

Alma Mater Studiorum – Università di Bologna

In collaborazione con LAST-JD consortium:

Università degli studi di Torino

Universitat Autònoma de Barcelona

Mykolas Romeris University

Tilburg University

e in cotutela con

THE University of Luxembourg

DOTTORATO DI RICERCA IN

**Erasmus Mundus Joint International Doctoral Degree in Law,
Science and Technology**

Ciclo 29 – A.Y. 2013/2014

Settore Concorsuale di appartenenza: 12H3

Settore Scientifico disciplinare: IUS20

**Contextual Integrity and Tie Strength in Online Social
Networks: Social Theory, User Study, Ontology, and Validation**

Presentata da: Javed Ahmed

**Coordinatore
Prof. Monica Palmirani**

**Relatore
Leon van der Torre
Antonino Rotolo**

**Co- Relatore
Guido Governatori
Serena Villata**

Esame finale anno 2017

Alma Mater Studiorum – Università di Bologna
in partnership with LAST-JD Consortium
Università degli studi di Torino
Universitat Autònoma de Barcelona
Mykolas Romeris University
Tilburg University
and in cotutorship with the
THE University of Luxembourg

PhD Programme in

Erasmus Mundus Joint International Doctoral Degree in Law,
Science and Technology
Cycle 29 – A.Y. 2013/14

Settore Concorsuale di appartenenza: 12H3

Settore Scientifico disciplinare: IUS20

**Contextual Integrity and Tie Strength in Online Social
Networks: Social Theory, User Study, Ontology, and Validation**

Submitted by: Javed Ahmed

The PhD Programme Coordinator
Prof. Monica Palmirani

Supervisor (s)
Leon van der Torre
Antonino Rotolo

Co-Supervisor (s)
Guido Governatori
Serena Villata

Year 2017

DISSERTATION

Defence held on 29/09/2017 in Bologna

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN INFORMATIQUE

AND

DOTTORATO DI RICERCA
IN LAW, SCIENCE AND TECHNOLOGY

by

Javed AHMED

Born on 10th October 1979 in Ghotki (Pakistan)

CONTEXTUAL INTEGRITY AND TIE STRENGTH IN ONLINE
SOCIAL NETWORKS: SOCIAL THEORY, USER STUDY,
ONTOLOGY, AND VALIDATION

Dissertation defence committee

Dr Leon van der Torre, dissertation supervisor
Professor, Université du Luxembourg

Dr Antonino Rotolo
Professor, Università degli Studi di Bologna

Dr Burkhard Schäffer, Chairman
Professor, University of Edinburgh

Dr Adrian Paschke
Professor, Freie Universität Berlin

Dr Harko Verhagen, Vice Chairman
Professor, Stockholm University

Acknowledgement

First of all, I am thankful to LAST-JD consortium and European Commission for funding and support. I have been fortunate to win prestigious Erasmus Mundus doctoral fellowship that helped me to complete my interdisciplinary doctoral degree from four different European Universities. I would like to take this opportunity to express my heartfelt gratitude to my main supervisor professor Dr. Leon van der Torre for his invaluable guidance and support throughout the long and complicated journey. In addition to his academic guidance, he has shown a great deal of patience towards me and supported my research in a way that goes beyond a professional relationship. One such instance is that he hosted me for several days at his home during my mobility to Luxembourg. His caring attitude has always given me the strength to continue my research endeavour.

I have been very lucky to have co-supervisors like professor Dr. Guido Governatori and Dr. Serena Villata. They provided constructive feedback and kept me focused on my goals. They have been a reliable source of sound advice throughout my doctoral studies. It would have been difficult to navigate the challenges of PhD process without their guidance. This PhD has been one of the most challenging experiences of my academic life and it was not possible to finish this journey without the extensive and unconditional support of LAST-JD administration especially professor Dr. Monica Palmirani, professor Dr. Guido Boella and Dr. Dina Ferrari.

Furthermore, I would like to thank all coordinators from consortium partner universities and their staff for their proactive administrative and academic support to make our mobility plan successful during LAST-JD PhD program. I spent multiple brainstorming sessions with Dr. Silvio Peroni and would like to thank him for his kindness. Thanks to all colleagues who helped me unconditionally to complete this doctoral program, in particular, Livio Robaldo, Cedric Pruski, Aldo Gangemi, Samir Ouchani, and Pouyan Ziafati.

Finally, I am deeply grateful to my family and Sukkur IBA for supporting me during this long journey. I was blessed with a lovely daughter in the middle of the journey and this dissertation is dedicated to her. I had very little chance to spend time with her while she was growing older, but I missed her every moment during my stay in Europe. I thank my parents for their unfaltering love and trust, and my brother for inspiring me by his perseverance and courage. Words cannot express my feelings of gratefulness to my family for all their love, kindness, support and prayers that sustained me thus far.

Abstract

Online Social Networks (OSNs) have become an important part of daily digital interactions for more than half billion users around the world. Unconstrained by physical spaces, OSNs offer to social web users new means to communicate, interact, and socialize. Online social networks exhibit many of the characteristics of human societies in terms of forming relationships and sharing personal information. However, current OSNs mainly assume binary, static, and symmetric relationship of equal value between the connected users. In human societies, social relationships are of varying tie strength, dynamic, and asymmetric in nature. The lack of an effective mechanism to represent diversity in social relationships leads to undesirable consequences of users personal information leakage to the unwanted audience and raises privacy concerns. The issue of privacy has received significant attention in both the research literature and the mainstream media.

In this dissertation, we conduct a user study to analyze users' attitude towards personal information disclosure in online social networks. The study gives insight into user's information sharing behavior and interaction patterns in online social networks. The findings reveal that personal information disclosure depends on the quality of relationship among the users and it can be easily inferred from user interaction pattern in online social networks.

We propose a theoretical framework that addresses the aforementioned issue from a social science perspective and exploits existing social theories of Goffman, Granovetter, and Nissenbaum to model social privacy for OSNs users. Based on this theoretical framework, we developed SOCPRI (SOCial PRiVacy) ontology to represent diversity in social relationships in online social networks. This model regulates personal information disclosure on the basis of the social role and the relationship quality between the OSNs users. The model is evaluated by translating competency questions into description logic (DL) queries to demonstrate the applicability of our approach. The results of ontology evaluation demonstrate the appropriateness of our ontology against proposed requirements. Based on this model a privacy-friendly online social networking environment can be developed to address some of the existing issues such as context collapse and user control.

Contents

1	Introduction	1
1.1	Introduction	1
1.1.1	Background	2
1.1.2	Motivation and Problem Description	5
1.1.3	Research Questions and Objectives	8
1.2	Research Methodology	9
1.3	Our Contribution	11
1.4	Scope and Limitations	13
1.5	Structure of the Dissertation	14
1.5.1	Background and Related Literature	14
1.5.2	Privacy: A Theoretical Framework for OSNs	15
1.5.3	Privacy: OSNs User Perspective	15
1.5.4	Ontology Based Privacy Modeling for OSNs	16
1.5.5	Evaluation of SOCPRI Ontology	16
1.5.6	Conclusion and Future Work	16
2	Background and Related Literature	17
2.1	Social Networks	17
2.1.1	Difference between Online & Offline Social Networks	19
2.2	Defining Online Social Networks	21

2.2.1	Evolution of Online Social Networks	24
2.2.2	Features of Online Social Networks	26
2.2.2.1	Personal Space Management	26
2.2.2.2	Social Connection Management and Networking Features	28
2.2.2.3	Communication and Interaction Management	28
2.2.2.4	Social Search Features	29
2.2.2.5	Privacy and Security Features	29
2.2.3	Classification of Online Social Networks	30
2.2.3.1	Personal Online Social Networks	30
2.2.3.2	Professional Online Social Networks	31
2.2.3.3	Interest Oriented Online Social Networks	31
2.2.3.4	Functionality Oriented Online Social Networks	31
2.2.3.5	Connection Oriented Online Social networks	32
2.2.3.6	Content Oriented Online Social Networks	33
2.2.3.7	Context-based Online Social Networks	34
2.2.4	Data Collection by Online Social Networks	35
2.3	Privacy in Online Social Networks	41
2.3.1	Attitude of Users towards Privacy	41
2.3.2	Data Disclosure Method of OSNs	43
2.3.3	Identification of Privacy Threat	44
2.3.3.1	User-related Privacy Threats	45
2.3.3.2	Service Provider-related Privacy Threats	46
2.3.4	Attack Spectrum for OSNs Users	48
2.3.4.1	Stalking	48
2.3.4.2	Digital Dossier Aggregation	49
2.3.4.3	Profile Squatting	49
2.3.4.4	Image Tagging and Cross-profiling	49

2.3.4.5	Phishing	50
2.3.4.6	Spamming	50
2.3.4.7	Cross-site Scripting	50
2.3.4.8	Corporate Espionage	50
2.4	State of Art	51
2.4.1	Semantic Relationship Modeling	51
2.4.2	Relationship Strength Prediction	53
2.4.3	Social Identity Management	55
2.5	Concluding Remarks	56
3	Privacy: A Theoretical Framework for OSNs	59
3.1	Privacy: A Multifaceted Concept	60
3.1.1	Legal Perspective of Privacy	60
3.1.1.1	Current Legal Framework	61
3.1.2	Social Perspective of Privacy	64
3.1.3	Technical Perspective of Privacy	64
3.2	Defining Privacy for OSNs	65
3.3	Shifting Privacy Research Paradigm for OSNs	71
3.3.1	Role and Relationship-based Self-Presentation	72
3.3.2	Realizing Contextual Integrity	73
3.3.3	Modeling Relationship Strength	75
3.4	Proposed Privacy Framework for OSNs	78
3.4.1	Context Segregation	79
3.4.2	Contact Segregation	80
3.4.3	Content Segregation	81
3.5	Concluding Remarks	81

4 Privacy: OSNs User Perspective	83
4.1 Background and Purpose of User Study	84
4.2 User Study Methodology	87
4.2.1 User Study Design	87
4.2.1.1 Survey Content	88
4.2.1.2 Data Collection	89
4.3 Analysis of Results	90
4.3.1 User Attitude Towards Online Privacy	91
4.3.2 Social Relationships Formation in Online Social Networks	94
4.3.3 User Interaction Pattern and Relationship Strength	99
4.3.4 Profile Information and Interactions Ranking	101
4.4 Discussion and Implications	105
4.5 Limitations of User Study	106
4.6 Concluding Remarks	107
5 Ontology Based Privacy Modeling for OSNs	109
5.1 Towards the Social Semantic Web	110
5.1.1 Representing Social Data with Semantic Web	111
5.2 Knowledge Representation and Ontologies	115
5.2.1 Ontologies in Computer Science	116
5.2.2 Languages for Encoding Ontologies	118
5.2.3 Knowledge Representation using OWL DL	120
5.3 Introduction to the SOCPRI Ontology	121
5.3.1 Contribution of the SOCPRI Ontology	127
5.4 Ontology Development Methodology	130
5.4.1 Requirement Specification for SOCPRI Ontology	132
5.4.1.1 Identification of Purpose, Intended Uses and Users	135

5.4.1.2	Motivating Scenarios	136
5.4.1.3	Competency Questions	139
5.4.1.4	Implementation Language	141
5.4.2	Reusing Existing Ontologies	142
5.5	Core Conceptual Elements of SOCPRI Ontology	143
5.5.1	Social Web User Modeling	146
5.5.2	Role and Relationship based Social Context Modeling	156
5.5.3	Social Interaction-based Tie Strength Modeling	163
5.5.4	User Resource Modeling	169
5.5.5	Contextual Privacy Modeling	173
5.6	Defining the Classes and Properties of SOCPRI	178
5.7	Comparative Analysis with Social Web Ontologies	201
5.8	Concluding Remarks	204
6	Evaluation of the SOCPRI Ontology	205
6.1	Overview of Ontology Evaluation Approaches	206
6.2	Validation of SOCPRI Ontology	209
6.2.1	W3C RDF Validation Service	210
6.2.2	Reasoning with SOCPRI Ontology	211
6.3	SOCPRI Evaluation Metrics	213
6.3.1	Concept Hierarchy	214
6.3.2	Property Structure	217
6.3.3	Property Restrictions	220
6.3.4	Domain / Range Definition of Properties	221
6.3.5	Disjointness Restrictions	222
6.3.6	Documentation/Visualization	223
6.3.7	Naming Conventions	224

6.4	Pitfall Scanning of SOCPRI Ontology	226
6.4.1	Ontology Pitfall Scanning Tool	226
6.4.2	Catalogue of Common Pitfalls	227
6.4.3	Evaluation Results for SOCPRI	238
6.4.3.1	Pitfall based Consistency Evaluation	242
6.4.3.2	Pitfall based Completeness Evaluation	243
6.4.3.3	Pitfall based Conciseness Evaluation	244
6.4.3.4	Pitfall based Structural Dimension Evaluation	244
6.4.3.5	Pitfall based Functional Dimension Evaluation	245
6.4.3.6	Pitfall based Usability-Profilng Dimension Evaluation	246
6.5	Evaluation of SOCPRI Using OntoClean	247
6.5.1	OntoClean Meta Properties	247
6.5.2	OntoClean Constraints	249
6.5.3	OntoClean Process	250
6.5.4	Applying OntoClean to SOCPRI Ontology	251
6.6	Query Based Evaluation of SOCPRI Ontology	255
6.7	Concluding Remarks	271
7	Conclusion and Future Work	273
7.1	Summary of Contribution	273
7.2	Research Questions Revisited	276
7.3	Outlook on Further Research	278
	Appendices	281
A	List of Publications	283

List of Figures

1.1	Research Process Summary	12
2.1	Timeline Online Social Networking Sites [1]	25
2.2	Online Social networks as Popular Activity on Internet [2]	27
2.3	OSN User Data Disclosure [3]	40
2.4	Privacy Implication of Data Disclosure in OSNs [4]	41
2.5	Google+ Data Disclosure Method	43
2.6	Facebook Data Disclosure Method	45
2.7	Privacy Threat Matrix [5]	48
4.1	Attitude of OSNs users towards online privacy	92
4.2	User opinion about user friendliness of privacy interface and its usage	93
4.3	Attitude of OSNs users towards formation of online relationships	96
4.4	Attitude of users towards personal information disclosure in OSNs	97
4.5	Relationship strength based interaction patterns of OSNs users	100
4.6	Personal data disclosure attitude and misuse concerns of OSNs users	103
4.7	Ranking of OSN users' profile information and interactions	103
4.8	Disparity between high privacy concerns and data disclosure practice	106
4.9	Disparity between high privacy concerns and online relationship formation	107
5.1	Semantic Web Layered Stack [6]	117

5.2	SOCPRI Ontology Metrics	123
5.3	Taxonomic Representation of SOCPRI ontology	125
5.4	Methontology ontology development process [7]	131
5.5	ORSD Template [8]	134
5.6	Competency Questions about OSN Users	136
5.7	Competency Questions about Digital Resources and User Profile	137
5.8	Competency Questions about User Roles and Relationships	139
5.9	Competency Questions about User Interactions and Privacy Policy . . .	140
5.10	Competency Questions about Tie Strength and Predictive Indicators . .	142
5.11	Overview of Classes and Object Properties Relations in SOCPRI	145
5.12	Graphical Representation of Social Web User Modeling	150
5.13	Graphical Representation of User Roles in SOCPRI	157
5.14	Contextual Representation of User Role in SOCPRI	162
5.15	Representation of User Relationships in SOCPRI	164
5.16	Representation of User Interactions and Tie Strength in SOCPRI	165
5.17	Representation of User Digital Resource in SOCPRI	170
5.18	Representation of User Contextual Privacy in SOCPRI	174
5.19	Description of main SOCPRI classes	183
5.20	Description of main SOCPRI classes	190
5.21	Description of main SOCPRI classes	194
5.22	SOCPRI Ontology Implementation in Protege	201
6.1	Checklist for Modeling Errors [9]	207
6.2	SOCPRI Results from RDF Validation Service	210
6.3	Preferences Set for Reasoning SOCPRI Ontology	212
6.4	Inferred Ontological View of the SOCPRI with Hermit Reasoner	213
6.5	Class Hierarchy Evaluation of SOCPRI ontology	217

6.6	Property Hierarchy Evaluation of SOCPRI ontology	218
6.7	Evaluation of SOCPRI for PropertyRestrictions	222
6.8	Evaluation of SOCPRI against Multiple Metrics	223
6.9	Evaluation of SOCPRI for Naming Conventions	225
6.10	Web Interface for Ontology Pitfall Scanner	228
6.11	Classification of pitfalls by level of importance [10]	229
6.12	Classification of pitfalls against various evaluation aspects	239
6.13	Evaluation results for SOCPRI ontology using Pitfall Scanning Tool . .	240
6.14	P19 Pitfall Identification and Correction in SOCPRI	240
6.15	Evaluation Results for Fixed Version of SOCPRI	243
6.16	Overview of various Individuals of SOCPRI Ontology	255
6.17	Queries to retrieve users related with a certain individual	258
6.18	Queries to retrieve working individuals with different criteria	260
6.19	Queries to retrieve users associated with certain group and organization	262
6.20	Queries to retrieve members of different groups with certain roles	264
6.21	Queries to retrieve parentage facts	266
6.22	Example queries to retrieve parentage facts	268

List of Tables

4.1	Demographics of the Participants	91
4.2	Gender wise Usage Frequency of Participants	91
4.3	Age group wise Usage Frequency of Participants	92
4.4	Age group wise Privacy Policy Awareness	94
4.5	Gender wise Privacy Policy Awareness	94
4.6	Age group wise Privacy Concerns of Participants	94
4.7	Gender wise Privacy Concerns of Participants	95
4.8	Age group wise Size of Friend Networks	98
4.9	Gender wise Size of Friend Networks	98
4.10	Age group wise Stranger addition in Friend Network	98
4.11	Gender wise Stranger addition in Friend Network	99
4.12	Descriptive Statistics for Interactions and Data Disclosure	101
4.13	Age group wise Personal Data Disclosure	104
4.14	Gender wise Personal Data Disclosure	104
4.15	Age group wise Misuse Concern about Personal Data	104
4.16	Gender wise Misuse Concern about personal Data	105
5.1	Comparison of SOCPRI ontology with existing Social Web ontologies . .	204
6.1	OntoClean Meta-Properties Summary	250

6.2	Applying Meta-Properties to SOCPRI Ontology-I	252
6.3	Applying Meta-Properties to SOCPRI Ontology-II	253
6.4	Applying Meta-Properties to SOCPRI Ontology-III	254

Chapter 1

Introduction

The main objective of this chapter is to introduce the research problem that we addressed in this dissertation. This chapter highlights need for the research on the topic by quoting existing problems in the domain. We discuss our motivation and present a set of research questions that are answered in this dissertation. We describe our research methodology in detail. We also list contributions and limitations of our research work. Finally, we explain the structure of this dissertation.

1.1 Introduction

Online Social Networks (OSNs) represent a virtual society that exhibits many of the characteristics of real human societies in terms of forming relationships and socializing with friends. However, existing online social networks lack an effective mechanism to represent diversity in social relationships. This leads to undesirable consequences of context collapse. The collapse of social contexts together has emerged as an important problem with the rise of online social networks [11]. It is a root cause for many privacy problems in online social networks. Online social networks collapse multiple audiences into a single context, making it difficult for people to use the same techniques online that

they do to handle multiplicity in an offline conversation. Context collapse often blurs the public and private, professional and personal, many different selves and situations in which individuals find themselves [12, 13].

The objective of this dissertation is to make a contribution towards modeling privacy for OSNs users from a social science perspective. Social sciences address this issue by exploiting existing social theories such as those of Goffman, Granovetter, and Nissenbaum to model self-presentation and contextual privacy on the basis of tie strength among OSN users [14, 15, 16]. We propose a theoretical framework for privacy in online social networks. We also develop an ontology using Semantic Web technologies to represent diversity in social relationships. Our developed ontological model is a step towards privacy friendly social web that uses contextual roles and relationship strength to make personal information disclosure decisions.

1.1.1 Background

Online social networks experienced exponential growth in last decade. Facebook and Google+ are top most visited sites on the Internet ¹ and fourth most popular activity of the Internet users. ² Nearly half of the internet surfers are active online social network users. ³ Online social networks are one of the most popular fora for self-presentation, social interactions and promote the vision of human-centric web [1]. These sites are an easy and cost-effective way for people to reach out to their friends, family, colleagues, classmates, acquaintances and even strangers from across the globe. A large percentage of success of these social networking sites can be attributed to a fact that they give users an opportunity to create their own space and a great way to connect with like-minded people, and share personal information. Web users spend an unprecedented amount of time using social networking sites, and upload a large amount of personal

¹Alexa <http://www.alexa.com/topsites>

²Nielsen <http://www.nielsen.com/>

³PewResearchCenter <http://www.pewglobal.org/2010/12/15/global-publics-embrace-social-networking/>

information. According to Zephoria, 4.75 billion pieces of content shared daily on Facebook and 300 million photos uploaded per day.⁴ In online social networks, the uploader of the data must decide which of his/her friends should be able to access the data. This resulted in a fundamental shift in the status of end-users. An individual end-user becomes the content creator and manager instead of just being the content consumer. The responsibility of managing appropriate privacy settings for every single piece of data shared put a cognitive burden on the users and hence most of the users end up using default privacy settings [17]. The default privacy settings are very permissive in nature leading to undesirable consequences of users' personal information disclosure to unintended audiences [18]. This poses a serious privacy threat to end-users, as a result, the issue of privacy in online social networks has received significant attention in both the research community and mainstream media.

Current online social networks provide a multitude of privacy controls to manage access to uploaded content. However, privacy control interfaces are too complicated to most of the normal users. The current interfaces have limited visual feedback and promote a poor mental model of how the controls affect the profile visibility [19, 20, 21]. Even after modifying settings, users can experience difficulty in ensuring that their settings match the actual desired outcome. Madejski et al. [22] show that privacy settings for uploaded content are often incorrect, failing to match users' expectations. According to Liu et al. [18], current Facebook privacy settings match users' expectations only 37% of the time. The authors further emphasized that incorrect settings tend to be more open and expose the content to more users than expected. A number of papers report that users have trouble with existing privacy settings. The vast majority of users do not utilize privacy settings to customize their accessibility [23, 24].

Some of the social networking sites provide features of lists and circles, in order to help users to share content selectively with their friends [25, 26]. Each friend-list

⁴Zephoria <https://zephoria.com/top-15-valuable-facebook-statistics/>

contains a subset of a users' friends, and then allow a user to share content only with members of the friend-list. Unfortunately, the usefulness of this feature is overshadowed by the cognitive burden that is placed on users. It is the responsibility of the users to populate their lists and grouping several hundred friends into different lists can be a laborious task. Another challenge is maintaining the appropriate membership of these lists over time because the relationships in everyday life evolve with time, whereas, these lists remain stagnant after their creation. As a result, it is unsurprising that many users do not use the friend list feature and this is a step towards a more unusable mechanism for controlling privacy in online social networks. Recently, the smart list feature is introduced by Facebook to reduce the user effort required to create these lists [27, 28]. The lists are generated automatically based on profile similarity attributes of the users such as workplace, city, school etc. The majority of the user profiles lack necessary attribute information required by this feature to function properly. Therefore, the usefulness of this feature is also questionable. It is important to note that smart lists do not take into consideration relationship strength, but the only function on profile similarity attributes.

Current tools for managing personal information disclosure in online social networks are effective in managing outsider threat to some extent, but these are unsuitable for mitigating concern over insider threat [23]. The insider threat is dynamic and deals with appropriate sharing of content with members of the friend network. The friend network of online social network users is heterogeneous in nature. Many users fear that boss or romantic partner will see something awkward on the OSN profile that was not intended for them. Online social network users attempt to avoid these situations. Research shows that current approaches for managing privacy are fundamentally flawed and cannot be fixed because of their implicit design assumptions about the social organization of the users. All friends are created equal and online social networks are unable to distinguish friends easily and automatically on the basis of relationship quality between a user and

his/her friends [29]. Providing additional privacy controls will not solve the privacy issue until this fundamental problem is fixed within online social networks.

1.1.2 Motivation and Problem Description

Despite the multitude of privacy controls, current online social networks fail to provide an effective mechanism to manage access to uploaded content about the users [22]. The main reason for this failure is the shortcoming of the online social networks to represent diversity in social relationships. Online social network users communicate and interact with people representing various facets of their life such as work, family, education, etc. In such a scenario, it is essential for users to be able to distinguish between these different types of contacts and form various virtual relationships. It is also important for users to understand and acknowledge these different relationships and take them into account when disclosing personal information. Unfortunately, major online social networks have been found to be falling short of appropriately accommodating these diverse social relationships in their privacy controls. Most online social networks employ “friendship” as the only type of bidirectional relationship. Friendship is a binary relationship in OSNs, which includes family members, close friends, colleagues, classmates, acquaintances, and even strangers. This provides only a coarse indication of the nature of the relationship. In human societies, relationships are much more complicated than a single binary relational tie. Various aspects of privacy such as what to reveal and whom to reveal are controlled by context and strength of the relationship among people, whereas such mechanism does not exist in contemporary online social networks.

We conclude that there is a disparity between the desired and the actual relationship representation of current online social networks. The main reason for this disparity is that online social networks assume a binary, static and symmetric relationship of equal value between all the directly connected users [29, 30]. In reality, social relationships are of varying tie strength [31] (how close two individuals are to one another), dynamic

(change over time), and asymmetric in nature (one person pays attention to another, it does not mean the latter will reciprocate). This problematic assumption in the implicit design of the OSNs projects a single unified image of the user to the world. There is no easy way for the user to establish and maintain many separate and sometimes overlapping social spheres that characterize real life.

In real life, people play diverse roles in different social contexts and disclose their personal information according to the contextual roles. Each individual has several role-based identities to preserve the contextual integrity of the information that is being disclosed. The notion of privacy as contextual integrity can be compromised by online social networks. For example, one may self-present in significantly different ways when in a business meeting versus when on a date. Online social networks place employers and romantic partners on the same communication plane, make it more difficult for users to segment audiences and present varied versions of the self. Difficulty in disclosing information selectively to various life facets lead to “context collapse” [11, 12]. The collapsing of social contexts has emerged as an important problem with the rise of online social networks. Context collapse makes it difficult for people to use the same techniques online that they do to handle multiplicity in face-to-face conversations. The features of lists and circles fall far short of preserving context integrity and reflecting many complex and overlapping spheres of offline life [32]. It is a challenging task to model social relationships for online social networks that reflect diverse relationships of offline life.

Social theories of Erving Goffman and Helen Nissenbaum can be exploited to address the issue of context collapse in online social networks [33, 32, 34, 35]. Privacy in online social networks revolves around the person’s ability to keep the audience separate and compartmentalize his social life. According to Goffman, each individual performs multiple and possibly conflicting roles in everyday life and it needs to segregate the audiences for each role, in a way that people from one audience cannot witness a role

performance, that is intended for another audience and thereby keeping a consistent self-presentation [14]. Nissenbaum’s theory of contextual integrity supports Goffman concept of audience segregation. She argues that audience segregation revolves around contextual integrity and goal of audience segregation can be accomplished by preserving contextual integrity. The theory of contextual integrity also provides a framework for understanding privacy implications of recent developments in OSNs and offer a useful conceptual apparatus for designing solutions to mitigate this problems [36].

The quality of the relationship can be determined by Granovetter’s concept of tie strength [37]. Tie strength is one of the most influential concepts in sociology. Granovetter characterized two types of ties: strong and weak. Strong ties are the people we really trust and their social circles tightly overlap with our own. Often, they are also the people most like us (i.e., homophily). Weak ties, conversely, are merely acquaintances. The existing literature of sociology suggests seven dimensions of tie strength. According to Petroczi et al. [38], the relationship indicators in online social networks are similar to those in offline communities. All tie strength dimensions can be easily inferred from user interactions pattern and profile similarity attributes in existing online social networks [39, 40]. Research proves that mixture of contextual grouping and tie strength could allow appropriate sharing of personal information [41].

The main motivation for this research study is to identify a theoretical framework for privacy and self-presentation in the online social network. This framework takes into consideration existing literature from sociology [] to identify dimensions of contextual integrity and relationship strength. The framework redefines privacy from the social perspective that addresses the concerns of the social web users. This theoretical framework for privacy is formalized through an ontological model. To better understand the social perspective of privacy in online social networks, we first conduct a user study to identify the link between personal information disclosure and user interaction patterns.

1.1.3 Research Questions and Objectives

The main research question for this study is: **How to model contextual privacy and self-presentation of the users in the dynamic environment of online social networks with diverse social relationships?** Additionally, we want to explore whether a user's interaction pattern with his/her friends can be used as a basis for inferring relationship strength among users. We also examine the link between profile similarity attributes and relationship context of the users. The strength and context of a relationship are key factors to control personal information disclosure of the users in online social networks. Self-presentation management in online social networks requires segregation of information and audience as per contextual role of the user. What are the key parameters for information and audience segregation is crucial question that needs to be answered for management of self-presentation in OSNs. We break main research question into following sub-questions:

1. How to redefine privacy for user generated content in the social web environment?
2. How do interaction patterns and profile similarity attributes reveal context and quality of relationships among OSNs users?
3. How to model diverse social relationships of OSNs users based on the relationships' quality and context?
4. How to evaluate the resulting model?

We define four main research objectives to address these questions. The first research objective is to develop a theoretical framework for privacy in online social networks from the social sciences perspective. We benefited from existing literature of sociology to understand social aspects of privacy. The work of well-known sociologists such as Goffman, Nissenbaum, and Granovetter provided insights into the social behavior of the individuals about privacy and presentation of self. Our theoretical framework

for privacy is inspired by their social theories about self-presentation, contextual integrity, and tie strength. The second research objective is to conduct a user study to investigate the attitude of social web users towards privacy in online social networks. The study examines information sharing and relationship forming behavior of online social network users. The main goal of this study is to determine whether there exists a relationship between the interaction patterns and the tie strength of the users. We explore the possibility of using users' interaction patterns with their friends as a criteria for making personal information disclosure decision. We assume that relationship strength is directly proportional to the frequency of interactions among users and personal information disclosure depends on the relationship strength. The third research objective is to formalize the theoretical framework of privacy into an ontological model. The goal of developing this model is to represent diversity in social relationships of OSN users. The innovative aspect of this ontological model is that it is based on most influential social theories of Goffman, Nissenbaum, and Granovetter. The last objective of this research is to perform the evaluation of the ontological model to show its validity.

1.2 Research Methodology

We reviewed the sociological literature in first phase to develop a theoretical framework for privacy from the social sciences perspective that suits the needs of social web with user generated content. Our theoretical framework addresses the issue of privacy from three aspects such as context segregation, contact segregation, and content segregation. The contact segregation provides users ability to keep audience separate and compartmentalize their social life. The content segregation plays a vital role in the classification of user data available in their profiles on the basis of information sensitivity. Context segregation controls access to sensitive personal information on the basis of role and relationship based social context.

We conducted a user study in the second phase to examine user attitude towards online privacy. A research questionnaire was designed and disseminated to collect user data. In total 323 participants took part in the study out of which 245 were male and 81 were female (leading to male bias). This online survey serves dual purpose: on the one hand, it gives insights into users' behavior towards online privacy and their information sharing pattern; on the other hand, it identifies the relationship between personal information disclosure and tie strength. Additionally, the findings facilitate categorization of profile information and user interactions on the basis of sensitivity and frequency respectively.

We developed an ontological model in third phase to represent diverse social relationships of online social networks users. The ontology design methodology used to develop this model is "Methontology" [42]. This model is based on well-founded social theories of Goffman, Nissenbaum, and Granovetter. The purpose of developing this ontological model is to enhance the management of user privacy in online social networks. Our ontology models the role and relationship based self-presentation of users in the dynamic environment of the social web. It represents tie strength dimensions and relates these dimensions to users' social interactions in OSNs. SOCPRI also models contextual privacy of OSNs users which takes into consideration contextual norms for appropriate information disclosure within and across contexts. Finally, we evaluated different aspects of the ontological model. We used three different reasoners to check logical consistency of our ontological model which includes Fact++⁵, Hermit⁶, and Pellet⁷. The model is evaluated against well-established evaluation metrics resulted from ontology-summit of 2007 [43]. We checked our model against various common ontological pitfalls that were introduced during the ontology development process. We also used OntoClean to evaluate our ontological model on the philosophical level. The

⁵Fact++, <http://owl.cs.manchester.ac.uk/tools/fact/>

⁶Hermit, <http://www.hermit-reasoner.com>

⁷Pellet, <https://www.w3.org/2001/sw/wiki/Pellet>

evaluation of the model is also carried out at assertional level by translating competency questions in DL queries and retrieving satisfactory answers from A-Boxes of the ontology. The query based evaluation also demonstrates consistency between T-Boxes and A-Boxes of the SOCPRI ontology. The summary of the research process is depicted in figure 1.1.

1.3 Our Contribution

Privacy in online social networks is hotly debated topic in the computer science research literature, but existing literature ignored the social aspects of privacy. The innovative aspect of our approach is that we address this issue from the social science perspective and exploit existing social theories of Goffman, Nissenbaum, and Granovetter to model privacy using Semantic Web language and standards. Our proposed model is based on the refined abstraction of contexts that embodies the philosophy of contextual integrity, tie strength and presentation of self. We argue that this model better captures users' privacy expectations and mimic real life personal information disclosure patterns. We summarize the main contribution of this dissertation as follows:

1. We developed a conceptual framework for social web privacy taking into consideration shift in the status of the user from content consumer to content manager. This framework of privacy is inspired from the most influential theories of sociology about self-presentation, contextual integrity, and tie strength.
2. We conducted a user study to examine the behavior of online social network users about information sharing and forming relationships. The study reveals user communication and interaction patterns with people representing various facets of their life such as work, family, and friends.
3. We developed an ontological model to represent diverse social relationships of online social network users. The model is based on the conceptual framework of

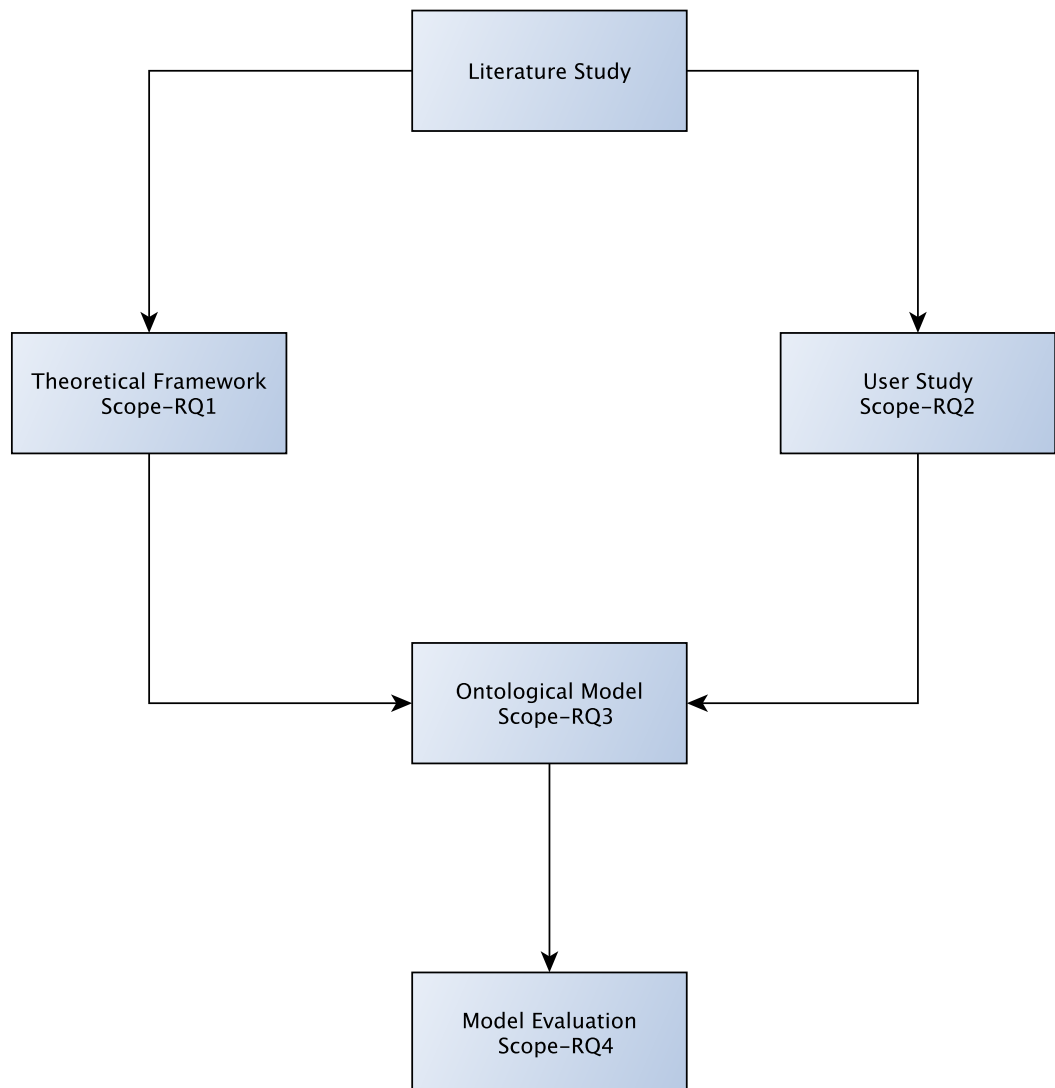


Figure 1.1: Research Process Summary

privacy that is inspired from the well-founded social theories of Goffman, Nissenbaum and Granovetter

4. We performed an evaluation of the ontological model.

1.4 Scope and Limitations

The work presented in this dissertation is multidisciplinary in nature. The basis for our ontological model come from social sciences. The social theories of Goffman, Nissenbaum, and Granovetter model various aspects of our theoretical framework. According to Erving Goffman each individual performs multiple and possibly conflicting roles in everyday life, and he/she needs to segregate the audience for each role in a fine-grained manner. Helen Nissenbaum argues that privacy revolves around contextual integrity and it gives an individual ability to keep audience separate. The social theories of Goffman and Nissenbaum converge at a single point that preserving the contextual integrity play vital role in compartmentalizing social life and presenting varied versions of self to diverse social relationships. The Granovetter gives insight on nature of relationships (tie strength) between individuals whether it is strong or weak. Tie strength is a quantifiable social networks concept that indicates the quality of relationships. The existing social science literature suggests seven dimensions of tie strength. The research proves that relationship indicators in online social networks are similar to those in offline communities [44]. We formalize our theoretical framework using Semantic Web languages and standards and perform an evaluation of the formal ontological model. The nature of this research work is interdisciplinary and it covers computer science, social sciences, and law. Therefore, we also critically analyzed emerging social web tools from legal perspective and presented some of the flaws in implicit design of the tools that violates existing data protection legislation.

There are certain inherent limitations in conducting multidisciplinary research. The

major limitation in our case is formalizing the entire social theories of Goffman, Nissenbaum, and Granovetter. We formalize some parts of these theories with the ontological model. Our ontological model will be able to reflect various offline relationships between users, one between users and resources, and one between resources. The model also represents various role-based identities of the online social network users. The quality of relationships between users is inferred on the basis tie strength dimensions and their predictive variables. Another limitation for this research concerns the evaluation of such ontological model. The best option for evaluation of such an ontology is human expert (social theorist) related to the domain of social science who tries to assess how well the ontology meets a set of predefined criteria and requirements implicit in the social theories of Goffman, Nissenbaum, and Granovetter. We do not opt this option for evaluation of the ontological model due to various constraints. Our evaluation mechanism checks the technical quality of the ontology from two different aspects. First we evaluate SOCPRI ontology for its correctness and consistency. The evaluation of this aspect is carried out using various reasoners, some Protege evaluation plugins, web based validation and verification tools, and the OntoClean methodology. The second aspect assesses our ontology for its appropriateness against the proposed requirements using queries.

1.5 Structure of the Dissertation

1.5.1 Background and Related Literature

In this chapter, we introduce the concept of online social networks and present their basic building blocks. We define the important terminology that is used throughout this dissertation. This chapter begins with the definition and evolution of social networks. We present historical background and discuss main differences between offline and online social networks. We also present the future perspective of online social networks

where semantic is needed that transforms these networks into semantic social networks. Finally, we analyze privacy risks associated with usage of online social networks and provide an extensive review of existing literature about privacy issues in online social networks.

1.5.2 Privacy: A Theoretical Framework for OSNs

In this chapter, we present multifaceted notion of privacy and describe its various dimensions. Our main focus is to provide legal, social, and technical perspectives of privacy. We also define the notion of privacy in the context of online social networks and identify various challenges that are posed to information privacy by unprecedented rise of online social networks. We introduce a new research paradigm for privacy in online social networks that inherits some properties from classical social theories of Erving Goffman, Helen Nissenbaum, and Mark Granovetter. The paradigm presents an agreed understanding about the nature and scope of privacy problem in the social web domain. Finally, we give an overview of our proposed privacy framework for online social networks. The framework addresses the multidimensional issue of privacy from multidisciplinary perspective and benefits from classical social theories.

1.5.3 Privacy: OSNs User Perspective

In this chapter, we present the result of user study conducted with online social networks. We introduce the purpose of the study and discuss the design of research questionnaire and its content. The user study provides insights into user interaction patterns and personal information disclosure practices. Finally, we conclude this chapter by presenting some of the important results of the study.

1.5.4 Ontology Based Privacy Modeling for OSNs

The objective of this chapter is to present the SOCPRI ontology which represents diverse social relationships of users in online social networks. The purpose of developing the SOCPRI ontology is to enhance the management of user privacy in online social networks. We present our methodology to build the SOCPRI ontology and highlight contributions of the SOCPRI ontology. We introduce core conceptual elements of our ontological model along with their definitions and relations. Finally, we present a comparative analysis of the SOCPRI ontology with other existing ontologies that represent data for the social web.

1.5.5 Evaluation of SOCPRI Ontology

In this chapter, we describe the evaluation of SOCPRI ontology. We focused mainly on two different aspects of the evaluation. First evaluation aspect deals with checking consistency and correctness of our ontology. The second aspect of evaluation assesses the appropriateness of the ontology against the proposed requirements.

1.5.6 Conclusion and Future Work

In this chapter, we conclude the dissertation by summarizing its contributions and reporting limitations of this research. We also revisit the research questions and discuss how our work addressed these questions. Finally, we present future research directions that stem from this dissertation.

Chapter 2

Background and Related Literature

In this chapter, we introduce the concept of online social networks and present their basic building blocks. We define important terminology that is used throughout this dissertation. This chapter begins with definition and evolution of social networks. We present historical background and discuss main differences between offline and online social networks. We also present the future perspective of online social networks where semantic is needed that transforms these networks into semantic social networks. Finally, we analyze privacy risks associated with usage of online social networks and provide an extensive review of existing literature about privacy issues in online social networks.

2.1 Social Networks

The term “social network” is pervasive nowadays. It is commonly used and understood by millions of the internet users. With the advent of Web 2.0, the notion of social networks started to refer to web-based communities which enable their users to perform

various types of interactions and collaborations. As a matter of fact, the term originates from the field of sociology and refers to social actors and their relationships. According to Wasserman [45], a social network represents “a social structure made up of a set of social actors and a complex set of the dyadic ties between these actors.” The concept of social actor refers to various types of entities such as a person, group, and organization. A tie is the set of all relationships that exist between two social actors. Another recent perspective on the social network is presented by Aggarwal [46], the author defines a social network as a network of interactions or relationships, where the nodes consist of actors, and the edges consist of the relationships or interactions between these actors. The concepts of social actors, their ties, and their interactions has always been central to the study of social networks. Many studies investigated ties between friends and relatives in order to distinguish between strong and weak ties. Granovetter [37] distinguishes between strong and weak ties on the basis of the relationship duration, intimacy between actors, and their emotional intensity towards a relationship. Strong ties often compose thick communities, whereas weak ties are related to thin communities. Thick communities contain group of actors who personally know each other and maintain frequent communication and interaction. On the other hand, socially and physically distant actors form thin communities. The detailed discussion on predictive variables for identifying strong and weak ties will be presented in next chapter.

The raise of social web paradigm given broader meaning to the notion of the social network. The term is no more restricted to sociology and generic face to face interaction between social actors. Now, it also refers to the web-based services that incorporate sociological aspects into world wide web and build a virtual community for online interactions and collaborations. Rheingold introduced the concept of virtual community and describes it as an alternate reality to their counterpart in everyday life. The author considers that these communities on the internet have capacities to transform society and can compensate the decline in the real-world communities [47]. The virtual com-

munity on the web cannot be regarded as similar to the traditional offline community. It has led to some ambiguity in the use of term “social network”. Scientific literature differentiates between two communities and uses different terminology to refer these communities. Traditional real world communities are termed as offline social networks, whereas web-based communities are referred as online social networks.

2.1.1 Difference between Online & Offline Social Networks

The social web paradigm introduced the possibility for users to interact and collaborate with each other using web-based services such as Facebook, Google+ etc. It has increased fraction of online interactions occurring through social media and given birth to the new form of self-expression and interaction between people. In order to understand how technology supports social web paradigm, we discuss technological aspects which distinguish online social networks from offline social networks. According to Boyd [48], the following four characteristics separate online social networks from their offline counterpart.

Persistence: It refers to the fact that communication and interactions in online social networks are recorded for posterity. This enables continued availability of the digital content beyond the temporal moment of its creation.

Searchability: It refers to the ability of vast search and discovery tools for online social networks that provide instant access to the recorded online expressions and the identities that have been generated through OSNs using the multimedia digital content.

Replicability: It refers to ease of duplication of the digital content. The online expressions can be copied from one place to another verbatim such that there is no way to distinguish the original from the copy.

Scalability: It refers to huge visibility of personal content in online social networks.

The users often share personal content with their large and diverse friend network (audience), who can further share it with their own friend networks, expanding its reach far beyond the interaction situation.

Papacharissi et al. [49] adds a feature of shareability to this list. The shareability reflects the tendency of online social networks to encourage sharing of personal content. According to Stutzman [50], shareability is perhaps the reason for widespread use of online social networks. Due to these features, online social networks differ drastically from their offline counterparts. The shared content persists beyond the ephemeral moment of sharing. It is easily searchable and replicable. It is possible for shared content to be viewed more broadly. Online social networks make the sharing of digital content increasingly effortless. They provide novel opportunities to connect with other people and to stay in touch even over a distance. In online social networks, an individual is speaking to all people across all space and all time due to the technological aspects of these services. These characteristics transformed classical social networks and resulted into new dynamics for online interactions.

Invisible Audiences: It refer to the obscured viewership for one’s self-presentation and content creation. Although users often act as though their audiences are bounded, they are in fact, potentially limitless [51]. All audiences are not visible and co-present at the moment an individual user is generating digital content for online social networks [48].

Collapsed Contexts: It refers to how a user, information, and norms from one context seep into the bounds of another [12, 11]. The technical features of online social networks obfuscate temporal, spatial, and social boundaries and make it difficult to maintain distinct social contexts [48].

Blurring of Public and Private: It refers to the lack of control over maintaining the distinction between public and private spheres in online social networks. The

disclosure boundary is compromised and personal content is increasingly available for public interactions [48].

These three dynamics differentiate online social networks from more classical offline social networks and have a huge impact on user's self-presentation and personal information disclosure. In offline settings, the audience is typically visible, if not completely know, and an individual adapts his self-presentation according to the given audience. It is quite straightforward to manage social context in offline social networks due to the fact that temporal, spatial, and social boundaries are very clear during the course of interactions. Self-disclosure made through offline social networks also preserve the distinction between public and private spheres. The selective self-presentation in the context of online social networks requires a new framework that takes into consideration these dynamics. We will introduce this privacy framework for online social networks in the third chapter of this dissertation.

2.2 Defining Online Social Networks

Today, online social networks are becoming de facto a predominant service on the web and revolutionizing the way people socialize in the modern world. They have become a mainstream cultural phenomenon and most popular activity on the web. Many social networks services have sprung in recent years due to their popularity. However, these services are heterogeneous in their functionalities, focus, content, and use, and therefore many different definitions of social network services exist in the scientific literature. The key elements of any online social network are captured by the widely used definition of Boyd and Ellison [52].

