

[97] SIMPSON, A. On the need for user-defined fine-grained access control policies for social networking applications. *Proceedings of the workshop on Security in Opportunistic and SOCial networks - SOSOC '08* (2008), 1–8. pages 122

[98] SINGH, A., AND LIU, L. Trustme: Anonymous management of trust relationships in decentralized p2p systems. In *Peer-to-Peer Computing* (2003), N. Shahmehri, R. L. Graham, and G. Caronni, Eds., IEEE Computer Society, pp. 142–149. pages 105

[99] SKANTZE, G. *Error Handling in Spoken Dialogue Systems-Managing Uncertainty, Grounding and Miscommunication. Doctoral dissertation, KTH*. PhD thesis, Department of Speech, Music and Hearing, 2007. pages 4

[100] SKYRMS, B. Pragmatics, logic and information processing. In *Language, games, and evolution*. Springer, 2011, pp. 177–187. pages 92, 93

[101] SPERBER, D., AND WILSON, D. *Relevance: communication and cognition*. Harvard University Press, Cambridge, MA, USA, 1986. pages 51, 89

[102] SQUICCIARINI, A. C., SHEHAB, M., AND WEDE, J. Privacy policies for shared content in social network sites. *The VLDB Journal—The International Journal on Very Large Data Bases 19*, 6 (2010), 777–796. pages 122

[103] SQUICCIARINI, A. C., AND SUNDARESWARAN, S. Web-traveler policies for images on social networks. *World Wide Web Internet And Web Information Systems 12*, 4 (2009), 461–484. pages 126

[104] SRIVASTAVA, N., AND SALAKHUTDINOV, R. Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems* (2012), pp. 2222–2230. pages 106

[105] STACH, C., AND MITSCHANG, B. Privacy management for mobile platforms–a review of concepts and approaches. In *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on* (2013), vol. 1, IEEE, pp. 305–313. pages 38

[106] TUFEKCI, Z. Facebook, youth and privacy in networked publics. In *ICWSM* (2012), J. G. Breslin, N. B. Ellison, J. G. Shanahan, and Z. Tufekci, Eds., The AAAI Press. pages 17, 22, 50

[107] TURNER, R. M. A model of explicit context representation and use for intelligent agents. In *Proceedings of the Second International and Interdisciplinary Conference on Modeling and Using Context* (London, UK, UK, 1999), CONTEXT '99, Springer-Verlag, pp. 375–388. pages 89

[108] VAN DIJK, T. A. Context models in discourse processing. *The construction of mental representations during reading* (1999), 123–148. pages 4

[109] VAN DIJK, T. A. Discourse and context. *A Sociocognitive Approach, Cambridge University* (2008). pages 4, 86

[110] VERBRUGGE, R., AND MOL, L. Learning to apply theory of mind. *Journal of Logic, Language and Information 17*, 4 (2008), 489–511. pages 93

[111] VOJNOVIC, M., XU, F., AND ZHOU, J. Sampling based range partition methods for big data analytics. Tech. rep., Microsoft Research, 2012. pages 31

[112] WANG, C., AND LEUNG, H.-F. A secure and private clarke tax voting protocol without trusted authorities. In *Proceedings of the 6th international conference on Electronic commerce* (New York, NY, USA, 2004), ICEC '04, ACM, pp. 556–565. pages 125

[113] WANG, Y., NORCIE, G., KOMANDURI, S., ACQUISTI, A., LEON, P. G., AND CRANOR, L. F. "i regretted the minute i pressed share": a qualitative study of regrets on facebook. In *SOUPS* (2011), p. 10. pages 17, 22, 49

[114] WARD JR, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association 58*, 301 (1963), 236–244. pages 62

[115] WARREN, J. *Self-imposed violations of privacy in virtual communities*. University of Texas, College of Business, 2008. pages 98

[116] WARREN, S., AND BRANDEIS, L. The right to privacy. *Harward Law Review 4*, 5 (1890), 193—220. pages 117

[117] WIRTH, L. *The ghetto*. Transaction Publishers, 1928. pages 4

[118] WITTKOWER, D. *Facebook and Philosophy: What's on Your Mind?* Open Court Pub Co, 2010. pages 98

[119] WOOD, A. F., AND SMITH, M. J. *Online communication: Linking technology, identity, & culture*. Routledge, 2004. pages 94, 97

Publications

International Conference Papers

- Rula Sayaf, Sören Preibusch, and Dave Clarke. Contextual healing: privacy through interpretation management. In Proceedings of the 8th IEEE International Conference on Social Computing, IEEE Computer Society 2015.
- Rula Sayaf, Dave Clarke, and James B. Rule. The other side of privacy: Surveillance in data control. In Proceedings of the 2015 British HCI Conference (New York, USA, 2015), British HCI 2015, ACM, pp.184–192
- Rula Sayaf, Dave Clarke, and Richard Harper. CPS^2: a contextual privacy framework for social software. In SECURECOMM 2014 (2014), Springer, pp. 25–32.
- Rula Sayaf, James B. Rule, and Dave Clarke. Can users control their data in social software? an ethical analysis of data control approaches. In IEEE S&P Workshops (SPW) 2013, pp. 1–4.

Book Chapter

- Rula Sayaf, and Dave Clarke. Access control models for online social networks. In Social Network Engineering for Secure Web Data and Services, L. Caviglione, M. Coccoli, and A. Merlo, Eds. IGI, 2012, pp. 32–65.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
DISTRINET
Celestijnenlaan 200A box 2402
B-3001 Leuven
first.last@cs.kuleuven.be