Definition 1 *An online social network is a web-based service that allows individuals to:*

1. *construct a public or semi-public profile within the service,*

2. *articulate a list of other users with whom they share a connection,*
3. *view and traverse their list of connections and those made by others within the service.*

According to this definition, every ONS user can create his/her own *profile*. A profile is a digital representation of an OSN user. Information about each social network user is maintained in a user profile which contains a number of attributes related to the demographics of users, their personal and professional addresses, their interests and preferences, as well as different types of user-generated contents (e.g., posts, photos, and videos) [53]. According to Grimmelmann [54] Facebook knows an immense amount of sensitive personal information about its users. A fully filled-out Facebook profile contains about 40 pieces of recognizably personal information, by the time you are done, Facebook has a reasonably comprehensive snapshot both of who you are and of who you know. Another important feature offered by almost all OSNs is *connections* which is a list of people to indicate existing social relationships of the users. Confusingly, many online social networks refer to connections as “friends”, however, connections are established between many users whose relationship may be better described by a different label such as family, friends, colleagues, acquaintances, etc. Labeling connections as “friends” is problematic because it treats all connections within the network equally; precluding differentiation for selective information sharing. In addition, the term “friends” carries connotations of friendship and trust which do not exist for all relationships within the network [55]. The final feature of OSNs as per above definition is *traversing connection* which allows users to find each other and construct a networked community within which they can share information. However, this definition lacks emerging services that become apparent when observing the use of online social networks: The communication of users through message exchange, commenting on the profiles of others or annotation of profiles and enabling of third party applica-

tions featuring advanced social interactions between users ranging from simple poking of another user to a variety of gifts and likeness applications for interactions between users. Boyd and Ellison offered an update to their original definition to cover new functionalities of contemporary online social networks [56].

Definition 2 *A social network site is a networked communication platform in which participants:*

1. *have uniquely identifiable profiles that consist of user-supplied content, content provided by other users, and/or system-provided data;*
2. *can publicly articulate connections that can be viewed and traversed by others;*
3. *can consume, produce, and/or interact with streams of user generated content provided by their connections on the site.*

This new version of the definition is much broader and articulates in a more nuanced fashion the different types of content that users' profile represents. Furthermore, it also highlights the possibility that profiles may contain content provided not only by the profiles' owners themselves but also by other users which demonstrate the influence of other users over the online identities of social networks users. Both definitions provided by Boyd and Ellison, completely ignore the privacy and security concerns that are faced by contemporary online social networks. This motivated Ho [57] to propose a definition for online social networks which encompass privacy and security concerns that appear in an online community.

Definition 3 *A social networking service is a website that allows users to:*

1. *connect with other users by befriending (Facebook), following (Twitter), subscribing (Youtube), ...*
2. *interact with content posted by other users, for example by commenting, replying or rating,*

3. restrict their own content to authorized users only.

This definition focuses on the restrictions imposed on social media content and limiting the use of posted content with social connections. Another definition highlighting the user generated content perspective of online social networks is given by Kaplan and Haenlein [58]. In their view, “A group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of User Generated Content”. In this definition, there is focus on emerging status of an individual user as the content producer or manager rather than content consumer. With respect to the user generated content dimension of online social networks, The concept of self-presentation becomes very important and usually, such a presentation is done through self-disclosure. It is a conscious or unconscious revelation of personal information. Self-disclosure is critical step that requires some kind of restriction on user generated content which is advocated by Ho [57] in his definition for online social networks. From this discussion, it is clear that there is not one clear and unambiguous definition of online social networks. Online social networks started off as websites such as Facebook, LinkedIn, and Twitter. However, nowadays these are ubiquitous in nature and accessed through many devices and platforms. We consider OSNs as services with intention of emphasizing how broadly and tightly these are interwoven into everyday activities and provide a coherent medium through which people can be interactive and socialize.

2.2.1 Evolution of Online Social Networks

It is challenging task to locate the origin of online social networks. The way social networking has evolved can be linked with the evolution of the internet itself. Many early online services such as bulletin boards, made some efforts to support social networking paradigm through computer-mediated communication. The commercial online services such AOL and its brethren promoted the development of niche communities in the early

and mid 90s. Many dating and community websites included the use of profile as early as the 1990s. The concept of a unidirectional list of friends was supported by instant messaging services such as AIM and ICQ. Although concepts of profile and friend list were not implemented as it is nowadays in online social networks.

The first recognizable online social networking site according to the definition of Boyd and Ellison was SixDegrees.com. This site was launched in 1997 and allowed users to create profiles, list their friends and traverse friend lists. Its name originates from the six degrees of separation concept. Six degrees of separation is the theory that anyone can be connected to any other person through a chain of acquaintances that has no more than five intermediaries. Although this site attracted million of users, it could not evolve into a sustainable business and closed down in 2000. The founder of SixDegrees.com believes that it was ahead of its time. From 2003, we witnessed a revolution and uptake of OSN sites that established most of nowadays most popular OSN sites. This revolution has brought a dramatic shift in the business, the cultural and the research landscape of the world wide web. Figure 2.1, presents a timeline that shows the evolution of OSN sites during the last decade. Today online social networks

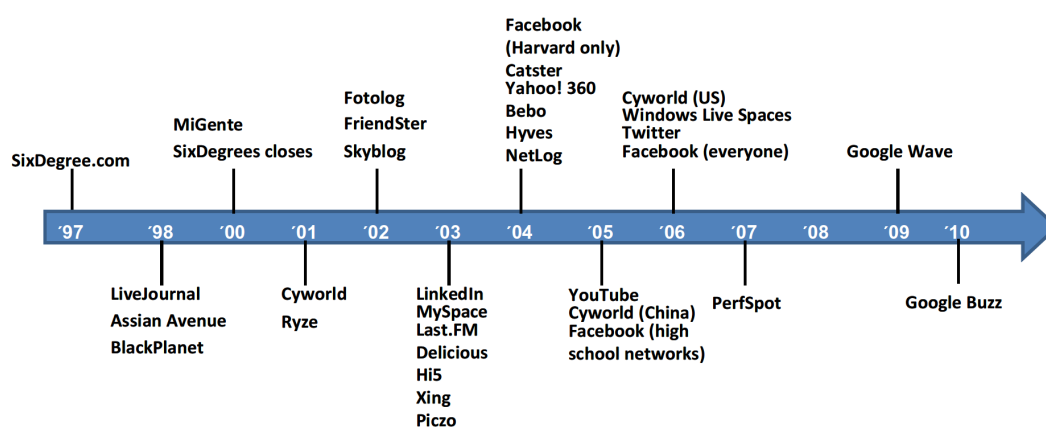


Figure 2.1: Timeline Online Social Networking Sites [1]

are most popular activity on the web. According to Nielsen Online's research ¹, online social networks and blogging sites are nowadays the fourth most popular activity on the Internet; this means that more than two-thirds of the global on-line population visit and participate in social networks and blogs. In fact, social media have pulled ahead of e-mail in the rank of the most popular online activities. Another interesting finding is that social networking and blogging accounts for nearly 10% of all time spent on the Internet. These statistics suggest that OSNs have become a fundamental part of the online experience on the Web throughout the world. Figure 2.2, summarizes some of the major online social networks used nowadays, taking into account those with a number of users higher than 1 million, and/or known at a worldwide level, and/or with a good global rank according to Alexa², a well-known provider of free global Web metrics.

2.2.2 Features of Online Social Networks

There is vast diversity among existing online social networks depending on the type of interaction and activities promoted these services. However, it is widely accepted that all OSNs share some core features. Based on the recent definition of Boyd and Ellison, An online social network should provide following core features to facilitate self-presentation and online interactions.

2.2.2.1 Personal Space Management

The personal space of OSNs user includes profile, wall and upload multimedia digital content. The profile contains personal information about the user. A user can advertise himself via their own profile. The profile update and retrieval functions allow the OSN user to maintain and visit the profile. The wall is the main thread which represents

¹Nielsen <http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen-globalfaces-mar09.pdf>

²Alexa. The Web Information Company. <http://www.alexa.com>

Name	Description	Launched	Registered Users	Restrictions	Alexa Ranking
Academia.edu	Academics/researchers	September 2008	3,000,000+	Open	2,459
Bebo	General	July 2005	117,000,000	Open to people 13+	6,010
BlackPlanet	Black Americans	September 1, 1999	20,000,000	Open	2,278
Classmates	School, work and the military	1995	50,000,000	Open to people 18+	2,273
DeviantART	Art community	August 7, 2000	25,000,000	Open to people 13+	133
Facebook	General	February 2004	1,100,000,000+	Open to people 13+	2
Flixster	Movies	2007	63,000,000	Open to people 13+	13,426
Flickr	Photo-sharing/-commenting	February 2004	32,000,000	Open to people 13+	77
Foursquare	Location based mobile OSN	2009	20,000,000	Open	535
Friendster	General (Southeast Asia)	2002	90,000,000	Open to people 16+	16,305
Google+	General	June 28, 2011	500,000,000	Open to people 13+	NA
Habbo	General for teens	August 2000	268,000,000	Open to people 13+	19,386
hi5	General (not USA)	2003	60,000,000	Open to people 13+	997
Instagram	Photo/video-sharing	October 2010	100,000,000+	Open	39
LinkedIn	Business	May 2003	200,000,000	Open to people 18+	10
LiveJournal	Blogging (Russian world)	April 15, 1999	17,564,977	Open (OpenID)	141
Meetup	General. Meetings	2001	13,400,000	Open to people 18+	388
Mixi	General (Japan)	October 25, 2000	24,323,160	Open	693
MyHeritage	Family-oriented	2003	75,000,000	Open	5,010
MyLife	Locating friends and family	2002	51,000,000	Open	3,565
MySpace	General	August 2003	30,000,000+	Open to people 13+	628
Netlog	General (Europe, Arab world, Quebec)	July 2003	95,000,000	Open to people 13+	1,045
Orkut	General (Brazil and India)	January 24, 2004	33,000,000	Open	1,732
Odnoklassniki	Old classmates (Russian world)	March 4, 2006	45,000,000	Open	57
Pinterest	Organizing and sharing interests	2011	70,000,000	Open	33
Sina Weibo	Microblogging site (China)	August 14, 2009	500,000,000+	Open	36
SkyRock	General (French-speaking world)	December 17, 2002	22,000,000	Open	694
Sonico.com	General (Latin world)	July 28, 2007	50,000,000	Open to people 13+	4,940
Tagged	General	October 2004	100,000,000	Open	326
Tumblr	Microblogging and SN	February 2007	200,000,000	Open	23
Twitter	General	July 15, 2006	500,000,000	Open	11
Vkontakte	General (Russian world)	September 2006	123,612,100	Open	20
Viadeo	Business	May 2004	50,000,000	Open	863
WeeWorld	Teenagers - 10 to 17	2000	30,000,000	Open to people 13+	20,297
XING	Business (German-speaking Europe)	August 2003	11,100,000	Open	401

Figure 2.2: Online Social networks as Popular Activity on Internet [2]

activity stream of the user and his/her friends. In some online social networks, it is termed as a timeline. An OSN user willing to set up multimedia gallery typically calls the upload function, which transfers digital data from user's device to the OSN database. The functions such as like allow the users to evaluate the digital content published by other users. Using the comment function a user can articulate their point of view in a more explicit way. Usually, in most of the OSNs, a user can associate another user by tagging in case the content depicting him. According to Zhang [59], An online social network should support a user to perform following actions to manage his/her personal space.

- Create/cancel an account in online social network service.
- Create/edit user profile
- Upload/edit user-generated content

All the actions carried out by the user in his/her personal space will be reported to his/her social connection automatically.

2.2.2.2 Social Connection Management and Networking Features

Social connection management feature allows an OSN user to articulate their relationships with other users. Through friending function, the user can set up a new relationship with some other user. This function typically sends a friendship request to the other user, who in turn can accept or ignore or reject the request. If the request is accepted then the users are added to friend lists of each other and new edge representing their relationship is added to the social graph. The actions such as establish/maintain/revoke a social connection should be supported by an online social network. Besides friendship, some of the online social networks offer networking features to facilitate the of the group. The grouping feature help users find people with similar interest or engage in activities of information dissemination on a certain topic. The groups are called by other names depending on the social networking service. The term network is used by Facebook to represent groups.

2.2.2.3 Communication and Interaction Management

The communication and interactions with other users are central features offered by online social networks. The main reason for the widespread growth of OSNs is their socialization and interaction facilities. There are various set interaction modes which include posting, messaging, chatting, liking, tagging, sharing etc. Another attractive

feature of online social networks is that they offer social applications and APIs. Facebook was the first to release a social networking API for third party developers. The Google also launched OpenSocial API for social applications across multiple online social networks. Each online social network offers its own brand of social applications which provide an enriching experience to their users.

2.2.2.4 Social Search Features

This is a crucial feature of online social networks that supports for searching and traversing digital social space of other members of the service. Social search encourages to establish new social connections and enriching of the social graph of the users. There are two ways to search personal space of an unknown user in online social networks [60]

Global Keyword Search: It is useful to find an unknown user by conducting a global name search. A successful search would produce for the accessing user the search listing of the target user. A user may specify a search policy to allow only a subset of users to be able to reach her search listing through a global name search.

Social Graph Traversal: A user may traverse social graph by examining the friend lists of other users. More specifically, the friend list of a user is essentially the set of search listing of his/her friends. Social graph traversal is a very common activity to find and establish new connections. A user may restrict traversal by specifying a traversal policy, which specifies the set of users who are allowed to examine his/her friend list after her search listing is reached.

2.2.2.5 Privacy and Security Features

This feature can help users to customize how their personal information is visible and who can access it. The users are allowed to define their privacy settings through some control functions. In particular, an online social network user may have control over

the following items

- controlling the profile visibility
- controlling the friend list or social graph visibility
- controlling the access to user generated content

These are some of the common features support by online social networks in order to achieve the goals of usability and sociability. Note that this not an exhaustive list of features and there may exist other features such as efficiency and reconfigurability. In this section, we are focusing on those features that potentially align with above-mentioned definitions of online social networks.

2.2.3 Classification of Online Social Networks

The contemporary online social networks are diverse in nature and promote a wide variety of activities. The scientific literature suggests several classifications for online social networks depending on their functionality and features. Our main concern is to streamline personal information disclosure and address privacy issues in online social networks. Therefore, we present three classifications that distinguish OSNs from data-centric and social perspectives. The first classification is suggested by Ho [57] and the author based his classification on two criteria: how OSNs affect the privacy of their users and which type of personal content is exchanged among users. The author divided OSNs into four categories and for each category, some illustrative examples are provided.

2.2.3.1 Personal Online Social Networks

Online social networks of this category focus on providing the opportunity for users to connect with their family, friends, and acquaintances. Typical examples of personal online social networks include Facebook and Google+. The user profiles are one of

the core functionality of these OSNs. The profile contains huge amount of personal information

2.2.3.2 Professional Online Social Networks

The main purpose of this kind of social networks is to connect users with their business contacts and help them to find a job or look for employees. A typical example of professional online social networks is LinkedIn and Xing. The information available on these OSNs is of professional nature which includes details about expertise, recommendations, and job offers.

2.2.3.3 Interest Oriented Online Social Networks

These OSNs allow users to share their hobbies and interests. Typical examples of this kind of OSNs are Last.fm and Flixster. The information shared on these services cannot be used to directly identify a user. The problem of privacy is not their core issue and these services are considered to be less sensitive with respect to user privacy.

2.2.3.4 Functionality Oriented Online Social Networks

These online social networks are known for their specific functionalities such as photo sharing, social bookmarking, and microblogging. Typical examples of these OSNs includes Twitter, Flickr, Instagram and LiveJournal. These OSNs does not necessarily capture demographic information but rather a large amount of other personally identifiable information such as photos. Another attempt to classify online social networks on the basis of pseudo-scientific literature was made by Beye et al. [61]. The authors classify suggest two broad categories on the basis of connections and content focused on online social networks.

2.2.3.5 Connection Oriented Online Social networks

Connection OSNs focus on the social connections and interactions between users, by providing users with a social contact list, channels for interaction, or matching services. Their general purpose is usually to connect users to new or existing friends and acquaintances or to provide an easy way to maintain such relationships.

Business: These OSNs aim to provide professionals with useful business contacts.

Searching for profiles does not always require signing up. Profiles display a users capabilities and work field as well as a means to contact that user. This is usually done through the OSN via messages. Users can also add other users to their network (connection) so that other professionals can see who the user is working or has contact with. An example of this class is LinkedIn.

Socializing: Fitting the more traditional view of social networks. Here users can connect with current friends and find new ones. All types of information found in an OSN are also found in this class, often a lot of this information is public. The revenue for the OSN provider often comes from advertisements and selling information about the OSN, but can sometimes be combined with a subscription for additional functionalities (as with Hyves). In order to attract and keep users this type of OSN usually has a lot of additional functionalities such as social and competitive games. For a user the value of such an OSN is often largely determined by the number of friends on the OSN. Some well known examples of this class are Facebook, and Google+.

Dating: Dating sites are websites that aim to help users find the love of their life, many of which incorporate OSN aspects these days. Each user has login credentials and usually a profile to attract potential lovers. Connections are typically in the form of love interests, but friendship links are also common; groups may also exist. Traversing the OSN is often based on searching or recommendations rather

than through navigating existing connections. Messages exchanged between users are often kept private to these users, although in some cases comment sections, viewable by connections, are offered. Example dating sites are match.com ³.

2.2.3.6 Content Oriented Online Social Networks

Content OSNs focus more on content provided by or linked to by users. This content can be multimedia or information like knowledge, advice, or news. The social interactions with other users usually revolve around and are driven by a search for information or the exchanging of said media.

Multimedia Content Sharing: Sharing of user-generated content can happen within a selected group, such as friends or family, or a far wider audience. Content that is shared is usually multimedia; this is often of potential interest to a wide audience, and even for selected audiences, e-mailing such content is cumbersome and often impossible due to size of the data. Uploading content generally requires users to sign up and log in; sometimes viewing content also requires logging in. The content tagging and recommendation may be an integral part of the system. Examples are Instagram ⁴ and youtube ⁵.

News Sharing: Some OSNs focus on world news or gossip, but a multitude of micro-blogging OSNs provide a stage mainly for sharing personal news, opinions, and experiences. Examples are Twitter ⁶.

Hobbies/Entertainment: Many OSNs focus on audiences that have similar interests and hobbies. Such OSNs may involve multimedia uploads, recommendation, or advice sharing elements, but the main distinguishing feature is their homogeneous

³www.match.com

⁴www.instagram.com

⁵www.youtube.com

⁶www.twitter.com

audience. This means that the topic of the OSN mainly determines its character and appeal for users. Examples are Xbox Live ⁷

The latest classification of online social networks from semantic web perspective is given by irfan et al. [62]. The authors categorize online social networks into three broad categories which include context based OSNs, content based OSNs and media based OSNs. The content-based online social networks allow the text-based interactions among users such as microblogging, social news, etc. The media based online social networks provide user interaction through various multimedia formats such as video and audio. The integration of semantic web technologies with online social networks can be more useful and productive for development of the intelligent social communicational services. According to authors, the past literature focused on content and media based OSNs, whereas most integral and crucial perspective related to social semantic was overlooked. The lack of semantic analysis was major barrier for effective intelligent social communication services. We will describe the context based online social networks in following section, as far as content and media based OSNs are concerned these already addressed in the previous section.

2.2.3.7 Context-based Online Social Networks

The context based online social networks provide an appropriate platform for the integration of physical and logical contextual information that can be gleaned from various sources such as user profile, interaction and communication pattern of the users. An integration of the contextual information with interactive computing can be a promising solution for the development of effective intelligent social communication services. There are three different type of context-based online social networks. The brief description of these types is given below:

⁷www.xbox.com/en-us/live/

Social Semantic Web The social semantic web implements ontologies for context-based knowledge management.

Social Search The social search is shared effort of a group of users to obtain the relevant information. The collaboration is an important aspect of an social search where multiple users can participate by suggesting different keywords, query syntax, and query reformation.

Social Recommendations The social recommendation target social media domain and include online social relationships as an additional input parameter. There are two main categories of social recommendation systems which includes content-driven recommendation systems and collaborative-filtering based recommendation systems.

2.2.4 Data Collection by Online Social Networks

With the increased usage frequency and ubiquitous usage of online social networks, the quantity and sensitivity of user data that is stored on OSNs has grown tremendously as well. In this section, we aim at presenting diverse sensitive data disclosure possibilities and their implications towards the everyday life of OSN users. There has been several efforts by some researchers to provide taxonomy of data disclosed by the users in online social networks. Beye et al. [63] deduce from Boyd and Ellison's definition that following user-related data must exist in an OSNs:

Profile: A profile is tied to a user and is their representation to the outside world. Usually this is a self-description or the description of an alter ego (pseudonym, avatar). This may typically include a short biography, a picture and attributes like age, gender, location, and the like.

Connections: A connection exists between two users and can be of several types, like friend, colleague, fan, etc. A collection of connections can be represented by a

graph.

Login credentials: Most OSNs require the user to login to make use of the service.

A user account ties a profile to the user behind it, and to sign in, the user needs certain login credentials. Such credentials can also be found in traditional websites.

Messages: We view messages in the broadest sense of the word. Any piece of data exchanged between a user and another user or a group of users is a message; these may contain text or multimedia. Messages form the basis for additional OSN functionalities. Note that in some cases a message can be instantaneous and short-lived, as in an instant messaging setting. In other cases, messages may be stored for an indefinite time and be read long after being sent; think of blog posts or messages left on a user's "Wall" on Facebook.

Multimedia: Actual content that can be attached to messages but may also be uploaded to private or public data spaces (e.g. Facebook "Wall") or be attached to a profile.

Groups: A group is a collection of users, who usually share some common attributes, resources, or privileges, for example, similar preferences or backgrounds, a collaborative document, or access to a common virtual space.

Tags: We define tags in the broad sense, as in collaborative filtering systems: descriptive keywords (metadata) that are attached to content by users (either the uploader or other users). In Facebook terminology, "tagging" refers to the specific case where a user identifies the people depicted in a photo by tags the photo with their names, thus explicitly linking these people to the picture.

Preferences/ratings/interests: Many OSNs provide their users with some type of matching or recommendation functionality for either content or peers. In order to

provide relevant recommendations, information on a user's attributes or preferences is required. Often, users are asked to explicitly express their preferences or rate items. The resulting information may be publicly visible (interests on a profile page, ratings for an item shows along with who provided them) or restricted to the service provider only.

Behavioral information: By this we mean browsing history, profile settings, and any actions undertaken by the user while performing tasks within the OSN.

Schneier [64] describes following different types of the user data items are harvested by online social networks:

Service Data: It is the data you give to a social networking site in order to use it.

Such data might include your legal name, your age, and your credit-card number.

Disclosed Data: It is what you post on your own wall pages: blog entries, photographs, messages, comments, and so on.

Entrusted Data: It is what you post on other people's pages. It's basically the same stuff as disclosed data, but the difference is that you don't have control over the data once you post it?another user does.

Incidental Data: It is what other people post about you: a paragraph about you that someone else writes, a picture of you that someone else takes and posts. Again, it's basically the same stuff as disclosed data, but the difference is that you don't have control over it, and you didn't create it in the first place.

Behavioral Data: It is data the site collects about your habits by recording what you do and who you do it with. It might include games you play, topics you write about, news articles you access (and what that says about your political leanings), and so on.

Derived Data: It is data about you that is derived from all the other data. For example, if 80 percent of your friends self-identify as gay, you're likely gay yourself.

Richthammer et al. [3] propose taxonomy for the data disclosed in online social networks. The authors distinguish between the data disclosed to OSN service providers and OSN users. They further divided OSN user related disclosed data into semantically specified and semantically unspecified data. The semantically specified data refers to data instances that have a clearly defined meaning and its content is clearly understood. Examples include predefined attribute types of an OSN profile such as name, birthdate, and hometown, Whereas, semantically unspecified data types are provided to freely express some facets of one's personality, such as status posts whose content is not semantically predefined. The detail description of the data disclosure taxonomy is describe below.

Login Data: Login data is considered a data type that is required by the OSN service provider to provide evidence of a claimed identity. Common instances of this data type are identifiers such as user name and email address as well as passwords used to verify an identity.

Connection Data: Connection request leads to a variety of digital traces created by protocols on several layers of the OSI model. Especially browser related information and location are deemed sensitive and entail privacy implications when being available to OSN service providers, such as for acquiring detailed user information through cookies and browsing history or for creating a movement profile based on location data.

Application Data: Besides OSN platform usage, data originating from the use of third party services running within the boundaries of the OSN platform or having API access can be differentiated. Common examples are player statistics of OSN games, application usage statistics, or In-App purchase data such as credit card

information. Depending on the data instance, privacy implications may range from none to serious.

Mandatory Data: Mandatory data refers to personal information that needs to be provided by the user during the registration or profile creation process.

Extended Profile Data: This contains input fields for attribute types like address, education, favorite music, favorite films, hobbies, interests, etc. The profile picture, which is a common feature of OSNs, is also arranged in this category.

Ratings/Interests: This class of data covers expressed interests such as Facebook's Like and Google's +1 and the rating of photos shared by other users.

Network Data: This class of data includes information about their relationship with other users. The collection of all connections of a particular user is often referred to as his social graph.

Contextual Data: This class of data refers to a property of an existing item that is made explicit and provided with semantics, hence forming a new data type. Common examples include the tagging feature, allowing to make peoples' names (and eventually their identity) in an existing picture explicitly available to other OSN users. Further instances are the location of a picture and the relation of a shared item to an activity or an event.

Private Communication Data: This class covers data elements that originate from private communication between OSN users. While private communication may comprise text messages as well as other media formats, their content is not semantically specified.

Disclosed Data: A frequent user activity on OSNs is to post information on one's wall and this instance of this class.

Entrusted Data: In contrast, entrusted data refers to information that is both user-generated and user-published but in the domain of a contact. Consequently, once the data is shared, control passes over to the domain owner that is from then on able to define its visibility.

Incidental Data: Incidental data originates from a contact sharing a data element on the user’s wall, however the information is shared in the user’s domain.

Disseminated Data: The user generated data elements are considered that are further disseminated by a contact within his own domain.

The authors analyzed four major online social networks– Facebook, Google+, Twitter, and LinkedIn to demonstrate various aspects of data disclosure taxonomy. Figure 2.3 presents overview personal data disclosure online social network user. The average user

Data types	Facebook	Google+	Twitter	LinkedIn
Login data	Email, phone, password	Email, password	Email, username, password	Email, password
Connection data	Device information, log information, location information, cookies	Device information, log information, location information, cookies	Device information, log information, location information, cookies	Device information, log information, location information, cookies
Application data	Usage statistics, credit card information	Usage statistics, credit card information	Usage statistics	Usage statistics
Mandatory data	Name, email*, birthday*, gender*	Name, email*, birthday*, gender*	Name, email*	Name, email, job status
Extended profile data	Several general-purpose input fields	Several general-purpose input fields	Three single input fields (location, website, bio)	Several professionally-related input fields
Ratings/interests	Page, status/photo/video	Page, status/photo/video	Verified account, Tweet	Company, status
Network data	Unidirectional, bidirectional	Unidirectional	Unidirectional	Bidirectional
Contextual data	Tag in status/comment, on photo, at location	Tag in status/comment, on photo, at location	Mention in Tweet	n/a
Private communication data	Private message, video chat, poke	Private message, video chat	Private message	Private message
Disclosed data	Text post, photo (album), video, check-in	Text post, photo (album), video, check-in	Text post, single photo	Text post
Entrusted data	<i>See disclosed data</i>	<i>Restricted to comments on disclosed data</i>	n/a	<i>Restricted to comments on disclosed data</i>
Incidental data	<i>See disclosed data</i>	<i>Restricted to comments on disclosed data</i>	n/a	<i>Restricted to comments on disclosed data</i>
Disseminated data	<i>See disclosed data</i>	<i>See disclosed data</i>	<i>See disclosed data</i>	<i>See disclosed data</i>

Figure 2.3: OSN User Data Disclosure [3]

is not aware of the risks assumed when deciding to disclose personal information in the online social network. Users tend to introduce a high amount of sensitive data, without even being aware of the impact on their privacy. Banescu et al. [4] classify risk and threat tree for Facebook, figure 2.4 presents this information.

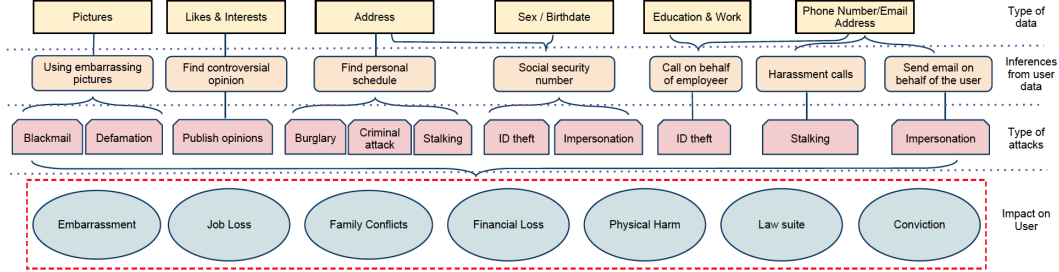


Figure 2.4: Privacy Implication of Data Disclosure in OSNs [4]

2.3 Privacy in Online Social Networks

Privacy in online social networks has become an important issue due to the fact that these platforms managed to gather a huge amount of personal information of their users. The reason for this growing concerns is social protocol and traditional way in which people regulate their privacy in offline social networks, are no longer applicable to online social networks because of technologically supported features of persistence, replicability, searchability, and scalability. Due to these features of online social networks, a great deal of personally identifiable information is out there, waiting to be structured, organized, and accessed by those looking for it. The various aspects of privacy issues associated with OSNs are widely reported in scientific literature. Initially, we identify threats associated with usage of online social networks.

2.3.1 Attitude of Users towards Privacy

A wide variety of research is conducted on privacy attitude of online social networks users. Analysis of these studies demonstrates that OSNs users claim that they are worried about their privacy and are aware of the many risks associated with uploading personal content on social networking sites, but at the same time their information sharing pattern reveals carelessness while uploading sensitive personal information on their profiles. Barnes [65] reported this phenomenon and defined it as a “privacy

paradox”. Similar findings were reported by Gross et al. [66], as per results of this research OSN users are not concerned about the privacy of their personal information and tend to share as much information as possible with their connections. According to the authors, many users reveal even the most sensitive personal information such as their sexual orientations or drug usage to everyone. Acquisti et al. [67] found that most of the users do not change the default privacy settings. The default settings are very permissive in nature and expose user’s profile to a large audience. A later research by Lewis et al. [68] also confirms the findings of Acquisti et al. The authors inspected the Facebook profiles of university students and found that only 1 out of 3 was set as private. Another study on MySpace by Thelwall [69] has similar results. The author analyzed more than twenty thousand profiles and found only 0.25% profile set as private. The attitude of OSN users has changed towards privacy in recent years due to the extensive discussion of privacy issues in mainstream media. This fact is also confirmed by the recent research studies of Dey et al. [70] and Boyd [71]. The authors noticed a significant change in the behaviour of OSN users towards their privacy. Dey et al. examined 1.4 million Facebook profiles to observe information revelation trends over a period of two years. Comparing his first observation results with second observation, the authors found that OSN users in the sample have become considerably less exposed throughout this period by hiding much of the profile information that they had previously revealed to the public. Boyd also conducted research survey to examine changes in the privacy attitude of OSN users. As per the results of users’ self-reported behavior their willingness to share personal information and to connect with new contacts has decreased due to privacy risk stories reported by media outlets. PEW Internet reports that in 2011, 63% of Facebook users had removed someone from their friend network [72], an increase compared to the 56% of users who reported to have unfriended someone in 2009. The same survey found deleting and untagging posts to be common among OSN users. These studies indicate privacy attitude of OSN users

is changing and users require more control over managing his identity in online social networks.

2.3.2 Data Disclosure Method of OSNs

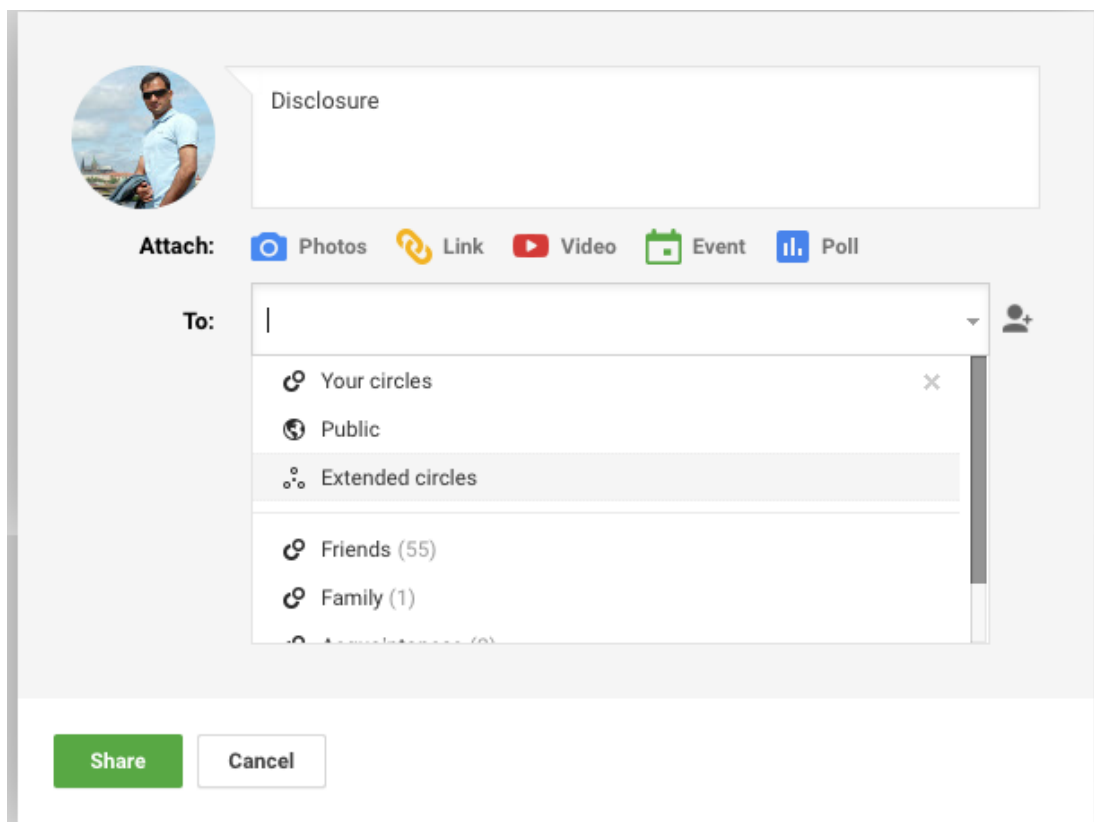


Figure 2.5: Google+ Data Disclosure Method

Current interfaces of online social networks support various data disclosure methods depending on user choice while sharing digital content. Most of the online social networks provide options of public data disclosure. Public disclosures are visible to every user in the entire social networking site, regardless of whether fellow users are added in the social connection of the user disclosing the data or how many degrees of separation they have between them. Another option provided by online social networks for data disclosure is with social connections. The term social connections refer to friends

in Facebook and circles in Google+. This is a set of contacts to which an individual user is connected in online social networks. Social connection disclosure is visible to all users directly connected to the user disclosing the data. Some of the online social networks offer extended social connections disclosure option which is only visible to a user's friends and friends of the friends. The Google+ support this option with term extended circles as illustrated in figure 2.5. Private disclosure option is supported by Facebook which is visible only to the user that upload the data. Facebook refers this data disclosure option as "only me" as illustrated by figure 2.6. Online social network like Facebook and Google+ offer features of friend lists or circles. The user disclosing data may select an appropriate friend list or circle to share the upload content with a member of that closed group. Some of the online social networks allow a user to customize the sharing of uploaded data. The main purpose of this customization is choosing a collection of users referred as a group for selectively disclosing uploaded data. The design mechanism of online social networks offer these coarse-grained data disclosure methods which limit user control over his disclosed information. By offer user fine-grained control over personal information disclosure, we equip them for better self-presentation and mitigate privacy threat encounter by the users.

2.3.3 Identification of Privacy Threat

In order to elicit privacy threats related to usage of online social networks, we rely on two main perspectives: the first stakeholder involved in the threat and second the way information is acquired. Beye et al. [63] distinguish between two main classes of privacy threats stakeholders which include user related privacy threats and service provider related privacy threats. A significant difference between each class is the type of information that can be accessed. By all means, users have limited access to the data as compared to the service provider. The tools required to protect user data from fellow users are ineffective in solving an issue related to service providers. Because the

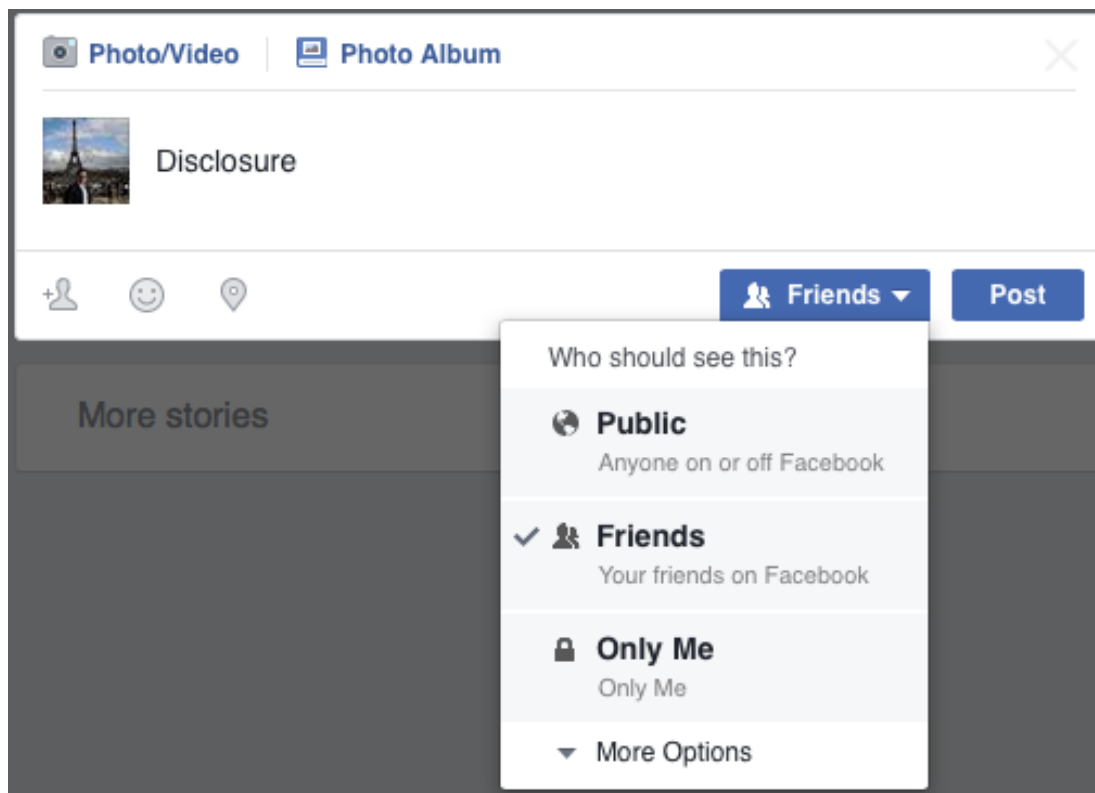


Figure 2.6: Facebook Data Disclosure Method

type of access differ greatly, both categories of threats require their own specific defense mechanisms. The detailed description of both types of privacy threats is given below:

2.3.3.1 User-related Privacy Threats

This is a very common type of privacy threat in online social networks. In many cases, the privacy of a user is breached by fellow OSN users. There various types of privacy threats that involve disclosure of personal information to other users:

Stranger Access Due to an invisible audience, sometimes personally identifiable information of the user is accessible by strangers. This can also happen because of design flaws on the part of the OSN service provider or lack of understanding or attention to privacy controls of the user himself. Some of the examples of

these flaws are beacon feature of Facebook and second-degree access support by some OSNs. The user control over who views the information is lost due to such features and user personal information reaches to unwanted audiences.

Unwanted Friend or Group Accesss Due to context collapse, The users do not have the control to act differently towards one user or a group, than towards others and not all OSN provides features to manage disclosure of personal information on a fine-grained level. Therefore, all friends have equal access to profile information of the user and it is quite difficult to hide information from unwanted friends and groups.

Information Disclosure due to Fellow Users Due to the phenomenon of interdependent privacy, the OSNs user is bound to be affected by the information disclosure decisions of his/her fellow users and his privacy could be out of his own control. The feature of photo tagging is a straightforward example of interdependence in online social networks.

2.3.3.2 Service Provider-related Privacy Threats

This type of privacy threat involves a relationship between user and service provider. In contrast to a user related privacy threat, The OSNs service provider can view all data provided to the central system, no matter what personal information disclosure settings are applied to the data. There are also various types of service provider related privacy threat which includes data retention, data selling, and targeted marketing.

Data Retention It is often impossible or very difficult to remove data after posting it to online social networks. Some of the online social networks do not provide users with means to delete their profile, even if the profile is apparently erased still backup copies of the data resides elsewhere on the OSNs. Data retention is discouraged by EU regulations. The general data protection regulations (GDPR)

forces controllers (service providers) to give users right to erasure of the data.

Data Selling Online social networks gather a huge amount of personally identifiable information about their users. This information can be very useful for several stakeholders such as government organizations, marketing agencies, and other third parties. There are some accidents reported in mainstream media which confirming selling of this personal information of the users.

Targeted Marketing Online social networks also contain information about user preferences and behavior. This information can be highly valuable for marketing purposes and can be exploited for targeted marketing to the vast user base of OSNs.

Another perspective to identify privacy threats is based on the way information is acquired. Chen et al. [73] describe two type of privacy threats related to this perspective; intentional privacy violations and accidental privacy violations. The first discusses threats where information is obtained through force, deception or foul play, whereas accidental privacy violations are made through the normal use of OSNs while disclosing personal information without realizing who might be able to see it, or without the ability to control who can see it. Jones [5] combine these two methods of classification to produce a matrix of “privacy threat model” which illustrated in figure 2.7. According to his privacy threat model, user related intentional privacy threats are posed by online predators, hackers, and posting activities of other users, whereas user related accidental threats are due to poor usability, permissive defaults, and lack of awareness. As far as service provider related accidental privacy threats are concerned, these are the result of design flaws, security leaks, and bugs. Service providers pose intentional privacy threat due to retaining data for longer than allowed, selling data for profit or allowing employees to browse private information.

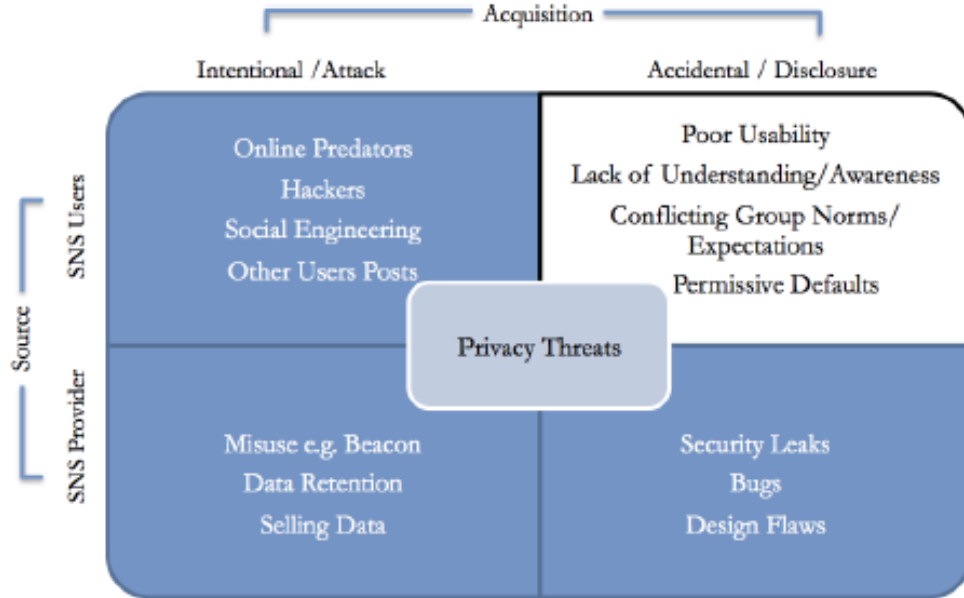


Figure 2.7: Privacy Threat Matrix [5]

2.3.4 Attack Spectrum for OSNs Users

The usage of online social networking platforms opens doors for a variety of attacks on the privacy of the users. The users are often not aware of nature of attacks launched by third parties to gain access to their personal information, even they are ignorant about the size of audience accessing their profile information. In this section, we will describe main attack types mounted by third parties against OSNs users, however, no meaningful protection appears possible if the attacks are launched by the provider itself. European network and information security agency prepared a comprehensive report on such attacks and their negative consequences [74].

2.3.4.1 Stalking

Cyberstalking is a serious invasion of online privacy. An individual user frequently posts status updates, photos and location information in online social networks. This

information may pose danger for the user to become a potential victim of stalking without even being aware of this risk. The impact of stalking range from mild intimidation and loss of privacy to serious physical harm and psychological damage.

2.3.4.2 Digital Dossier Aggregation

Online social networks play a vital role in the availability of a large amount of personally identifiable information online. This information can be used for purposes and in a context different from the ones the profile owner had considered. Third parties can create a digital dossier of personal data by downloading and storing OSNs user profile over time and incrementally. The negative consequences associated with this attack are exploiting the personal information out of context for embarrassing, blackmailing, or even damaging the image of profile holder.

2.3.4.3 Profile Squatting

This attack involves the creation of fake profile to impersonate a renowned person or a brand. It requires minimal effort to establish fake profile due to weak identity verification process by online social networks. Profile squatting can do a significant damage to the reputation of a person or any brand which may, in turn, lead to the financial social embarrassment.

2.3.4.4 Image Tagging and Cross-profiling

Photo-tagging is widely used feature of online social networks. This allows OSN users to tag images with metadata such as name of the person in the photo or link to his/her profile page. The feature can be used negatively to slander some well-known personality or a brand.

2.3.4.5 Phishing

Online social networks host large amount of reliable personal information which may be useful to launch a highly personalized phishing attack. The success rate of phishing attack has improved significantly due to exploiting the information available on social networking sites [75]. The negative consequences associated with phishing includes compromised logins, identity theft, financial and reputation damages.

2.3.4.6 Spamming

The spammers are benefiting from exponential growth of online social networks and affect OSNs with the same scale of spams which was affecting email communication. Social network spams produced huge traffic overload in online social networks. The negative consequences of SN spamming are similar to those for email spamming.

2.3.4.7 Cross-site Scripting

Online social networks offer third party developers' platforms for external programmers to produce widgets that are integrated with social networking sites. These third party applications are weakly verified and make OSNs vulnerable to cross-site scripting attacks. Some of the negative consequences of this attack are loss of privacy, identity theft and zombification of OSN account.

2.3.4.8 Corporate Espionage

Online social networks are used as an effective tool to attack enterprises using organized social engineering attacks. Online social network based social engineering attacks are growing and often underrated risk to corporate IT infrastructure. Some of the negative consequences of corporate espionage are loss of corporate intellectual property, causing damage by hacking corporate networks and even access to physical assets.

2.4 State of Art

A substantial amount of work has been done on the problem of preserving privacy in online social networks. We focus our discussion of related literature on aspects of privacy issues faced by online social networks users.

2.4.1 Semantic Relationship Modeling

With the emergence of the semantic web, ontologies have provided new potential for enhancing expressiveness, formal semantic, and reasoning capabilities of several approaches. Ontologies together with rules can be exploited to develop an underlying privacy platform for online social networks. In this section, we compare our proposed approach with some of the other relevant initiatives in this area. FOAF (Friend of a Friend)[76] is one of the first semantic models to grasp social interconnections between people. Persons, their activities, and relationships to other people or objects are modeled in this ontology. FOAF is a lightweight and simplified model. FOAF has a "knows" property that defines a social relationship. However, representing relationship using such RDF property fail to accommodation rich context information and diverse social relationships. FOAF realm [77, 78] quantifies the knows relations in the context of FOAF ontology as a trust metric and support rules that control access of friend to resources in online social networks by stating maximum distance and minimal friendship level. RELATIONSHIP ontology⁸ also model user relationships in online social networks in a precise manner. This ontology specializes the "knows" property of FOAF to characterize various user relationships (personal, professional, sentimental and family). The AMO (Access Management Ontology) [79] is another approach that allows annotating the resources and modeling the access control policy. These existing ontologies do not take into consideration contextual integrity framework for preserving the privacy of OSNs users. Elahi et al. [80] propose ontologies to represent relationships among

⁸RELATIONSHIP, <http://vocab.org/relationship/>

individuals and the community in order to enforce access restrictions to the resources. Carminati et al. [81, 82] propose conceptually similar, but much richer OWL ontology for modeling various aspects of online social networks. The use of semantic ontology allows the model to infer the relationships among users and resources. The authors define three types of policies, namely, access control policy, filtering policy, and admin policy. Access control policies are positive authorization rules; filtering policies can limit someone's access to information by him/herself; admin policies can be used to express who are authorized to define those policies. Although the authors outline an access control framework, lack of formal descriptions and implementation leaves behind many ambiguities. A more detailed approach is developed by Masoumzadeh et al. [83, 84], which proposes the ontology-based social network access control model, encompassing two ontologies; the Social Networking systems Ontology (SNO), capturing the information semantics of a social network, and the Access Control Ontology (ACO), which allows for expressing access control rules on the relations among concepts in the SNO. This model takes into account intricate semantic relationships among different users, data objects, and between users and data objects. The model enables expressing much more fine-grained access control policies on a social network knowledge base than already discussed by Carminati et al. These models assume that online social networks provide a rich social model that is capable of representing different types of social relationships. Unfortunately, this is not true in current OSNs, as they usually only consider a friend as the only type of possible bidirectional relationship. Barkhuus [85] investigates the application of contextual integrity to the consideration of privacy in HCI research. The qualitative study of Shi et al. [86] provide preliminary insights in understanding user's interpersonal privacy concerns from the perspective of contextual integrity. Lipford et al. [87] claim that failure of privacy management on online social networks reflects the nuanced and contextual nature of privacy in the offline social world. The authors argue that online social networks ought to be designed from the perspective of contextual

integrity to preserve the privacy of their users. Among the existing work, the approach proposed by kayes et al. [88, 89] is most similar to our proposal. The author claims that their system (Aegis) implemented contextual integrity, whereas, their example contexts and policies reflect basic access controls akin to UNIX access controls. The Aegis system express limited forms of norms and contexts and ignore underlying principles of contextual integrity such as norms with roles and attributes. Compared to kayes et al., we provide a much richer set of contexts and roles to model diverse aspects of users' social relationships in online social networks. Compared to existing approaches, we take into consideration contextual integrity as a key factor to perform context segregation. Furthermore, tie strength dimensions and their suggested predictive variables are utilized to infer relationship quality between users. We exploit rich scientific literature from sociology that makes this approach unique from other existing approaches.

2.4.2 Relationship Strength Prediction

Gilbert et al. [39] proposed a model that predicts tie strength among users of Facebook. The authors selected a group of variables available on Facebook. They evaluated their model with the participation of Facebook real users. The researchers achieved an accuracy of 84%. One of the limitations of this model is the huge amount of information it requires to predict tie strength. Gilbert et al.[40] expands his previous work [39] and proposes a model to infer the tie strength of relationships on Twitter. In order to evaluate the new model, Gilbert developed a tool called We Meddle that predicted tie strength. Users were asked to try We Meddle and evaluate its predictions. This work showed that the model proposed in [39] can somehow be generalized and adapted to different OSNs. Kahanda et al. [90] propose using a transactional information to predict tie strength in social networks. The model is constructed using 50 variables. The study lacks an evaluation with humans; therefore, the accuracy of the results may have been affected as the tie strength is a purely human-dependent concept. Xiang

et al. [91] proposed a model to infer relationship strength based on profile similarity, with the goal of automatically distinguishing strong relationships from weak ones. The model proposed by Xiang et al. uses the concept of homophily to infer the tie strength between two individuals. Xiang et al. test their model with proprietary data of the LinkedIn and data from students of Purdue University on Facebook. The main difficulty of this model to work accurately is that it relies on profile information, which tends to be incomplete or of low quality. Fogues et al. [44] introduce a tool called Best Friend Forever (BFF) that automatically groups and assigns a tie strength value for the contacts of a user. In order to infer tie strength values, BFF follows an approach similar to [39] and [90]. However, BFF uses a much smaller set variables, only 11. The reduction in the number of variables makes the variable collection task faster and less costly, thus increasing the utility of the tool. The authors made an experimental evaluation of the tool and compared the results obtained by their tool with the preferences of 17 participants. Despite the reduction of variables for tie strength prediction, the tool performed accurately. Lerone et al. [29] introduced interaction count based approach to determine relationship strength. In this approach, the authors simply take into consideration three types of interactions and count them in order to calculate relationship strength. This model is not based on semantic web approach. It is very simple and does not differentiate between interactions on the basis of their role in developing relational ties. Waqar et al. [30] extend work of Lerone et al. by applying data mining approach to calculate relationship strength for online social networks, Whereas, this data mining model is not validated on real OSNs data. Xiang et al. [91] propose a model to infer relationship strength based on profile similarity and interaction activity. The authors compute three features to determine profile similarity. These features are a common group, common friends, and logarithms of the normalized counts of common networks. In addition to profile similarity features, the authors consider wall posting and photo tagging for interaction activity. Lizi et al. [92] propose interaction rank-

ing based trustworthy friend recommendation model. This model is able to effectively recommend trustworthy friends to community members by taking into consideration four interaction attributes: reply frequency, comment length, time difference, and domain similarity. Another interesting work by the authors [93] propose trust ranking based recommendation model for suggesting the most trustworthy community members. The authors investigate four new interaction attributes that influence trust in virtual communities. These interaction attributes are interaction quality, seriousness in interactions, consistency over a long period, and common interest. A recent work related to friend recommendation is done by Zhao et al. [94], this model propose scalable and explainable friend recommendation model for social network systems. This model takes multiple relationship factors into accounts such as common friends, common followed users, common followers, and common joined groups of the target user and the candidate for friend recommendation. Another interesting fact demonstrated by Frank et al.[95] that more users are willing to divulge personal details to an adversary if there is a mutual friend connected to the adversary and the user. Christo et al. [96] show that users tend to interact mostly with a small subset of friends, often having no interactions with up to 50 percent of their friends. The authors suggest a model for representing user relationships based on user interactions. Existing research literature supports our idea that all friends should not be given equal access to user personal information, but access to personal information should be administrated based on relationship strength among online social network users.

2.4.3 Social Identity Management

One of the first research studies on audience segregation was conducted by Leenes et al. [35, 34]. The authors develop an experimental online social network prototype known as the Clique. The Clique is inspired from Goffman's theory of presentation of self and offers the mechanism for audience segregation. The Clique required the users to create

faces (profiles) and collections (friend lists) for various contexts. The Clique required the users to invest energy and time to perform audience segregation. Some of the online social networks also offer a very simple model of audience segregation by providing features of friend lists and circles. The majority of the users ignore such features due to an extra cognitive burden of configurations. There are also research studies which propose automatic techniques to perform audience segregation. Adu-Oppong et al. [97] have proposed partitioning a user's friends into lists based on communities extracted automatically from the network, as a way to simplify the specification of privacy policies. Squicciarini et al. [98, 99] propose an approach to facilitate online social network users to group their contacts into social circles with common interests. The authors design a multi-criteria model that takes into account multiple aspects of user's profiles, and automatically groups each user's contacts into social circles with common characteristics. Fang et al. [100, 101] propose the privacy wizard for automatically grouping of the user contacts, the wizard considers community, profile, and activity features. Personal information disclosure is managed according to the groups created by the wizard. Kelley et al. [102] have done preliminary work towards investigating how users create friend groups on Facebook. They have examined four different methods of friend grouping and their results show that the type of mechanism used, affects the groups created. Their findings lead to a number of recommendations for designing group-based privacy controls for online social networks.

2.5 Concluding Remarks

The conceptual background of online social networks is presented in this chapter. The purpose served by this introductory information is that it familiarizes the readers with important terminology, which is used throughout the dissertation. We start the chapter with various definitions of OSNs available in scientific literature. We also discuss

historical developments and evolution of the online social networks. The description of the differences between offline and online social networks are explained in detail. The social semantic web is explained due to the reason that it is future of online social networks. Finally, we present extensive state of art on privacy risks associated with widespread usage of OSNs.

Chapter 3

Privacy: A Theoretical Framework for OSNs

In this chapter, we present multifaceted notion of privacy and describe its various dimensions. Our main focus is to present legal, social, and technical perspectives of privacy. We also define the notion of privacy in the context of online social networks and identify various challenges that are posed to information privacy by unprecedented rise of online social networks. We introduce a new research paradigm for privacy in online social networks that inherits some properties from classical social theories of Erving Goffman, Helen Nissenbaum, and Mark Granovetter. The paradigm presents an agreed understanding about nature and scope of the privacy problem in social web domain. Finally, we give an overview of our proposed privacy framework for online social networks. The framework addresses the multidimensional issue of privacy from multidisciplinary perspective and benefits from classical social theories. We also address a first research question in this chapter that deals with redefining privacy for social web users.

3.1 Privacy: A Multifaceted Concept

The concept of privacy has broad historical roots in sociology and anthropology. The scholarly contribution of American anthropologist Barrington Moore [103] are the basis of privacy in social and anthropological sciences. According to his concepts, the creation of private sphere ultimately results from the need to transgress social rules in a safe and socially accepted way, which is not disruptive to the whole society. In psychology, Freud distinguishes between the private and public realms that are used as one of the strategies by human civilization to deal with the burden of contemporary society. In political philosophy, the core concept of privacy lies in the negotiation of the boundary between internal and external spheres of human existence. The contemporary notion of privacy is associated with the concept of autonomy. With the emergence of online social networks, privacy is extensively debated topic in computer science. The concept of privacy is so intricate that there is no universal definition for it. Many different definitions have been put forth depending on the context of its use, but there is no consensus as to its meaning or value. Some researchers have defined privacy as a function of accessibility to persons, whereas others have defined it in terms of control over personal information. Some of the researchers even claimed that it no longer exists due to the invasion of mankind by modern technologies. To better understand the concept of privacy, it should be analyzed from different perspectives. There are three main perspectives from which the notion of privacy are commonly described and analyzed. In the following sections, we describe the notion of privacy from these perspectives.

3.1.1 Legal Perspective of Privacy

From the legal perspective, privacy is viewed as a “right” of an individual and as a matter of personal “freedom”. The origin of the right to privacy can be traced back to the

nineteenth century. In 1890, Warren and Brandeis [104] published “The Right to Privacy” an influential article that postulated a general common law right of privacy. The authors characterized privacy as “the freedom to be let alone”. This right to privacy relates to the modern concept of “informational self-determination” which emphasizes an individual’s right to control the collection and use of personal data. The information self-determination also reflects Westin’s description of privacy [105]. According to the author, privacy deals with the right of the individual to decide what information about himself should be communicated to others and under what circumstances. The author also describes four states of privacy: solitude, intimacy, anonymity, and reserve. Due to remarkable growth experienced by online social networks in the last decade, the personal information of an individual is easily accessible to corporations, governments, and other individuals. One of the reasons for rising concerns over privacy is the way that information is being handled by service providers and third parties. It undermines user privacy and cast many doubts about Fair Information Practices (FIPs). The overall purpose of FIPs is to ensure that a user will maintain control over his personal information when it is in the hands of organizations. Many of the data protection laws were inspired from FIPs and impose a complex set of data management requirements and end user rights. In the following section, we present a brief overview of the current legal frameworks that ensures user’s constitutional right to privacy and protects the liberty of individuals to make certain crucial decisions regarding sharing of their personal information.

3.1.1.1 Current Legal Framework

The Legal framework for privacy differs around the world. The idea of privacy in the European Union has been legislated to a great degree. The European Commission enacted data protection directive (Directive 95/46/EC) in 1995 ¹. The EU data pro-

¹Data Protection Directive 95/46/EC <http://eur-lex.europa.eu/legal-content/EN/LSU/?uri=celex:31995L0046>

tection directive establishes a regulatory framework that strikes a balance between a high level of protection for the privacy of individuals and the free movement of personal data within the European Union. The directive is the key regulatory instrument adopted by the institutions within EU to regulate issues related to privacy. The Directive only applies to data controllers that either process personal data in the context of the activities of an establishment within the European Union, or makes use of equipment situated in the European Union. The data protection directive was supplemented by the ePrivacy Directive (Directive 2002/58/EC) to regulate the issues in the field of electronic communications. The ePrivacy Directive was drafted specifically to address the requirements of new digital technologies and ease the advance of electronic communications services. The ePrivacy Directive was amended by Directive 2009/136, which introduced several changes, especially in what concerns cookies, that are now subject to prior consent.

It is important to understand that data protection directive was enacted before the emergence of Web 2.0 and some issues with respect to online social networks were uncovered. The European Commission proposed replacing the 1995 directive with general data protection directive (GDPR) ². This new framework intends to strength and unify data protection for individuals within the European Union and reaches companies that target EU consumers from outside of the European Union. It will be enforced after a two years transition, beginning on May 25, 2018, replacing venerable 1995 EU data protection directive. The GDPR largely retains the principles and terminology of the 1995 directive. It also adds some new principles to better address contemporary privacy challenges posed by online social networks, cloud computing, big data, etc. The GDPR places onerous accountability obligations on data controllers to demonstrate compliance. This includes requiring them to implement data protection by design and by default which is missing in contemporary online social networks. Our proposed privacy

²GDPR: EU General Data Protection Regulation <http://www.infolawgroup.com/2016/05/articles/gdpr/gdpr-getting-ready-for-the-new-eu-general-data-protection-regulation/>

framework also emphasized the need for data minimization. One of the main ambitions of European Commission in proposing a new data protection framework was to bolster the rights of individuals. One of the rights that received the most attention is the “right to be forgotten” that allows individuals to require the data controller (OSNs) to erase their personal data without undue delay. As discussed earlier that online social networks provide the feature of persistence due to this personal information of the users are accessible beyond its temporal bounds. The new obligation imposed in GDPR will facilitate user to manage their temporal boundary in an efficient manner.

Privacy in the United States is not governed by legal writ to the same degree as in the European Union. There is no single regulatory authority dedicated to overseeing data protection law in the United States. The United States does not have a dedicated data protection law. The United States data protection framework resembles a patchwork quilt. One of the such regulations is privacy act of 1974 ³ which established a code of fair information practice that governs the collection, maintenance, use, and dissemination of personally identifiable information about individuals that is maintained in systems of records by federal agencies. Another example is the electronic communication privacy act (ECPA) ⁴ which imposes criminal sanctions for interception of electronic communications. However, the loopholes are so large as to render the act effectively meaningless. As a matter of fact, privacy practices are policed in the United States in a reactive manner by agencies which investigate corporate privacy behaviors that are potentially unfair or deceptive. Yao [106] studies data protection legislation in China and according to his findings only data protection laws in China had been the local laws until recently. A research study conducted by Kumaraguru reveals that India lacks legal protections for privacy [107]. Caruana et al. conducted a comparative study about data protection legislation between Islamic countries and EU. The authors found that EU is substantially more privacy protective than that of those countries

³Privacy Act of 1974 <https://www.justice.gov/opcl/privacy-act-1974>

⁴ECPA 1986 <http://www.it.ojp.gov/PrivacyLiberty/authorities/statutes/1285>

[108]. In a world with around 200 countries, more than forty countries have enacted major data protection laws [109]. We conclude that the data protection framework of European Union is up to date and covers most of the contemporary issues related to end-user privacy.

3.1.2 Social Perspective of Privacy

From the social perspective, privacy is viewed as a “socially constructed” behavior of an individual during their everyday social interactions. The social aspects of privacy have psychological and cultural roots. Privacy manifests itself differently in different cultures depending on social norms and cultural value [110]. This dimension of privacy focuses on managing social relationships and the boundaries between private and public life. According to Altman [111], the process of privacy management is a “dialectic and dynamic boundary regulation process” which allows an individual to exercise “selective control of access to the self”. The Goffman [14] conceptualizes privacy as an attempt to “impression management” where identities are constructed depending on the social role of the individual. An individual performs multiple and possibly conflicting social roles in everyday life, and he attempts to present consistent and coherent self-image in the specific social context. Nissenbaum [16] describes privacy as a “contextual integrity”. She argues that personal information is published within a well-defined social context and privacy is all about respecting the contextual boundedness of the shared personal information. Privacy is breached if the personal information is available outside its intended social context. These social theories form the basis of our theoretical framework and their detail description is given in the section on privacy research paradigm.

3.1.3 Technical Perspective of Privacy

From the technical perspective, privacy is viewed as a functional characteristic of digital systems. This aspect of privacy is concerned with how the legal and social considera-

tions can be represented formally and implemented practically in an operational system. The social norms and privacy policies are translated into technical specifications. The instances of such technical specification include code for fair information practice and platform for privacy preference project. These are popular examples which enhance the ability of an individual to control personal information disclosure by technical means. These three privacy perspectives are not mutually exclusive but interdependent. Privacy legislation can be enacted on the basis of social norms and social interactions can be altered due to changing technology.

3.2 Defining Privacy for OSNs

With the emergence of online social networks, a new debate started about the meaning and value of privacy. According to some researchers, privacy has been undermined by online social networks, even some of them claim that it no longer exist.⁵ The notion of privacy is hard to define particularly for online social networks where users voluntarily disclose personal information as part of their social activity. Due to the exponential growth of online social networks, sharing personal content on the web has gained acceptance and become a routine behavior for millions of the users. The level and depth of personal information disclosure has raised serious concerns about privacy. The migration from physical to digital environments has changed the traditional approach to privacy altogether. Current definitions of privacy focus on privacy as process or capability for social situations rather than privacy being an attribute attached to the particular information. In this section, we present various definitions of information privacy from the online social network perspective. According to Kang [112], the information privacy is an individual's claim to control the terms under which personal information is acquired, disclosed or used. Palen et al. [113] view information privacy as a state

⁵Do Social Networks Brings the End of Privacy? <http://www.scientificamerican.com/article/do-social-networks-bring/>

of social withdrawal which is quite undesirable in today's information society. More privacy is not necessarily better because it leads to isolation which is not the desirable state. In fact, information privacy is a delicate act of balancing between disclosure and concealment that allows users to interact with one another socially. The recent focus of privacy research shifted towards the struggle to control one's own self-disclosure. The authors also suggest three boundaries of information privacy with which OSNs users are struggling.

Disclosure Boundary Managing the tension between what is private and what is public

Identity Boundary Managing self-presentation with specific audience

Temporal Boundary Managing past action with future expectations

The users have a scope in mind when they upload personal information in online social networks. This scope is defined by disclosure, identity, and temporal boundaries. The privacy is breached when information is moved beyond its intended scope either accidentally or maliciously. Simply a breach can occur when information is shared with a party for whom it was not intended, it can also happen when information is abused for a different purpose than was intended, or when information is accessed after its intended lifetime. The detailed discussion of these boundaries is presented in the section on shifting privacy research paradigm for online social networks. Anwar [114] puts forward relatively similar definition for information privacy, where author expands on widely perceived notion of privacy as control over personal information. According to him, information privacy boils down to control over three aspects of personal information: flow, boundary, and persistence. The flow is defined as the act of sharing information with multiple stakeholders. The boundary of information is defined as the scope or realm within which shared information to be used. Persistence of information is defined as the period of time shared information is available to or usable by the

stakeholders with whom it is shared. The flow of information is a unique characteristic of Anwer's definition and boundary and persistence of information are concerned for privacy only when the flow of information takes places.

Gurses et al. [115] view privacy in online social networks from a different perspective. According to the author's negotiation of boundaries address only one type of the privacy problems in online social networks, whereas the researchers highlight two other type of privacy problems which fall out of the scope of boundary regulation. Contemporary online social networks raise three types of privacy problem which are described as follows:

Surveillance Privacy: This problem arises when the personal information and social interactions of OSN users are leveraged by government and service providers.

Social Privacy: This problem emerges through the necessary renegotiation of boundaries as social interactions get mediated by OSN services.

Institutional Privacy: This problem is related to users losing control and oversight over the collection and processing of their information in OSNs.

The surveillance and institutional privacy problems are out of the scope of any technical solutions and require a strict legal framework which ensures data protection of the OSNs users. The main focus of this research is social privacy problem which relates to the concerns that users raise and harm that they experience when technologically mediated communication disrupt social boundaries. The authors stress on enabling appropriate privacy practices to respect identity, disclosure, and temporal boundaries suggested by Palen et al [113]. A comprehensive characterization of social web privacy from multidisciplinary and multi-party perspectives is proposed by Netter et al. [116]. The authors break the concept into a set of characteristics that are subsequently used to conduct privacy impact analysis. The detail description of each characteristic is given below:

Audience Segregation This characteristic describes that each individual performs multiple and possibly conflicting roles in everyday life and it needs to segregate the audience for each role, in a way that people from one audience cannot witness a role performance that is intended for another audience. In current online social networks almost all friends are treated equally, As a result, privacy is threatened because a large audience might have access to personal information.

Data Sovereignty It describes to what extent an individual is able to control the processing of its personal data. In the case of online social networks, personal data is available in a structured manner. It can easily be copied, linked, aggregated, and transferred. The problem increases as OSNs typically lack the spatial, social, temporal boundaries of the real world which limits the flow of personal information by default.

Data Transience This characteristic revolves around the loss of personal information over the time. In computer-mediated communication permanency of personal information poses the great challenge to privacy, whereas data transience can be considered as typical characteristic of real world communication.

Transparency It describes transparency of processing and dissemination practices. Taking the social point of view, transparency implies an individual's possibility to recognize contextual boundaries, which is important to contextual integrity.

Protection against profiling It describes an individual's ability to prevent an adversary from collecting, aggregating and linking personal data in order to create a digital dossier. The current landscape of online social networks poses this threat at large scale.

Privacy Awareness It describes that an individual's awareness of privacy risk is a prerequisite for privacy-preserving behavior.

Enforcement It describes an individual's means to bring privacy preference into force.

The careful study of characteristics such as audience segregation, data sovereignty, data transience, and transparency reveals that lack of self-presentation, spatial, social, temporal, and contextual boundaries are the root causes of all privacy problems in online social networks. The definition of privacy adopted within this dissertation is inspired by work of Pfitzmann et al. [117]. The authors view privacy as a three component concept which includes data minimization, user control, and contextual integrity. The authors term this concept as a Privacy 3.0 suitable for Web 3.0. The detail description of each component is given below:

Data Minimization: Data minimization is one of the main motivations for the development of privacy-enhancing technologies which aim to limit collection and processing of personal data by data controllers.

User Control: User control of personal information disclosure supports users in deciding which personal information is released to whom and in which situation.

Contextual Integrity: Contextual integrity provides a new quality of privacy by making the original context in which particular personal data have been generated easily accessible to all entities that are aware of that particular personal data.

The authors argue that traditional approach of data minimization is not always feasible and certainly not in every situation. The user control of personal information disclosure is also not suitable for ubiquitous computing. Thus, the objective of contextual integrity is to ensure the protection of communicated information from decontextualization. The differentiation into three component is mainly driven by the constraints of the historical evolution of information technology. The data minimization is to be understood as the traditional driving concept of the field of privacy enhancing technologies (PET) in

the early 1980s. In 1990s, social interaction supporting technology achieved a level of sophistication and mutual interdependence that it required more disclosure of personal data and needed fine-grained control over disclosed data. Thus, the user control over personal data disclosure was inevitable. None of the two mentioned characteristics can be fulfilled in the field of ubiquitous computing, thus contextual integrity protects users from embarrassment by controlling disclosure of personal information out of the communicated context.

We customize this definition of privacy to suit the needs of social web users. The social web users face three major problems in their effort to manage their privacy in on-line social networks. These privacy problems are an invisible audience, interdependent privacy and context collapse. The invisible audience refers to the fact that all audience are not visible and co-present at the moment an individual user is generating digital content for online social networks. Interdependent privacy refers to the phenomenon that privacy of individual user could be out of their own control and affected by the decisions of other connected users. Context collapse refers to difficulty in disclosing personal information selectively to various life facets. Context collapse makes it difficult for people to use the same techniques online that they do to handle multiplicity in face-to-face conversation

Our definition of privacy revolves around the three-component approach of Pfizmann et al. We address the issue of the invisible audience by minimizing disclosure to personal information of the user to first-degree contacts. This is the component in our definition termed as a disclosure minimization. The second component addresses the issue of interdependent privacy which means enhancing the ability of the users to control access to their content residing into the spaces of their friends. Finally, the issue of context collapse is resolved by preserving the contextual integrity of the users in the social web environment. These three components (disclosure minimization, user control, and contextual integrity) can be traced back to the philosophical discussion

of famous social theorists such as Goffman, Altman, and Nissenbaum. In the light of above definitions, we conclude that privacy in social web revolves around three parameters: self-presentation management, boundary regulation, and disclosure minimization. In the following section, we present a privacy research paradigm which is based on these features.

3.3 Shifting Privacy Research Paradigm for OSNs

In this section, we propose the paradigm shift in privacy research of online social networks. Most of the research in this area has ignored the importance of social aspects of privacy and existing solutions are struggling between isolation and crowding. The existing solutions trade off between sociability and privacy. The individuals compromise on their sociability to enhance their privacy. The contemporary online social networks encourage sociability and it is one of the main reasons for their exponential growth during last decade. Sociability is design goal of traditional online social networks and it has some inherent design conflicts with privacy and security. We identify three main factors for privacy research paradigm which brings privacy without compromising sociability in online social networks. This new privacy research paradigm is based on a set of agreed assumptions about the nature and scope of a privacy problem in social science. As we already discussed in chapter two that online social networks are the modern form of their offline counterpart with computer-mediated communication. The main principles for self-presentation and disclosure control in the online world remains the same as in the offline world. Thus, we benefit from principles of well-founded social theories about self-presentation, contextual integrity and boundary regulation to minimize disclosure of personally identifiable information of the users without compromising sociability. In the following subsections, we describe main factors of our paradigm in the context of online social networks. Hence, these three factors are not exhaustive,

but rather provide us with a way to approach privacy research in online social networks from a social perspective. For each of these factors, we describe agreed assumptions it relies on, with reference to relevant social theory.

3.3.1 Role and Relationship-based Self-Presentation

Erving Goffman’s social theory about the presentation of self [14] describes a process in which individuals make a series of conscious decisions regarding how to present themselves based on the audience (social context) with whom they are interacting at a given time. According to Goffman, each individual performs multiple and possibly conflicting roles during their interactions with others and it needs to segregate the audience for each role, in a way that people from one audience cannot witness a role performance that is intended for another audience. Goffman argues that an individual’s self-presentation varies based on their audience. Variations in self-presentation range from minor to significant depending on social context and relationship strength between an individual and his audience. One of the key elements of Goffman’s perspective on identity is the fact that individuals attempt to present self-images that both are consistent and coherent. To accomplish this, they engage in **audience segregation** that allows users to be “round characters” in different roles, rather than “flat ones” in a conflated context.

The concept of audience segregation was coined as part of a perspective on the way in which identities are constructed and expressed in interactions between human beings in everyday contexts. The users in online social networks lack such mechanism to separate and manage the various audiences for whom they perform. Current online social networks conflate different social groups into the singular notion of a friend. The information which is suitable for one social group may be entirely unsuitable for another social group. Difficulty in disclosing information selectively to various life facets can lead to context collapse. The collapsing of social contexts together has emerged as an

important problem with the rise of online social networks that so often blurs the public versus private, professional versus personal and many different selves and situations in which individuals find themselves. Context collapse is a serious privacy issue in online social networks. To address the issue of collapsing context without sacrificing sociality requires to model role and relationship based notion of social context for OSNs users.

The findings of number of researchers [118, 119, 48] suggest that Goffman’s original framework is of a great usefulness for understanding identity construction and presentation of self in online social networks. Applying role and relationship based audience segregation would improve the quality of interactions and self-presentation in OSNs. While Goffman’s idea of audience segregation didn’t originally relate directly to privacy, it is easy to see that audience segregation and privacy are closely linked. In fact, privacy revolves around person’s ability to keep audience separate and to compartmentalize his social life. With segregated audiences for the presentation of specific roles, each individual holds multiple partial identities (profiles) for different social contexts. A partial identity (profile) should adequately represent an individual in a specific role in a specific social context. SOCPRI model enables users to be selective in sharing personal information through analyzing role and relationship of information seeker and justifying contextual norms for information flow activity. The model defines three types of user-centric roles which are commonly performed by each individual user on daily basis and models appropriate profile subset to accommodate information needs of the specific user-centric role. This model classifies, infer and store rich social context information.

3.3.2 Realizing Contextual Integrity

Nissenbaum’s theory of contextual integrity supports Goffman concept of audience segregation. She argues that audience segregation revolves around contextual integrity and goal of audience segregation can be accomplished by preserving contextual integrity. The theory of contextual integrity also provides a framework for understanding privacy

implications of recent developments in OSNs and offer useful conceptual apparatus for designing solutions to mitigate these problems [36]. Nissenbaum’s account of privacy as contextual integrity is based on two non-controversial facts. First, every transfer of personal information happens in a certain social context and all areas of life are governed by context-specific norms of information flow. For example, it is appropriate to tell one’s doctor all about one’s medical history, but it is most likely inappropriate to share that information with strangers. Second, people move among a plurality of distinct social contexts every day, as they move between different social contexts, they have to alter their behaviors to correspond with norms of those contexts, aware to the fact that information appropriately shared in one context becomes inappropriately shared into a context with different norms. For example, it is appropriate to share romantic details with friends, but sharing such details with the employer is out of place.

According to Nissenbaum, there is no such thing as context-free information; the information is always associated with the context in which it is revealed. She also emphasizes that there is no such thing as universal privacy norms; the scope of privacy norms is always internal to a context. On the basis of these facts, she suggests that contextual integrity is maintained when informational norms are upheld and compromised when such norms are breached. She proposes contextual integrity as benchmarks for privacy. Informational norms are of two types: norms of appropriateness and norms of distribution. Norms of appropriateness determine whether a given type of personal information is either appropriate or inappropriate to divulge within a particular context. Norms of distribution restrict the flow of information within and across contexts. The four parameters are proposed to determine informational norms: **social contexts** are characterized by assemblages of roles and set of behaviour guiding norms; **actors** are stakeholders such sender, recipient, and subject of the information (whom the information is about); **attributes** are defined as the type (category, nature, class) of

information in the information flow; **transmission principles** are the constraints to the information flow from one party to another in a given social context.

Nissenbaum also stresses that these norms are implicitly engrained into everyday life and negotiating such norms is a normal part of everyday life. However, online social networks make such negotiation complicated because of binary nature of friendship relationship. The friendship is a binary relational tie, which provides only a coarse indication of the nature of the relationship. In everyday life, relationships are much more complicated than a single binary relational tie. Online social network users interact with people representing various facets of their life such as family, friends, colleagues, classmates etc. In such a scenario, it is essential for users to be able to distinguish between these different types of contacts and determine the quality of relationships. The quality of the relationship can be easily inferred by relationship context and strength. In the following section, we describe in detail the concept of relationship strength and how to model it for online social networks.

3.3.3 Modeling Relationship Strength

Tie strength is one of the most influential concepts in sociology. Mark Granovetter introduced the concept of tie strength in 1973 and defined tie strength as follows [37]:

“The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.”

Granovetter characterized two types of ties: strong and weak. Strong ties are the people you really trust, people whose social circles tightly overlap with your own. Often, they are also the people most like you (i.e., homophily). Weak ties, conversely, are merely acquaintances. Granovetter identified four dimensions of tie strength: amount of time, intimacy, emotional intensity, and reciprocal services [37]. Subsequently, several

researchers have also attempted to identify other dimensions of the tie strength between people, such as structural dimension [120, 52, 121], emotional support dimension [122], and social distance dimension [120, 123]. In total, there are seven dimensions suggested in previous research studies that can be used to examine the strength of ties among individuals.

The findings of Petroczi et al. [38] suggest that tie strength indicators for online social networks are similar to those in offline social networks. Like offline ties, strong online ties have the same dimensions of tie strength. All tie strength dimensions can be easily inferred from user interactions pattern in online social networks. Research also suggests that tie strength affects the nature and frequency of online interactions between OSNs users [124, 121]. The individuals with strong ties interact more frequently and intimately [15, 125], sharing more information, revealing more about them, supporting each other emotionally and reciprocally, and committing more time for each other [15, 125, 122]. Whereas, weak ties interact less frequently and are less intimate than strong ties [15, 125]. Online social network users have a finite amount of time to use in forming and maintaining relationships. They invest this time towards relationships that they consider important [126]. Therefore, the nature and frequency of online interactions between users depend on the strength of their relationship [124, 121].

The user's interaction pattern in online social networks can be utilized to examine tie strength between users. According to Gilbert et al. [39], interaction activities on social network sites may help in predicting tie strength. Several attempts have been made to find a valid set of predictive variables to infer tie strength on the basis of available interaction methods in contemporary online social networks. Gilbert et al. [39] identified 74 predictive variables, Kahanda et al. [90] identified 50 variables, Spiliotopoulos et al. [127] identified 18 variables, and Xiang et al. [121] identified three variables for tie strength prediction. The findings of Stutzman et al. [128], explicitly mention that strong ties are composed of family members and close friends, whereas

weak ties include casual friends and acquaintances. These studies support the validity of our idea about establishing a link between relationship strength and user interactions pattern in online social networks.

The focus of this dissertation is how to model tie strength for online social networks. We also explore whether modeling tie strength can play some role in preserving the privacy of OSNs users, specifically how tie strength can be useful for regulating the flow of information across strong and weak ties. As opposed to real life, every interaction is recorded in online social networks. We show that one can extract meaningfully strong and weak ties from recorded interaction histories of users in OSNs. We reviewed a large body of research literature on tie strength and concluded that each of the seven dimension of tie strength (describe earlier in this section) can be identified by several interaction methods available in current online social networks. Given below is the detailed description of all tie strength dimension along with predictive variables related to each tie strength dimension.

The amount of time dimension refers to the duration of the relationship between friends. The history of communication among friends plays a vital role to infer this duration in online social networks. The predictive variables for inferring this duration are first communication (message, comment or tag to approximate the duration of friendship) and last time interaction among friends. The dimension of intimacy refers to deep affection between two friends acting as a sense of reliance and security. The level of intimacy among online social network users can be inferred from predictive variables such as the recency of communication, relationship status in common, appearances together in photos (tagged in photos), and listed in the same check-in with friends. The dimension of intensity refers frequency of interaction between friends. People with high intensive relationships will spend more time together and produce longer communication history than people with less intensive relationships. The predictive variables related to this dimension are wall posts frequency, comments frequency, and

private message frequency.

The reciprocal services dimension refers communication-related to sharing information and resources with relational ties. The strong ties include more reciprocal services in exchange than weak ties. The reciprocal services dimension on the online social network can be inferred by links exchanged through wall post and applications. The social distance dimension is the difference in socio-economic status, educational level, political affiliation, race, and gender. The profile information such as identity information, language setting, political and religious affiliations, work and educational history can be used to infer the social distance dimension. Finally, the structural dimension referred as a function of social homogeneity, shared affiliations, and overlap of social circles. The predictive variables used to infer this dimension are mutual friends, common interests, listed in the same group, and the relation between the participant and the friend. So far, we described tie strength dimensions and their predictive variables for current online social networks. SOCPRI models these dimensions and facilitates characterization of strong and weak ties by identifying interaction pattern and profile similarity with referring to these predictive variables associated with current online social networks.

3.4 Proposed Privacy Framework for OSNs

The social perspective of privacy in online social networks focuses on managing self-presentation, preserving the contextual integrity and maintaining the balance between privacy and publicity. We propose a 3C segregation privacy framework which addresses these aspects of social privacy. Main problems faced by online social networks today are collapsed context, conflated contacts, and co-joined content (public and private). Contemporary online social networks provide their users with single Timeline/Wall for all the contacts which represent single universal context for all kinds of contacts. The

single user profile represents all kinds of personal information irrespective of their sensitivity. The social graph of the user represents all kinds of contacts without taking into consideration tie strength among the user and his contacts. Our proposed framework performs contexts, contacts, content segregation. It also addresses aforementioned issues of online social networks related to collapsed context, conflated contacts, and co-joint content.

3.4.1 Context Segregation

Context collapse is a serious privacy issue in online social networks. The notion of contextual integrity resolves the issue of privacy if applied to online social networks. Contemporary online social networks conflate different social groups into the singular notion of a friend. The information which is suitable for one social group may be entirely unsuitable for another social group. To address the issue of collapsing context without sacrificing sociality requires to perform context segregation in online social networks. Context segregation gives a user ability to compartmentalize their social life. The relationship network of OSNs users is diverse in nature and users play several roles across different social contexts. Ozenc et al. [129] identified that three social contexts are very common among all OSNs users and needed segregation of these social contexts for better social experience in online social networks.

1. **Family:** This context refers to relatives and can be inferred by analyzing profile attributes such as relationship status.
2. **Work:** This context refers to professional circle and can be inferred by analyzing profile attributes such as present and past work affiliations.
3. **Social:** This context refers to friends and can be inferred by analyzing profile attributes such as educational background, interests etc.

Context segregation allows coarse-grained separation of audiences on the basis of relationship context. Each context includes a large number of friends of varying relationship strengths; further segregation within the context is performed on basis of relationship quality, which provides fine-grained separation of audience. The main factor which facilitates this type of segregation is user role at a particular point in time. Several roles are associated with each individual users in the online environment and switching between these roles is management by a role in time concept.

3.4.2 Contact Segregation

Most online social networks employ "friendship" as the only type of bidirectional relationship. The friendship is a binary relational tie which provides only a coarse indication of the nature of the relationship. In human societies, relationships are much more complicated than a single binary relational tie. This demonstrates that OSNs carry problematic assumptions in their implicit design of forming relationships. All friends are created equal which means they have equal access to all personal information of the user. Online social network users interact with people representing various facets of their life such as family, friends, colleagues, classmates etc. In such a scenario, it is essential for users to be able to distinguish between these different types of contacts and determine the quality of relationships. Granovetter's concept of tie strength [37] can play a vital role to determine the quality of relationships. Tie strength is one of the most influential concepts in sociology. The existing literature of sociology suggests seven dimensions of tie strength: communication intensity, intimacy, relationship duration, social distance, emotional support, reciprocal services and structural dimension. According to Petroczi et al. [38], the relationship indicators in online social networks are similar to those in offline communities. All tie strength dimensions can be easily inferred from user interactions pattern and profile similarity attributes in existing online social networks [39, 40]. Research proves that mixture of contextual grouping and

tie strength could allow appropriate sharing of personal information [41]. Our privacy framework also adopts contexts segregation to perform coarse-grained segregation of users' social graph into contextual grouping. Further, fine-grained segregation is performed on the basis of tie strength which minimizes disclosure of personal information

3.4.3 Content Segregation

Online social network users upload huge amount of personal information in their social networks sites. Mostly, this information is accessible to all friends without taking into consideration factors such as social context and tie strength. Our framework proposes content segregation that allows identification of sensitive personal information and restricts its disclosure to only intimate friends. We divided this content into a context free and context sensitive categories. Context-free content is accessible to all friends of a user without access restrictions, whereas, the disclosure of context sensitive content is managed through contextual norms. The contextual norms are divided into two categories: norms of appropriateness and norms of distribution. As discussed earlier, norms of appropriateness determine whether a given type of personal information is either appropriate or inappropriate to divulge within a particular context. Norms of distribution restrict the flow of information within and across contexts. We also divided user profile in default profile and contextual profiles. The default profile contains context-free content. The contextual user profiles contain context-sensitive content and regulated through contextual norms.

3.5 Concluding Remarks

In this chapter, we explain the multifaceted notion of privacy for online social networks from legal, technical, and social perspectives. We identify various challenges posed to information privacy by unprecedented rise of online social networks. We present

a theoretical framework for privacy that addresses the emerging issues in online social networks such as context collapse, disclosure minimization, and user control. The framework addresses the multidimensional issues of privacy from multidisciplinary perspective and benefits from classical social theories. We address a first research question in this chapter that deals with redefining privacy for social web users.

Chapter 4

Privacy: OSNs User Perspective

In this chapter, we present the study that was conducted to understand the behavior of users about personal information disclosure and relationship formation in online social networks. The first section discusses the background and purpose of the study and formulates research questions. We describe the methodology of the user study in the second section. The section presents the content of research questionnaire along with data collection method. The results of the user study are presented in the third section. These results give insights into some of the current trends on online relationship formation and personal information disclosure of the OSNs users. The fourth section revisits research questions in the light of the results and highlights some of the implications. Finally, we conclude this chapter by presenting some of the limitations of this user study. The contribution of this chapter is to provide answers for second research question that deals with identifying the link between OSNs users interaction patterns and personal information disclosure practices.

4.1 Background and Purpose of User Study

Various surveys have been conducted on privacy concerns of OSN users. One of the first qualitative studies on privacy issues in online social networks was conducted at Carnegie Mellon University which analyzed 4,000 profiles of CMU students. As per result of this research user of online social networks disclose the huge amount of personal information in their profiles. Online social network users are less concerned to use site's privacy settings to control the visibility of the profile; only 0.06% out of 4,000 users changed the default profile visibility in Facebook [66]. Most of the online social networks provide very permissive default settings and only very few users change their default settings as per the results of above-mentioned research at CMU.

A qualitative research conducted by the Office of Communication of United Kingdom showed that concerns about privacy and safety are not top of mind for most of the users. This research also demonstrated that users create well-developed profiles as the basis of their online presence, and share personal information with a wide range of friends [130]. Gross [66, 67] showed in a case study that most users do not change the default privacy settings as provided by the OSN. Furthermore, these users share a large amount of information on their profile. Madejski [22] shows that privacy settings for uploaded content are often incorrect, failing to match user's expectations. Tufecki [131] investigated the relationship between user's privacy concerns and their level of disclosure on an OSNs and found no relationship. Even users who expressed many privacy concerns divulged large amounts of personal information on their profiles. However, the study only asked about the relatively static fields of a profile like age, sex, gender, interests, and favorite books, rather than concerns over dynamic content (e.g., status updates, comments, etc.). He concludes in his case study that privacy-aware users are more reluctant to join social networks. However once a privacy-aware user joins he is willing to disclose a lot of information and a user's privacy is regulated mostly through

visibility, i.e. the privacy settings of the OSN. This privacy-aware user aims to remain in control. Furthermore, users are more pre-occupied with the current visibility of their information and do not look towards future implications. It seems that users implicitly trust social network providers to handle user data in a fair and conscientious way.

Krasnova [132] held focus groups with university students to identify categories of privacy concerns about their Facebook use. The most frequent theme was concern over unwanted audiences viewing shared content, where the list of audiences mentioned included future employers, supervisors, family members, peers, and subordinates. Participants also frequently mentioned organizational threats related to the collection and use of their data by the OSNs provider and third parties. Concerns about social threats were another common theme for concerns including people purposefully posting content to harm the individual and general concern over a lack of control over the actions of other users. Lampinen [133] conducted 20 semi-structured interviews in order to understand user behavior to manage heterogeneous friends network. He reported that many users fear that a boss or acquaintance might see something embarrassing that was not intended for them, and that users attempt to avoid these situations through self-censorship and using context to carefully selecting a suitable communication medium. Skeels [134] also studied the dynamics of group co-presence, but focused on SNS usage in the workplace, and found that users have trouble coping with the co-presence of coworkers and other contacts in an OSNs friends network. Many participants noted the burden associated with constantly maintaining an awareness that the two groups are present in their audience. Participants also noted the need of limit access to select content based on relationship.

Online social network users apply several strategies for mitigating their privacy concerns. Young [135] identified boundary regulation mechanisms that include deleting tags, and using direct messages to limit audiences. Stutzman [128] found that users who employed supplemental privacy preserving behaviors, like curating the posts on their

wall and collaboratively adjusting OSNs behavior among friends, were more likely to have friends-only profile. Several papers have reported that users cope with conflicting social spheres by maintaining separate profiles, limiting access to subsets of the friend network, carefully selecting a communication medium, or using separate OSNs for different audiences [136, 134]. PEW Internet reports that in 2011, 63% of Facebook users had removed someone from their friend network [72], an increase compared to the 56% of users who reported to have unfriended someone in 2009. The same survey found deleting and untagging posts to be common among all user demographics.

The prior works leave an important question unanswered how social interactions of users determine relationship strength and how relationship strength can be utilized to control information disclosure in online social networks. We conduct a user study that is focused on identifying the relationship between user interaction patterns and personal information disclosure practices. More specifically, we want to explore whether a user's interaction with his/her friends can be used as a basis for making data access decision for that user. More specifically, we analyzed how interaction frequency and choice of interaction type reveal the relationship strength and how it plays a vital role in controlling personal information disclosure in online social networks. We explored whether a user's interaction with his/her friends can be used as a basis for making data access decision for the user. The results were used to validate the following hypotheses about privacy and interaction patterns of the social web users.

- H.1.** Personal information disclosure depends on relationship strength among the users.
- H.2.** Relationship strength depends on the frequency of interactions among the users.
- H.3.** Choice of the interaction type for communication with audience depends on relationship strength

The majority of users has a large friends network (social graph) in online social networks consisting of more than few hundred friends, but their interaction network (interaction

graph) is found to be very small. The users prefer to share their personal information with their interaction network rather than a large friends network. Users are very selective in choosing the type of interaction in online social networks. The findings reveal that the choice of interaction type also gives an indication of the relationship strength. Additionally, the findings facilitate categorization of profile information and user interactions on the basis of sensitivity and frequency respectively.

4.2 User Study Methodology

This section outlines our method for conducting the user study and presents the detailed content of the questionnaire. Subsequently, we describe the method for recruiting participants for the user study. The findings of our study are discussed in next section.

4.2.1 User Study Design

The online survey was designed to examine privacy concerns and interaction patterns of users in online social networks. It is the most widely used method to reach a large audience easily. Moreover, It helped us to collect data globally. GoogleForms were used to develop, disseminate and collect user data. It is important to note that all data collected for this study was based on users' informed consent. The participants had the choice to skip any part of the survey, in case, they felt it was asking for sensitive personal information, though in designing the questionnaire, we took special consideration not to ask any personally identifiable information from participants. The questionnaire is divided into six sections. The first section asks the qualifying question to be able to continue filling the online survey, including getting informed consent and collecting demographic data. The second section collects data concerning the privacy of online social network users. The third section gets data concerning user social relationships in online social networks. The fourth section collects data about user communication and

interaction patterns with his/her online connections. The fifth section of the study is designed to collect data to rank various social interactions on the basis of their weight in developing relational ties. The last section collects data to rank user profile information on the basis of information sensitivity.

4.2.1.1 Survey Content

The survey contained 37 questions of various types which include ordinal-scale, close ended and open ended questions. The demographic section of the survey contains six questions. The first question in the survey is about informed consent of the user to participate in the study. The user is informed about the purpose of the study and personal information required to participate. The user can proceed with the questionnaire only if his response is positive. Demographic information (age, gender, nationality) is the only personally identifiable information that it is not required, but “asked” by the user in the questionnaire. The study is designed for active online social network users. The participants with a negative response to the question about having at least one OSN account are diverted to the end of the survey. The last question in this section checks the active involvement of the users by asking about their usage frequency of online social networking sites. The actual survey starts with a section on privacy. This section is designed to examine privacy concerns of OSN users and their attitude towards using existing privacy controls. The questions asked in this section are about reading the privacy policy, using the privacy settings, user friendliness of the privacy controls, and Likert scale question about the privacy concern of OSN users.

The section on social relations is designed to study relationship forming behavior of OSN users. We categorized relationships into six types such family, friends, colleagues, classmates, acquaintances, and strangers. The questions asked in this section are focused on studying user relationship forming patterns and information sharing practices with the six categories of people. The social interactions section is focused on studying the

frequency of the user interactions with people from the six categories. Additionally, we divided social interactions into eight categories such as messaging, posting, tagging, commenting, liking, birthday wishing, chatting, and playing games. Questions intended to examine users' preferred interaction type with people from the six categories of relationships are also included. The next section investigates weightage for various types of interactions based on their frequency of the usage.

The profile section of the questionnaire is designed to check the accuracy of personal information provided in user profiles in online social networks. The final section of the questionnaire is about ranking the profile information on the basis of information sensitivity. We sliced profile information of OSN users into nine categories such as basic information, education and employment information, activity streams, photos and videos, family and relationship information, affiliations information, events information, preference information, and religious and political views. The users were asked to provide a ranking for these categories depending on the information sensitivity.

4.2.1.2 Data Collection

Data were collected through an online questionnaire which was circulated via numerous university mailing lists and postings in popular OSN groups. The survey targeted Facebook and Google+ users. Participant had to be an active OSNs user and over 18. The main requirement to participate in the study was at least having one account with any social networking site. The responses were collected from May to August 2015, with an overall gross sampling consisting of 334 participants. After deleting the responses that were unusable due to excessive missing data and their non-explicit answer to the question about informed consent to participate in the study, a final net sample of 323 participants was obtained. Young and active users of online social networks dominate our sample. Our sample consists of participants from 20 different countries of the world.

4.3 Analysis of Results

We analyze the demographics of the participants in this section. The results of this analysis are shown in Table 4.1. The total number of participants that took part in the study are 323, out of which 245 are men and 81 are women, that leads to male bias. 69% of the participants belongs to the age group of “between 20 to 30”, and 26% of the participants belongs to the age group of “between 31 to 40”. There are also few participants belonging to either “below 20” or “above 40” age groups. Most of the participants are active OSNs users either constantly logged into their account on social networking sites (35%) or check their OSN account several times per day (37%). The geographical location of the majority of participants is from the Indian subcontinent, whereas a small number of participants are from Europe. It is one of the few limitations of our study for the generalisability of our results.

The usage frequency of the participants is further analyzed on the basis of gender and age group. The analysis results are presented in Table 4.2 and 4.3 respectively. We identified slight variations in the behavior of male and female participants with respect to the usage pattern. As discussed earlier, the majority of the participants are either constantly logged-in or login several times per day. Mostly, the fewer female participants are constantly logged-in as compared to male participants. The age group wise usage frequency reveals that 50% participants “below 20” are constantly logged-in, whereas, the smallest number of participants from age group “above 40” are constantly logged-in on social networking sites. The participants from age group “between 20 and 30” prefer to check their online social networking account several times a day. The usage frequency of participants from age group “between 30 and 40” reveal that they also check their OSNs accounts on daily basis. The usage frequency analysis either gender wise or age group wise demonstrate that number of the participant using OSNs rarely is negligible. The small number of participants login into their OSNs accounts

on weekly basis. The vast majority of the participants is active online social network users.

Table 4.1: Demographics of the Participants

Demographics	Category	Percentage
Gender	Male	75%
	Female	25%
Age	Below 20 years	2%
	Between 20 and 30 years	69%
	Between 31 and 40 years	26%
	Above 40 years	3%
Geographical Location	Asia	82%
	Europe	8.66%
	Africa	1.2%
	United States of America	0.61%
	Australia	0.30%
	Undeclared	6.81%
Frequency Of Usage	Constantly Logged-In	35%
	Several Times per Day	37%
	Nearly Everyday	23%
	At least Once a Week	3%
	Rarely	2%

4.3.1 User Attitude Towards Online Privacy

Our empirical study contains a section of four questions about user's attitude towards online privacy. In question 1, participants are asked to choose a response about reading the privacy policy of social networking site. 48% of the participants selected that they don't read privacy policy at all, whereas, 16% of the participants selected that they read

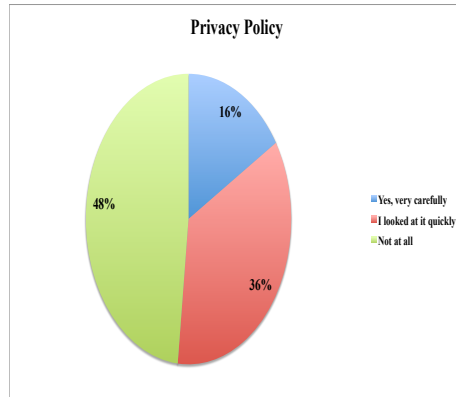
Table 4.2: Gender wise Usage Frequency of Participants

Item		Usage Frequency				
		Constantly Logged-In	Several Times per day	Nearly Everyday	At least, Once a week	Rarely
Gender	Female	33.33%	40.30%	23.07%	1.28%	1.28%
	Male	35.26%	36.09%	22.82%	4.14%	0.41%

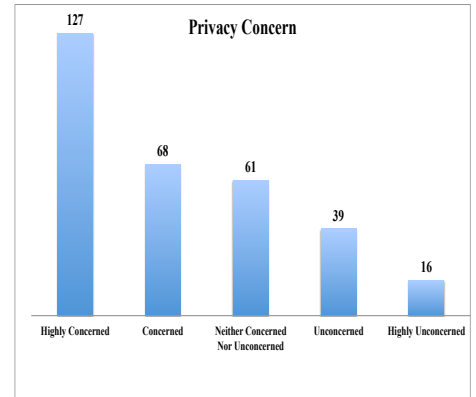
Table 4.3: Age group wise Usage Frequency of Participants

Item		Usage Frequency				
		Constantly Logged-In	Several Times per day	Nearly Everyday	At least, Once a week	Rarely
Age Group	Below 20	50%	16.66%	18.18%	0%	0%
	Between 20 and 30	34.97%	40.80%	21.07%	2.69%	0%
	Between 30 and 40	34.24%	29.26%	28.04%	4.87%	0%
	Above 40	18.18%	45.4%	18.18%	9.09%	9.09%

carefully and remaining 36% of the participants indicated that they just looked at it quickly (Fig.4.1a). The results of age group wise privacy policy awareness demonstrated that only small percentage of people from all age groups carefully read the privacy policy. Almost 50% participants from all age groups do not read the privacy policy at all. Remaining vast majority of people from age group “above 40” quickly look at the privacy policy of OSNs. The age group wise privacy policy awareness results are shown in Table 4.4. The results of gender wise privacy policy awareness reveal that female participants are more aware of the privacy policy as compared to male participants. The gender wise privacy policy results are shown in Table 4.5.



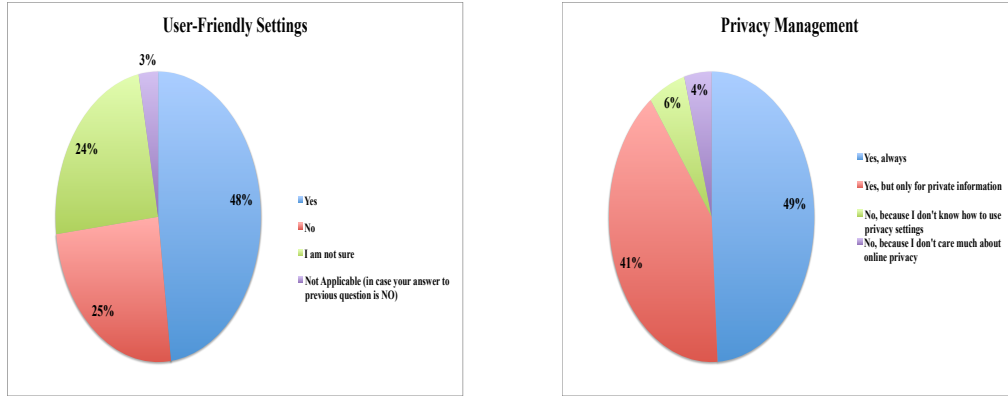
(a) Privacy policy awareness of OSNs users



(b) Online privacy concerns of OSNs users

Figure 4.1: Attitude of OSNs users towards online privacy

In the question about how concerned the participants are regarding the privacy of their profile information. 60% of the participants showed high concern, 17% of



(a) User friendliness of privacy settings

(b) Privacy management pattern of users

Figure 4.2: User opinion about user friendliness of privacy interface and its usage

the participants are unconcerned, and 23% of the participants are neither concerned nor unconcerned (Fig.4.1b). The results of age group wise privacy concern reveal that people from age group “above 40” and “below 20” shown slightly higher concerns related to privacy. Only highly unconcerned small minority is from age groups “between 20 and 30” and “between 30 and 40”. The results of age group wise privacy concerns are shown in Table 4.6.

In questions about usage and user friendliness of privacy settings, 49% of the participants responded that they always use privacy settings, 41% of the participants use privacy settings to protect their private information only, and 4% of the participants don't care about online privacy (Fig.4.2b). As far as user-friendliness of the privacy interface is concerned, 48% of the participants acknowledged the ease of use of the interface, whereas, 25% of the participants feel that the interface is not easy to use. The remaining participants are not sure about the user-friendliness of privacy management interfaces (Fig.4.2a). The findings reveal that female participants are more concerned about reading privacy policy as compared to male participants. The total percentage of

Table 4.4: Age group wise Privacy Policy Awareness

Item		Privacy Policy Awareness		
		Yes,very carefully	I looked at it quickly	Not at all
Age Group	Below 20	16.66%	33.3%	50%
	Between 20 and 30	18.38 %	34.52%	46.63%
	Between 30 and 40	10.97%	36.58%	50%
	Above 40	9.09%	45.4%	45.4%

Table 4.5: Gender wise Privacy Policy Awareness

Item		Privacy Policy Awareness		
		Yes, very carefully	I looked at it quickly	Not at all
Gender	Female	15.38%	44.87%	39.74%
	Male	16.59%	32.36%	50.20%

female participants that read privacy policy is 23%, whereas the negligible percentage of male participants read the privacy policy. The results of gender wise privacy concerns are a little bit unusual due to the strange response of female participants. The female participants are less concerned about privacy as compared to the male participants. Table 4.7 shows the results of gender wise privacy concerns of the participants.

4.3.2 Social Relationships Formation in Online Social Networks

A section of the research questionnaire is focused on understanding the users' attitude towards forming online relationships. For the question about what kind of people they have in their online social circle. We developed six categories of friends that include: family, friends, colleagues, classmates, acquaintances and strangers. This categorization is inspired by existing research studies on the classification of OSN users social circles. Some of the research studies on social categorization are discussed below to highlight

Table 4.6: Age group wise Privacy Concerns of Participants

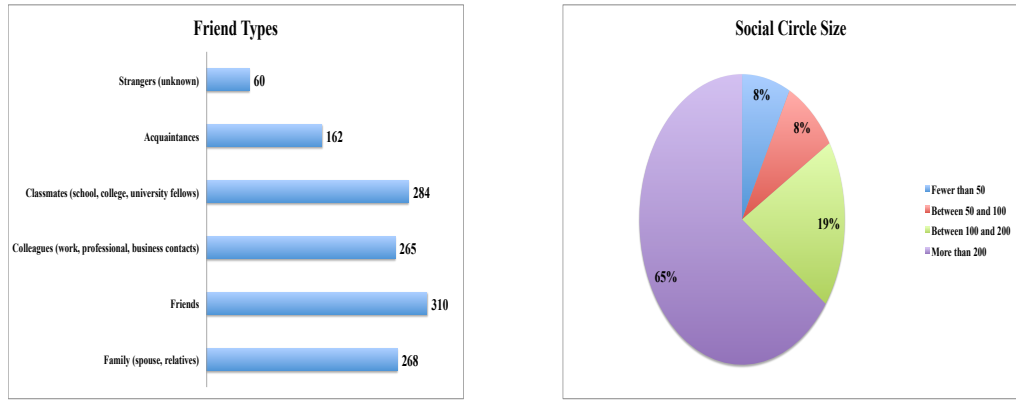
Item		Privacy Concerns				
		Highly Concerned	Concerned	Neutral	Unconcerned	Highly Unconcerned
Age Group	Below 20	50.0%	33.3%	16.6%	0%	0%
	Between 20 and 30	41.4%	18.5%	20.4%	12.8%	6.6%
	Between 30 and 40	40.2%	25.9%	18.1%	12.9%	2.5%
	Above 40	45.4%	36.3%	9.0%	9.0%	0%

Table 4.7: Gender wise Privacy Concerns of Participants

Item		Privacy Concerns				
		Highly Concerned	Concerned	Neutral	UnConcerned	Highly Unconcerned
Gender	Female	39.18%	17.56%	17.56%	14.86%	10.81%
	Male	41.66%	22.80%	20.17%	11.84%	0.35%

the reasons why we adopted this classification for our study. Spencer et al. [137] identified eight categories of friendship, which includes: associates, useful contacts, fun friends, favor friends, helpmates, comforters, confidants, and soul-mates. Kelly et al. [102] carried out a user study to find out how individual users group their Facebook friends using four different methods: card sorting, tagging, hierarchical file organization and using Facebook friend-list interface. The findings of this study suggest categories such as general friends, college friends, other educational friends, family, church, and don't know. The category of general friends is further divided into location-based friends, generic friends, and friends of friends. Recently, Zhang et al. [138] conducted a study on social categorization in online social networks. The main interest of the authors is to investigate how people group online friends into different categories. The findings suggest that main categories of friends on OSNs are school friends, work related friends, friends sharing similar interests/activities and family members. Some studies discuss the diverse nature of online relationships and emphasize that vast majority of OSNs users add friends, family, colleagues, classmates, and acquaintances into their social circle [139]. One of the studies claims that OSNs users also add strangers as friends while forming online relationships [140]. Apart from these research studies, we also explored current design options provided by Facebook and Google+ for social categorization. Facebook introduced the friend-list feature for social categorization in 2007. It allows users to group their friends into specific lists and assist users to share content selectively with their friends [25, 27, 28]. Google+ also introduced a similar concept in the form of circles, which enable users to easily group and classify their online friends [26]. Default friend-lists provided by Facebook include close friends,

family, acquaintances, restricted, and Google+ provides default circles such as public, friends, family, and acquaintances. These research studies motivated our grouping of online relationships into the aforementioned six categories. Our findings presented in figure 4.3a also show that friends circle of OSNs users consists all of these categories. The main purpose of this study is to identify users' willingness to disclose personal information with all friends equally.

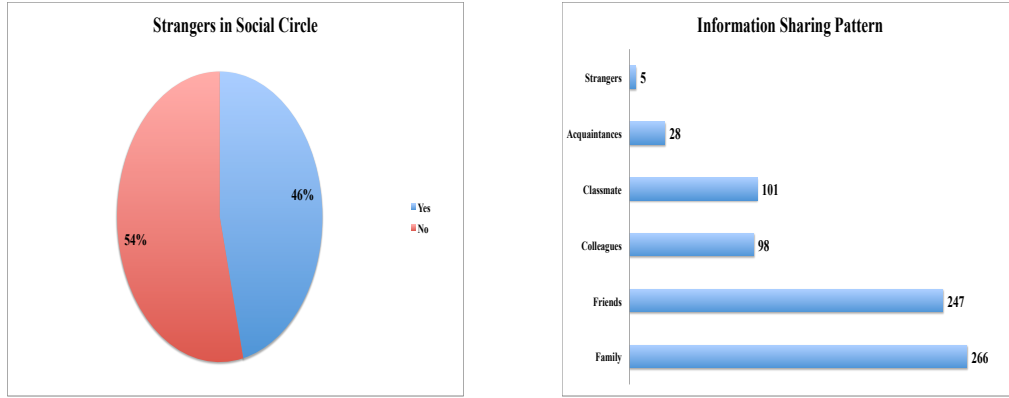


(a) Diversity of friendship in OSNs

(b) Density of friendship in OSNs

Figure 4.3: Attitude of OSNs users towards formation of online relationships

The results of our study confirm this fact that social circle of OSN user is diverse in nature. It includes people with strong ties as well as with weak ties. Moreover, the participants of the study also acknowledge adding strangers into their friend circle (Fig.4.3a). Adding strangers is quite common activity among all age groups. 66% participants of age group “below 20” add strangers as a friend. The same practice is followed by participants belong to age group “Above 40”. The strangers added by this group is around 54%. The age has very limited impact on stranger adding behavior of the OSNs users. The results of age group wise stranger addition are shown in Table 4.10. The gender wise stranger addition behavior is slightly different. 38% female



(a) Adding strangers as online friends

(b) Personal information disclosure pattern

Figure 4.4: Attitude of users towards personal information disclosure in OSNs

agreed to the fact that they add strangers in the friend network. Around 50% male participants add strangers to their friend network. Table 4.11 shows the results of gender wise stranger addition in friends network. This behavior of participants is in contradiction to their high concern about privacy of their profile information. Profile information of users is shared equally among all friends (by default) in existing access control mechanisms of online social networks. Once strangers are added as a friend, they have complete access to the profile information of the user.

In response to the question about how many friends a participant has in his/her friends circle in online social networks. 65% of the participants have more than 200 friends in their friend circle (Fig. 4.3b). The results of age group wise friend network size reveal that people below the age of 20 are careful in added people in their friend network. Only 33% from this age group has more than 200 friends. The results of age group wise friend network size are shown in Table 4.8. The results of gender wise friend network size reveal that 71% male participants have more than 200 friends on their social graph, whereas, the percentage of female participants is much lower than

Table 4.8: Age group wise Size of Friend Networks

Item		Size of Friend Network			
		Fewer than 50	Between 50 and 100	Between 100 and 200	More than 200
Age Group	Below 20	16.6%	33.3%	16.6%	33.3%
	Between 20 and 30	6.72%	7.62%	19.28%	65.47%
	Between 30 and 40	7.31%	7.31%	19.51%	64.6%
	Above 40	27.2%	18.1%	0%	54.5%

Table 4.9: Gender wise Size of Friend Networks

Item		Size of Friend Network			
		Fewer than 50	Between 50 and 100	Between 100 and 200	More than 200
Gender	Female	12.82%	14.10%	26.92%	44.87%
	Male	6.22%	6.63%	15.76%	70.95%

that. Table 4.9 shows the results of gender wise friend network size. Also, this behavior is in contradiction to users' relationship forming pattern in offline social networks. People tend to have a limited number of friends in offline social networks. Controlling disclosure of personal information in offline social networks is quite easier as compared to online social networks. Offline social networks lack features such as data permanence, invisible audience, searchability, and recordability. These features are inherent in online social networks and make controlling information disclosure a difficult task. Controlling privacy becomes more complicated with huge friend networks that include people from different facets of life.

In response to the question about the willingness of participants to share their profile information with strangers, 98.5% of the participants are unwilling to share their profile information with strangers, whereas 46% of the participants are adding strangers to their friend circle (Fig. 4.4a & b). This is also an example of the disparity between

Table 4.10: Age group wise Stranger addition in Friend Network

Item		Strangers added Friend Network	
		Yes	No
Age Group	Below 20	66.6%	33.3%
	Between 20 and 30	46.63%	52.46%
	Between 30 and 40	40.24%	57.31%
	Above 40	54.5%	45.4%

Table 4.11: Gender wise Stranger addition in Friend Network

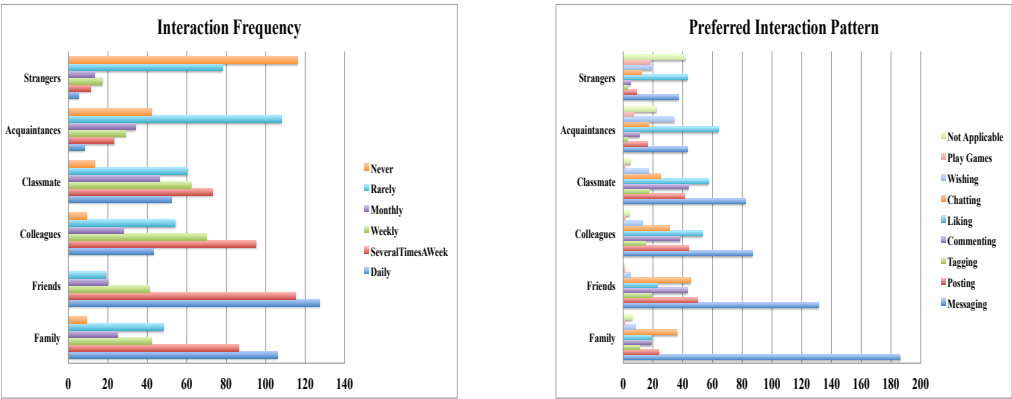
Item		Strangers added Friend Network	
		Yes	No
Gender	Female	38.46%	60.25%
	Male	48.54%	50.62%

privacy concerns and social relationship forming behavior of OSNs users.

4.3.3 User Interaction Pattern and Relationship Strength

The frequency of user interactions is a vital factor in determining the relationship strength. Our questionnaire contains a section to analyze the interaction pattern of the participants with their friends circle. We developed questions about the frequency of interactions and preferred interaction type. In the question about the frequency of interactions, grid row items represent a different type of people in friend circle (family, friends, colleagues, etc.) and grid column items represents the frequency of interactions such as daily, several times a week, weekly, monthly, rarely, and never. According to responses of the participants, their interaction frequency with family and friends is on a daily basis. They interact with colleagues and classmates several times a week or on a weekly basis, whereas their interaction pattern with acquaintances and strangers is rarely and never respectively (Fig.4.5a). The results give an indication of the link between relationship strength and interaction frequency and validate our hypothesis H2. The hypothesis assumes a relationship between the frequency of interactions and the relational tie. In the question about preferred interaction type to communicate with different types of people in friends circle, grid row items represent different types of relationships and grid column items represent different types of interactions such as messaging, posting, tagging, commenting, liking, etc. According to the results, messaging is preferred interaction type for the family. Preferred interaction types for friends include messaging, posting and commenting. In case of classmates and colleagues

messaging, posting, commenting, and liking are preferred interaction types. Liking is preferred interaction type for a relationship such as acquaintances and strangers (Fig. 4.5b).



(a) Interaction frequency with online ties (b) Preferred interaction with online ties

Figure 4.5: Relationship strength based interaction patterns of OSNs users

The responses demonstrated that the participants prefer to use private messaging to communicate with family, whereas public modes of communication are frequently used to interaction with friends, colleagues, and classmates. The liking and wishing are preferred types of interactions for weaker relational ties, whereas messaging is preferred interaction type for stronger relational ties. Our hypothesis H3 is also validated by these results. The hypothesis assumes the relationship between the choice of interaction type and relational tie.

Table 4.12 presents descriptive statistics for preferred user interaction types and profile data disclosed on OSNs. Descriptive statistics help to describe the features of specific data set, by giving short summaries about the sample and measures of the data. Descriptive statistics are useful to determine measures of central tendency and measures of dispersion. Measures of central tendency include the mean, median

Table 4.12: Descriptive Statistics for Interactions and Data Disclosure

Items	N	Minimum	Maximum	Mean	Std. Deviation
Privacy Concern	305	1	5	2.19	1.245
Messaging	317	1	10	3.09	2.495
Wall Posting	313	1	10	5.01	2.853
Photo Tagging	312	1	10	6.12	3.050
Commenting	312	1	10	4.28	2.505
Liking	308	1	10	3.18	2.506
Chatting	310	1	10	3.55	2.429
Playing Games	308	1	10	7.33	3.396
Basic Information	306	1	10	2.68	2.213
Educational and Employment Info	304	1	10	4.35	2.792
Activity Streams	304	1	10	4.28	2.645
Photos and Videos	301	1	10	3.03	2.295
Family and Relationship Info	297	1	10	3.03	2.384
Affiliations Information	301	1	10	4.10	2.533
Events Information	292	1	10	4.19	2.843
Preferences Information	300	1	10	5.58	2.943
Religious and Political Views	302	1	10	4.38	2.982

and mode, while measures of dispersion include the standard deviation, minimum and maximum variables. Table 4.12 presents N, minimum, maximum, mean and standard deviation. The value of N refers to sample size. The values of minimum and maximum refer to the largest observation and smallest observation of the sample. Both of these two variables are used to calculate the range, which is simply the difference between the maximum and minimum. Arithmetic mean is the most common method to describe central tendency. Dispersion the data is shown by the standard deviation. It refers to the spread of the values around the central tendency.

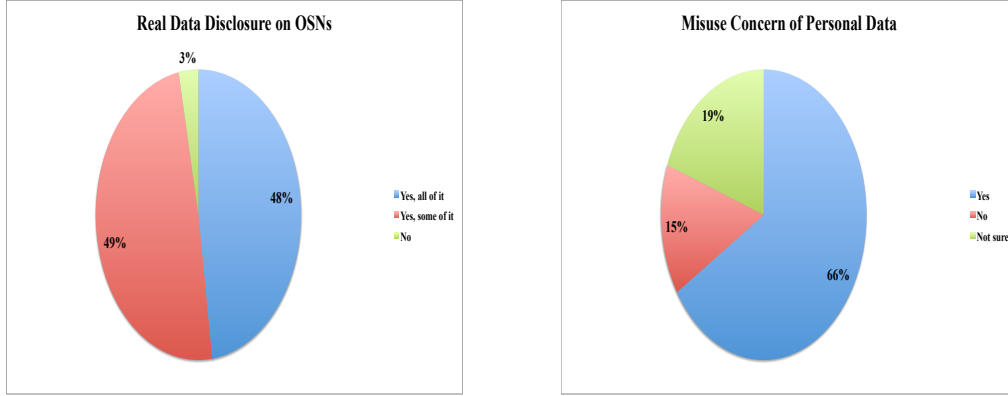
4.3.4 Profile Information and Interactions Ranking

The online social network users share a huge amount of personal information. What is the accuracy of shared information and how concerned the participants are about misuse of their real personal data? Our survey includes questions that measure these aspects of profile information. In response to the question about the accuracy of shared profile information, 97% of the participants provide real personal information on OSNs

(Fig.4.6a). The age group wise analysis of the data demonstrate that 18% of participants from age group “above 40” do not provide their real personal information, whereas, almost all other groups share their personal data on OSNs either all of it or some of it. Table 4.12 presents the results of age group wise personal information disclosure on OSNs. The gender wise personal information disclosure pattern is shown in Table 4.13. More than 50% male participants responded that they provide their complete real personal information on OSNs, whereas, only 38% female participants disclose their complete personal information on OSNs.

In response to the question about misuse concern of the disclosed personal data, 66% of the participants are concerned about any misuse of their personal information shared on OSNs (Fig.4.6b). The results of age group wise analysis of the data suggested that vast majority of the participants have misuse concern about their shared personal data. 83% participants of the age group “below 20” shown their concern regarding misuse of the data. As discussed earlier this group share more personal information than any other group. The least concerned is shown by participants belong to the age group “above 40” about the misuse of their shared personal information. The results of age group wise misuse concern about personal data are shown in Table 4.14. The results of gender wise misuse concern do reflect any noticeable difference between the responses of male and female participants. Both are highly concerned about the misuse of their personal data and vast majority responded positively to the question “Are you concerned about any kind of misuse of your personal information on online social networks?” Table 4.15 shows the results of gender wise misuse concern about personal data.

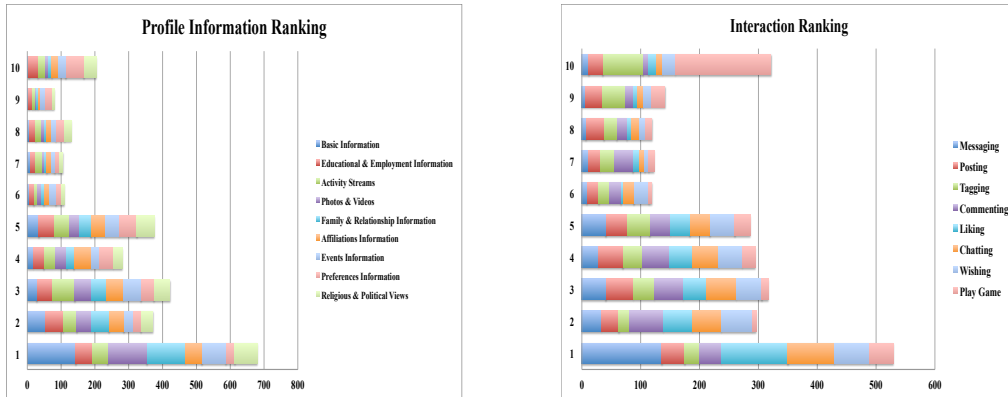
One of the main purposes of the empirical study is to provide the ranking for profile information and user interactions on the basis of sensitivity and usage frequency respectively. We developed questions to determine the most sensitive and the least sensitive profile information. According to the results the most sensitive profile information ele-



(a) Data disclosure attitude of users

(b) Misuse concerns about personal data

Figure 4.6: Personal data disclosure attitude and misuse concerns of OSNs users



(a) Sensitivity based information ranking

(b) Frequency based interactions ranking

Figure 4.7: Ranking of OSN users' profile information and interactions

ments are basic information (date of birth, phone, email, addresses), photos and videos, family and relationship information (family members, relationship status), and event information (events attended, GPS, exact locations), whereas, the least sensitive profile information elements are preference information (books, movies, music, TV shows),

Table 4.13: Age group wise Personal Data Disclosure

Item		Real Personal Information Disclosure		
		Yes, All of it	Yes, Some of it	No
AgeGroup	Below 20	66.67%	33.33%	0%
	Between 20 and 30	47.08%	48.43%	2.24%
	Between 30 and 40	43.90%	52.43%	2.43%
	Above 40	63.64%	18.2%	18.2%

Table 4.14: Gender wise Personal Data Disclosure

Item		Real Personal Information Disclosure		
		Yes, all of it	Yes, some of it	No
Gender	Female	38.46%	56.41%	3.84%
	Male	51.04%	44.81%	2.48%

religious and political views, educational and employment information (Fig.4.7a). We also developed questions to determine most and least frequently used interaction types. The results demonstrate that messaging, liking, chatting, commenting, wishing, and posting are the most frequently used interaction types, whereas playing games and tagging are the least frequently used interaction types (Fig. 4.7b). Ranking of profile information and users' interactions can be useful to segregate information and friends in online social networks.

Table 4.15: Age group wise Misuse Concern about Personal Data

Item		Misuse Concern about Personal Data		
		Yes	No	Not Sure
AgeGroup	Below 20	83.3%	16.7%	0%
	Between 20 and 30	64.57%	15.24%	18.83%
	Between 30 and 40	64.63%	10.97%	21.95%
	Above 40	54.5%	27.3%	18.2%

Table 4.16: Gender wise Misuse Concern about personal Data

Item		Misuse Concern about Personal Data		
		Yes	No	Not Sure
Gender	Female	60.25%	17.94%	20.51%
	Male	66.39%	13.69%	18.67%

4.4 Discussion and Implications

In this section, we identify conflicts between privacy concerns and information sharing practices of OSNs users. The results demonstrate that majority of the users share their real personal data and they are highly concerned about misuse of this data. Apart from being highly concerned about misuse of their personal data, the majority of the users ignore to read the privacy policy of OSN service providers (Fig.4.8a & b). Another indication of how concerned OSNs users are about their privacy is a large friend circle that includes even strangers (Fig.4.9a & b). We conclude with these examples that there are contradictions between privacy concerns and information sharing practices of OSNs users. The purpose of our study was to answer the research questions about identification of conflicts between privacy concerns and information sharing practices of OSN users. Also, we wanted to demonstrate that personal information disclosure and relationship strength are directly proportional to each other. The interaction patterns of OSN users play a vital role to identify relational strength among them. Our three hypotheses are validated to be true by this empirical study. According to the results, only 3% of the participants agreed to share their personal information with strangers and 12% with acquaintances. Whereas, a vast majority of them is interested in sharing their personal information with family (82%) and friends (76%). Also, a reasonable number of the participants are willing to share their personal information with classmates (31%) and colleagues (30%). These results completely validate our hypothesis H1. The results also validate our hypothesis H2. A vast majority of the participants interact with friends and family on daily basis and classmates and colleagues on weekly

basis. Whereas, their interaction frequency with strangers and acquaintances is either rarely or never. It is interesting to notice that participants preferred interaction types for family and friends are messaging or chatting. Preferred interaction type for acquaintances or strangers is liking. This choice of interaction types also validates our hypothesis H3.

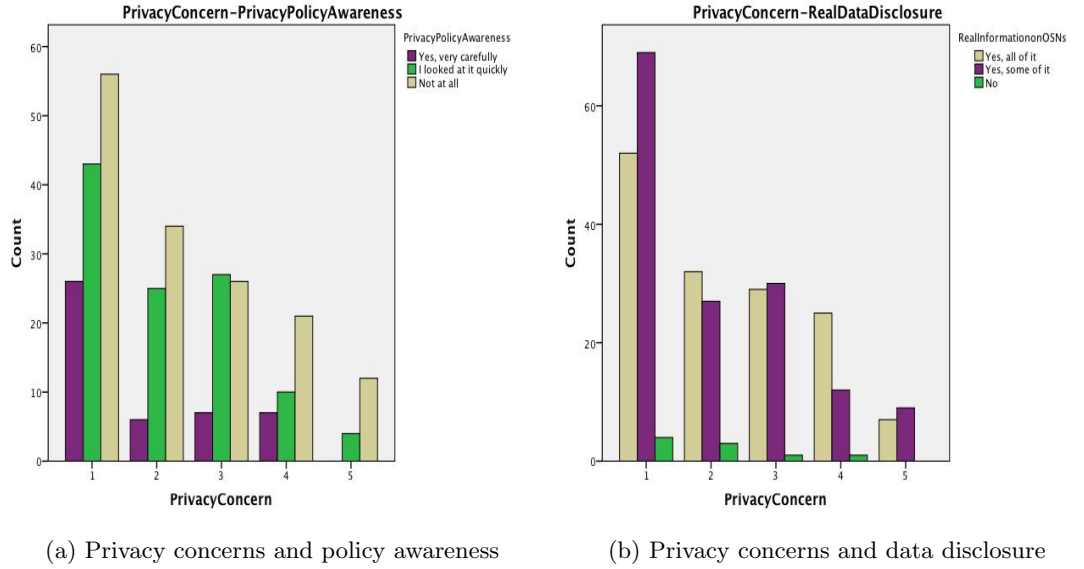


Figure 4.8: Disparity between high privacy concerns and data disclosure practice

4.5 Limitations of User Study

Our user study has some limitations. The first limitation deals with a relatively small number of users participating in this study. As OSNs are the current craze and their user base is in millions. Therefore, the response of less than 400 hundred users cannot be conclusive. The second limitation deals with gender-biased nature of our user study. The majority of the participants in this study are male. Only 25% female participants recorded their responses. The results of the study are male biased and cannot be generalized. The third limitation deals with geographical distribution of the participants. The majority of the participants belong to Indian sub-continent and their responses

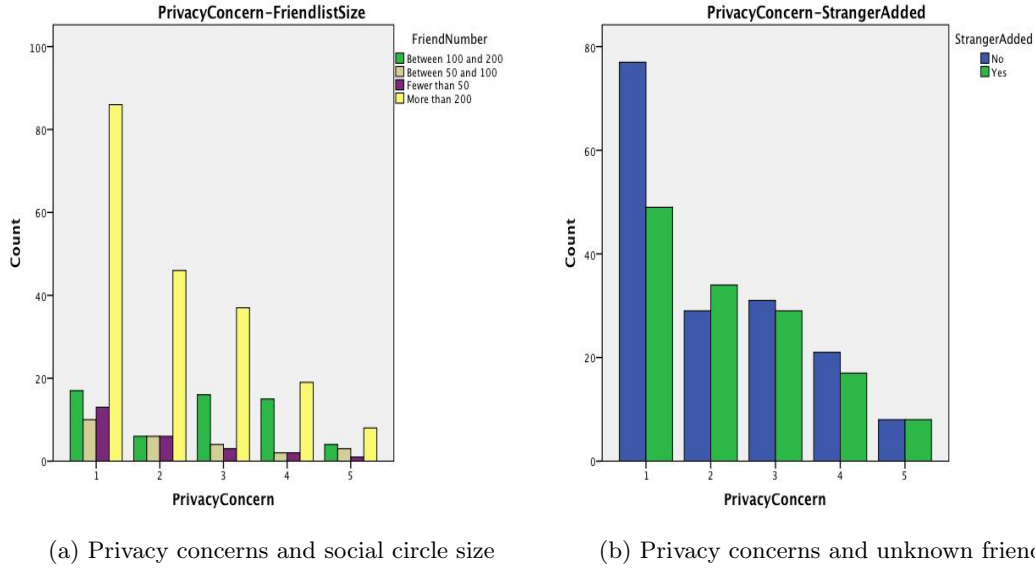


Figure 4.9: Disparity between high privacy concerns and online relationship formation

lack cultural diversity. The results of the user study cannot be generalized in cross-cultural scenarios. It is important to mention that we approached many users across the globe and successfully got responses from some of them, but the percentage of European, American and Australian participants is less than 10%. The cultural diversity is good ideas for the user study, but for this user study, we fail to recruit participants across the globe.

4.6 Concluding Remarks

We presented the results of the user study that demonstrate the disparity between privacy concerns and information sharing behavior of online social network users. The results also establish the link between personal information disclosure and relationship strength. The relationship strength among online social network users can be estimated via their interaction pattern. The choice of interaction type used for communication has also significance in identifying relational tie. Additionally, the findings facilitated categorization of profile information and user interactions on the basis of sensitivity

and frequency respectively. The main contribution of this study is that it provides answers for the second research question.

Chapter 5

Ontology Based Privacy Modeling for OSNs

The objective of this chapter is to present the SOCPRI ontology which represents diverse social relationships of users in online social networks. The purpose of developing the SOCPRI ontology is to enhance the management of user privacy in online social networks. Our ontology models the role and relationship based self-presentation of users in the dynamic environment of the social web. It also models tie strength dimensions and relates these dimensions to social interactions of the users in OSNs. The SOCPRI models contextual privacy of OSNs users which takes into consideration contextual norms for appropriate information disclosure within and across contexts. Before introducing the SOCPRI ontology, we describe social semantic web that represents online social networks using semantic web technologies. We describe the role of ontologies in knowledge representation and present various ontologies that represent social data. We present our methodology to build the SOCPRI ontology and highlight contributions of the ontology. We introduce core conceptual elements of our ontology along with their definitions and relations. Finally, we present a comparative analysis of the SOCPRI ontology with other existing ontologies that represent data for the social web.

5.1 Towards the Social Semantic Web

The social and semantic web describe two different diversifications of the world wide web. The social web provides new ways of communication and collaboration to its users that enable them to simulate real life social interactions by establishing object centered social networks. The object centered social networks form connections between people and their objects of interest and facilitate people meet through things they have in common. As these object centered social networks grow bigger and more diverse, more intuitive methods are needed for representing and navigating the information in these networks. This requires representation mechanism to interconnect people and objects on the web in an interoperable and extensible way. The semantic web provides such representation mechanism. The goal of the semantic web is to support data sharing and data interoperability on the web.

Both the social and the semantic web overlap in the goal to support sharing of information. The social web from the user point of view and semantic web from technological point view. The goal of the social web is to enable users staying in control of their data and allowing sharing it. The semantic web can support this with a set of tools that allows us to describe information in a machine readable form. This can be achieved using social semantic web which is the implementation of the semantic web technologies in a social networking knowledge base. The social semantic web implements ontologies for the common understanding of the information. FOAF is an example of a popular social semantic web based ontology that provides logical and machine readable information about the social relations and friends using RDF and XML structure. Another well-known example of social semantic web based ontology is a SIOC that is used to interconnect online community information. The concept of social semantic web has been discussed extensively in the scientific literature [141, 142, 143]. In the following section, we describe some of the popular ontologies that play a vital role in representing

social data with semantic web tools to realize the vision of social semantic web.

5.1.1 Representing Social Data with Semantic Web

Online social networks contain a lot of personal data about users, their relationships, their interests, and their online activities. Some researchers represent this data by graphs, whereas others prefer ontologies to represent social data. Graph based representation allows structuring concepts and relations between the concepts. It also supports better visualization, however, graph lacks semantic representation of the concepts. Ontology based representation is more expressive and less ambiguous. The use of ontologies offer better knowledge representation and keep the semantic relationships between concepts. Moreover, ontologies support inference mechanism that can be used to enhance semantic matching.

Several ontologies are found in the scientific literature of computer science that represent data related to social web users. One of the most popular is FOAF (Friend Of A Friend)¹ ontology. The FOAF initiative provides a way to represent online social networks data in a shared and machine readable way. FOAF is light weight and very simplified RDF vocabulary to describe users' profiles, relationships, affiliations, and other online activities. FOAF is one of the first semantic models to grasp social interconnections between people [76]. The “knows” property is used to connect people and to build a social network. Other properties are also available to represent users' profile and his web usage such as “nick”, “interest”, “online account”, “membership”, etc. The main features of FOAF ontology are high data interoperability which refers to the flexible integration with other systems. Due to the adoption of this ontology by web 2.0 platforms, millions of FOAF profiles are now published on the web [144]. However, representing relationships using such RDF property fail to accommodation rich context information and diverseness in social relationships. RELATIONSHIP² on-

¹FOAF, <http://xmlns.com/foaf/spec/>

²RELATIONSHIP, <http://vocab.org/relationship/>

tology also models user relationships in online social networks in more precise manner. It extends the “knows” property of FOAF to characterize various user relationships for example the relation “livesWith” specializes the relation “knows”. The RELATIONSHIP ontology offers a large set of properties that are used to represent most of the personal, professional, sentimental and family relationships.

SIOC³ (Semantically Interlinked Online Community) is another well-known ontology for modeling social community sites such as blogs, wikis, online forums, etc. SIOC extends FOAF for the specific description of users’ activities. The examples of the primitives of FOAF extended by SIOC ontology are “OnlineAccount” and “hasOnlineAccount”. SIOC provides the basis for defining a user, the content a user produces and the actions of other users on this content. The key concepts in SIOC ontology are a user, sites, posts, tags, forum, etc. Initially, SIOC was designed to define content publishing activities in online communities and interaction with published content. SIOC has been extended to support wide kind of social media by adding several extension modules. SIOC types ontology⁴ extends “sioc:item” to specify different types of resources produced online. SCOT⁵ ontology is used to model tagging activities. Another extension of SIOC ontology is SIOC services⁶ that allows to describe a service available on given site and binding it to its interface.

The SKOS⁷ (Simple Knowledge Organization System) designed for representation of taxonomies, classification schemes, or any other type of structured controlled vocabulary [145]. The main objective of SKOS is to enable easy publication and use of such vocabularies as linked data. The key elements in SKOS ontology are concepts, labels, notations, semantic relations, mapping properties, and collections. We can specify the meaning of tags and posts topics using SKOS ontology. The SKOS ontology is part of

³SIOC, <http://www.w3.org/Submission/sioc-spec/>

⁴SIOC Types, <http://rdfs.org/sioc/types>

⁵SCOT, <http://scot-project.org/>

⁶SIOC Services, <https://rdfs.org/sioc/services>

⁷SKOS, <http://www.w3.org/TR/skos-reference/>

the semantic web family of standards like FOAF and SIOC.

Apart from FOAF, there are several other ontologies to represent users and their social media interactions. The GUMO (General User Modelling Ontology) is a comprehensive user model that intends to cover a wide range of user related information such as demographics, contact information, personality etc. GUMO was created by Heckman et al. [146] and represents the most generic user model. However, it falls short of representing user interests, which makes it unsuitable for the social web. The SWUM (Social Web User Model)⁸ ontology is designed to overcome the shortcomings of GUMO ontology. Plumbaum et al. [147] derived a number of user model dimensions required for social web by analyzing 17 social web applications. Their taxonomy of user dimensions includes demographics, interests and preferences, needs and goals, mental and physical state, knowledge and background, user behavior, context, and individual traits. A key shortcoming of SWUM ontology is its lack of grounding in other ontologies. UBO (User Behaviour Ontology)⁹ is another ontology designed by the authors. It builds semantic user behavior model. It has been used to model user behavior for online social networks such as Twitter. It has classes that model the impact of posts, user behavior, user roles, and other interaction information [148]. Recently, SemSNI (Semantic Social Network Interactions) [149] is designed as an extension of SIOC ontology that models home pages, private messages, discussion topics and documents that do not exist in SIOC types with required semantics. SemSNI defines the classes “UserHome”, “PrivateMessage”, “Topic” and “Document” as subclasses of “sioc:Item”. SemSNI introduces new elements to express interactions such as visits and private messages. It allows to gather the visits made by a user to a resource and also the private messages that the user shared with another user on the social web. The SemSNI ontology also introduces the notion of the user profile page used in social media such as Facebook or Google+.

⁸SWUM, <http://swum-ontology.org>

⁹UBO, <http://ubo-ontology.org>

Social tagging is one the most popular activity on online social networks. OSNs users tag resources of the social web such as pictures, videos, blog posts etc. Numerous ontologies have been designed to capture and exploit the activities of social tagging. The MOAT¹⁰ (Meaning Of A Tag) allows users to define the semantic meaning of a tag through linking open data [150]. The ontology describes tagging by (tag, resource, user, meaning) to specify the local meaning of a tag because there are several terms with a multitude of global meaning. The ontology defines two kinds of tags: global (across all content) and local (particular tag on a given resource). The MOAT uses the SIOC ontology extensively to describe online communities. As discussed earlier in this section, the SCOT¹¹ (Social Semantic Cloud of Tags) ontology gives a semantic structure of tagging data for social interoperability among the different social media [151]. The main goal of SCOT ontology is to represent collaborative tagging activities. The SCOT ontology focuses on a way to share tags by modeling tagclouds and also provides various properties to link tags together. The SCOT uses concepts and properties of SIOC and SKOS ontologies with main objective to aggregate tags used by the same persons in the clouds. The MUTO¹² (Modular Unified Tagging Ontology) is an extensible tagging ontology that unifies existing ontologies for tagging purpose [152]. The MUTO ontology supports different forms of tagging such as common, semantic, group, private and automatic tagging. There are numerous other ontologies in the research literature that support tagging activities. The social tagging ontologies presented in this section offer most useful representation for collaborative tagging systems that are prevalent in web 2.0 applications.

Finally, we describe an ontology that enables users to create fine-grained privacy preferences for their data freely accessible through current open linked data environment. Sacco et al. [153] proposed this lightweight ontology known as PPO¹³ (Privacy

¹⁰MOAT, <http://moat-project.org/ontology>

¹¹SCOT, <http://rdfs.org/scot/spec/>

¹²MUTO, <http://muto.socialtagging.org/core/v1.html>

¹³PPO, <http://vocab.deri.ie/ppo>

Preference Ontology). PPO enables linked data creator to describe privacy preferences for restricting access to specific data at a fine-grained level. For instance, PPO can be used to restrict part of a FOAF user profile only to users that have similar interests. PPO vocabulary is platform-independent. Thus, it can be used by any system for describing restrictions as long as it is based on semantic web technologies.

5.2 Knowledge Representation and Ontologies

The Knowledge Representations (KR) deals with the basic problem of making knowledge explicit that can be used by machines to automatically process knowledge and share knowledge unambiguously. The main purpose of explicitly representing knowledge is to be able to reason about the knowledge, to make the inference and assert new knowledge. There are a number of knowledge representation formalisms that provide the precise notation for representing knowledge such as first order logic, propositional logic, description logic, ontologies etc. A key trade-off in the design of a knowledge representation formalism is that between expressivity and practicality. The ultimate knowledge representation formalism in terms of expressive power and compactness is first order logic. In this dissertation, we will use ontologies for representing the protected knowledge in online social networks and in some access control decision information. The reason to choose ontologies is due to the fact that ontologies are the core of the Semantic Web and suitable for developing larger and modular knowledge bases that can communicate and integrate with each other.

The term **Ontology** is derived from Greek words “Onto”, which means being, and “Logia”, which means written or spoken discourse. In the Oxford dictionary, an ontology is defined as “The science or study of being; that branch of metaphysics concerned with the nature or essence of being or existence; or a theory or conception relating to the nature of being.” The ontology is originally a philosophical term meaning “the

study of the nature and relations of existence”. In the beginning of the 20th century, the German philosopher Edmund Husserl coined the term **Formal Ontology** which deals with formal aspects of objects irrespective of their particular nature. Formal Ontology is a philosophical discipline that aims at developing a system of general categories and their ties, which can be used in the development of scientific theories and domain-specific common sense theories of reality. Formal ontological theories developed and applied to solve the problems in the field of computer and information sciences are known as **Applied Ontology**. It is this kind of ontology with which this chapter is mainly concerned. We describe the role of ontologies in computer science and more specifically in the semantic web in the following sections.

5.2.1 Ontologies in Computer Science

Ontologies have been applied in a multitude of areas in computer science such as artificial intelligence, information systems, semantic web, etc. Accordingly, the definitions of ontologies have evolved with reference to technical means for developing ontologies. We present some of these definitions from the computer and information science perspective. The term “Ontology” in computer and information science literature appeared for the first time in 1967 in a work on data modeling by S.H. Mealy. The field of data modeling has been a fruitful ground for the applications of ontological theories, either implicitly or explicitly. One of the widely used definitions of ontology in computer science literature is defined as “a formal, explicit specification of a shared conceptualization”[154]. Conceptualization refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. Explicit means that the type of concepts used and constraints on their use are explicitly defined. Formal refers to the fact that ontology should be machine readable. Shared reflects that ontology should capture consensual knowledge accepted by the communities.

Gruber’s definition of the ontology is abstract, a more elaborated definition for on-

tology is given by Fensel [155]. The author defines ontology as “a shared and common understanding of a domain that can be communicated between people and heterogeneous and distributed systems”. This definition describes how the philosophical nature of ontologies could be incorporated into computer science. Huhns et al. [156] defines technical aspects of ontology by saying that “a computer model of some portion of the world”. We adopted recent redefinition of the ontology by Gruber for this dissertation. The author redefines ontology as “a set of representational primitives with which to model a domain of knowledge or discourse, where the representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members)” [157]. This definition provides the representational primitives for modeling information and developing a knowledge base to manage such information. The representational primitives for modeling an ontology are five: concepts, properties, relations, axioms, and instances. Basically, the role of ontology is to construct a domain model using these primitives. It is widely recognized that constructing a domain model or ontology is an important step in the development of knowledge based systems.

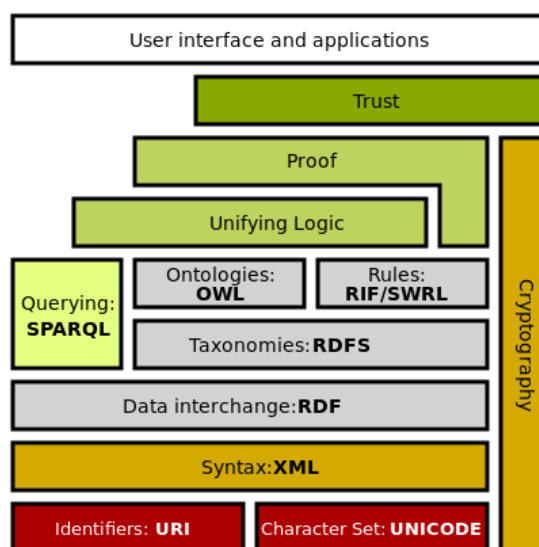


Figure 5.1: Semantic Web Layered Stack [6]

Ontologies play an important role in the **Semantic Web**. The Semantic Web is

“an extension of the current web in which information is given well-defined meaning, better-enabling computers and people to work in cooperation” [158]. The semantic web is a technology driven approach with the key challenge to ensure a common understanding of information allowing it to be shared between applications and interpreted by machines as well as humans. The Semantic Web allows machines to present logical connections among the information using different ontologies. Ontologies are an attractive way of describing semantics since they allow for a much more formal and less ambiguous definition of what certain concepts mean than for example natural language descriptions. Basically, most ontologies consist of the definition of a number of concepts, and some sort of hierarchy between these concepts that define relations that can exist between concepts. The semantic web significantly improves visibility and extensibility aspects of knowledge sharing. W3C standardized a set of technologies that build the Semantic Web. The technologies are illustrated in the semantic web stack presented in figure 5.1. In the semantic web stack, the semantic part is enabled by a stack of evolving languages such as RDF, RDFS, OWL, etc. The rapid evolution of the languages was enabled by inheriting and extending some of their useful features. In the following section, we cover some languages that can potentially be used to express semantic of data explicitly.

5.2.2 Languages for Encoding Ontologies

In this section, we give a brief overview of three encoding languages used to enable semantic part in semantic web stack [6] presented in figure 5.1. These languages evolved with the evolution of the semantic web. According to evolutionary trends observed in the development of the semantic web, RDF was proposed in 1998 as simple graph model, followed a year later by RDF schema and finally evolved into OWL which was drafted in 2002 and become W3C standard in 2004. The rapid evolution of encoding languages was enabled by learning from the experiences in developing existing KR formalisms and

database conceptual models.

The Resource Description Framework (RDF)¹⁴ is a W3C recommended language for describing information about resources on the world wide web. The RDF offers a simple graph model which consists of nodes and binary relations. The RDF data model allows structured representations of web resources to be made, by expressing statements about web resources in form of triples. The syntactical representation of triple is subject-predicate-object, where subject identifies the resource, predicate defines the property, and object represents the value of the property [159]. These triples can be coded in RDF/XML¹⁵, Notation 3(N3)¹⁶ or Turtle¹⁷ syntaxes. The major drawback of RDF is that it cannot define very strong semantics. In order to make the semantics much clearer RDF Schema was designed.

The RDF(S)¹⁸ Schema is a semantic extension of RDF. It provides a data modeling vocabulary for RDF data [160]. The RDF Schema class and property system are similar to the type systems of object oriented model. This model allows us to define simple class hierarchies and relations between the classes in the hierarchy, and to instantiate this hierarchy with resources. The authors can create simple class hierarchies using the **rdfs:Class** and **rdfs:subClassOf** properties. The authors can create instances of the classes by means of the **rdfs:type** property. The RDFS allows authors to specify the domain and range of properties using **rdfs:domain** and **rdfs:range**. The RDFS also allows declaring sub-properties of other properties by means of the **rdfs:subPropertyOf** property. The expressive power of RDFS is fairly limited. There are no ways to declare that two resources are identical to each other or specifying transitive and inverse properties etc.

Mainly because of limited expressiveness of RDF Schema, researchers have worked

¹⁴RDF, <https://www.w3.org/RDF/>

¹⁵RDF/XML, <https://www.w3.org/TR/REC-rdf-syntax/>

¹⁶N3, <https://www.w3.org/TeamSubmission/n3/>

¹⁷Turtle, <https://www.w3.org/TR/turtle/>

¹⁸RDFS, <http://www.w3.org/TR/rdf-schema/>

on developing more powerful alternatives. One of the most widely known alternatives is the Web Ontology Language (OWL). The OWL is a revised version of the DAML+OIL¹⁹ language and standardized by W3C²⁰. The OWL is based on the Description Logic (DL)²¹ and brings with it the expressive and reasoning power of the DL. A key of feature of OWL is that it allows building a shared understanding of a domain and its concepts in well defined semantic [161]. The OWL has three distinct species with different computational complexity and the expressiveness. The OWL Full has more expressive power than OWL DL, which has, in turn, more expressive power than OWL Lite. In terms of ease of reasoning, OWL Lite is considerably easier than reasoning over OWL DL or OWL Full. These three species are layered in a sense that every OWL Lite ontology is a legal OWL DL ontology, every legal OWL DL ontology is a legal OWL Full ontology. The inverse of these relations generally does not hold. In this dissertation, we use OWL DL for modeling and reasoning about online social network knowledge base.

5.2.3 Knowledge Representation using OWL DL

The OWL DL is based on description logic which is a family of formal knowledge representation languages. A Description Logic models concepts, roles, individuals, and their relationships. The fundamental modeling concept of DL is the axiom which a logical statement relating roles and/or concepts. The design of OWL DL is grounded on SHOIN(D) description logic. The OWL DL supports maximum expressiveness while retaining computational completeness and decidability. It provides additional language constructs such as complex set operations, enumerated classes, etc. The OWL DL also offers subsumption and consistency to enhance the inference capability of the reasoning systems. The OWL DL uses all OWL ontology constructs with some restrictions such

¹⁹DAML+OIL, <https://www.w3.org/TR/daml+oil-walkthru/>

²⁰OWL, <https://www.w3.org/OWL/>

²¹Description Logic, <https://en.wikipedia.org/wiki/Description-logic>

as type separation.

A typical DL knowledge base comprises of two components, namely the **TBox** and the **ABox**. A TBox is an abbreviation for a terminological box which contains intensional knowledge in form of a terminology. The basic form of declaration in a TBox is the definition of new concepts in terms of other previously defined concepts. For example, a woman can be defined as a female person and concept of woman belongs to TBox. The ABox contains extensional knowledge that is specific to the individual of the domain of discourse. The extensional knowledge is also called assertional knowledge and the assertional box is abbreviated as an ABox. The example of an ABox entities is an individual Alice that is a woman defined as a female person in a TBox. Another important feature of description logic knowledge bases is that it is based on **Open World Assumption(OWA)** which means that lack of knowledge about a fact does not immediately imply knowledge of the negation about the fact. We use OWL DL for specifying ontologies in this dissertation. We owe an explanation to choose OWL DL sub-language on other species of OWL language. The reason to reject OWL Lite is obvious that it has less expressive power as compared to OWL DL. When using OWL Full instead of OWL DL, reasoning support becomes less predictable [162].

5.3 Introduction to the SOCPRI Ontology

The goal of the SOCPRI ontology is to develop a privacy framework to facilitate fine-grained access to user generated content in the social web environment. The rise of online social networks introduced an important problem of **Context Collapse** which leads to undesirable consequences of merging multiple audiences into single social context. This often blurs the public versus private, professional versus personal and many different selves and situations in which individuals find themselves. This leads to undesirable consequences of disclosing personal information of the OSNs users with the

unintended audience. As matter of fact, context collapse is the root cause of many of privacy issues in online social networks. We address this issue from the social perspective and exploit existing social theories of Goffman, Nissenbaum, and Granovetter. The SOCPRI ontology is inspired by these most influential social theories and model privacy for social web users using the refined abstraction of contexts that embodies the philosophy of contextual integrity and presentation of self. The detailed conceptualization of social theories is already presented in section 3.3 of the third chapter. In this section, we first describe key features of SOCPRI ontology and motivate the need for the development of this privacy framework in the context of online social networks. The technical description of core conceptual element and details about classes and properties of SOCPRI ontology is presented later in sections 5.5 and 5.6 respectively.

We adopt an ontology-based approach to model social privacy on the basis of diverse social contexts and varied relationship strengths among online social network users. Ontology-based approaches have been evaluated as most promising for representing social web data due to their advantages of reusability, sharing, and reasoning. The main motivation for using an ontology is the possibility to model privacy of OSNs users in a flexible way and independently from the concrete online social network. Another advantage is the possibility to integrate and reuse other existing ontologies. The core conceptual elements of our ontological model are user's profiles, social relationships, social contexts, social interactions, digital resources, and privacy policies. Our ontological model has been defined in OWL2-DL that is a sub-language of OWL2. The OWL2-DL has been the most practical choice for most ontological applications due to its maximum expressive power without losing computational completeness and decidability. The OWL ontologies are described in terms of classes of individuals as well as the properties of those individuals. The properties can connect different individuals or they can relate data attributes to an individual. The relations are described in a formal way with strictly defined semantics that allows applying inference rules to infer

Metrics

Axiom	1161
Logical axiom count	551
Declaration axioms count	253
Class count	102
Object property count	58
Data property count	14
Individual count	75
Annotation Property count	9
DL expressivity	ALCROIQ(D)

Class axioms

SubClassOf	86
EquivalentClasses	52
DisjointClasses	22
GCI count	0
Hidden GCI Count	32

Object property axioms

SubObjectPropertyOf	11
EquivalentObjectProperties	0
InverseObjectProperties	5
DisjointObjectProperties	0
FunctionalObjectProperty	6
InverseFunctionalObjectProperty	1
TransitiveObjectProperty	0
SymmetricObjectProperty	10
AsymmetricObjectProperty	1
ReflexiveObjectProperty	0
IrreflexiveObjectProperty	0
ObjectPropertyDomain	54
ObjectPropertyRange	52
SubPropertyChainOf	0

Data property axioms

SubDataPropertyOf	2
EquivalentDataProperties	0
DisjointDataProperties	0
FunctionalDataProperty	6
DataPropertyDomain	11
DataPropertyRange	13

Individual axioms

ClassAssertion	75
ObjectPropertyAssertion	114
DataPropertyAssertion	30
NegativeObjectPropertyAssertion	0
NegativeDataPropertyAssertion	0
SameIndividual	0
DifferentIndividuals	0

Annotation axioms

AnnotationAssertion	357
AnnotationPropertyDomain	0
AnnotationPropertyRangeOf	0

Figure 5.2: SOCPRI Ontology Metrics

implicit facts from existing ones. Basically, the role of ontology is to construct a domain model using these modeling elements. It is widely recognized that constructing a domain model or ontology is an important step in the development of knowledge based systems.

We develop the SOCPRI ontology because we could not find another ontology in the literature able to capture the diverse social contexts, varied relationship strengths and rich social interaction information of social web users. To highlight this fact, we have presented a detailed analysis of the existing ontologies for representing social web data in Section 5.1.1. The existing ontologies design various aspects of the social web, but the SOCPRI is specialized ontology to model privacy for online social networks from the social perspective. We must acknowledge that the SOCPRI is integrating and reusing some classes and properties of the existing ontologies such as FOAF. The details about reusable classes and properties of other ontologies are discussed later in subsection 5.4.2. The current version of the SOCPRI ontology consists of 102 classes, 58 object properties, and 14 datatype property. The total axioms count is 1161, which includes 551 Logical axioms, 253 Declaration axioms, 86 SubclassOf axioms, 52 Equivalent classes axioms, 22 Disjoint classes axioms, 11 SubObjectPropertyOf axioms, 5 InverseObjectProperty axioms, 6 FunctionalObjectProperty, 10 SymmetricObjectProperty axioms, 1 AsymmetricObjectProperty axioms, 52 ObjectPropertyDomain axioms, 51 ObjectPropertyRange axioms, 2 SubDataPropertyOf axiom, 6 FunctionalDataProperty, 11 DataPropertyDomain axiom, 14 DataPropertyRange axiom, 73 Class Assertions, 109 ObjectProperty Assertion, 14 DataPropertyAssertion and 355 AnnotationAssertion axiom. The snapshot of SOCPRI ontological metrics is shown in figure 5.2. The taxonomic relationships of SOCPRI ontology are shown in figure 5.3. The OSN user is the central concept in the ontology that is represented as a subclass of “foaf:Agent”. The SOCPRI ontology differentiates between roles and relationships of the user. There are two main types of roles associated with OSN user: (1) user-centric roles and (2)

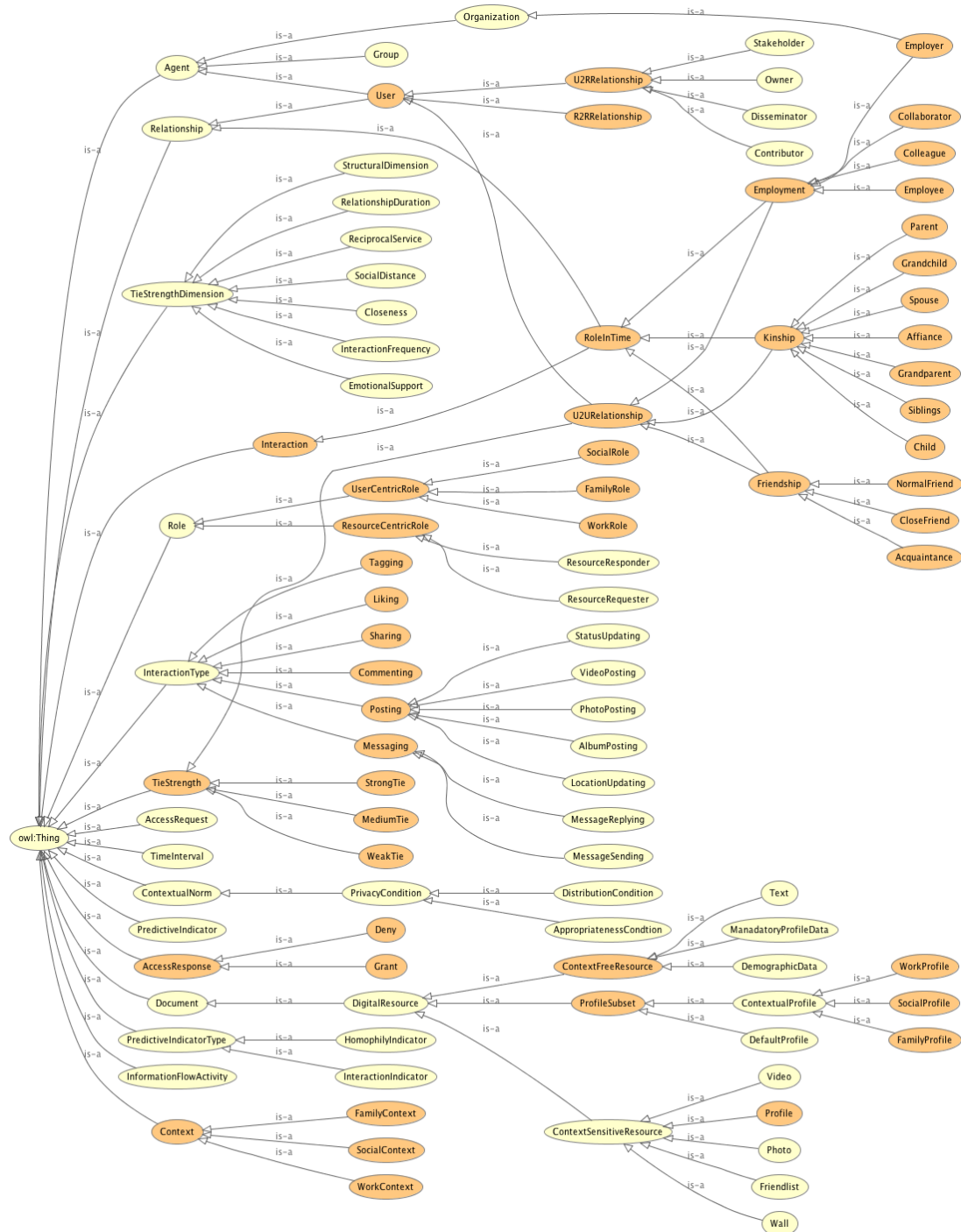


Figure 5.3: Taxonomic Representation of SOCPRI ontology

resource-centric roles. The user-centric roles are further sub-divided into family, social and work roles, whereas, resource-centric roles are either requestor or responder for the resource being accessed. The SOCPRI ontology models temporal dimension of the roles which is represented by *TimeInterval* and *RoleInTime* classes. The technical details of modeling temporal aspects in roles are discussed later in subsection 5.5.3. In the SOCPRI ontology, we represent three different types of the user relationships such as user-to-user (U2U) relationships, user-to-resource (U2R) relationships and resource-to-resource (R2R) relationships. The user-to-user relationships are further subdivided into Kinship, Friendship, and Employment relationships. All family relationships are extended from Kinship class, whereas work and social relationships are extended from Employment and Friendship classes respectively.

The users become the content creator in current online social networks. They share an unprecedented amount of digital resources in the social web environment. SOCPRI ontology models various digital resources associated with OSNs users. The ontological model categorizes user's resource into context sensitive and context free resources. The access to context sensitive resource is regulated through contextual norms. The upper class for all user's resources is *DigitalResource* class. The *DigitalResource* class is extended from "foaf:Document". The context sensitive resources represented by this class are Wall, Profile, Friendlist, Photo and Video. Apart from Profile class, SOCPRI also models contextual profiles which include profile subsets for social, work and family contexts. The demographic data and mandatory profile data are treated as context free resources in the digital resource modeling. As discussed in chapter 3 section 3.4, Three social contexts are very common among all OSNs users. Our ontological model provides coarse grained context segregation and main factor which facilitates this is user role at a particular point in time. Temporal dimension elements of the SOCPRI ontology play the vital role to infer different social contexts of OSNs users. The main classes that facilitate modeling and reasoning about diverse social contexts are Context

class along with its subclasses and ContextualNorm class.

The online social networks allow users to perform different types of social interactions. To classify such social interactions the SOCPRI ontology provides Interaction-Type class which extends various types of user interactions supported by the current online social network. To define the concept of social interaction we offer a separate class by the name of Interaction. The SOCPRI also models information flow activities and takes into consideration norms of appropriateness and norms of distribution for access management of a digital resource of an online social network user. The relationship strength among online social network users is represented by TieStrength and its dimension which includes concepts such as InteractionFrequency, SocialDistance, Closeness, RelationshipDuration, etc. The classes like PredictiveIndicator and PredictiveIndicatorType play an important role with their related object properties to give an idea about profile similarity and interaction pattern. The SOCPRI ontological model represents all key concepts that are needed to define privacy in online social networks from the social perspective. Besides semantically enhanced representation of these concepts, inferred axioms allow extracting more implicit information. The reasoning capability makes the SOCPRI ontological model so attractive to apply in the context of privacy where the flow of information is controlled by explicit and implicit contextual norms. The SOCPRI ontology is fairly complex structure and it is more practical to elaborate core conceptual elements with more technical details. Before starting the technical discussion, we describe ontology developing methodology for extracting core conceptual elements and their relationships.

5.3.1 Contribution of the SOCPRI Ontology

The main contribution of the SOCPRI ontology is modeling privacy for online social networks from the social perspective. The social perspective of privacy is ignored by all of the existing ontologies developed for representing social data. Social perspective of

privacy takes into consideration existing rich research literature of the sociology about contextual integrity, presentation of self, and tie strength. These social theories give insight on how to manage diverse social relationships and what are interaction patterns and personal information disclosure practices between strong and weak ties. With this conceptual background from sociology, the SOCPRI models privacy framework that addresses the issues of context collapse, invisible audience, and user control. These are the major issues faced by OSNs users to protect their personal information from other users in the social web environment.

Another contribution of SOCPRI ontology is modeling role and relationship based notion of social context. Erving Goffman's social theory about the presentation of self plays the vital role in conceptualizing social context. According to the sociological approach, social context is a synonym of the social environment in which users interact with each other and perform their roles. The self-presentation of the user varies based on social environment and audience. For example, a user may self-present in significantly different ways when in a business meeting versus on a date. Online social networks have introduced an entirely new method of self-presentation where the employer and romantic partners are placed on same communication plane and make it difficult for users to segment audiences and present varied versions of the self. The modeling of role and relationship based social context is a step towards context segregation and managing privacy friendly presentation of self.

One of the contributions of SOCPRI ontology is modeling relationship strengths among OSN users. Granovetter social theory about tie strength plays an important role to conceptualize strong and weak relationships among OSNs users. Current online social networks assume static, binary, symmetric relationship of equal value between all the directly connected users. In reality, social relationships are of varying tie strength (how close two individuals are to one another), dynamic (change over time), and asymmetric in nature (one person pays attention to another, it does not mean the latter

will reciprocate). It is a challenging task to capture strength, dynamism, and asymmetry in user relationships in contemporary online social networks. Most online social networks employ "friendship" as the only type of possible relationship. The friendship is a binary relational tie which provides only a coarse indication of the nature of the relationship. In human societies, relationships are much more complicated than a single binary relational tie. The online social networks are kind of virtual societies that exhibit many of the characteristics of human societies in terms of forming relationships and how those relationships are utilized. In human societies, the relationship strength is a crucial factor for individuals while deciding the boundaries of their privacy. The SOCPRI ontology models tie strength and infers relationship strength from predictive indicators associated with various dimensions of tie strength. The existing literature of sociology suggests seven dimensions of tie strength. Each dimension of tie strength can be identified by several relationship predictive indicators. Current literature on online social networks proves that relationship indicators for online ties are similar to those for offline ties. The detail description of dimensions of tie strength and relationship predictive indicators is given in section 3.3.3 of the third chapter.

Finally, SOCPRI ontology models temporal dimension of the roles and relationships. The reuse of **TimeInterval**²², Time-indexed Value in Context (**TVC**²³) ontology patterns along with Publishing Role Ontology (**PRO**²⁴) facilitate representation of dynamic roles and relationships that change over time. The role can be specified to exist over a defined period of time, and within a specific context, and with respect to a particular relationship. We also differentiate between user-centric and resource-centric roles of online social networks users. The further technical details about temporal modeling are discussed in subsections 5.5.2 and 5.5.3. In the following section, we describe in detail ontology development methodology used to build the SOCPRI ontology.

²²TimeInterval, <http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl>

²³TVC, <http://www.essepuntato.it/2012/04/tvc>

²⁴PRO, <http://www.sparontologies.net/ontologies/pro/source.html>

5.4 Ontology Development Methodology

The ontological engineering has evolved into a discipline of its own and few researchers have proposed a series of steps and methodologies for developing ontologies. These methodologies cover various aspects of ontology development process such as defining scope and requirements of an ontology, specifying an ontology life cycle model, describing methods, techniques, tools that can be used to support the development process. The first ontology development methodology was proposed by Uschold and King in 1995 [163]. This methodology is composed of four distinct stages: identification, construction, evaluation, and documentation. The first stage identifies the purpose and scope of the ontology. The ontology construction stage is composed of three phases. In the first phase concepts and their relationships are defined. The focus of the second phase is the formalization of the concepts and relationships defined in the earlier phase. The sole purpose of integration phase is to find the possibility of reusing existing ontologies and this activity can be carried out in parallel with capture and coding phases. Evaluation stage of ontology development process uses technical criteria to verify the specification using competency questions or real-world validations. Developing the documentation of the ontology is the final stage of this methodology. This methodology has been criticized for offering little support in identifying ontology classes and relationships.

Gruninger and Fox proposed a methodology [164] that was derived from their own experience in developing ontologies for the domain of business processes and activities modeling. The authors used motivating scenarios to describe the problem that was not addressed by methodology proposed by Uschold and King [163]. This methodology proposed six stages for ontology development. Description of motivating scenarios is the first stage which describes problem and carries the informal semantics of the concepts and relations to be included in the ontology. The second stage deals with the formulation of informal competency questions based on the motivating scenarios.

These questions are considered as requirements that an ontology must be able to meet. They are also used to evaluate the ontological commitments. The next stage is the specification of ontological terms using a formal representation. The terminology is specified using a knowledge representation language, such as FOL (First Order Logic) or KIF (Knowledge Interchange Format). Once the terminology of the ontology is formalized, it is quite easy to describe the competency questions into a formal language. This stage is also referred as the formulation of formal competency questions. The last two stages are dealing with axiom specification and verification of the ontology completeness respectively. This approach has a major problem that it derives the concepts of an ontology from motivating scenarios alone. In fact, the scenario technique is best for the identification of dynamic entities, rather than static entities [165].

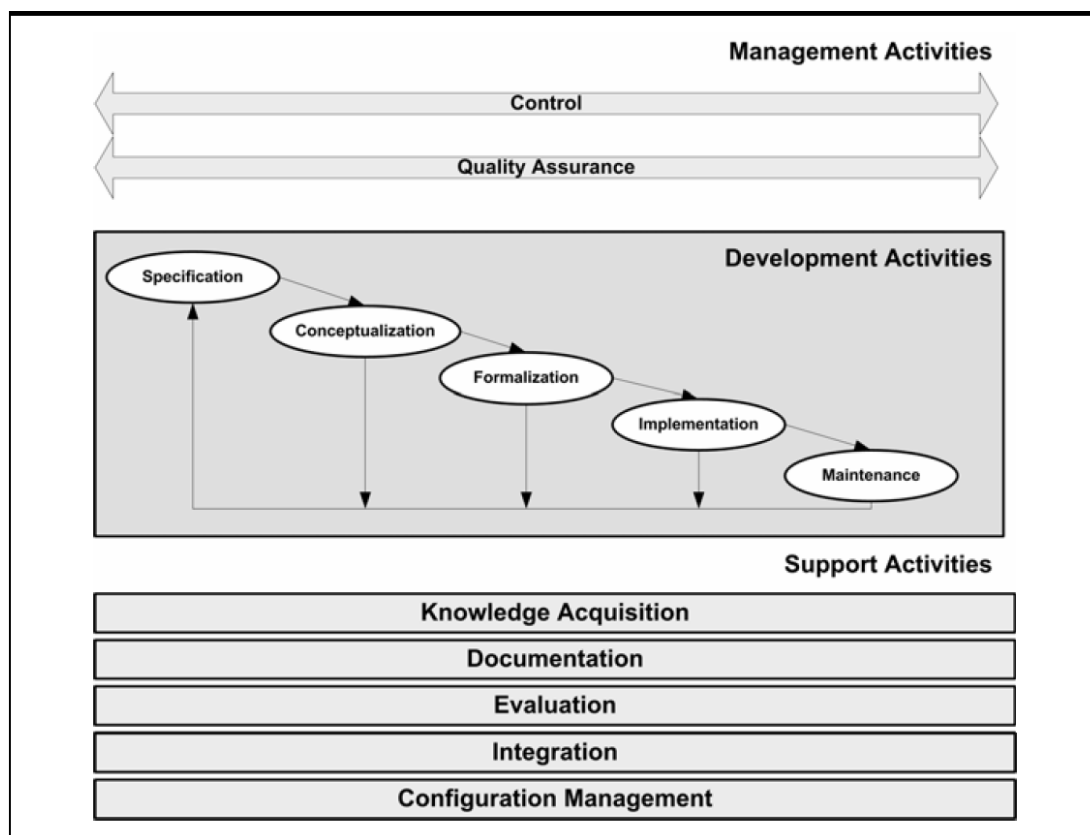


Figure 5.4: Methontology ontology development process [7]

With the maturity of ontology engineering discipline, few other ontology development methodologies are proposed by researchers. Some of these methodologies are methontology [42], on-to-knowledge methodology [166], neon methodology [167], and simplified methods like ontology development 101 [168]. In this dissertation, we adopt **Methontology** methodology [42] for the development of the SOCPRI ontology. Methontology is a well-structured methodology used to build ontologies from the scratch. It is based on the IEEE standard for software development [169] and supported by ODE [170]. The methodology suggests three types of activities for building an ontology. These activities are ontology management activities, ontology development activities, and ontology support activities. The tasks associated with management activities are control, and quality assurance. The development activities include specification, conceptualization, formalization, implementation, and maintenance. The tasks related to ontology support activities are knowledge acquisition, documentation, evaluation, integration, etc. The management and support activities can be performed in parallel with development activities. From software engineering perspective, the recommended life cycle for building ontologies with methontology is evolving prototypes. The ontology development process based on methontology is presented in figure 5.4. In the following subsections, we discuss the development of SOCPRI ontology using methontology development process.

5.4.1 Requirement Specification for SOCPRI Ontology

The main goal of requirement specification phase is to state why SOCPRI ontology is being created and what specific requirement this ontology must fulfill. The analysis of most of the methodologies for building ontologies reveals that ontology requirement specification activity is carried out in all methodologies in one way or the other. Although the aforementioned methodologies propose methods for carrying out the ontology requirement specification activity consisting of high-level steps, they do not

provide detailed guidelines that explain how to carry out each step. The Methontology [42] identifies the goal of ontology requirement specification activity that is to produce either an informal, semi-formal or formal ontology specification document written in natural language. The Neon methodology call this document **ORSD (Ontology Requirement Specification Document)** and propose a template for writing the ORSD as shown in figure 5.5. This template contains information about the purpose, scope, implementation language, intended end-users, intended uses, requirements, and pre-glossary of terms of the ontology which is being built.

The main purpose of an ORSD is to serve as an agreement among ontology engineers, users, and domain experts on what requirements the ontology should cover. It plays a key role during ontology development process because it can be used for speeding up the development process. It facilitates reuse of existing ontologies or ontology design patterns. The document also permits the verification of ontology along the development process with respect to requirements that the ontology should fulfill. We prepared ORSD before building the SOCPRI ontology. The document helped to search for existing ontology resources for re-engineering them into the SOCPRI ontology. We are reusing some classes and object properties from Friend of A Friend (FOAF ²⁵) and Publishing Role Ontology (PRO ²⁶). We also reused some concepts from TVC ²⁷ and TimeInterval ²⁸ ontology design patterns. During requirement identification phase, we identified in total 96 competency questions which are described in detail in subsection 5.4.1.3. These competency questions facilitated extraction of the SOCPRI ontology terminology and eased the verification the SOCPRI ontology with respect to requirements. The following subsections describe different parts of ORSD. We followed the ORSD template shown in figure 5.5.

²⁵FOAF, <http://xmlns.com/foaf/spec/>

²⁶PRO, <http://www.sparontologies.net/ontologies/pro>

²⁷TVC, <http://www.essepuntato.it/2012/04/tvc>

²⁸TimeInterval, <http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl>

Ontology Requirements Specification Document Template	
1 Purpose	
	<i>The main general goal of the ontology. In other words, the main function or role that the ontology should have.</i>
2 Scope	
	<i>The general coverage and the degree of detail that the ontology should have.</i>
3 Implementation Language	
	<i>The formal language that the ontology should have.</i>
4 Intended End-Users	
	<i>The intended end-users expected for the ontology.</i>
5 Intended Uses	
	<i>The intended uses expected for the ontology.</i>
6 Ontology Requirements	
a. Non-Functional Requirements	
	<i>The general requirements or aspects that the ontology should fulfil, including optionally priorities for each requirement.</i>
b. Functional Requirements: Groups of Competency Questions	
	<i>The content specific requirements that the ontology should fulfil, in the form of groups of competency questions and their answers, including optionally priorities for each group and for each competency question.</i>
7 Pre-Glossary of Terms	
a. Terms from Competency Questions	
	<i>The list of terms included in the competency questions and their frequencies.</i>
b. Terms from Answers	
	<i>The list of terms included in the answers and their frequencies.</i>
c. Objects	
	<i>The list of objects included in the competency questions and in their answers.</i>

Figure 5.5: ORSD Template [8]

5.4.1.1 Identification of Purpose, Intended Uses and Users

In this section, our objective is to present the main goal of SOCPRI and establish who the main intended end-users of the ontology are. We also identify intended uses of SOCPRI ontology. The main goal of developing the ontology is modeling diverse social relationships of online social networks users. This ontological model is useful in representing OSN users relationships on the basis of social context and relationship strength among them. The SOCPRI ontology focused on realizing contextual integrity of OSN users in the environment where context collapse is inherent. It also models role based presentation of self and relationships strength of OSNs users. The ontological model is inspired from well-founded social theories about self-presentation, contextual integrity, and tie strength. These social theories provide insights on how to manage diverse social relationships and what are interaction patterns and personal information disclosure practices between strong and weak ties. Based on this model a privacy friendly online social networking environment can be developed to address existing issues of privacy, therefore, the intended end-users for SOCPRI ontology are people using social networking sites such as Facebook and Google+. The intended uses of this ontological model are:

Use 1. Modeling social relationships of online social networks users

Use 2. Modeling digital resources of online social networks users

Use 3. Modeling social interactions of online social networks users

Use 4. Modeling social contexts of online social networks users

Use 5. Inferring relationship strength of online social networks users

Use 6. Inferring privacy requirements of online social networks users

Use 7. Inferring privacy policy of online social networks users

The scope of the SOCPRI ontology can be determined by the list of competency questions that it should be able to answer. These questions can be easily extracted from the scenarios motivating the development of the ontology.

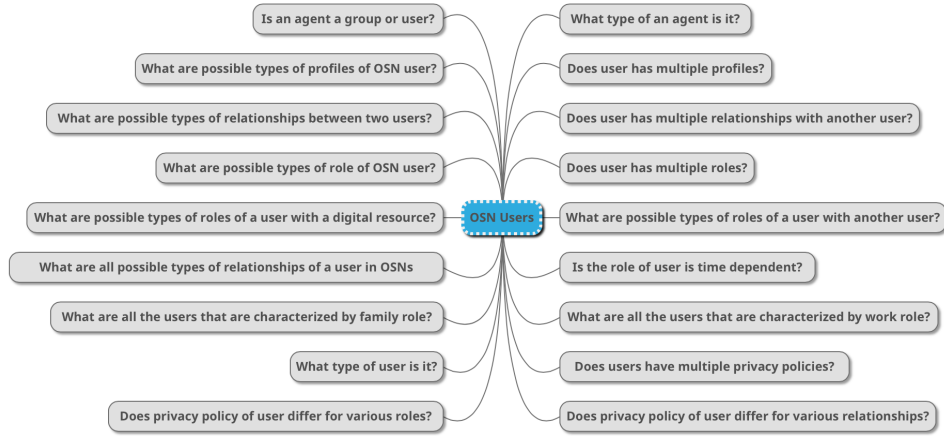


Figure 5.6: Competency Questions about OSN Users

5.4.1.2 Motivating Scenarios

In this section, we present some scenarios that are representative of real situations in current online social networks to demonstrate the shortcomings of existing privacy solutions. One of the prevailing problems is context collapse that makes it difficult for people to disclose personal information selectively to various life facets. Let us walk through a scenario to illustrate this problem. We consider, Bob, as the primary actor in this scenario, he is connected with several people from different facets of his life. Bob is the father of Alice and he wants to project himself as a dad through his online identity. Bob is the supervisor of Charlie and he wants to present his consistent and coherent personality as a researcher and academician. Bob is also connected with his old university friend Eve and she always shares old memories by tagging him in pictures taken

during their university life. Bob wants to present himself in significantly different ways when interacting with Alice, Charlie and Eve. Current OSNs places family members, colleagues and romantic partners on same communication plane, make it more difficult for Bob to segment audiences and present varied versions of the self. Bob wants to avoid any embarrassing situation caused by revealing inappropriate information to unintended audiences. He wants to reveal only family related information with Alice and work related information with Charlie. Bob intends to share more colorful aspects of his social life with Eve. Currently, all profile information of Bob is freely available to all his friends equally by default. There is no such thing as context related information; all the available information is context free and easily accessible by all friends until tediously managed through cumbersome privacy controls or by creating lists and circles. It has been already highlighted that the responsibility of maintaining the appropriateness of these settings and lists lies solely on the user, and the vast majority of users do not use these features. Moreover, social relationships are dynamic; therefore, maintaining the appropriateness of audience is highly contextual. In this case, Bob faces the problem of maintaining competing social spheres with context centric information sharing in an online social network.

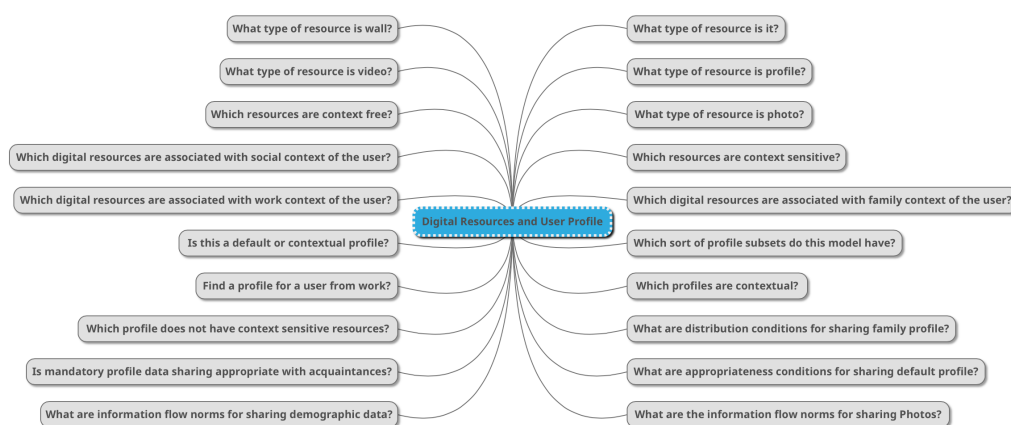


Figure 5.7: Competency Questions about Digital Resources and User Profile

Another scenario deals with privacy problem emerging due to the absence of norms of flow that deals with the transfer of information from one party to another party. Consider a scenario where a user Bob adopts a policy that conceals his friends from the public. On the other hand, Eve adopts a weaker policy that allows any user to view her friends. In this case, Bob's relationships with Eve can still be learned through Eve. Each relationship carries potentially sensitive information that either user may not wish to reveal. OSNs provide a mechanism to conceal a user's friend-list, a user can only control one direction of an inherently bi-directional relationship. Eve's weaker privacy policy causes the inappropriate distribution of information about Bob to strangers.

Photo tagging is another example of a common situation for inappropriate distribution of information. Photo tagging allows users to share contextual information about themselves or their friends by linking a user to certain content on OSNs. In this scenario, Dave is a friend of Elena and Fred in both real life and on OSNs, whereas, Fred and Elena are not friends. Elena and Fred met each other during a party at Dave's residence. Fred realized that Elena is a member of selection board at the company where he already applied for the position of software engineer. Fred presented professional facet of his life to Elena during this meeting to enhance his chances of short listing and later on selection for the job. Dave posted some pictures of himself and Fred on OSN and tagged him. These pictures were taken at some party couple of weeks ago. Fred did not mind sharing such pictures with friends, but he has concerns about Elena reaction to these pictures from the perspective of professionalism. Elena as a friend of Dave can view all the pictures posted by him. Fred is worried about out of context disclosure of these pictures that can jeopardize his chances of selection in Elena's company.

These scenarios show that consistent and coherent self-presentation of the OSN users is under constant threat in contemporary online social networks. Bob can employ many methods to address the issue faced by him, but he must be constantly vigilant before sharing every piece of sensitive personal information and follow several steps necessary

to ensure that his online information may not propagate to unintended audiences. Nevertheless, following unnecessary steps put the extra cognitive burden and complicate the online social experience, especially, when he hardly understands how information flows on current online social networks. So we have a healthy skepticism about the safe utility of such online social networks.

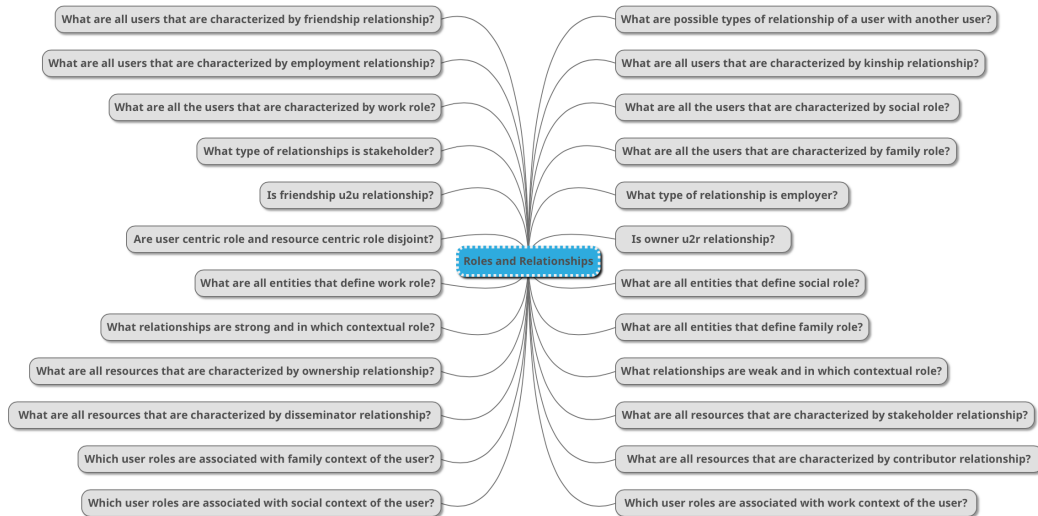


Figure 5.8: Competency Questions about User Roles and Relationships

5.4.1.3 Competency Questions

One of the ways to determine set of requirements for an ontology is to sketch a list of questions that the ontology should be able to answer. These are called competency questions and play an important role in the ontology development life cycle. Given a set of scenarios related to a domain of discourse, the ontology engineer should be able to place a set of questions which represent the user requirement and limits the scope of ontology. These questions are expressed in natural language and supports the development process in different ways. From the competency questions and their answers, ontology engineer can manually extract the terminology that will be formally represented in the ontology by means concepts, attributes, and relations. These ques-

tions will serve as the litmus test later in the ontology evaluation phase. With a set of competency questions at hand, it is possible to determine whether an ontology was created correctly. If the ontology is capable to answer correctly the competency questions with its necessary and sufficient axioms then the ontology is validated against its requirements.

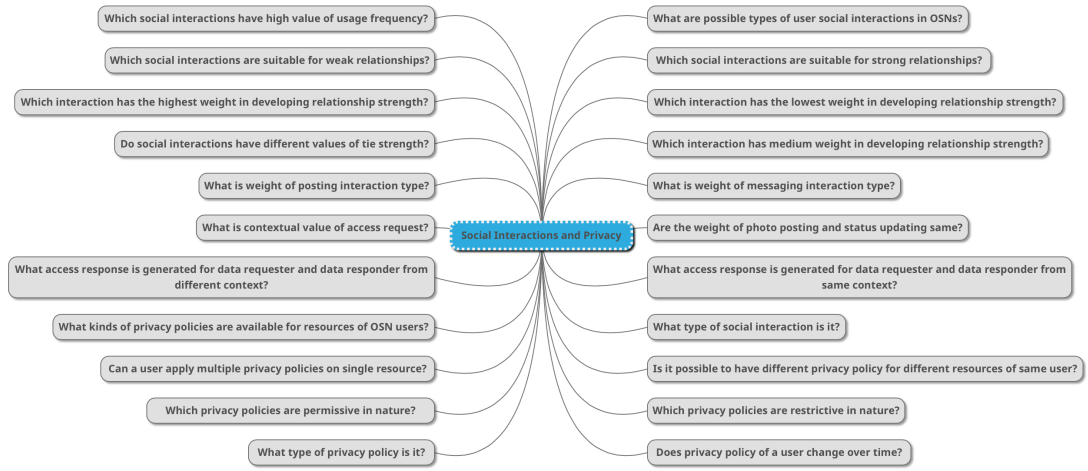


Figure 5.9: Competency Questions about User Interactions and Privacy Policy

Most of the existing methodologies suggest the identification of competency questions as the technique for establishing the ontology requirements. However, current practices of ontology engineering make a superficial use of competency questions. One of the main reasons for their superficial use is the lack of tools that facilitate ontology engineers to check if competency questions are fulfilled by the ontology being defined. In this dissertation, we translate natural language competency questions into queries with the aim of validating an ontology. This approach ensures appropriate use of competency questions in ontology development life cycle.

We identified 104 competency questions for the SOCPRI ontology and manually grouped these into five groups such as OSN user, roles and relationships, tie strength and predictive indicators, digital resources and profile, and social interactions and pri-

vacy. We used mind map tools to represent these questions that are used to generate, visualize, structure and classify ideas. Another advantage of these tools is that requirements visualization in form of a hierarchy is very intuitive and easy to understand and manage. The figures from 5.6 to 5.10 represent five groups of competency questions in mind map diagrams. We adopted middle out approach to identify competency questions that mix both top-down and bottom-up approaches. We started with important questions that were composed and decomposed later on to form abstract and simple questions respectively. The goal of developing competency questions was to extract terminology that will be formally represented in the ontology by means of classes, properties, and their relations.

5.4.1.4 Implementation Language

The SOCPRI ontology has been defined in OWL (Web Ontology Language) that is W3C recommendation to express meanings and semantics. The OWL ontologies are described in terms of classes of individuals as well as the properties of those individuals. The properties can connect different individuals or they can relate data attributes to an individual. The relations are described in a formal way with strictly defined semantics that allows to apply inference rules to infer implicit facts from existing ones. The ontology has been developed with the aim to conform to the OWL-DL subset of the OWL ontology language, in order to enable broad tool support and to ensure computational completeness and decidability. We used Protege as an ontology editor. We used various reasoners for checking consistency of the ontology such as Fact++²⁹, Hermit³⁰, Pellet³¹, etc. We used OWLViz³² and OntoGraf³³ Protege plugins for visualization of our ontology. We also used an open source tool Graffoo³⁴ (Graphical Framework for

²⁹Fact++, <http://owl.cs.manchester.ac.uk/tools/fact/>

³⁰Hermit, <http://www.hermit-reasoner.com/>

³¹Pellet, <https://www.w3.org/2001/sw/wiki/Pellet>

³²OWLViz, <https://protegewiki.stanford.edu/wiki/OWLViz>

³³OntoGraf, <https://protegewiki.stanford.edu/wiki/OntoGraf>

³⁴Graffoo, <http://www.essepuntato.it/graffoo/>

OWL Ontologies) to present the classes, properties and restrictions in SOCPRI ontology in clear and easy to understand way. The DL query ³⁵ Protege plugin is used to query SOCPRI ontology.

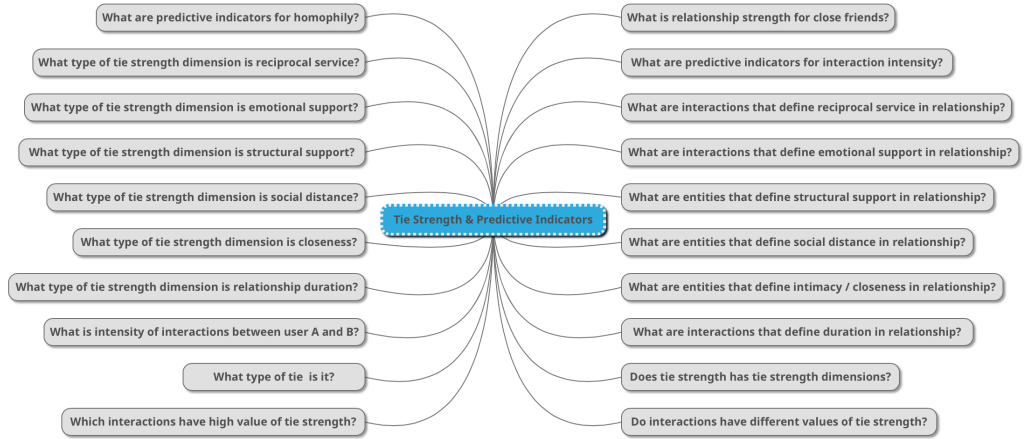


Figure 5.10: Competency Questions about Tie Strength and Predictive Indicators

5.4.2 Reusing Existing Ontologies

One of the advantages of using an ontology based social relationships modeling is that it provides an opportunity to integrate and reuse classes and properties other existing ontologies. In this model, we are reusing some concepts from **FOAF (Friend Of A Friend)** and **PRO (Publishing Roles Ontology)** ontologies. We are also reusing **TVC (Time indexed Value in Context)** and **TimeInterval** ontology design patterns to model temporal roles of OSN users. FOAF ontology describes persons, their activities and their relations to other people and objects. We reuse three classes from FOAF ontology: Agent, Group, and Document. The foaf:Agent class defines any kind of agents, such as a person, a group, an organization or a software agent. The foaf:Group class represents a collection of individual agents. The foaf:Document class

³⁵DL Query, <https://protegewiki.stanford.edu/wiki/DLQueryTab>

is an abstract entity that extends all kind of digital resources in online social networks. We also reuse foaf:member object property that relates a Group to an Agent that is a member of that group. The User is main concept in the SOCPRI ontology. We extend User from foaf:Agent class and User is connected with foaf:Group via memberOf object property.

The publishing roles ontology is used to specify how an agent has a role relating to a contextual entity, and the period of time during which that role is held. We reuse two classes and two object properties of PRO for our ontology. The pro:Role class model a role an agent has. Individual members of this class or its subclasses are used to specify particular roles. The pro:RoleInTime class can be used to describe a role an agent may have in a particular situation that can be restricted to a particular time interval. The pro:holdsRoleInTime object property relates an agent to a role that he holds and pro:withRole object property connects an agent's role in time to a definition of the type of role held by this agent. The reuse of these elements of PRO in SOCPRI ontology facilitates to specify how a user has a role relating to a contextual entity, and the period of time during which that role is held. The ontology design patterns TVC and TimeInterval along with PRO make it easy to extend the set of specified roles, simply by adding new individuals to the classes **SocialRole**, **FamilyRole**, and **WorkRole**. The TVC ontology pattern used for the description of scenarios that involve entities having some value during a particular time and within a particular context. Presentation of self in online social networks requires a user to demonstrate some face at a particular time and within a particular context.

5.5 Core Conceptual Elements of SOCPRI Ontology

The main purpose of developing the SOCPRI is to represent diverse social relationships of OSNs users. The ontology is inspired by well grounded social theories of Goffman,

Nissenbaum, and Granovetter. The SOCPRI ontology comprises of 102 classes, 58 object properties, 14 datatype properties and 1161 axioms. The detailed ontology metrics are shown in figure 5.2. The ontology is populated with 75 individuals and 114 object property assertions about these individuals. The instantiation of ontology provides an opportunity to evaluate the functional aspects of the model by extracting related data against various queries. These queries are composed of competency questions which provided the basis for the development of the model. The detail description of query based ontology evaluation is provided in section 6.6 of the sixth chapter.

In this section, we focus our discussion on technical details of the various parts of the SOCPRI ontology. The SOCPRI ontology can be divided into five different parts depending on the modeling aspects of the ontology. The first part discusses different modeling elements of the social web user. The main entities in this part are *User*, *Agent*, *Group*, *Wall*, *Profile*, etc. The second part describes various modeling elements related to user roles and relationships. The conceptual elements of this part are *Role*, *Relationship*, *RoleInTime*, etc. The third models tie strength and its various dimensions. Social interactions of OSN users play the vital role to infer tie strength among users. Our ontological model represents various types of interactions supported by current social web environment. The key entities contributing in this part are *Interaction*, *InteractionType*, *TieStrength*, etc.

Modeling digital resources owned by the social web user is the focus of fourth part. It divides digital content into a context free and context sensitive resources. The access to context sensitive content is managed by contextual norms. The key entities associated with this part are *DigitalResource*, *ContextFreeResource*, *ContextSensitiveResource*, *ContextualProfile*, etc. Finally, Our ontology models contextual privacy of social web users. The contextual privacy deals with restricting the disclosure of digital resources out of the context. Figure 5.11 shows the graphical overview of the SOCPRI ontology.

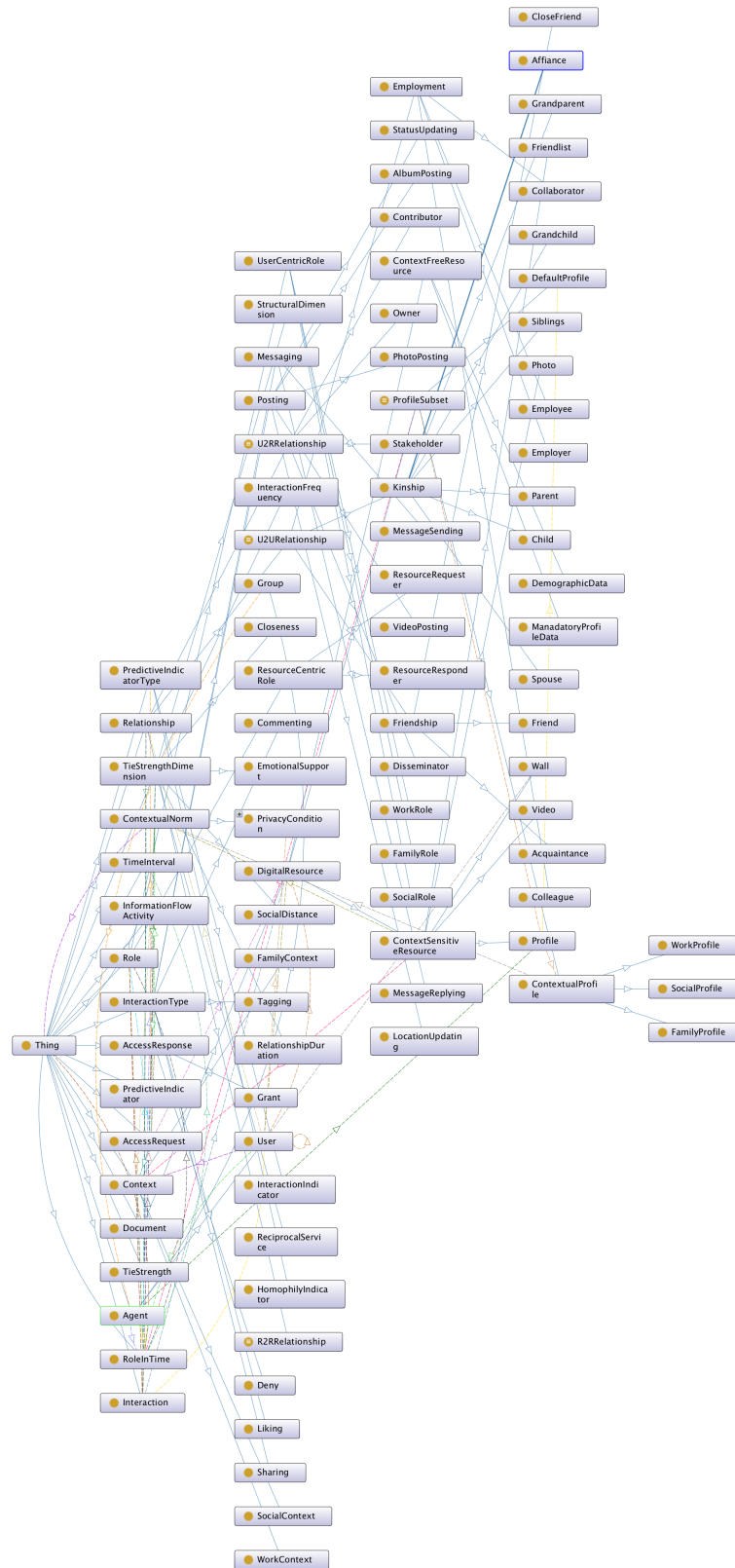


Figure 5.11: Overview of Classes and Object Properties Relations in SOCPRI

5.5.1 Social Web User Modeling

The core conceptual element in SOCPRI ontology is social web user. This is represented by a **User** class which is derived from **foaf:Agent** class. Social web user is connected to other users with an object property known as **relates**. This property also connects a social web user to a **DigitalResource** class. This property is sub-property of **has-Relationship** object property. The modeling details of the **User** class and related object properties are given below in Manchester syntax. In this dissertation, we are using Manchester OWL syntax to express different elements of the SOCPRI ontology. It is a new syntax designed to write OWL descriptions which is derived from the OWL Abstract Syntax, but is less verbose which means it is quick and easy to read and write.

Class : User

EquivalentTo :

Agent

and (withRole some UserCentricRole)

and (hasContext some Context)

and (hasProfile some Profile)

and (hasRelationship some Relationship)

and (holdsResource some DigitalResource)

and (ownsWall some Wall)

and (performs some Interaction)

and (hasCurrentContextualNorm some AppropriatenessCondition)

and (hasCurrentContextualNorm some DistributionCondition)

ObjectProperty : hasRelationship

Domain :

User

Range :

```

        Relationship
ObjectProperty: relates
    Domain:
        User
    Range:
        User ,
        DigitalResource
SubPropertyOf:
    hasRelationship

```

A social web user participates in different kind of relationships which are modelled by classes such as **U2URelationship**, **U2RRelationship** and **R2RRelationship**, etc. The object property **hasRelationship** connects **User** and **Relationship** classes. The social context of the user is represented by **Context** class. This is generic class which provide general features of the user context. The more specific classes are derived from the context class such as **FamilyContext**, **WorkContext**, etc. The detailed technical discussion of diverse social contexts of social web user is presented in section 5.5.3. The object property **hasContext** establishes connection between **User** and **Context** classes. This property is functional which means a user can have only one social context at a particular time. The manchester OWL description of **U2URelationships** and their subclasses is given below along with some related object properties.

```

Class: U2URelationship
    EquivalentTo:
        relates exactly 2 User
SubClassOf:
    Relationship ,
    hasTieStrength some TieStrength ,

```

```

        hasFrequencyOfInteractions some xsd:integer
DisjointWith:
    R2RRelationship , U2RRelationship
Class: U2RRelationship
    EquivalentTo:
        relates exactly 1 (DigitalResource
            and User)
    SubClassOf:
        Relationship
DisjointWith:
    R2RRelationship , U2URelationship
Class: R2RRelationship
    EquivalentTo:
        relates exactly 2 DigitalResource
    SubClassOf:
        Relationship
DisjointWith:
    U2RRelationship , U2URelationship
Class: Employment
    EquivalentTo:
        U2URelationship
        and (withRole some WorkRole)
        and (hasContext some WorkContext)
    SubClassOf:
        U2URelationship
DisjointWith:
    Friendship , Kinship

```

Class: Friendship

EquivalentTo:

U2URelationship

and (withRole some SocialRole)

and (hasContext some SocialContext)

SubClassOf:

U2URelationship

DisjointWith:

Employment, Kinship

Class: Kinship

EquivalentTo:

U2URelationship

and (withRole some FamilyRole)

and (hasContext some FamilyContext)

SubClassOf:

U2URelationship

DisjointWith:

Employment, Friendship

ObjectProperty: hasContext

Domain:

User

Range:

Context

Characteristics:

Functional

ObjectProperty: withRole

Domain:

RoleInTime

Range :

Role

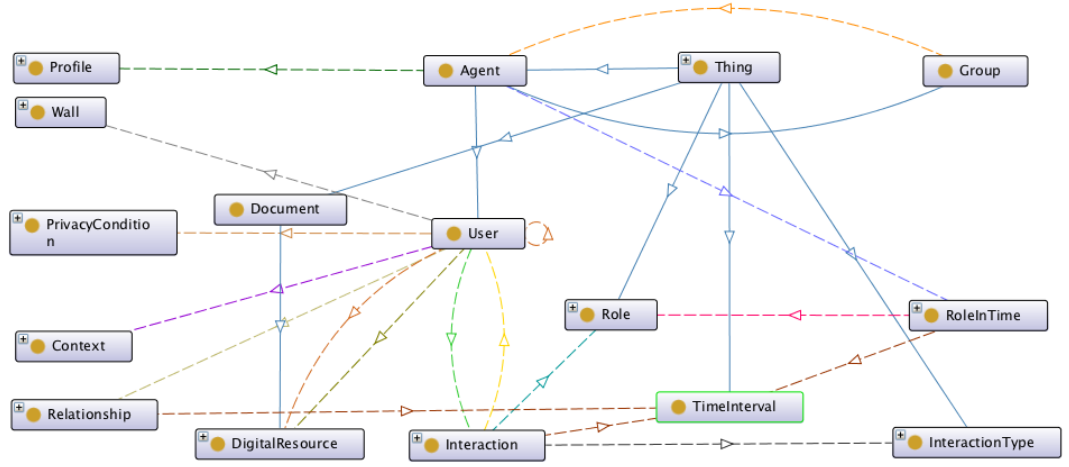


Figure 5.12: Graphical Representation of Social Web User Modeling

The user engage in wide variety of social interactions in social web environment. These social interactions are represented by classes called **Interaction** and **InteractionType**. The object properties **performs** and **performedWith** connect the user with **Interaction** class. The relationship between **Interaction** and **InteractionType** classes is established through **hasInteractionType** object property. The Interaction class is also connected with **Role** class which is superclass for diverse social roles of the user. The object property that establishes relation between Role and Interaction classes is known as **associatedRole**. The manchester OWL description of these classes and object properties is given below:

Class: Interaction

EquivalentTo:

(associatedResource some DigitalResource)

and (associatedRole some Role)
 and (generates some InformationFlowActivity)
 and (hasInteractionType some InteractionType)
 and (performedWith some User)

ObjectProperty: performs

Domain:

User

Range:

Interaction

InverseOf:

performedWith

ObjectProperty: performedWith

Domain:

Interaction

Range:

User

InverseOf:

performs

ObjectProperty: hasInteractionType

Domain:

Interaction

Range:

InteractionType

ObjectProperty: associatedRole

Domain:

Interaction

Range:

Role

ObjectProperty: associatedResource

Domain:

Interaction

Range:

DigitalResource

ObjectProperty: generates

Domain:

Interaction

Range:

InformationFlowActivity

As discussed earlier the status of an individual end-user is changed by social web environment. The individual end-user acts as a content producer rather than a content consumer. There is a huge amount of digital content that is generated by an individual end-users. The user generated content is represented by **foaf:Document** class. This is a generalized representation of the digital content. The SOCPRI ontology uses **DigitalResource** class to represent context free and context sensitive digital content of the user. The classes **User** and **DigitalResource** are connected through **relates** and **holdsResource** object properties. The activity stream of the user is represented by a class called **Wall**. Each user is connected to his/her wall through an object property known as **ownsWall**. All the users inherit a profile from the **Agent** class. The object property **hasProfile** relates a user with their profile. The Manchester syntax for these classes and object properties is given below:

Class: ContextFreeResource

EquivalentTo:

DemographicData or ManadatoryProfileData or Text

```
SubClassOf:
    DigitalResource
Class: ContextSensitiveResource
SubClassOf:
    DigitalResource ,
    belongsToContext some Context
Class: Profile
EquivalentTo:
    ContextSensitiveResource
    and (hasProfileSubset some ProfileSubset)
Class: Wall
EquivalentTo:
    ContextSensitiveResource
    and (hostsResource some DigitalResource)
    and (wallOwnedBy some User)
ObjectProperty: hasProfile
Domain:
    Agent
Range:
    Profile
Characteristics:
    Functional
ObjectProperty: holdsResource
Domain:
    User
Range:
    DigitalResource
```


ObjectProperty: ownsWall

Domain:

User

Range:

Wall

The SOCPRI ontology also reuses **Group** class of FOAF ontology. A social web user can be member of multiple groups at a time. This feature is inherited by the **User** class from **foaf:Agent** class. Apart from participating in various groups the user try to access the digital content of other users. The digital content of users is accessible if and only if privacy conditions are met. The privacy conditions are modelled using **PrivacyCondition** and its subclasses. This class is derived from **ContextualNorm**. The relation between the **User** and **PrivacyCondition** classes is established using an object property known as **hasCurrentContextualNorms**. The domain of this property is User class and range PrivacyCondition class. Our ontology also models diverse social roles of the user which includes social, family and work roles. The class **Role** is top level class for modeling various roles of the user. The classes **UserCentricRole** and **ResourceCentricRole** are derived from the **Role** class. The detailed description of user centric and resource centric roles is given in following section. The social web user performs different roles at different times to model this feature in our ontology, we reuse TVC and TimeInterval ontology patterns. The classes **Role**, **RoleInTime**, **TimeInterval** and **Agent** along with object properties **atTime**, **withRole** and **holdsRoleInTime** model temporal roles. The object property **withRole** connects **Role** and **RoleInTime** classes. The class **User** is connected to **RoleInTime** classes via its parent class **Agent**. The object property **holdsRoleInTime** connects these two classes. The temporal aspect is incorporated with **TimeInterval** class and **atTime** object property. The manchester syntactical description of some important classes and object properties is given below:

Class: ContextualNorm

SubClassOf:

hasContextualValue some
(AppropriatenessCondition or DistributionCondition)

ObjectProperty: hasCurrentContextualNorm

Domain:

User

Range:

PrivacyCondition

Class: TimeInterval

SubClassOf:

hasIntervalEndDate max 1 xsd:dateTime ,
hasIntervalStartDate max 1 xsd:dateTime

ObjectProperty: atTime

Domain:

RoleInTime ,
Relationship ,
Interaction

Range:

TimeInterval

The user is the central concept of the SOCPRI ontology. It is represented by *User* class. It is defined a class with necessary and sufficient conditions to model social web user. The main difference between primitive and defined classes is that primitive class provides only necessary conditions, whereas, the defined class contains at least one set of necessary and sufficient conditions. In this section, we have given a general overview of the various aspects of the social web user. The graphical representation shown in figure 5.12 contains main concepts, attributes and their relations about the social web

user.

5.5.2 Role and Relationship based Social Context Modeling

Role and relationship are important factors for modeling self-presentation of the social web user. As discussed in the earlier section, a user plays diverse roles in the social web environment. The modeling of user roles is a quite complicated problem. We benefited from existing ontology design patterns for modeling user roles. The figure 5.13 shows the graphical representation of user roles in SOCPRI ontology implemented in protege. The class **Role** is a super class of all classes related to social roles of the user. We further divided roles into two categories user centric and resource centric roles. These roles are represented by **UserCentricRole** and **ResourceCentricRole** classes. Three classes are extended from **UserCentricRole** class which includes **FamilyRole**, **WorkRole** and **SocialRole**. The resource centric roles identified are either **ResourceRequester** or **ResourceResponder**. The interaction pattern of the user changes with change in the role to model this feature the object property **associatedRole** links **Interaction** and **Role** classes. The role of the user is also taken into consideration for information access requests. The object property **requesterRole** contains **AccessRequest** as a domain and **Role** as a range.

The SOCPRI ontology characterizes diverse social role of the user in the social web environment. It allows users to specify roles that exist within a specific context and with a particular relationship over a defined period of time. Our ontology reuses some concepts from publishing roles ontology that is based on the time-indexed value in context ontology pattern. This ontology design pattern allows an agent to perform different contextual roles at different period of time. The SOCPRI ontology takes into consideration the Goffman's theory about the presentation of self which advocates multiple roles of an individual during their interaction with the different audience (context). This theory is also supported by another contemporary social theorist known as

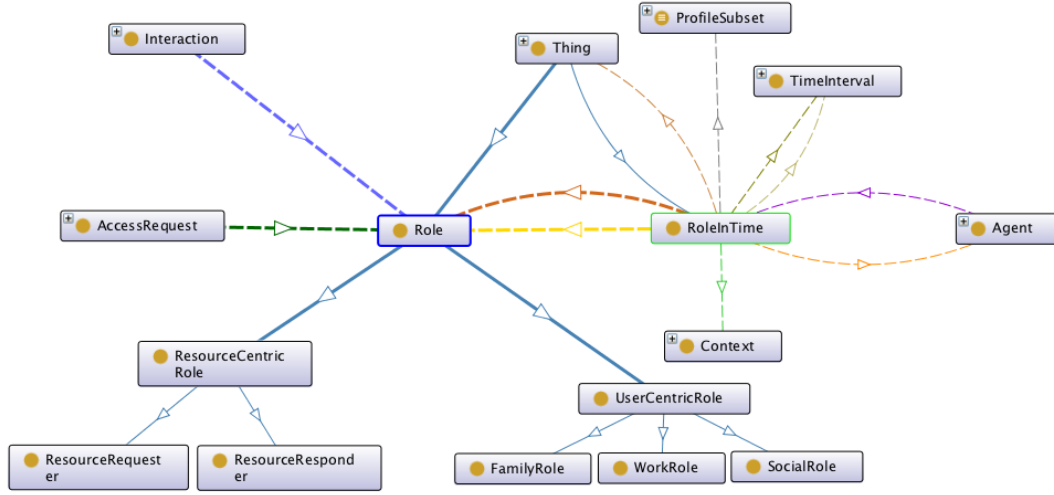


Figure 5.13: Graphical Representation of User Roles in SOCPRI

Helen Nissenbaum. According to her, preserving contextual integrity revolves around person's ability to keep audience separate and to compartmentalize his/her social life. With regards to the presentation of context specific roles, our ontology models multiple partial identities (profiles) for different contexts. The class **ProfileSubset** represents a variable subset of the user profiles depending on the context and role. The classes **DefaultProfile** and **ContextualProfile** are extended from this class. The contextual profiles of the user are further divided into different categories and access to these profiles is managed on the basis of a user's role at a particular time. The class **ContextualProfile** is linked to **RoleInTime** class indirectly through **ProfileSubset** class with object property **relatesToProfile**. The modeling of **RoleInTime** class and its relations with other entities is an important element of whole design of a temporal aspect of the user roles. The Manchester syntactical description of these classes and object properties is given below:

Class : RoleInTime

EquivalentTo :

(relatesToEntity some owl:Thing)

```

    and (withRole some Role)
    and (atTime some TimeInterval)
    and (atTime only
        (TimeInterval
            and (hasIntervalDate only xsd:dateTime)))
    and ( inverse (holdsRoleInTime) exactly 1 Agent)

```

Class: ProfileSubset

```

    EquivalentTo:
        isSpecificTo exactly 1 (Role
            and Context)

```

```

    SubClassOf:
        DigitalResource

```

Class: ContextualProfile

```

    SubClassOf:
        ProfileSubset ,
        hasCondition some ContextualNorm

```

```

    DisjointWith:
        DefaultProfile

```

Class: FamilyProfile

```

    EquivalentTo:
        ContextualProfile
        and (withRole only FamilyRole)
        and (hasContext only FamilyContext)

```

Class: SocialProfile

```

    EquivalentTo:
        ContextualProfile
        and (withRole only SocialRole)

```

and (hasContext only SocialContext)

Class: WorkProfile

EquivalentTo:

ContextualProfile

and (withRole only WorkRole)

and (hasContext only WorkContext)

ObjectProperty: relatesToProfile

SubPropertyOf:

relatesToEntity

Domain:

RoleInTime

Range:

ProfileSubset

ObjectProperty: relatesToContext

SubPropertyOf:

relatesToEntity

Domain:

RoleInTime

Range:

Context

ObjectProperty: isSpecificTo

Characteristics:

Functional

Domain:

ProfileSubset

Range:

Role

and Context

The figure 5.14 highlights key conceptual elements linked to the class. The class **RoleInTime** is connected to **TimeInterval** class through object property **atTime** that keeps track of certain temporal interval with help of two data properties. The user is represented by a sub-class of **foaf:Agent** class that is connected to **RoleInTime** class through object property **holdsRoleInTime**. Each user participates in different relationships during his / her life cycle. These relationships are formed and changed with time. To model these aspects, our ontology relates **User** class with **Relationship** class via object property **hasRelationship** and **Relationship** class with **TimeInterval** class through object property **atTime**. There is also indirect link between **Relationship** and **RoleInTime** classes through **relatesToEntity** object property. Everyday, a user performs multiple roles in different social contexts and present himself in accordance with audience, to model this contextual representation of the user, the SOCPRI ontology uses classes such as **User**, **Context**, **Role**, **RoleInTime**, **Interaction**, and **ProfileSubset**. These classes related to each other with object properties such as **hasContext**, **withRole**, **holdsRoleInTime**, **relatesToContext**, **relatesToProfile**, **isSpecificTo** and **associatedRole**. The detail graphical representation of the model is given in figure 5.14. The manchester OWL description of some important classes and object properties is given below:

Class : ResourceCentricRole

EquivalentTo :

ResourceRequester or ResourceResponder

SubClassOf :

Role

DisjointWith :

UserCentricRole

Class: UserCentricRole

EquivalentTo:

WorkRole or SocialRole or FamilyRole

SubClassOf:

Role

DisjointWith:

ResourceCentricRole

Class: FamilyRole

EquivalentTo:

{brother , daughter , father , husband ,
mother , sister , son , wife}

SubClassOf:

UserCentricRole

DisjointWith:

WorkRole, SocialRole

Class: SocialRole

EquivalentTo:

{boyFriend , casualFriend , collegeFriend , girlFriend ,
partyFriend , schoolFriend , stranger , uniFriend}

SubClassOf:

UserCentricRole

DisjointWith:

WorkRole, FamilyRole

Class: WorkRole

EquivalentTo:

{director , phdStudent , postDoc , professor , projectCoordinator ,
rector , researchAssociate , seniorResearcher , supervisor}

SubClassOf:

UserCentricRole

DisjointWith:

SocialRole , FamilyRole

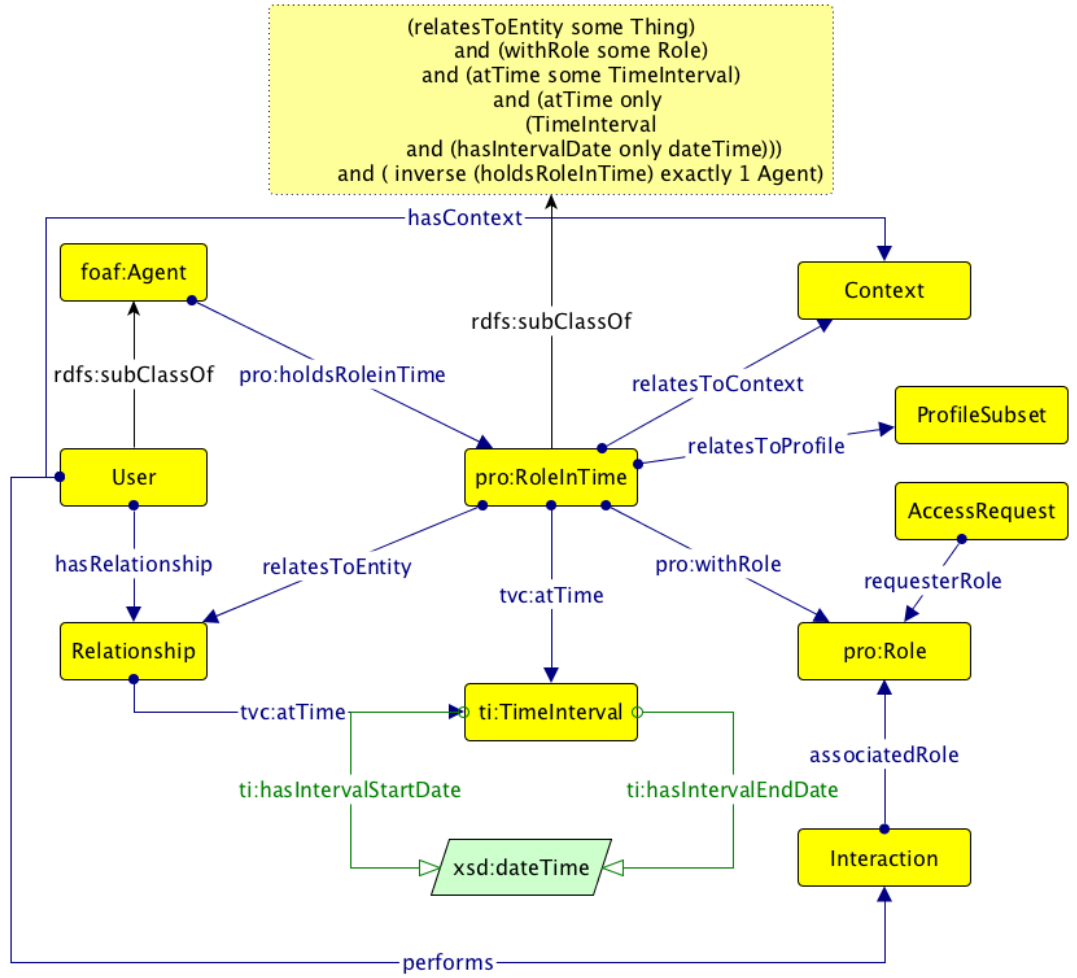


Figure 5.14: Contextual Representation of User Role in SOCPRI

The SOCPRI ontology models three types of the user relationships which includes user-to-user, user-to-resource, and resource-to-resource relationships. These relationships are represented by **U2URelationship**, **U2RRelationship**, and **R2RRelationship**

classes. These classes are a generic representation of relations in the social web environment and are further divided into more specific relationships. The **U2URelationship** class has three sub-classes which includes **Friendship**, **Kinship** and **Employment**. The **Friendship** is a generalized class to represent social relations such as friends, close friends, acquaintances. The **Kinship** class represent the most of the common family relationships such as a child, parent, siblings, spouse, grandparent, etc. The **Employment** class deals with the most common kind of work relations which includes employee, employer, colleague, and collaborator. The classes representing user to resource relationships are **Contributor**, **Stakeholder**, **Owner**, and **Disseminator**. The syntactical details of these classes are given in section 5.5.1. These relationships also incorporate temporal and strength aspects. The class **TimeInterval** is connected to **Relationship** class with object property **atTime**. The **Tie Strength** class is linked to **Relationship** class with object property **hasTieStrength**. The detailed description of the user social interactions and tie strength is given in the following section. Figure 5.15 shows the graphical representation of the user relationships in the SOCPRI ontology.

5.5.3 Social Interaction-based Tie Strength Modeling

Social interactions are an essential part of our social web usage. The interaction pattern of social web users can be utilized to examine relationship strength between them. The users have a finite amount of time to use in forming and maintaining online relationships. They invest this time towards a relationship that they consider important. Therefore, The nature and frequency of social interactions between users depend on the strength of their relationship. The SOCPRI ontology models diverse interaction set and predictive indicators for relationship strength between users. The core conceptual element is **Interaction** class which is connected to **User**, **Role** and **Digital-Resource** classes directly. The temporal aspects of social interactions are model with

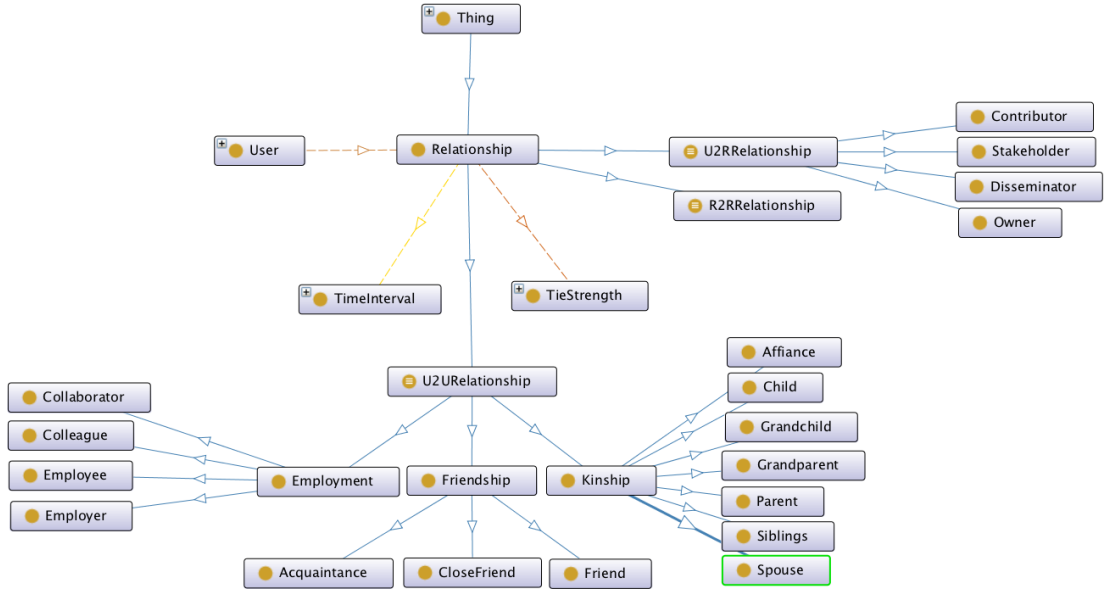


Figure 5.15: Representation of User Relationships in SOCPRI

help of **RoleInTime** and **TimeInterval** classes. The object property **atTime** relates **TimeInterval** class with **Interaction** class. The **RoleInTime** class is also connected with **Role** and **TimeInterval** classes using the same property. This modeling facilitates temporal role based user interactions in the social web environment. An individual user performs various interactions and holds multiple resources. The SOCPRI ontology models relation between digital resources and social interactions of users. The object property **associatedResource** links a digital resource with a particular social interaction of the user performed at certain time interval.

Current social web environments offer diverse set of interactions to their users. Our ontology also models various types of social interactions. The **InteractionType** class is upper class for all types of interactions represented by the SOCPRI ontology. The derived classes are **Messaging**, **Posting**, **Tagging**, **Sharing**, **Liking**, and **Commenting**. The classes **MessageSending** and **MessagingReply** takes into consideration the interaction initiation factor. The class **Posting** represents a generic type of a social interaction. It is further specialized with five child classes which in-



The relationship strength and its dimensions are represented by **TieStrength** and **TieStrengthDimension** classes. The object property **hasTieStrength** connects **Relationship** and **TieStrength** classes. The relation between **TieStrength** and **TieStrengthDimension** is established using **hasTieStrengthDimension** object property. The class **TieStrengthDimension** represents generic features of relationship strength. It is further categorized into seven classes which represents different aspects of user relationships. The extended classes includes **Closeness**, **RelationshipDuration**, **InteractionFrequency**, **SocialDistance**, **EmotionalSupport**, **ReciprocalService** and **StructuralDimension**. Each tie strength dimension has predictive indicators to reveal the nature of relationship. The link between **TieStrengthDimension** and **PredictiveIndicator** classes is established using **hasPredictiveIndicator**

object property. There are two types of predictive indicators which are represented by **HomophilyIndicators** and **InteractionIndicators** classes. These classes are derived from **PredictiveIndicatorType** class that is connected to **PredictiveIndicator** class through **hasPredictiveIndicatorType** object property. The figure 5.16 shows graphical representation of interaction and tie strength modeling in the SOCPRI ontology. The manchester Owl syntax of some important classes and object properties is given below:

```

Class: InteractionType
    SubClassOf:
        hasWeight some xsd:float
Class: Commenting
    EquivalentTo:
        InteractionType
        and (annotatesWith some Posting)
        and (hasWeight value 0.8f)
Class: Liking
    EquivalentTo:
        InteractionType
        and (annotatesWith some Posting)
        and (hasWeight value 0.4f)
Class: Messaging
    EquivalentTo:
        InteractionType
        and (annotatesWith some Photo)
        and (annotatesWith some Video)
        and (annotatesWith some Text)
        and (hasWeight value 0.9f)

```

Class: Posting

EquivalentTo:

(InteractionType
and (hasWeight value 0.6 f))
and (annotatesWith some Photo)
and (annotatesWith some Video)
and (annotatesWith some Text)

Class: Sharing

EquivalentTo:

(InteractionType
and (hasWeight value 0.5 f))
and (annotatesWith some Photo)
and (annotatesWith some Video)
and (annotatesWith some Text)

Class: Tagging

EquivalentTo:

(InteractionType
and (hasWeight value 0.7 f))
and (annotatesWith some Photo)

Class: TieStrength

EquivalentTo:

MediumTie or StrongTie or WeakTie

SubClassOf:

hasTieStrengthDimension some TieStrengthDimension

Class: TieStrengthDimension

SubClassOf:

hasPredictiveIndicator some PredictiveIndicator

Class: PredictiveIndicator

SubClassOf:

hasPredictiveIndicatorType some PredictiveIndicatorType

ObjectProperty: hasTieStrength

Characteristics:

Functional,

Asymmetric

Domain:

Relationship

Range:

TieStrength

ObjectProperty: hasTieStrengthDimension

Domain:

TieStrength

Range:

TieStrengthDimension

ObjectProperty: hasPredictiveIndicator

Domain:

TieStrengthDimension

Range:

PredictiveIndicator

ObjectProperty: hasPredictiveIndicatorType

Characteristics:

Functional

Domain:

PredictiveIndicator

Range:

```

    PredictiveIndicatorType
DataProperty: hasWeight
    Domain:
        InteractionType
    Range:
        xsd:float

```

5.5.4 User Resource Modeling

Social web changed the status of an end-user. An individual end-user becomes content producer instead of being content consumer. Each individual user uploads a huge amount of personal content in their social networking sites. It is challenging task to model all the aspects related to the user content. The SOCPRI ontology models typical digital content provided by social networking sites such as wall, profile, photos, videos, etc. The figure 5.17 shows graphical representation of user resource modeling in the SOCPRI ontology. Our ontology models resources as a class, beginning with a generic **DigitalResource** class which represents any object with digital, usually presentable content. It extends from **foaf:Document** class. The **DigitalResource** class is specialized by subclasses such as **ContextFreeResource**, **ContextSensitiveResource** and **ProfileSubset**. The context sensitive resources of the user are represented by classes such as **Wall**, **Profile**, **Friendlist**, **Photo**, etc. The classes extended from **ProfileSubset** are **DefaultProfile** and **ContextualProfile**. The **ContextualProfile** class is further specialized by social, work and family profiles. The **Profile** class is connected with **ProfileSubset**, **DefaultProfile** and **ContextualProfile** classes through object properties **hasProfileSubset**, **hasDefaultProfile**, and **hasContextualProfile** respectively.

The contextual role of a user at particular time is key factor to decide the profile subset for the user. The relation between **RoleInTime** and **ProfileSubset** classes

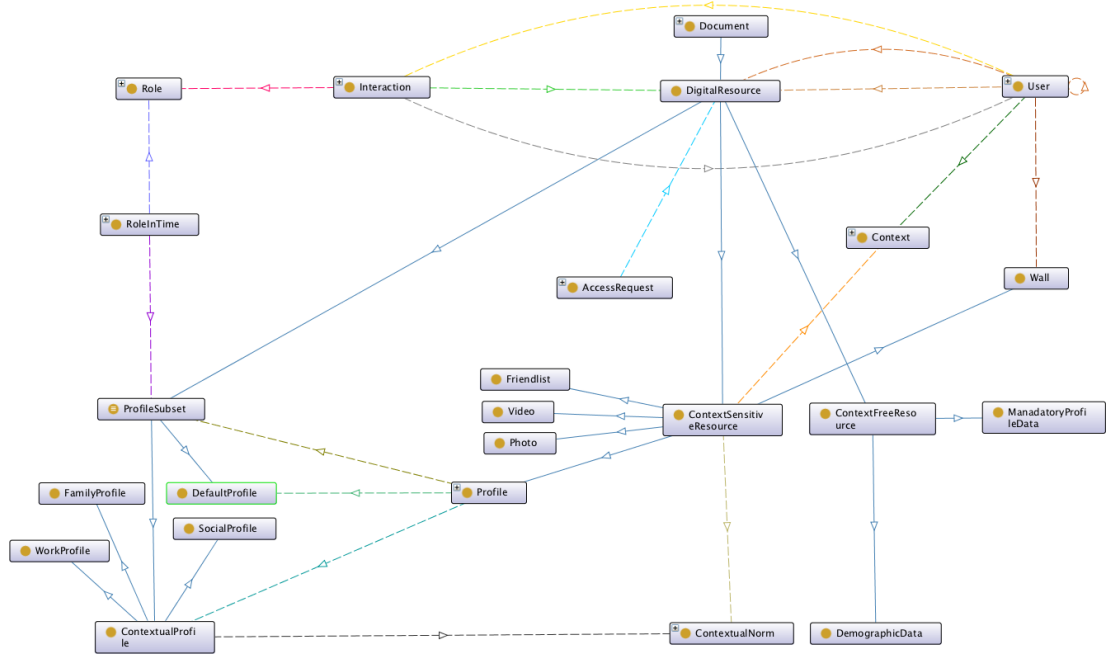


Figure 5.17: Representation of User Digital Resource in SOCPRI

is established using **relatesToProfile** object property. The major difference between default and contextual profiles of the user is that contextual profile is governed by contextual norms. The **ContextualNorm** class linked to contextual profiles as well as context sensitive resources. The relation between **ContextualProfile** and **ContextualNorm** is established using **hasCondition** object property. The classes **ContextSensitiveResource** and **ContextualNorm** are connected through **hasInformationFlowNorm** object property. The context sensitive resources of the user belongs to certain context and this is modeled using two object properties **belongsToContext** and **hasContext**. The classes **ContextSensitiveResource** and **Context** are connected through **belongsToContext** object property. The relation between **Context** and **User** class is established through **hasContext** object property.

All the users own a wall in social networking sites which is kind of context sen-

sitive resource. The object property **ownsWall** connects **User** and **Wall** classes. A user's wall is governed by contextual norms because of its inheritance relationship with **ContextSensitiveResource** class. The user hold some digital resources and make access request for other resources. The relationship between **DigitalResource** and **AccessRequest** classes is established via **requestResource** object property. The object properties **relates** and **holdResource** connects **User** and **DigitalResource** classes. Given below is manchester Owl syntax for important classes and object properties related to resource modeling of social web user.

Class: FamilyProfile

EquivalentTo:

ContextualProfile

and (withRole only FamilyRole)

and (hasContext only FamilyContext)

Class: SocialProfile

EquivalentTo:

ContextualProfile

and (withRole only SocialRole)

and (hasContext only SocialContext)

Class: WorkProfile

EquivalentTo:

ContextualProfile

and (withRole only WorkRole)

and (hasContext only WorkContext)

Class: Photo

SubClassOf:

ContextSensitiveResource ,

hasContent some xsd:hexBinary

Class: Video

SubClassOf:

ContextSensitiveResource ,

hasContent some xsd:hexBinary

Class: Text

SubClassOf:

ContextFreeResource ,

hasContent some xsd:string

ObjectProperty: requestedResource

Domain:

AccessRequest

Range:

DigitalResource

ObjectProperty: hasProfileSubset

Domain:

Profile

Range:

ProfileSubset

ObjectProperty: hasContextualProfile

Domain:

Profile

Range:

ContextualProfile

SubPropertyOf:

hasProfileSubset

ObjectProperty: hasDefaultProfile

Domain:

```

    Profile
Range:
    DefaultProfile
SubPropertyOf:
    hasProfileSubset
ObjectProperty: hasCondition
Domain:
    ContextualProfile
Range:
    ContextualNorm
ObjectProperty: hasInformationFlowNorm
Domain:
    ContextSensitiveResource
Range:
    ContextualNorm

```

5.5.5 Contextual Privacy Modeling

The SOCPRI ontology models contextual privacy for social web users. The digital resources are always associated with the context in which they are revealed. The disclosure of resources depends on contextual norms. These norms are always internal to the context and divided into norms of appropriateness and norms of distribution. Norms of appropriateness determine whether a given type of personal information is either appropriate or inappropriate to divulge within a particular context. Norms of distribution restrict the flow of information within and across contexts. There are four parameters to determine these norms: actor, social context, attributes, and information flow constraints.

The **User** class models the actor which is connected with **DigitalResource** class

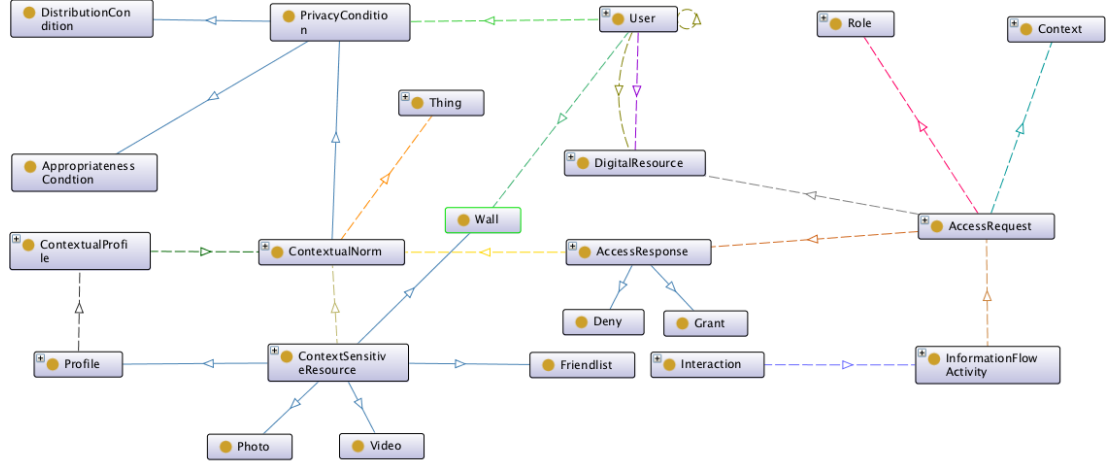


Figure 5.18: Representation of User Contextual Privacy in SOCPRI

through **holdsResource** object property. The social context of the requested resources is modeled by **context**, **AccessRequest** and **DigitalResource** classes. The **requestedResource** object property links **DigitalResource** and **AccessRequest** classes. The relation between **Context** and **AccessRequest** classes is established using **requestedInContext** object property. Apart from the context in which resource is requested, our ontology takes into consideration the role of requesting actor for deciding the access response. The relation between **Role** and **AccessRequest** is managed by **requesterRole** object property. The access to context sensitive digital resource generates a response that is either grant or deny. The access response always depends on contextual norms. The class **AccessResponse** has two subclasses by the name of **Grant** and **Deny**. The relation between these classes and **ContextualNorm** class is established using **dependsOn** object property. The class **ContextSensitiveResource** is connected to **ContextualNorm** class through object property **hasInformationFlowNorms**. The norms of appropriateness and distribution are modelled using **AppropriatenessCondition** and **DistributionCondition** classes. The **PrivacyCondition** is super class of these classes which connected to **User** class with object property **hasCurrentContextualNorm** and extends from **ContextualNorm**

class.

The user profile is one of the context sensitive digital resource and access management of contextual profiles depends on contextual norms. The relation between these two classes is established using **hasCondition** object property. Social interactions of the user generate information flow activities and require digital resources of other users. The information flow activities are generated due to social interaction of the user. The relations between **Interaction**, **InformationFlowActivity** and **AccessRequest** are modeled through **generates** and **hasAccessRequest** object properties. The figure 5.18 shows the graphical representation of contextual privacy modeling in the SOCPRI ontology. The Manchester Owl syntax of some important classes and object properties is given below:

Class: InformationFlowActivity

SubClassOf:

hasAccessRequest some AccessRequest

Class: AccessRequest

SubClassOf:

requestedInContext some Context ,
requestedResource some DigitalResource ,
requesterRole some Role

DisjointWith:

AccessResponse

Class: AccessResponse

EquivalentTo:

Deny or Grant

SubClassOf:

dependsOn some ContextualNorm

DisjointWith:

AccessRequest

Class: Deny

EquivalentTo:

AccessResponse

and (hasAccessDecision value "deny")

DisjointWith:

Grant

Class: Grant

EquivalentTo:

AccessResponse

and (hasAccessDecision value "grant")

DisjointWith:

Deny

ObjectProperty: hasContextualValue

Domain:

ContextualNorm

Range:

Context

Characteristics:

Functional

ObjectProperty: dependsOn

Domain:

AccessResponse

Range:

ContextualNorm

ObjectProperty: generateResponse

Domain:

AccessRequest

Range:

AccessResponse

ObjectProperty: hasAccessRequest

Domain:

InformationFlowActivity

Range:

AccessRequest

ObjectProperty: hasInformationFlowNorm

Domain:

ContextSensitiveResource

Range:

ContextualNorm

ObjectProperty: requestedInContext

Domain:

AccessRequest

Range:

Context

ObjectProperty: requesterRole

Domain:

AccessRequest

Range:

Role

5.6 Defining the Classes and Properties of SOCPRI

This section presents a brief description of the all the classes and object properties of the SOCPRI ontology. The purpose of defining classes and their relations with other entities is to understand conceptual modeling of the SOCPRI ontology. It also validates functional requirements that are expressed in form of competency questions. As a matter fact, this terminology is extracted from competency questions. The definition of terms also makes the purpose of their creation explicit and determines the scope of the ontology. In following section, we define classes, describe their inheritance relationship and elaborate their connections with other classes using object properties. The detailed relationship modeling of some core classes in SOCPRI ontology is shown in figures 5.19, 5.20, and 5.21. The figure 5.22 shows overall implementation of SOCPRI ontology using protege ontology editor.

Agent: An abstract class defining any kind of agents, such as a person, a group, an organization or a software agent. The SOCPRI reuse this class of FOAF ontology. It is super class of *Group* and *User* classes. It is related to *Profile*, *Grroup* and *RoleInTime* classes through object properties *hasProfile*, *member* and *holdsRoleInTime* respectively.

Group: Our ontology also reuses this class of FOAF ontology which represents a collection of individual agents. It is child class of *Agent*. It is also related to *Agent* class through an object property *member*

Organization: SOCPRI ontology also reuse this class of FOAF ontology which represents a kind of Agent corresponding to social institutions such as companies, societies etc. It is also derived from *Agent* class. The two object properties establish link between *User* and *Organization* classes. These object properties are *isEmployedBy* and *isEmployerOf*.

User: It is core conceptual entity of the SOCPRI ontology which represents an online social network user. It extends from *foaf:Agent* class. It is related to *Relationship*, *Context*, *Interaction*, *DigitalResource*, *Wall* and *PrivacyCondition* classes through various object properties. Figure 5.19(a) shows implementation of User class in Protege.

Role: This class represent user roles in OSN with other users and objects. The SOCPRI reuse this class of PRO ontology. It is super class of *UserCentricRole*, *ResourceCentricRole* classes. It is disjoint with *Relationship* class. It is related to *RoleInTime*, *Interaction* and *AccessRequest* classes through *withRole*, *associatedRole* and *requesterRole* object properties respectively.

RoleInTime: This class represents a particular situation that describe a role an agent may have, that can be restricted to a particular time interval. Our ontology reuses this class of PRO ontology. It is related to *Agent*, *TimeInterval*, *Role*, *Context* and *ProfileSubset* classes through various object properties. Figure 5.19(b) shows implementation of RoleInTime class in Protege.

UserCentricRole: This class represents user roles in OSN with other users. It is child class of *Role* and further categorized into three subclass which instantiate specific roles users play in family, work and social environments. It is disjoint with *ResourceCentricRole* class.

ResourceCentricRole: This class represents user roles in OSN with digital resources. It is child class of *Role* and disjoint with *UserCentricRole* class. It has two subclasses to model requester and responder status of the user.

TieStrength: This class models relationship strength between OSN users. It has two associated object properties which relate this class with *Relationship* and *TieStrengthDimension* classes.

StrongTie: This is subclass of *TieStrength*. It has one quantifier restriction that describes relation between frequency of interactions and strong tie. If the frequency of interactions is greater than 500 then the tie is considered to be strong.

WeakTie: This is subclass of *TieStrength*. It has one quantifier restriction that describes relation between frequency of interactions and strong tie. If the frequency of interaction is less than 100 then the tie is assumed to be weak.

MediumTie: It is also derived from *TieStrength*. It carries one quantifier restriction about the frequency of interactions. Medium tie fall between strong and weak tie in terms of frequency of interactions. If the interaction frequency is between 100 and 500 then the tie is assumed to be medium tie.

TieStrengthDimension: This class represents generic concept of tie strength dimension. It is further specialized by seven subclasses which represent various dimensions of relationship strength. This class is associated with *PredicativeIndicator* and *TieStrength* classes through two object properties.

Closeness: It is child class of *TieStrengthDimension* class and represents intimacy between OSN users.

EmotionalSupport: It is derived from *TieStrengthDimension* and represents emotional attachment between OSN users.

InteractionFrequency: This class represents intensity of interactions between OSN users. It is also derived from *TieStrengthDimension* class.

ReciprocalService: This class represents reciprocity services between OSN users. It is subclass of *TieStrengthDimension* class.

RelationshipDuration: This class represents duration of tie between OSN users. It is also subclass of *TieStrengthDimension* class.

SocialDistance: This class represents difference in socioeconomic status between OSN users. It is extended from *TieStrengthDimension* class.

StructuralDimension: This class represents social homogeneity between OSN users. It is also extended from *TieStrengthDimension* class.

TimeInterval: This class defines a particular period of time with starting and ending points. It is related to *Relationship*, *Interaction* and *RoleInTime* classes through *atTime* object property.

PredictiveIndicator: This class represents predictive variables for relationship strength. It relates to *TieStrengthDimension* and *PredictiveIndicatorType* classes through *hasPredictiveIndicator* and *hasPredictiveIndicatorType* object properties.

PredictiveIndicatorType: This class represents various types of predictive variables. It has two subclasses which modeled more specific types of predictive indicators related to user interaction pattern and profile similarity. It relates to *PredictiveIndicator* class using object property *hasPredictiveIndicatorType*.

HomophilyIndicator: This class represents profile similarity variables between OSN users. It is extended from *PredictiveIndicatorType* class.

InteractionIndicator: It is subclass of *PredictiveIndicatorType* and represents interaction frequency indicators between OSN users.

Context: This class represent an abstract context of an OSN user. It has three subclasses which represent more concrete contexts of the user. The class also relates to *User*, *RoleInTime*, *ContextSensitiveResource*, *ContextualNorm* and *AccessRequest* classes through different object properties.

FamilyContext: This class represents family related context of an OSN user. It is child class of *Context*. It is disjoint with *SocialContext* and *WorkContext* classes.

SocialContext: It is derived from *Context* class and represents social context of an OSN user including friends, classmates and acquaintances. The classes *FamilyContext* and *WorkContext* are disjoint with it.

WorkContext: It is also derived from *Context* class and represents work related context of an OSN user including all professional contexts. It is also disjoint with *SocialContext* and *FamilyContext* classes.

FamilyRole: It is child class of *UserCentricRole* and represents family roles of OSN users. The *WorkRole*, and *SocialRole* are disjoint classes with it.

WorkRole: It represent work roles of OSN users. It extends from *UserCentricRole* class and disjoint with *FamilyRole* and *SocialRole* classes.

SocialRole: It represents social roles of OSN users. It also extends from *UserCentricRole* class and disjoint with *FamilyRole* and *WorkRole* classes.

ResourceRequester: It is child class of *ResourceCentricRole* and represents user role as requester for digital resources. It is disjoint with *ResourceResponder* class.

ResourceResponder: It is also child of *ResourceCentricRole* and represents user role as responder to request for a digital resource owned by the user.

Relationship: This class represents user relationships with other users and digital resources. It is disjoint with *Role* class. It has three subclasses which includes *U2URelationship*, *U2RRelationship* and *R2RRelationship*. It is related to *User*, *TimeInterval* and *TieStrength* classes using *hasRelationship*, *atTime* and *hasTieStrength* object properties.

U2URelationship: This class represents generic relationships between two users. These relationships are further specialized with classes *Employment*, *Friendship*

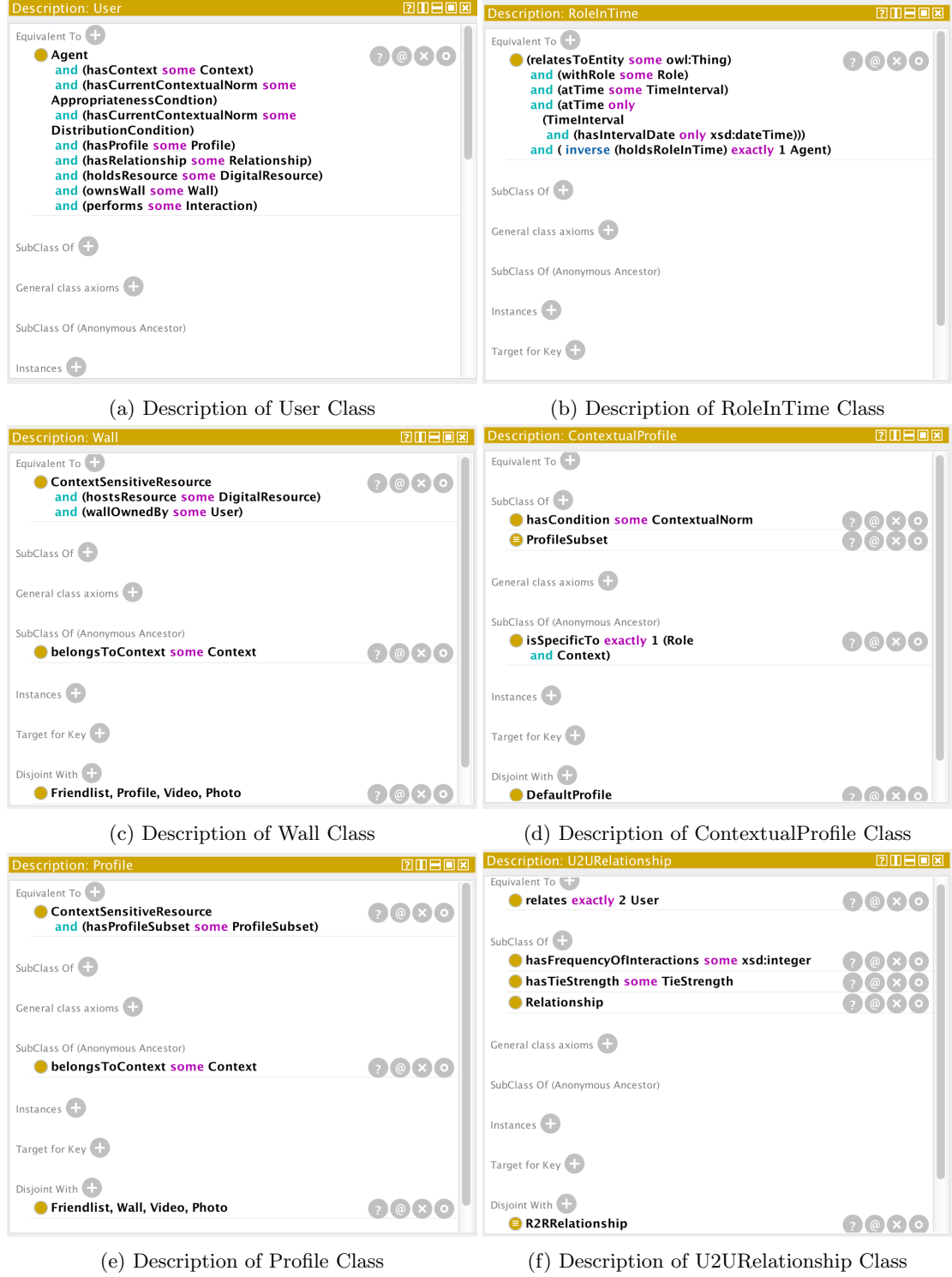


Figure 5.19: Description of main SOCPRI classes

and *Kinship*. Figure 5.19(f) shows implementation of *U2URelationship* class in Protege.

U2RRelationship: This class models relations between users and digital resources. It has four subclasses which includes *Contributor*, *Disseminator*, *Stakeholder* and *Owner*.

R2RRelationship: This class represents relationships between two digital resources. It also extends from *Relationship* class.

Kinship: This class represents user relationships in family context. It extends from *U2URelationship* class. It is further specialized with seven subclasses. Figure 5.20(c) shows implementation of *Kinship* class in Protege.

Employment: This class represents user relationships in work context. It is child class of *U2URelationship* and used as parent class of four kind of relationships that exist in work context. Figure 5.20(b) shows implementation of *Employment* class in Protege.

Friendship: This class represents user relationships in social context and extends three classes which also represent more specific social relations among OSN users. Figure 5.20(a) shows implementation of *Friendship* class in Protege.

Contributor: This class represents U2R relationship where OSN user is contributor to the resource. It is child class of *U2RRelationship*.

Disseminator: This class represents U2R relationship where OSN user is sharer of the resource. It is also child class of *U2RRelationship*.

Stakeholder: This class extends from *U2RRelationship* and represents relationship where OSN user is mutually owns the resource.

Owner: This class also extends *U2RRelationship* and represents relationship where OSN user is owner of the resource.

Affiance: It is subclass of *Kinship* class and represents a person to whom this person is betrothed / engaged.

Grandchild: This class represents U2U relationship where OSN user is grandchild to another user. It extends from *Kinship* class.

Parent: This class represents a person who has given birth to or nurtured and raised this person. It also extends from *Kinship* class.

Grandparent: It is child class of *Kinship* class and represents a person who has given birth to or nurtured and raised parent of this person.

Siblings: It represents user to user relation in which a person having one or both parents in common with this person. *Kinship* class is super class of it.

Spouse: It is also child class of *Kinship* class and represents one of the user to user relationship in which spouse is a person who is married to this person.

Employee: This class represents a person for whom this person's services have been engaged. It is derived from *Employment* class.

Employer: This class represent a person who engages the services of this person. It is also derived from *Employment* class.

Colleague: It is child class of *Employment* class and represents a person who is a member of the same profession as this person.

Collaborator: It is also subclass of *Employment* class and represents a person who works towards a common goal with this person.

Friend: This class represents a person who shares mutual friendship with this person.

It is kind of user to user relationship and derived from *Friendship* class.

CloseFriend: This class represents a person who shares a close mutual friendship with this person. It is also derived from *Friendship* class.

Acquaintance: This class represent a user to user relationship which demonstrate that a person having more than slight or superficial knowledge of this person but short of friendship. It is also subclass of *Friendship* class.

Interaction: This is generic class which represents user interactions in online social networks. It relates to more specific types of interactions using *hasInteractionType* object property. It is also related to *User*, *Role*, *TimeInterval*, *DigitalResource* and *InformationFlowActivity* classes through various object properties. Figure 5.21(e) shows implementation of Interaction class in Protege.

InteractionType: This class represents diverse set interactions offered by online social networks. It is related to *Interaction* class using *hasInteractionType* object property. It is also base class for the classes which represent more specific type of interactions.

Commenting: This class represents commenting interaction. It is child class of *InteractionType* and disjoint to *Liking*, *Messaging*, *Posting*, *Sharing* and *Tagging* classes.

Liking: This class extends from *InteractionType* and represents liking interaction. It is disjoint to *Commenting*, *Messaging*, *Posting*, *Sharing* and *Tagging* classes.

Messaging: This class represents messaging interaction. It is sub-class of *InteractionType* and disjoint to *Liking*, *Commenting*, *Posting*, *Sharing* and *Tagging* classes.

MessageReplying: This class is derived from *Messaging* and represents message replying interaction. It is disjoint with *MessagingSending* class.

MessageSending: This class represent message sending interaction. It is subclass of *Messaging* and disjoint to *MessagingReplying* class.

Posting: This class represents generic posting interaction. It is sub-class of *InteractionType* and disjoint to *Liking*, *Commenting*, *Messaging*, *Sharing* and *Tagging* classes. Figure 5.21(f) shows implementation of Posting class in Protege.

AlbumPosting: This class represents album posting interaction. It extends from *Posting* and disjoint with *LocationPosting*, *PhotoPosting*, *StatusPosting* and *VideoPosting* classes.

LocationUpdating: This class extends from *Posting* and represents location posting interaction. It is disjoint with *AlbumPosting*, *PhotoPosting*, *StatusPosting* and *VideoPosting* classes.

PhotoPosting: This class represents photo posting interaction. It is subclass of *Posting* and disjoint with *LocationPosting*, *Album*, *StatusPosting* and *VideoPosting* classes.

StatusUpdating: This class represents status posting interaction. It is also subclass of *Posting* and disjoint with *LocationPosting*, *AlbumPosting*, *PhotoPosting* and *VideoPosting* classes.

VideoPosting: This class extends from *Posting* and represents video posting interaction. It is disjoint with *AlbumPosting*, *PhotoPosting*, *StatusPosting* and *PhotoPosting* classes.

Sharing: This class represents generic sharing interaction. It is sub-class of *InteractionType* and disjoint to *Liking*, *Commenting*, *Messaging*, *Posting* and *Tagging*

classes.

Tagging: This class represents tagging interaction. It is also sub-class of *InteractionType* and disjoint to *Liking*, *Commenting*, *Messaging*, *Sharing* and *Posting* classes.

InformationFlowActivity: This class captures the notion of information flow for every context sensitive resource. It relates to *Interaction* class using *generates* object property. The relation between this class and *AccessRequest* class is established using *hasAccessRequest* object property.

Document: Our ontology also reuses this class of FOAF ontology which is an abstract class and represents all kind of digital resources in online social networks.

DigitalResource: This is child class of *Document* and represents digital content of an OSN user. It is three subclasses which represent context free and context sensitive digital resources. It also relates to *User*, *Interaction* and *AccessRequest* classes through *holdsResource*, *associatedResource* and *requestedResource* object properties respectively.

ContextFreeResource: This class represents digital content of an OSN user that is not tagged to a particular context. It is child class of *DigitalResource* and extends two more specific classes.

DemographicData: It represent user context free content related to demographic profile information. It extends from *ContextFreeResource* class.

MandatoryProfileData: This class extends from *ContextFreeResource* and represents profile data that is context free and mandatory for creation of OSN user's account.

Text: It is also subclass of *ContextFreeResource* and represents textual resources of the OSNs users. The object property *annotatesWith* establish link between **InteractionType** class and this class.

ContextSensitiveResource: This class represents digital content of an OSN user that is tagged to a particular context. It has five subclasses which represent more specific context sensitive resources. It relates to *Context* and *ContextualNorm* classes through *belongsToContext* and *hasInformationFlowNorm* object properties.

Friendlist: This class represents social graph of OSN user. *ContextSensitiveResource* is parent class for it. It is disjoint with *Photo*, *Profile*, *Video* and *Wall*.

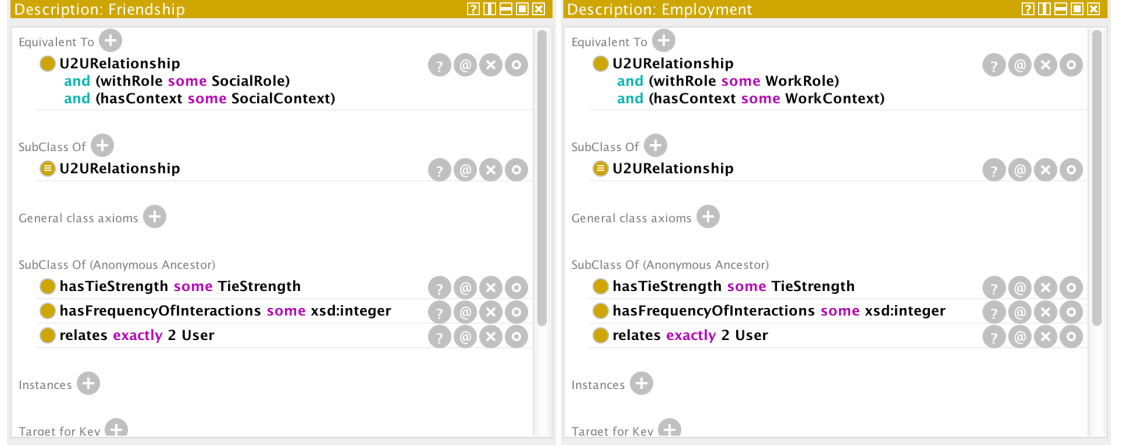
Photo: This class extends from *ContextSensitiveResource* and represents digital image. It is disjoint with *Friendlist*, *Profile*, *Video* and *Wall*.

Profile: This class represent profile page of OSN user. The base class for it is *ContextSensitiveResource*. It is disjoint with *Friendlist*, *Photo*, *Video* and *Wall*. Figure 5.19(e) shows implementation of Profile class in Protege.

Video: This class represents digital document containing video. It also extends from *ContextSensitiveResource* class and disjoint with *Friendlist*, *Photo*, *Profile* and *Wall*

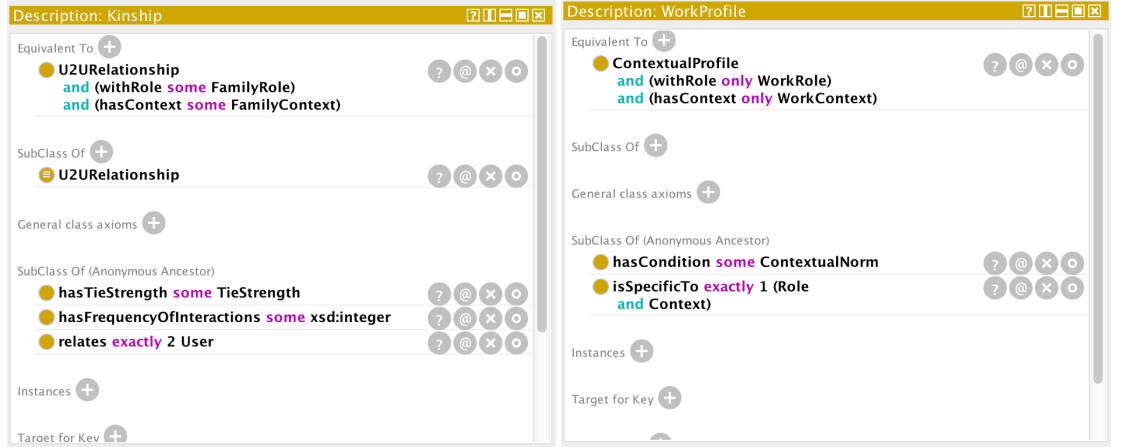
Wall: This class is child class of *ContextSensitiveResource* and represents newsfeed page of OSN user. It is disjoint with *Friendlist*, *Photo*, *Video* and *Profile*. Figure 5.19(c) shows implementation of Wall class in Protege.

ProfileSubset: This class represents various subset of OSN user profiles depending on the context. It has two subclasses *DefaultProfile* and *ContextualProfile*. It is also related to *RoleInTime* and *Profile* classes through *relatesToProfile* and *hasProfileSubset* object properties.



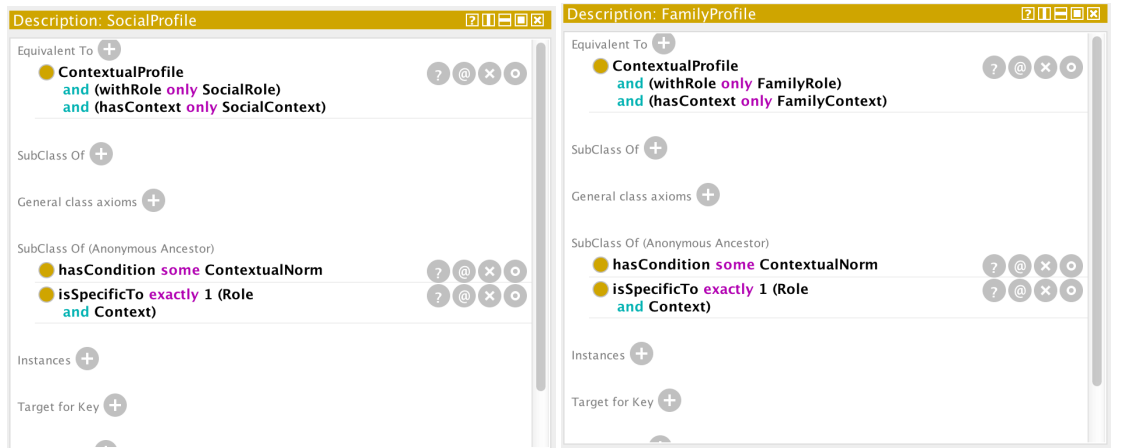
(a) Description of Friendship Class

(b) Description of Employment Class



(c) Description of Kinship Class

(d) Description of WorkProfile Class



(e) Description of SocialProfile Class

(f) Description of FamilyProfile Class

Figure 5.20: Description of main SOCPRI classes

DefaultProfile: This class represents default profile of an OSN user that is independent of any context. It is derived from *ProfileSubset* class and relates to *Profile* class using object property *hasDefaultProfile*.

ContextualProfile: This class represents various contextual profile of an OSN user. It has three subclasses *WorkProfile*, *SocialProfile* and *FamilyProfile*. It relates to *Profile* and *ContextualNorm* classes through object properties *hasContextualProfile* and *hasCondition* respectively. Figure 5.19(d) shows implementation of ContextualProfile class in Protege.

FamilyProfile: This class represents family profile subset of ONS user. It extends from *ContextualProfile* class and disjoint to *SocialProfile* and *WorkProfile* classes. Figure 5.20(f) shows implementation of FamilyProfile class in Protege.

SocialProfile: This class represents social profile subset of ONS user. It is subclass of *ContextualProfile* and disjoint to *SocialProfile* and *WorkProfile* classes. Figure 5.20(e) shows implementation of SocialProfile class in Protege.

WorkProfile: This class represents work profile subset of ONS user. It also extends from *ContextualProfile* class and disjoint to *SocialProfile* and *FamilyProfile* classes. Figure 5.20(d) shows implementation of WorkProfile class in Protege.

ContextualNorm: This class represents privacy condition under which a profile subset is valid to be displayed to the requester. It has a child class by the name of *PrivacyCondition*. It relates to *Context*, *ContextSensitiveResource*, *ContextualProfile* and *AccessRequest* classes through object properties *hasContextualValue*, *hasInformationFlowNorm*, *hasCondition* and *dependsOn* respectively. Every thing has contextual value that modeled using *Thing* class and *hasContextualValue* object property.

PrivacyCondition: This class models privacy condition for disclosure and distribu-

tion of the digital resources. It has two subclasses *AppropriatenessCondition* and *DistributionCondition*. The user is related to this class using object property *hasCurrentContextualNorm*.

AppropriatenessCondition: This class determine conditions for appropriateness flow of information within the context. It is subclass of *PrivacyCondition* and disjoint to *DistributionCondition* class.

DistributionCondition: This class determine conditions for appropriate distribution of information across the contexts. It also extends from *PrivacyCondition* and disjoint to *AppropriatenessCondition* class.

AccessRequest: This class models all information about a request to access a specific digital resource of OSN user. The object properties *requestedResource*, *requestedInContext* and *requesterRole* connect this class with *DigitalResource*, *Context* and *Role* classes. Each information flow activity is generated due to an access request. The classes *AccessRequest* and *InformationFlowActivity* are connected through *hasAccessRequest* object property. Figure 5.21(a) shows implementation of AccessRequest class in Protege.

AccessResponse: This class represents properties and set of obligations defining the access decision. It has two subclasses *Deny* and *Grant*. It relates to *AccessRequest* and *ContextualNorm* classes through object properties *generateResponse* and *dependsOn* respectively. Figure 5.21(b) shows implementation of AccessResponse class in Protege.

Deny: This class represents denial access response to the access request. It is child class of *AccessResponse* and disjoint to *Grant* class. Figure 5.21(c) shows implementation of Deny class in Protege.

Grant: This class represents positive access response to the access request. It also

extends from *AccessResponse* and disjoint to *Deny* class. Figure 5.21(d) shows implementation of Grant class in Protege.

Classes describe concepts in the domain of discourse and it is the focus of the most of the ontologies. But classes alone will not provide enough information to answer the competency questions. Once classes are defined then object properties are used to establish a relationship between these classes. The detailed description of object properties of the SOCPRI ontology is given below. This description includes various aspects of object properties such as their characteristics, their domains, and ranges, etc.

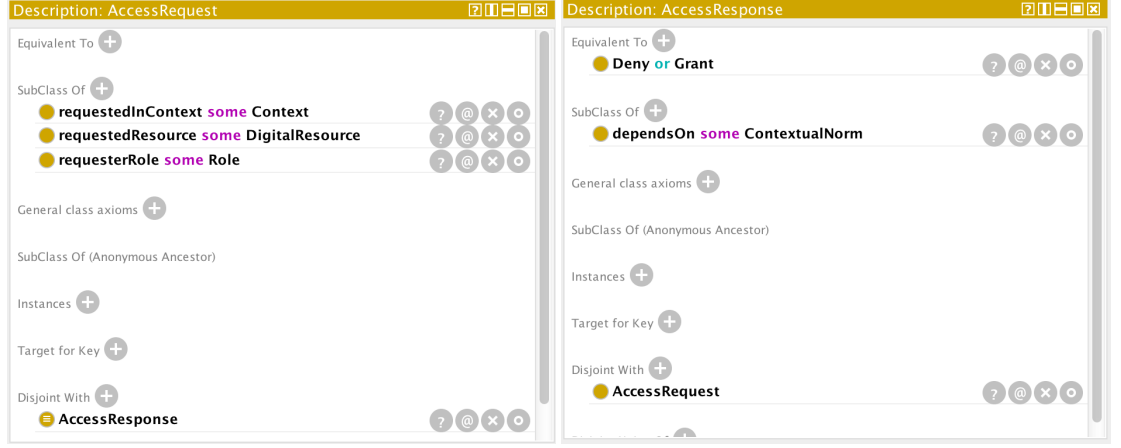
associatedResource: This object property links various digital resources with requesting user interaction. The domain of this property is *Interaction* class and range is *DigitalResource* class.

associatedRole: This object property links OSN user's performed interaction with their associated role. The *Interaction* class is the domain of the property. The range of the property is *Role* class.

atTime: This object property represents a time interval during which a role is held or a contribution is made by an agent. The domain classes for this property are *Relationship*, *Interaction* and *RoleInTime*. The range of the property is *TimeInterval* class.

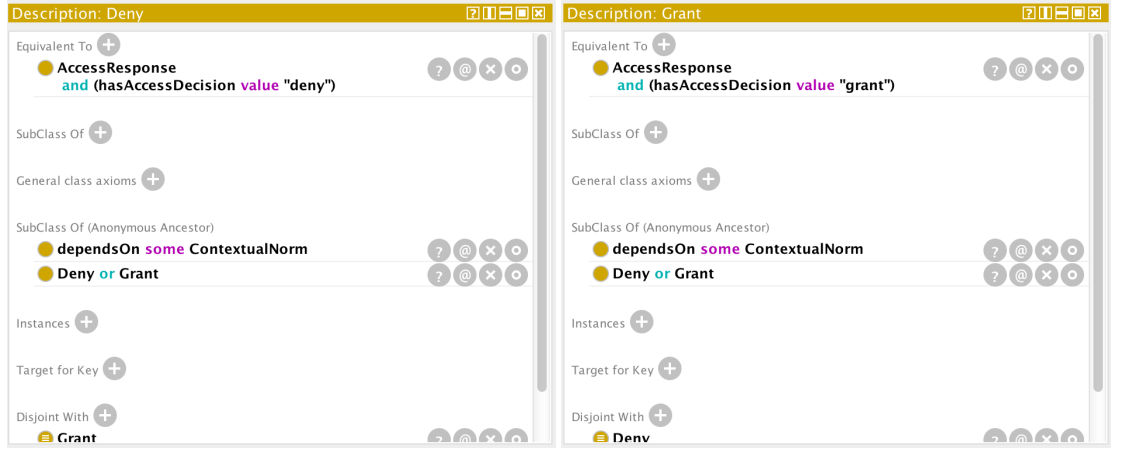
belongsToContext: This object property links context sensitive digital resources with their associated contexts. The domain class for the property is *ContextSensitiveResource* and range is *Context*.

annotatesWith: This object property establishes a link between digital resources and user interactions. The domain class for this property is *InteractionType* and range class is *DigitalResource*.



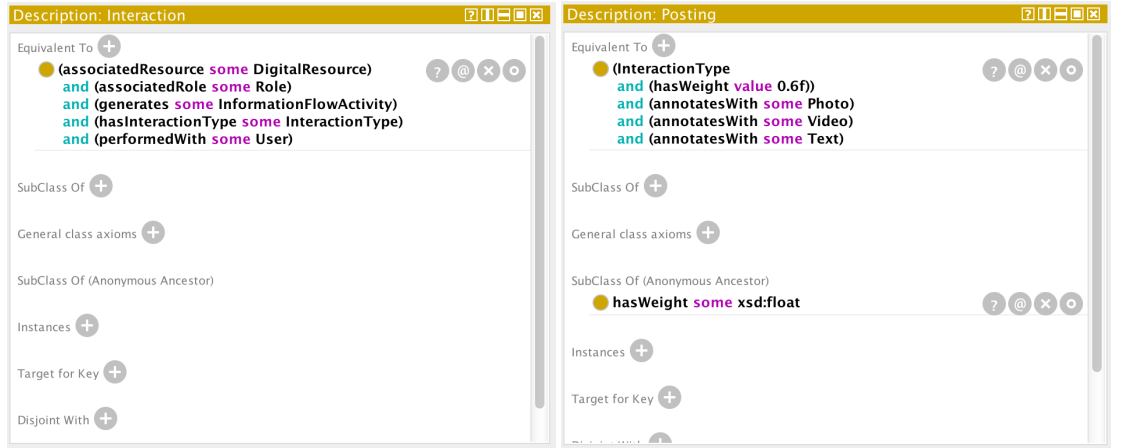
(a) Description of AccessRequest Class

(b) Description of AccessResponse Class



(c) Description of Deny Class

(d) Description of Grant Class



(e) Description of Interaction Class

(f) Description of Posting Class

Figure 5.21: Description of main SOCPRI classes

dependsOn: This object property links access response with their associated contextual norms. The domain and range classes for this property are *AccessResponse* and *ContextualNorm* respectively.

generateResponse: This object property links each access request with their appropriate access response. The domain class for the property is *AccessRequest* and range is *AccessResponse*.

generates: This object property links OSN user interactions with their generated information flow activity. *Interaction* class is domain of the property, whereas, range is *InformationFlowActivity*.

hasAccessRequest: This object property links information flow activity with the access request. The domain and range classes for this property are *InformationFlowActivity* and *AccessRequest* respectively.

hasCondition: This object property links contextual norms and with contextual user profiles. The *ContextualProfile* class is domain of the property and *ContextualNorm* is the range.

hasContext: This object property relates diverse social contexts with the online social network user. The domain class for this property is *User* and range is *Context* class.

hasContextualProfile: This object property establishes a relation between generic profile and contextual profiles of the user. It is child property of *hasProfileSubset* object property. *Profile* is the domain class and *ContextualProfile* is the range class for this property.

hasContextualValue: This is functional object property which links contextual norms with contexts of the user. The domain and ranges classes for this property are *ContextualNorm* and *Context* respectively.

hasCurrentContextualNorm: This object property links OSN user with current privacy conditions depending on the contextual norms. It has *User* as domain class and *PrivacyCondition* as range class.

hasDefaultProfile: It is child property of *hasProfileSubset* object property. It relates profile with a default profile that is suitable for weaker ties. The domain and range classes for this property are *Profile* and *DefaultProfile*.

hasInformationFlowNorm: This object property links context sensitive resources with their contextual norms. It has *ContextSensitiveResource* as a domain class and *ContextualNorm* as a range class.

hasInteractionType: It is a functional property of the ontology. This object property links various type of interaction performed by OSN users. The domain and range classes for this property are *Interaction* and *InteractionType* respectively.

hasPredictiveIndicator: This object property connects various dimensions of tie strength with their predictive indicators. It has *TieStrengthDimension* as a domain class and *PredictiveIndicator* as a range class.

hasPredictiveIndicatorType: It is a functional property of the ontology. This object property links predictive indicator for tie strength with various types of the indicators. *PredictiveIndicator* is a domain class for this property and range is *PredictiveIndicatorType* class.

hasProfile: It is functional object property that relates an agent with a profile. The domain and range classes for this property are *Agent* and *Profile* respectively.

hasProfileSubset: It relates a generic profile to its concrete profile subset depending on the contextual situation of the user. *Profile* is the domain class for this property and range class is *ProfileSubset*.

hasRelationship: This object property is symmetric and links a user with other users and resources. *User* is the domain class and *Relationship* is the range class for this property.

hasParent: This object property connects a user with his parents. This property has two sub-properties and one inverse property. The domain and range for this property is *User* class.

isParentOf: It is inverse object property of *hasParent* property. It connects the user to his / her children. The domain and range for this property is *User* class.

hasFather: It is sub-property of *hasParent* object property. It connects a user to his / her father. The domain and range for this property is *User* class.

hasMother: It is sub-property of *hasParent* object property. It connects a user to his / her mother. The domain and range for this property is *User* class.

hasGrandparent: It is connecting a user to his / her ancestors. It has also inverse object property. The *User* class is domain and range for this object property.

isGrandparentOf: It is inverse object property of *hasGrandparent*. It connects a user to his / her descendants. The *User* class is also domain and range for this object property.

hasSiblings: This object property connects users sibling to each other. It is symmetric object property by nature. The domain and range for this property is *User* class.

isSpouseOf: This object property is symmetric in nature and connects two persons who are married to each other. The *User* class is domain and range for this object property.

isFriendOf: This is symmetric object property. It connects two users who share mutual friendship with each other. The *User* class is domain and range for this

object property.

isBestFriendOf: This object property is symmetric in nature. It is connecting two users who share a close mutual friendship with each other. The *User* class is domain and range for this object property.

isColleagueOf: This is symmetric object property connecting two users who are a member of the same profession and working for the same organization. The domain and range for this property is *User* class.

isEmployerOf: This object property link a user with an organization which engages the services of this person. It has inverse object property by the label *isEmployedBy*. The domain and range classes for this property are *Organization* and *User*.

isEmployedBy: This object property is inverse of *isEmployerOf* property. It relates a user to an organization for whom this person's services have been engaged. The domain class of the property is *User* and range class is *Organization*.

collaboratesWith: This object property represents a person who works towards a common goal with this person. It is a symmetric object property. The *User* class is domain as well as the range of this property.

isEngagedTo: This object property represents a person to whom this person is betrothed. It is also symmetric object property. The *User* class is domain as well as the range of this property.

metWith: This object property represents a person who has met this person and they have some kind of acquaintance with each other. This object property has *User* as domain and range class. It is also a symmetric object property.

hasTieStrength: This object property is the functional and asymmetric property of the ontology. It links user relationships with relationship strength. The domain class for the property is *Relationship*, whereas, range class is *TieStrength*.

hasTieStrengthDimension: This property links tie strength with various dimensions of tie strength. The domain and range classes for this property are *TieStrength* and *TieStrengthDimension*.

holdsResource: This object property relating an OSN user to a digital resource that the user holds. The domain class is *User* and range class is *DigitalResource*.

hostsResource: This object property relating a user wall to a digital resource that the wall holds. The domain and range classes for this property are *Wall* and *Document* respectively.

holdsRoleInTime: This object property relating an agent to a role that the agent holds. Our ontology reuse this property from PRO ontology. *Agent* and *RoleInTime* are domain and range classes respectively.

isSpecificTo: It is functional object property that represents a profile subset suitable to specific context. The domain class for this property is *ProfileSubset*. The range classes are *Context* and *Role*.

member: This object property indicates a member of a group. We reuse this property from FOAF ontology. The domain and ranges classes for this property are *Group* and *Agent* respectively.

ownsWall: This object property connects an OSN user with specific digital resource representing user activity stream. The domain class for the property is *User* and range class is *Wall*. It is functional object property.

wallOwnedBy: This is inverse object property of *ownsWall*. The domain class for this property is *Wall* class and range is *User* class. It is inverse functional property.

performedWith: This object property connecting with target user with whom interactions are performed. It is inverse property of *performs*. *Interaction* is the domain class for the property. The range class is *User*.

performs: This object property represents various set of interactions performed by an OSN user. The domain and range classes for this property are *User* and *Interaction* respectively.

relates: This object property relates exactly two users in a relationship. It is sub-property of *hasRelationship* property. The domain class for this property is *User*. The ranges class for the property are *DigitalResource* and *User*.

relatesToEntity: A property relating a time-indexed situation to an entity representing the context for that situation. Our ontology reuse this property of PRO ontology. It has two sub-properties in the SOCPRI ontology.

relatesToContext: It is sub-property of *relatesToEntity*. It is relating a time-indexed situation to user context. The domain and ranges classes for this property are *RoleInTime* and *Context* respectively.

relatesToProfile: It is also sub-property of *relatesToEntity*. This property is relating a time-indexed situation to user profile. The domain and ranges classes for this property are *RoleInTime* and *ProfileSubset* respectively.

requestedInContext: This object property connecting access request to context in which access request is generated. The domain class for this property is *AccessRequest* and range is *Context*.

requestedResource: This object property connecting access request to the requested digital resource. The *DigitalResource* is range class for the property. The domain class for the property is *AccessRequest*.

requesterRole: This object property connecting access request to role of an accessor user. The domain and range classes for this property are *AccessRequest* and *Role* respectively.

withRole An object property connecting an agent's role in time to a definition of the type of role held by this agent. It is reused from PRO ontology. *RoleInTime* is domain class for this property and range class is *Role*.

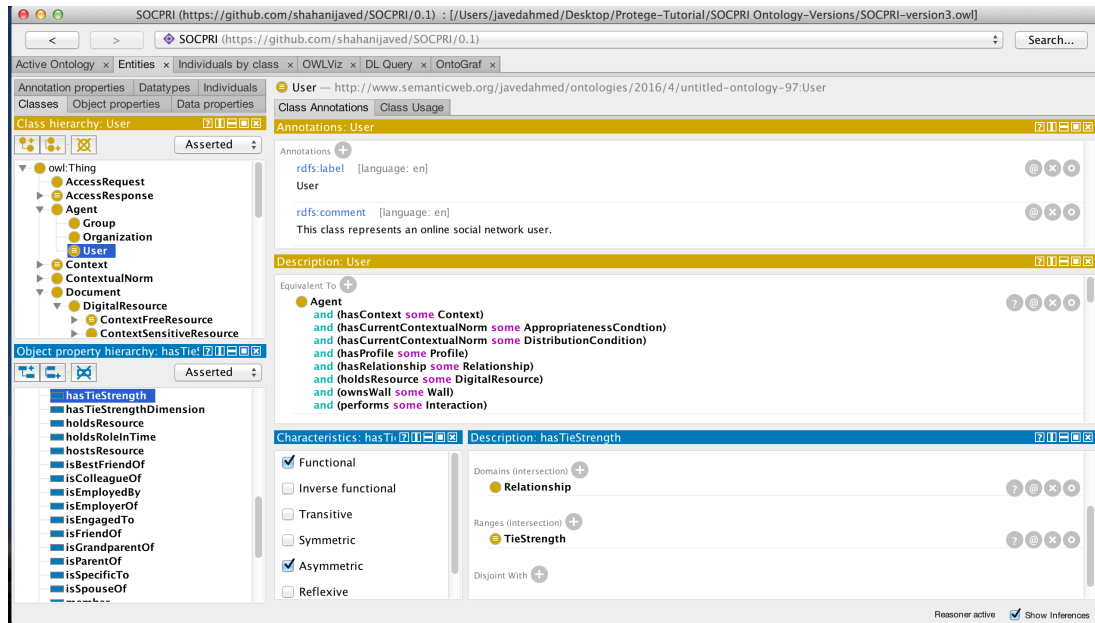


Figure 5.22: SOCPRI Ontology Implementation in Protege

5.7 Comparative Analysis with Social Web Ontologies

We have described various ontologies that represent social web data in section 5.1.1 of this chapter. This section is devoted to comparing these existing ontologies with

the SOCPRI ontology in terms of representing social data and access control decision information. The SOCPRI ontology is different from all of these ontologies in many aspects. First and far most, the SOCPRI ontology takes into consideration social perspective of privacy that is based on tie strength, context segregation, and role based friend segregation. Secondly, most of the existing ontologies ignore the temporal dimension of user relationship with other users and social media elements. The SOCPRI ontology addresses the temporal dimension especially when modeling changes over time in user relations, interactions, interests, etc. Thirdly, the SOCPRI ontology captures both people centric and object-centric relationships, whereas, the most of the existing ontologies considers either people-centric or object-centric relationships. Finally, we infer tie strength among users on the basis of profile similarity attributes and user interaction pattern, whereas none of the existing ontologies provide such capability through inference mechanism.

The existing scientific literature indicates some research efforts towards modeling social relationships by adopting and extending FOAF ontology [171]. However, representing relationships using RDF property suffers from lack of generality for two main reasons: (1) it does not allow specification of attributes such as strength and trust associated with relationship; (2) as a result of this more abstract and rich context information cannot be derived such as best friend, acquaintances, etc. Some researchers use FOAF in combination with the RELATIONSHIP ontology to represent many aspects of personal, professional, sentimental and family relationships. The RELATIONSHIP ontology misses aspects of contextualization of these relationships. For instance, family relations like `parentOf`, `childOf`, `siblingOf`, `spouseOf`, etc do not have the super property we need for all these relations: “family”. RELATIONSHIP ontology focuses only on people-centric relationships and it ignores representation of object-centric relations.

There are several other attempts in the research literature for social web user modeling. The GUMO ontology aims to cover a wide range of user-related information,

however, it falls short of representing user interests that make it unsuitable for the social web environment. Another key shortcoming of GUMO ontology is that it is very extensive and it might be complex to implement in a real system. So far the most comprehensive user model in the semantic web environment is proposed by SWUM ontology. The authors have derived a number of user model dimensions required for social web after extensive analysis of 17 social web applications. A key shortcoming of SWUM ontology is that it ignores tie strength as being one of the dimensions for user modeling. The primary aim of UBO and SemSNI ontologies is to model user interaction with online communities. These ontologies are not designed for purpose of representing object-centric and people-centric relationships.

Social tagging ontologies are specialized only for modeling tagging activities of users in social web environments. There is no direct comparison between SOCPRI and social tagging ontologies in terms of user relationship modeling and exploiting these relations to extract tie strength and rich contextual information. The detailed description of popular social tagging ontologies is presented in section 5.1.1 of this chapter. The primary goal of SIOC ontology is to describe content publishing activities and interactions with the published content. The SKOS is also describing concepts and their relations. The researchers exploited the combination of SIOC, FOAF, and SKOS to represent various aspects of online social networks. The representation of tie strength and diverse social contexts of online social networks users is also out the scope of this combination of ontologies.

In this section, we focused specifically on the comparative study of ontologies created to model different aspects of the social web such as user profile, online posting, tagging, liking, and other common user activities in online communities. We summarise the findings of our comparative study in table 5.1, which present ontologies for representing social media semantics alongside different dimensions these ontologies model. The last row of the table compares SOCPRI features against existing ontologies for the social

Table 5.1: Comparison of SOCPRI ontology with existing Social Web ontologies

Ontology	User	Relations	Interactions	Tags	Interests	Behaviour	Privacy	Tie Strength	Social Context	Temporal Dimension
SOCPRI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
FOAF	Y	Y	-	-	P	-	-	-	-	-
RELATIONSHIP	Y	Y	-	-	-	-	-	-	-	-
GUMO	Y					Y				
SWUM	Y				Y	Y				
UBO	Y		Y		Y	Y				P

web.

5.8 Concluding Remarks

In this chapter, we present core conceptual elements of the SOCPRI ontology. Our ontology models social web user along with his role and relationship-based social context. Our ontology models digital resources of users and represents various profile subsets depending on the contextual role of the user. Modeling social interaction based on tie strength is one of the contributions of this ontology. It also models contextual privacy of social web users which takes into consideration contextual norms for appropriate information disclosure within and across contexts. In this chapter, we also describe in detail the methodology used to develop the SOCPRI ontology. We highlight the role of ontologies in knowledge representation by presenting various ontologies that represent social data. We present a comparative analysis of the SOCPRI ontology with other existing ontologies that represent data for the social web. Briefly, we describe semantic social web before introducing our ontology. The main contribution of the SOCPRI ontology is also discussed in this chapter.

Chapter 6

Evaluation of the SOCPRI Ontology

In this chapter, we describe the evaluation of our ontology. We focused mainly on two different aspects of the evaluation. The first evaluation aspect deals with checking the consistency and correctness of our ontology. The second aspect of evaluation assesses the appropriateness of the ontology against proposed requirements. The consistency of our ontology is evaluated using various reasoners available for Protege. We also used some web based tools for validation of our ontology. These tools are W3C ontology validation service and Ontology Pitfall Scanner (OOPS). We also evaluated our ontology against well-established evaluation metrics resulted from ontology-summit of 2007.

The second aspect of the SOCPRI evaluation is focused on developing DL queries to answer various competency questions. The competency question represents ontology requirements and satisfactory answers to these questions demonstrate the appropriateness of the ontology against proposed requirements. We enriched our ontology with 75 individuals and established relationships between them by providing more than 100 object property assertions. Finally, we developed 24 different DL queries to check the consistency between A-Boxes and T-Boxes of SOCPRI ontology.

6.1 Overview of Ontology Evaluation Approaches

An ontology is a fairly complex structure and it is challenging task to evaluate ontologies. Gomez et al. [172] provide a set of initial and general ideas to perform the evaluation of ontologies. According to authors, ontology evaluation means to make a technical judgment of the ontologies, their associated software environments, and documentation with respect to a frame of reference. This work focuses on dealing with the problem of the three Cs: consistency, completeness, and conciseness. Consistency refers to the incapability of getting contradictory conclusions simultaneously from valid input data. Completeness refers to the extension, degree, amount or coverage to which the information in a user-independent ontology covers the information of the real world. Conciseness focuses on the question whether all the information gathered in the ontology is useful and precise. **Methontology** also provides the guideline for ontology evaluation and identifies different kinds of errors introduced by the ontology developers while modeling ontologies. The list of errors can be subdivided in inconsistency, incompleteness, and redundancy. Figure 6.1 shows detailed classification of these errors. It is important to mention that this list of errors was defined for taxonomies developed assuming frames as a modeling paradigm.

Gangemi et al. [173] identified three main types of dimensions for ontology evaluation in their work: structural, functional and usability profiling dimensions. Structural dimension focuses on the syntax and formal semantics. Functional dimension is related to the intended use of a given ontology and its components. Usability profiling dimension focuses on the ontology profile (annotations). Vrandevic proposed a framework for ontology evaluation [174] that is inspired by Gangemi's work. The framework enables us to assess the quality of an ontology for the web. According to the author, ontology evaluation is the task of measuring the quality of an ontology and is essential for a wide adoption of ontologies. In this work, six aspects are identified and a number

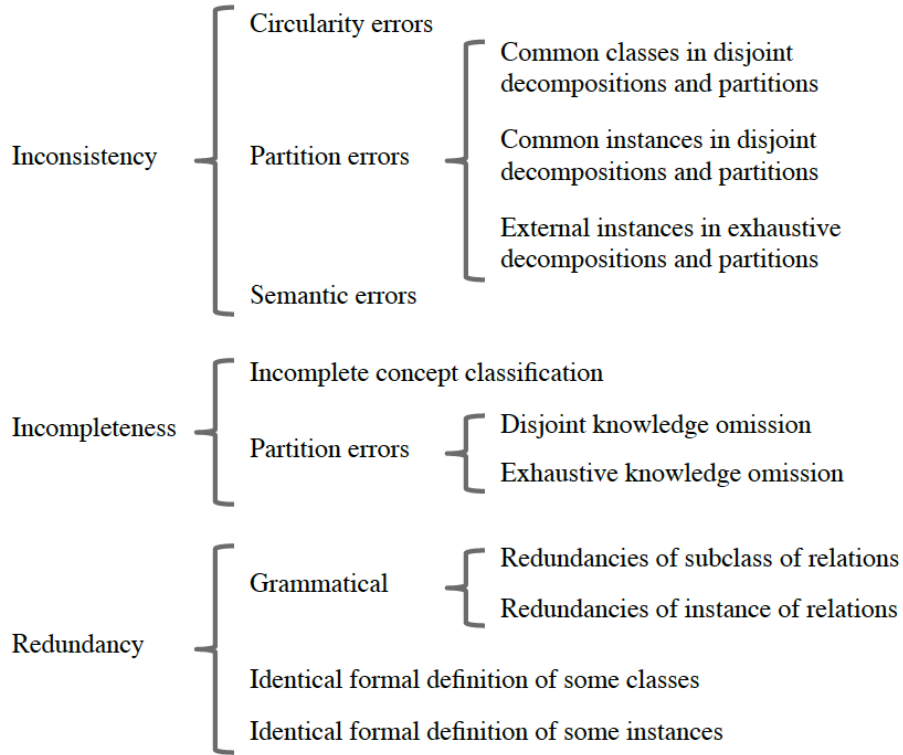


Figure 6.1: Checklist for Modeling Errors [9]

of evaluation methods are proposed in relation to each aspect.

Apart from the aforementioned approaches for ontology evaluation, existing ontology engineering literature mentions various other approaches. However, most of these approaches are tailored for specific application requirements and none of these perfectly fulfill all practical ontology evaluation needs single-handedly. A broad selection of these approaches are gathered by Brank et al. [175] and Sabou et al.[176]. In general, most of the evaluation approaches are classified into one of the following categories:

Qualitative Evaluation: There are several ways to perform a qualitative evaluation of the ontology. One way is to take a set of users and ask them to rate the ontology according to a number of criteria. However, It is difficult to select the

right set of users (ontologists, end-users or domain experts), and it is difficult to develop an actual scale on which to rate particular criteria of the ontology. For these reasons, we do not consider performing a qualitative evaluation of the SOCPRI Ontology.

Metric-based Evaluation: These techniques to evaluate ontologies offer a quantitative perspective of ontology quality. A number of ontology evaluation metrics can be derived automatically. We can distinguish these metrics into two categories: (1) structural metrics; (2) ontological metrics. Structural metrics evaluate the structure of the graph defining the ontology but not the ontology itself. Ontological metrics evaluate the actual model instead of just their underlying graph structure.

Task-based Evaluation: This approach evaluates an ontology based on the competency of the ontology in completing tasks. The disadvantage of this approach is that an evaluation for one application or task may not be comparable with another task.

Golden Standards: This approach compares an ontology with another ontology that is deemed to be the benchmark. We do not have such a gold-standard ontology, so this approach can be dismissed for evaluating the SOCPRI ontology.

In this dissertation, we focus on the metric-based evaluation that constitutes the basis of our evaluation methodology for SOCPRI ontology. We also carry out a partial task-based evaluation by developing various queries which evaluate important functional requirements of the SOCPRI ontology. The qualitative and golden standards based evaluation is not suitable for evaluating our ontology. The limitation in finding suitable domain experts for social theories of Goffman and Granovetter is the main reason to ignore qualitative evaluation for the SOCPRI ontology. We also don't have any golden standard ontology for these social theories, therefore golden standards based

evaluation of the SOCPRI ontology is out of the question. The evaluation metrics adopted for the SOCPRI ontology are described in detail in Section 6.3. The evaluation metrics cover structural aspects of the ontology which includes lexical, hierarchical, syntactical, etc. The ontological aspects of evaluation are covered in Section 6.4. This section evaluates the SOCPRI ontology against consistency, completeness, and conciseness. We also check various common errors introduced during ontology development process in this section. The OntoClean based evaluation of the SOCPRI ontology is presented in Section 6.5. The task-based evaluation of SOCPRI is described in Section 6.6.

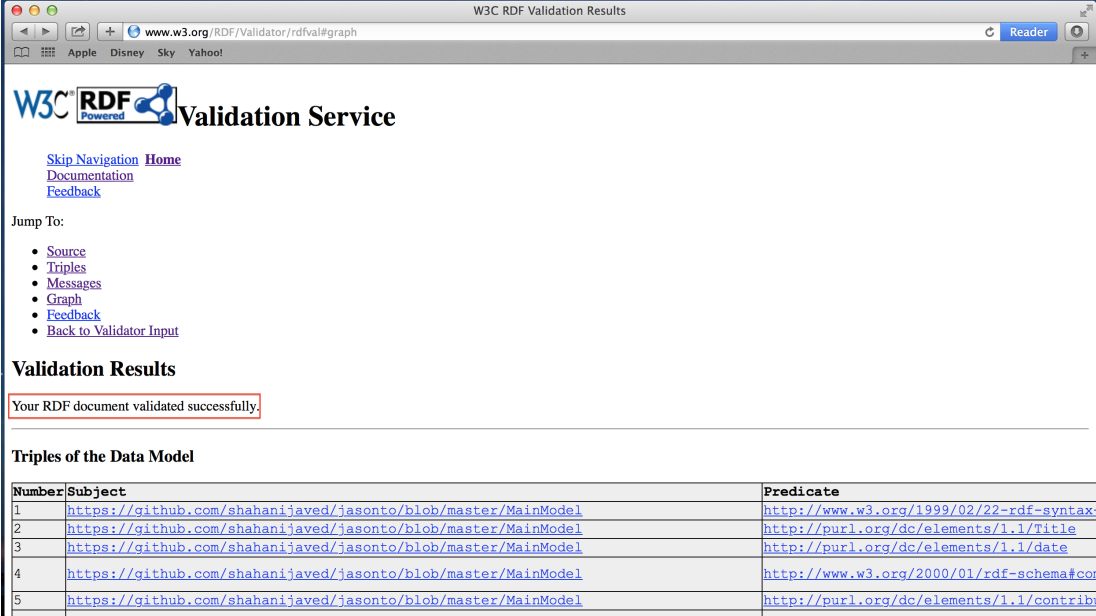
6.2 Validation of SOCPRI Ontology

Ontology evaluation methods cluster around validation and verification. Validation and verification are activities in ontology development process to check that an ontology satisfies its requirements. The ontology validation refers to a diagnostic process to check whether the meaning of ontology definition really represent the real world for which it was designed. Ontology verification refers to defining the correct structure of the ontology that satisfies its requirements. The developer has to verify ontology architecture, syntax, and content to ensure that ontology is well-structured. It is a challenging task for developer to judge the constructed ontology and figure out the anomalies introduced during the construction process and suggest some refinement steps for the errors. There are several automatic tools available to detect anomalies in the ontology definitions and rules. In this section, we present validation results for the SOCPRI ontology using such automatic tools. We used web-based W3C validation service along with three different ontology reasoners to validate the SOCPRI ontology. In section 6.2.1, we present results of W3C validation service for our ontology. We describe three different reasoners and their functionality along with their inference

results for our ontology in section 6.2.2.


6.2.1 W3C RDF Validation Service

W3C RDF Validation service is web based¹ ontology evaluation tool. The service displays the triple representation of the corresponding data model along with optional graphical visualization. There are two possibilities for providing RDF/XML document to the service either through direct input or through URI. The RDF validation service is based on ARP² approach. This W3C service supports the last call working draft specification³ issued by RDF core working group. It does not support deprecated elements and attributes of standard RDF model⁴. The SOCPRI ontology was successfully validated by the service and results are shown in figure 6.2.



W3C RDF Validation Results

www.w3.org/RDF/Validator/rdfval#graph

W3C[®] RDF[™] Powered by  Validation Service

[Skip Navigation](#) [Home](#)
[Documentation](#)
[Feedback](#)

Jump To:

- [Source](#)
- [Triples](#)
- [Messages](#)
- [Graph](#)
- [Feedback](#)
- [Back to Validator Input](#)

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate
1	https://github.com/shahanijaved/jasonto/blob/master/MainModel	http://www.w3.org/1999/02/22-rdf-syntax
2	https://github.com/shahanijaved/jasonto/blob/master/MainModel	http://purl.org/dc/elements/1.1/Title
3	https://github.com/shahanijaved/jasonto/blob/master/MainModel	http://purl.org/dc/elements/1.1/date
4	https://github.com/shahanijaved/jasonto/blob/master/MainModel	http://www.w3.org/2000/01/rdf-schema#cor
5	https://github.com/shahanijaved/jasonto/blob/master/MainModel	http://purl.org/dc/elements/1.1/contrib

Figure 6.2: SOCPRI Results from RDF Validation Service

¹Validation Service, <https://www.w3.org/RDF/Validator/>

²ARP, <https://jena.apache.org/documentation/io/arp.html>

³RDF Core, <http://www.w3.org/2001/sw/RDFCore/#documents>

⁴RDF Model, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

6.2.2 Reasoning with SOCPRI Ontology

One of the benefits of describing ontologies in OWL-DL is that they can be processed by reasoners. Reasoners are useful programs for automatic consistency checking of ontologies by inferring logical consequences from a set of explicitly asserted facts or axioms. The notion of a semantic reasoner generalizes that of an inference engine. The inference rules are commonly specified by the means of an ontology language, and often a description logic. By performing operations such as subsumptions, equivalence and instantiation checking, the reasoners compute the inferred ontology hierarchy and show the list of inconsistencies if they exist in the ontology structure. We are evaluating SOCPRI ontology for logical inconsistencies and design anomalies using three states of art description logic reasoners. These reasoners include Fact++, Pellet, and Hermit. We integrated these reasoners in Protege ontology editor as a software plugin. The preferences set for these reasoners to infer SOCPRI ontology are shown in figure 6.3 and the consistent ontological view is shown in figure 6.4. The detailed description of inferred models for our ontology is presented below.

Fact++ Reasoner was developed at University of Manchester and stands for Fast Classification of Terminologies. It was implemented in C++ language. It employs tableaux algorithms for SHOIQ description logic. It supports OWL DL and a subset of OWL 2 based ontology languages. The strategies used by this reasoner includes absorption, model merging, told cycle elimination, synonym replacement, ordering heuristics and taxonomic classification. It shows exceptional performance on expressive ontologies. The Fact++ Reasoner did not find any error or inconsistency in SOCPRI ontology. The reasoner is satisfied with the soundness and completeness of our ontology.

Hermit Reasoner was developed at University of Oxford. It was implemented in Java language. It employs hypertableau calculus which provides much more efficient

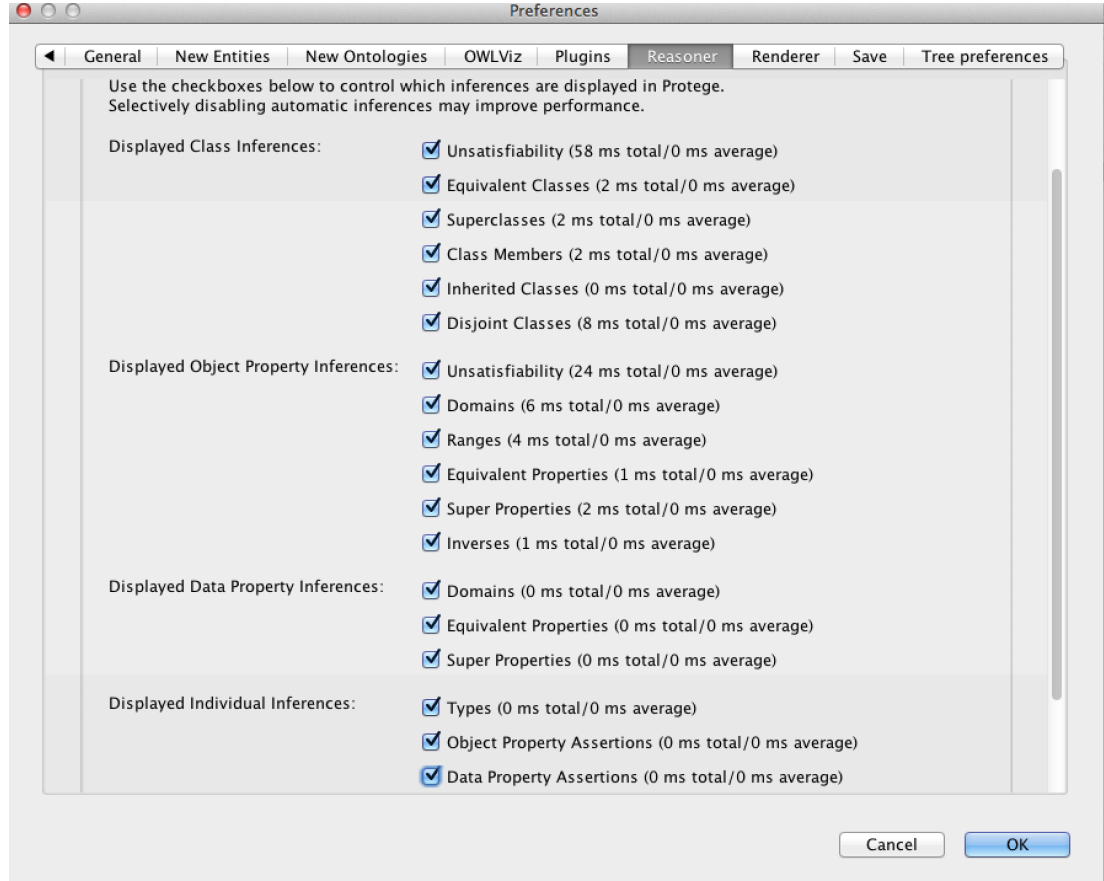


Figure 6.3: Preferences Set for Reasoning SOCPRI Ontology

reasoning than any previously known algorithm. It can determine whether or not a given ontology is consistent and identify subsumption relationships between concepts, among other features. HermiT uses direct semantics and passes all OWL 2 conformance tests for direct semantics reasoners. This reasoner also did not find any error or inconsistency in our ontology. It is also satisfied with soundness and completeness of SOCPRI ontology.

Pellet Reasoner is an open source OWL 2 reasoner. It was implemented in Java language. It employs tableaux algorithms. It was the first reasoner that supported all of OWL DL and has been extended to OWL 2. Pellet includes support for OWL 2 profiles including OWL 2 EL. Apart from its integration with protege, it

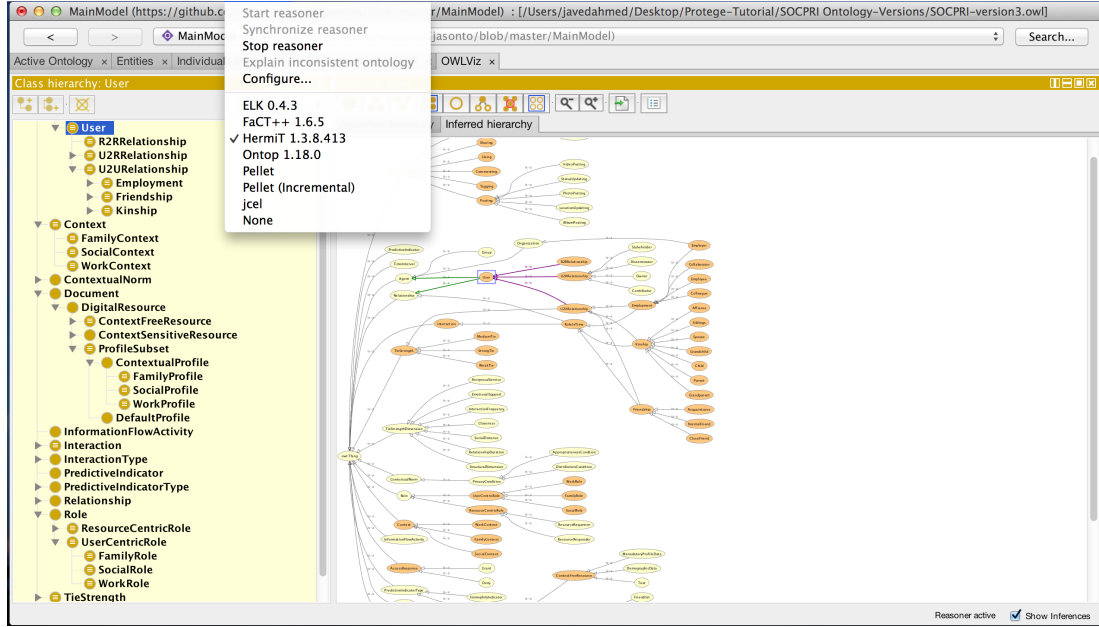


Figure 6.4: Inferred Ontological View of the SOCPRI with Hermit Reasoner

can be used in conjunction with both Jena and OWL API libraries. The inference result of this Reasoner also did not find any inconsistency in SOCPRI ontology. It is also satisfied with soundness and completeness of our ontology.

6.3 SOCPRI Evaluation Metrics

In this section, we present an ontology evaluation metrics for SOCPRI ontology. It is based on the layer-oriented approach which was introduced in ontology summit 2007 [43]. The approach describes a set of identified issues that should be taken into consideration while performing ontology evaluation. The approach distinguishes the layers or dimension into internal and external. Our work focuses on evaluating SOCPRI ontology against internal dimensions. The internal dimension deals with the structure of the ontology and comprises aspects such as concept hierarchy, property structure, disjoint restrictions, domain/range definitions of properties, and naming conventions. The detailed description of all these aspects is presented in following subsections along

with their evaluation metrics. We used ontology evaluation protege plugin⁵ to check SOCPRI ontology against these aspects. This plugin is based on ontology evaluation framework [177] inspired by layer-oriented approach proposed originally in ontology summit 2007. The plugin offers features that evaluate any given ontology against following aspects. The evaluation results for SOCPRI ontology are presented in following subsections.

6.3.1 Concept Hierarchy

Concept hierarchy belongs to the structural/architectural layer of the ontology. The structural/architectural layer characterizes the structural attributes of ontologies which include size, density, depth of the hierarchy, breadth of hierarchy, etc. The concept hierarchy indicates how well a specified taxonomy is structured. Some of the issues associated with inappropriate ontology structure are flat concepts, level of generality and too much depth. A flat concept hierarchy strongly suggests the existence of unexploited grouping possibilities for the concepts with similar semantics. The flat concept hierarchy lack modularity and depth in the ontology. The existence of branches with different structures is an issue that may result in too deep and unbalanced taxonomies. The level of generality is another issue that may result in an inappropriate ontology structure. All these issues need to be considered during the ontology design phase. The ontology evaluation protege plugin has predefined set of evaluation metrics for assessing concept hierarchy of any given ontology. It has following 13 parameters for evaluating concept hierarchy. We describe the results of SOCPRI against each parameter. Figure 6.5 shows concept hierarchy evaluation results for SOCPRI ontology.

C1: It refers to a total number of named classes in the given ontology. This parameter is used to calculate primitive classes in the ontology. SOCPRI has 102 named classes as per results of evaluation through protege plugin.

⁵Ontology Evaluation Plugin, http://protegewiki.stanford.edu/wiki/Ontology_Evaluation

- C2:** It refers to a number of primitive classes in the given ontology. The primitive classes in ontology have only necessary conditions. It can be calculated by subtracting defined classes from the named classes. SOCPRI ontology has 52 defined classes and 102 named classes. As per the results, our ontology has 50 primitive classes which satisfy the condition $C2 = C1 - C3$.
- C3:** It refers to a total number of defined classes in the given ontology. The classes with at least one set of necessary and sufficient conditions fall into this category. As shown earlier, this parameter is used to calculate primitive classes in any given ontology. As per the results, our ontology has 52 defined classes.
- C4:** This parameter of the evaluation tool refers to an average number of parents classes in the given ontology. The bigger the value of this parameter implies the denser the structure of the ontology. SOCPRI ontology is reasonably dense and the value for an average number of parent classes is 0.84.
- C5:** It refers to a maximum number of parent classes of all ontology classes. It is a structure related parameter and expresses the maximum number of is-a relation defined per class. Each class in SOCPRI ontology has maximum one parent class that is also demonstrated by the results shown in figure 6.5.
- C6:** It refers to an average number of siblings classes in the given ontology. The bigger value of the parameter reveals the dense nature of ontology structure. Our ontology has 2.78 average number of siblings that share the same parent node in the SOCPRI structure.
- C7:** It refers to a maximum number of siblings classes in the given ontology. The parameter also deals with dense nature of an ontology structure. The big value for this parameter shows that an ontology is dense with a huge number of child nodes per parent node. SOCPRI ontology has 7 maximum number of siblings

per parent node. It indicates that our ontology is not quite flat and also not too deep.

- C8:** This parameter computes the maximum depth of the ontology tree structure. A big value of this parameter shows a number of structure levels within the ontology. SOCPRI ontology has a maximum depth equal to 3.
- C9:** It refers to a total number of nodes in the ontology tree structure. This parameter also deals with dense nature of ontology structure. It can be computed with equation $C9 = C1 + 1$. SOCPRI ontology has 102 named classes($C1$) and a total number of nodes in SOCPRI is 103 that satisfied the aforementioned equation.
- C10:** It refers to a total number of root classes in the given ontology. It indicates the number of independent classes in the ontology tree structure. The parameter measures modularity of the ontology. Our ontology has 26 root classes.
- C11:** This parameter refers to a total number of internal nodes in the ontology. This metric indicates how dense is the ontology structure. It can be computed using $C11 = C1 - C13$ equation. SOCPRI ontology has 102 named classes($C1$) and a total number of external nodes is 83 ($C13$). Therefore, the total number of internal nodes is 19 which satisfies the aforementioned equation.
- C12:** It refers to a total number of children nodes in the ontology. This metric also expresses the dense nature of the ontology structure. Our ontology has 76 children nodes.
- C13:** This metric refers to a total number of external nodes in the ontology. The root nodes are also taken into account for the computation of this metric. SOCPRI ontology has 83 external nodes. This metric also deals with the density of ontology structure.

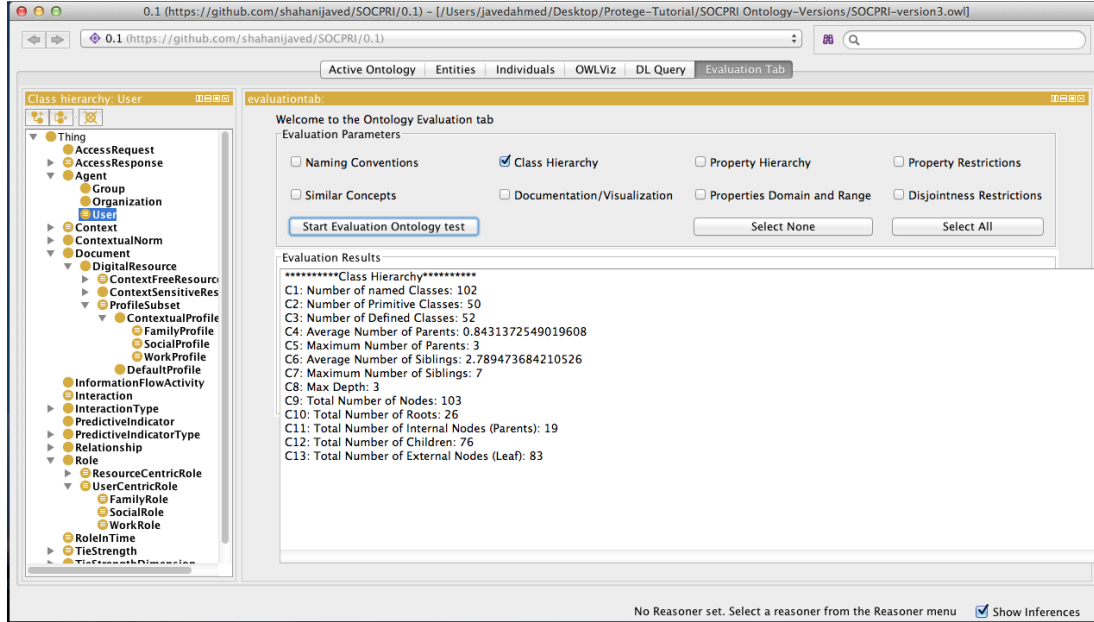


Figure 6.5: Class Hierarchy Evaluation of SOCPRI ontology

6.3.2 Property Structure

Property structure belongs to the structural/architectural layer of the ontology. The issue addressed by this criterion is a lack of well-structured properties in the ontology. The restructuring process is carried out to exploit grouping possibilities for properties with similar functions and equal domains/ranges. The introduction of one or more level of hierarchy between the properties results in more efficient representation and implies a more concrete and understandable ontology structure. Ontology evaluation plugin offers general property metrics that are used to measure a total number of properties (object, data, and annotation) in the ontology. It also offers metrics for measuring structural characteristics of properties. The assessment of the structural characteristics of properties is carried out using C6 to C13 parameters introduced for the assessment of concept hierarchy in the earlier subsection. Figure 6.6 shows results of property hierarchy evaluation for SOCPRI ontology. The detailed description of results is presented as follows:

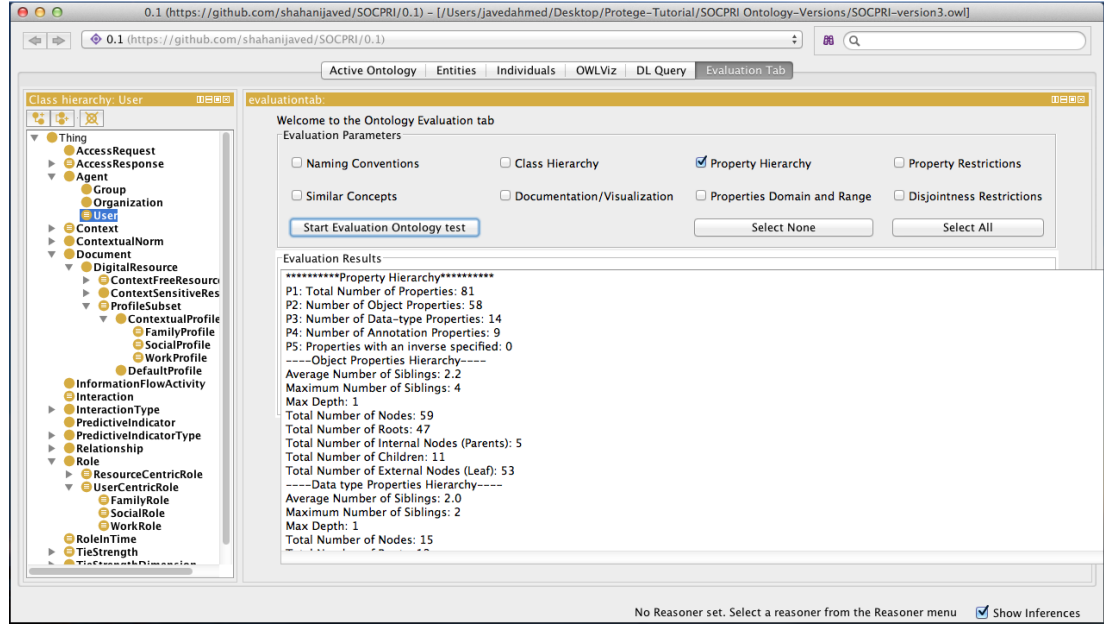


Figure 6.6: Property Hierarchy Evaluation of SOCPRI ontology

P1: It refers to a total number of properties in the ontology. It can be computed using equation $P1=P2+P3+P4$. The total number of properties for SOCPRI ontology are 81 which includes object, datatype, and annotation properties.

P2: It refers to a total number of object properties in the ontology. These properties provide associations between individuals of the same or different classes in ontology. Our ontology offer 58 object property.

P3: This property metric refers to a total number of data properties in the ontology. Data property used to associate individuals to RDF literals. Our ontology provides 14 data properties.

P4: It refers to a number of annotation properties in the ontology. The sole purpose of these properties to add metadata to classes and properties. SOCPRI ontology offers 9 annotation properties.

P5: It refers to the property with an inverse property specified. The SOCPRI ontology

contains five such properties that have their inverse property specified. The example of such properties includes `perform`, `performWith`, `ownsWall`, `wallOwnedBy`, etc.

Apart from these general property metrics, the structural characteristics of properties are treated in the way similar to that adopted for classes. The evaluation metrics discussed in this category measure the features such as depth, average number of siblings, the total number of internal and external nodes. The evaluation of these characteristic reveals that whether the properties of ontology are well structured or unstructured. The restructuring process is initiated on the basis of these results. The detailed description of structural characteristics for object properties of SOCPRI ontology is given below:

C6: It refers to an average number of siblings for object properties in SOCPRI ontology.

The value for this metric is 2.2 as per the results.

C7: It refers to a maximum number of siblings in the ontology. The value of this metric for object properties of SOCPRI ontology is 4.

C8: The maximum depth of the structure is reflected by this parameter. The depth level is 1 for object properties of SOCPRI ontology.

C9: The total number of nodes in the structure are represented by this metric. We have 59 nodes for object properties of our ontology.

C10: It refers to a total number of roots for object properties in SOCPRI ontology. We have 47 root nodes in object properties of our ontology.

C11: It refers to a number of internal nodes in the object properties of the ontology. The total number of internal nodes for object properties in our ontology are 5.

C12: This parameter represents a total number of children nodes in the object properties of the ontology. The value of this parameter for object properties of our ontology is 11.

C13: It refers to a number of external nodes in property structure of the ontology. SOCPRI ontology property structure has 53 external nodes.

We have already discussed in the earlier section about the importance of these structural characteristics and what issues they reflect with various values. The results show that property structure of SOCPRI ontology is quite balanced. It does require restructuring process to introduce further levels of hierarchy in the property structure.

6.3.3 Property Restrictions

The web ontology language supports property restrictions. This feature can be used to create restrictions that describe the constraints on relationships in which individuals participate for a given property. The restrictions describe anonymous classes that contain all of the individuals that satisfy these restrictions. The usage of restrictions in an ontology reflects that ontology has been designed carefully and concrete definitions of the ontology elements are provided. The total time taken by reasoner for checking the consistency of ontology also depends on the usage of these restrictions. The unnecessary use of restrictions would always result in poor performance for consistency checking algorithm.

The OWL restrictions fall into three main categories such as quantifier restrictions, cardinality restrictions, and hasValue restrictions. The quantifier restrictions are further subdivided into existential restrictions and universal restrictions. The SOCPRI contains several defined classes that are based on quantifier restrictions. The property restrictions can be evaluated by the number of various restrictions that exist in an ontology. The evaluation metrics for property restrictions used by protege plugin

are described below. The results of the property restrictions evaluation for SOCPRI ontology are shown in figure 6.7.

P6: It refers to the total number of restrictions applied on the given ontology. The absence of any property restriction indicates primitive nature of the ontology. A well-defined ontology contains a reasonable number of restrictions. The excessive usage of the property restrictions also increases the overhead for the ontology reasoner. The SOCPRI makes wise use of the property restrictions. The total number restriction applied on our ontology are 86.

P7: It refers to the number of existential restrictions applied on the given ontology. Our ontology contains 79 existential restrictions. The central concept of this ontology is **User** that contains eight such restricts which involves object properties such as `hasContext`, `hasProfile`, `hasRelationship`, `holdsResource`, `ownsWall`, etc.

P8: It refers to the number of universal restrictions applied to the given ontology. The SOCPRI ontology contains 7 universal restrictions. The universal restrictions are kind of necessary and sufficient conditions that accommodate the issue of open world assumption supported by OWL ontologies.

6.3.4 Domain / Range Definition of Properties

It is an important aspect that deals with the definition of the domain and range in ontology properties. It belongs to the data/application layer of the ontology. Poorly defined the domain and range in ontology properties results inconsistencies and prevent applications from properly consuming ontologies. The domain represents the objects to which the property can be applied, whereas, the range represents potential individuals to which domain objects are mapped. Object properties link individuals from their domain to individuals within their range. The protege plugin evaluates this aspect

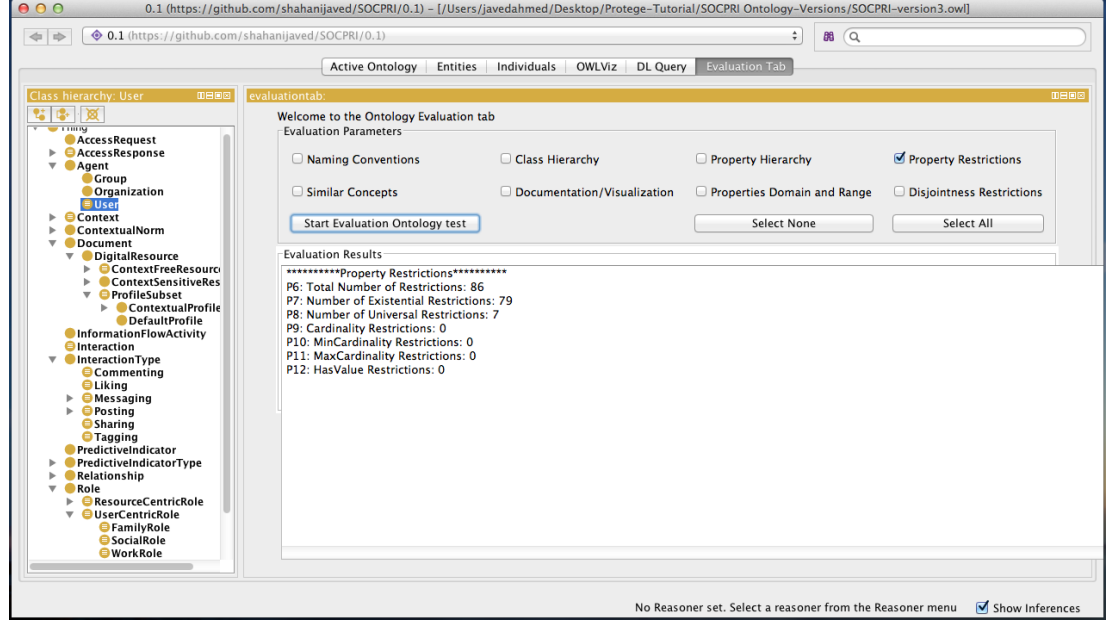


Figure 6.7: Evaluation of SOCPRI for PropertyRestrictions

against following two parameters. Figure 6.8 shows the results of domain and range definition for SOCPRI ontology.

R1: It refers to all properties in an ontology for which the domain attribute is defined as a valid non-empty entity. As per results of protege plugin for ontology evaluation, SOCPRI ontology contains 91.6% properties with their domain specified.

R2: This parameter refers to all properties in an ontology for which the range attribute is defined as a valid non-empty entity. Our ontology has 87.5% properties with their range defined as a valid conceptual entity from the ontology structure.

6.3.5 Disjointness Restrictions

Disjointness restrictions belong to the usability layer of the ontology. The restrictions are applied on ontological concepts and attributes in order to restrict the domain in which these entities are used. The proper definition of classes and properties with appropriate disjointness restrictions enhance their reusability by other applications. The

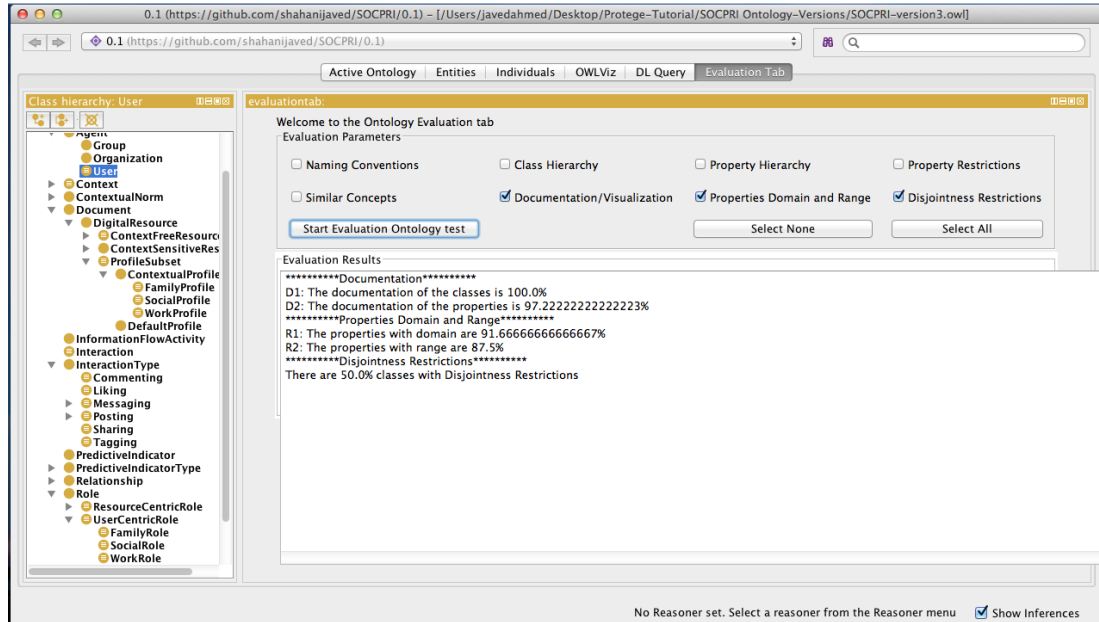


Figure 6.8: Evaluation of SOCPRI against Multiple Metrics

usage of disjointness restriction should be handled carefully during ontology development process. The OWL classes are assumed to overlap. In order to prevent this overlapping, the ontology designer must make the classes disjoint from one another. This ensures that an individual which has been asserted to be a member of one of the classes in the disjoint group cannot be a member of any other classes in that group. It is also learned from ontology development experiment that not all of the classes in the ontology should be disjoint. SOCPRI ontology applies disjoint restriction very carefully and only 50% classes in our ontology are with disjointness restrictions. Figure 6.8 shows the evaluation results of disjoint restrictions for SOCPRI ontology.

6.3.6 Documentation/Visualization

Documentation/Visualization belongs to the representational layer of the ontology. It focuses on the issue of representation of the ontology to the outside world. This metric deals with the activity of enriching an ontology with additional information such as

annotations, comments, metadata, etc. Particularly, it refers to anything that could be helpful to make the ontology more readable to the users. The experience demonstrated that ontologies are usually poorly documented which results in a lack of reusability and readability of the ontology. The goal of evaluation metrics is to assess the amount of information included in the ontology for documentation purposes. Figure 6.8 shows the results of documentation metrics for SOCPRI ontology. The detailed description of metrics is given below:

D1: It refers to the percentage of documented classes in the given ontology. The closer to 100% value of this parameter shows that an ontology is well-documented. As per the result of protege plugin used for SOCPRI evaluation, our ontology includes 100% documented classes.

D2: This parameter refers to the percentage of documented properties in the ontology. Similarly, the value closer to 100% reflects properly document ontology structure. Our ontology evaluation results show that 97% of the properties in the ontology are documented.

6.3.7 Naming Conventions

Naming conventions belong to the lexical/vocabulary layer of the ontology. It focuses on the formulation of well-formed terms and definitions. There are a number of useful conventions that can be applied in naming. According to this criterion, we adopted one common naming convention for all the classes and properties. Classes names used a pattern where words start with a capital letter such as *MyClassName*. The property names start with an initial small letter followed by words starting with capital letters such as *myPropertyName*. The evaluation metrics along with results for SOCPRI ontology are described below. Figure 6.9 shows the results of evaluation for naming convention of SOCPRI ontology.

N1: It refers to classes with same naming conventions in the ontology. The parameter is equal to the percentage of the majority of classes that adopt the same naming convention in the ontology. We described the pattern used to name classes in the SOCPRI ontology. As per the results, 95% classes follow the same described pattern.

N2: It refers to object properties with same naming conventions in the ontology. The parameter takes into account property names that begin with a lower case letter. As per results, the majority of the object properties in SOCPRI ontology follow same naming convention.

N3: This is the same parameter as discussed in the previous cases but it applies to data type properties in the ontology. The majority of the data properties in our ontology also follow the same naming convention.

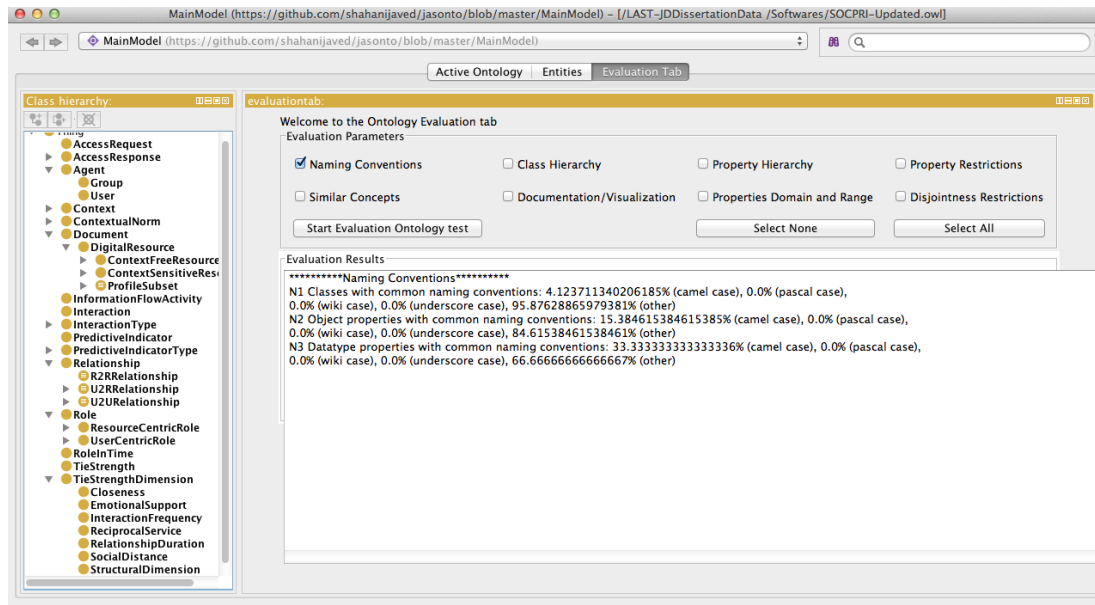


Figure 6.9: Evaluation of SOCPRI for Naming Conventions

6.4 Pitfall Scanning of SOCPRI Ontology

Ontology modeling is a complex activity that requires developers to tackle a wide range of difficulties while modeling ontologies. These difficulties can imply the appearance of anomalies in the ontologies. It is a crucial issue to identify these anomalies in the ontologies. Current literature in ontology engineering domain describes a set of common errors introduced by developers during ontology development process. One of the initial studies identifying a set common errors was carried by [172]. The authors categorize these errors into three types which include inconsistency, incompleteness, and redundancy. Rector et al. [178] also provided a list of common errors made by developers during ontology modeling.

In this section, we focus on work of Poveda [10] that identify an initial catalogue of common pitfalls for ontology evaluation. These pitfalls are bad practices followed by developers that could lead to errors in ontologies. This catalogue of pitfalls includes errors identified in existing literature as well as new set of common errors found after the manual review of more 30 ontologies [179]. Currently, the catalogue lists 41 pitfalls. It is not an exhaustive list and other pitfalls might be included in the future. A detailed description of these pitfalls is presented in section 6.4.2. A detailed description of a web-based tool is presented in section 6.4.1. This is developed by Poveda [179] to help ontology developers to evaluate a given ontology against these pitfalls. Finally, we evaluated SOCPRI ontology using this tool and evaluation results are discussed in detail in section 6.4.3.

6.4.1 Ontology Pitfall Scanning Tool

In this section, we present OntOlogy Pitfall Scanner (OOPS)⁶ a web-based application intended to help ontology developers to detect some of the most common errors introduced while modeling an ontology [179, 180]. The tool can be executed independently

⁶OOPS!, <http://oops.linkeddata.es>

of the ontology development platform. Apart from a web-based interface, the tool provides REST service for integration with third party applications. Ontology pitfall scanner supports a semi-automatic diagnosis of OWL ontologies. The tool checks the given ontology against a pitfall catalogue that contains 41 pitfalls and described in the following section. Web-based interface of the tool is compatible with some of the popular web browsers such as Firefox, Safari, Chrome, etc. It consists on a simple view that is shown in figure 6.10. A user can enter URL of the ontology or paste RDF code in the input box. The tool generates a comprehensive list of the pitfalls appearing in the given ontology as a result. The results provide a brief description about appearing pitfalls, their frequency, and list ontology elements affected by such a pitfall. It is worth mentioning that it can detect some of the pitfalls in automated ways, which means that they should be repaired; while other are detected in a semi-automated way, which means that they must be manually checked in order to discern whether the elements identified actually contains errors. The tool has been widely accepted by the semantic web community. Currently, it has been used by different organizations such as AtoS, TecNALIA, Departament Arquitectura, La Salle at Universitat Ramon Llull and Human Mobility and Technology Laboratory at CICtourGUNE [181]. Some of the advantages of this tool are: it is freely available on the web; it enlarges the list of errors detected by the most available tools such as Moki⁷, XD Analyzer⁸; finally, it is fully independent of any ontology development environment.

6.4.2 Catalogue of Common Pitfalls

This section describes the complete list of 41 pitfalls used by ontology pitfall scanner to detect anomalies in the given ontologies. This catalogue is developed by [10]. The catalogue provides a template for the description of pitfalls which includes fields such as title, description, elements affected, importance level, etc. The title provides a

⁷Moki, <https://moki.fbk.eu/website/index.php>

⁸XD Analyzer, <http://neon-toolkit.org/wiki/XDTools>

Figure 6.10: Web Interface for Ontology Pitfall Scanner

brief description of what the pitfall is about. The description contains the detailed explanation of what the pitfall consist on, bibliographical references and example. The “Element Affected” field points out specific ontology elements affected by the pitfall such as classes, object properties, datatype properties, etc. The pitfall description template also contains information about how critical each pitfall is. It is obvious that all the pitfalls are not equally important. The author has identified three levels of importance, namely: critical, important and minor. The critical status points out crucial nature of the pitfall that needs correction, otherwise, it could affect the ontology consistency, reasoning, applicability, etc. The important status points out that the pitfall is not critical for ontology function, but, it is important to correct this type of pitfall. Minor status points out that it is not a real problem, but, by correcting it we will make the ontology nicer. Figure 6.11 shows the complete list of pitfalls along with their level of importance. Following is a brief description of all pitfalls that are adapted from work of Poveda [10].

P01. Creating polysemous elements: This pitfall deals with an issue related to the inclusion of ontology element whose identifier has different meanings in the ontology and represents more than one conceptual idea. The importance level of this pitfall is critical and affects to classes, object and datatype properties

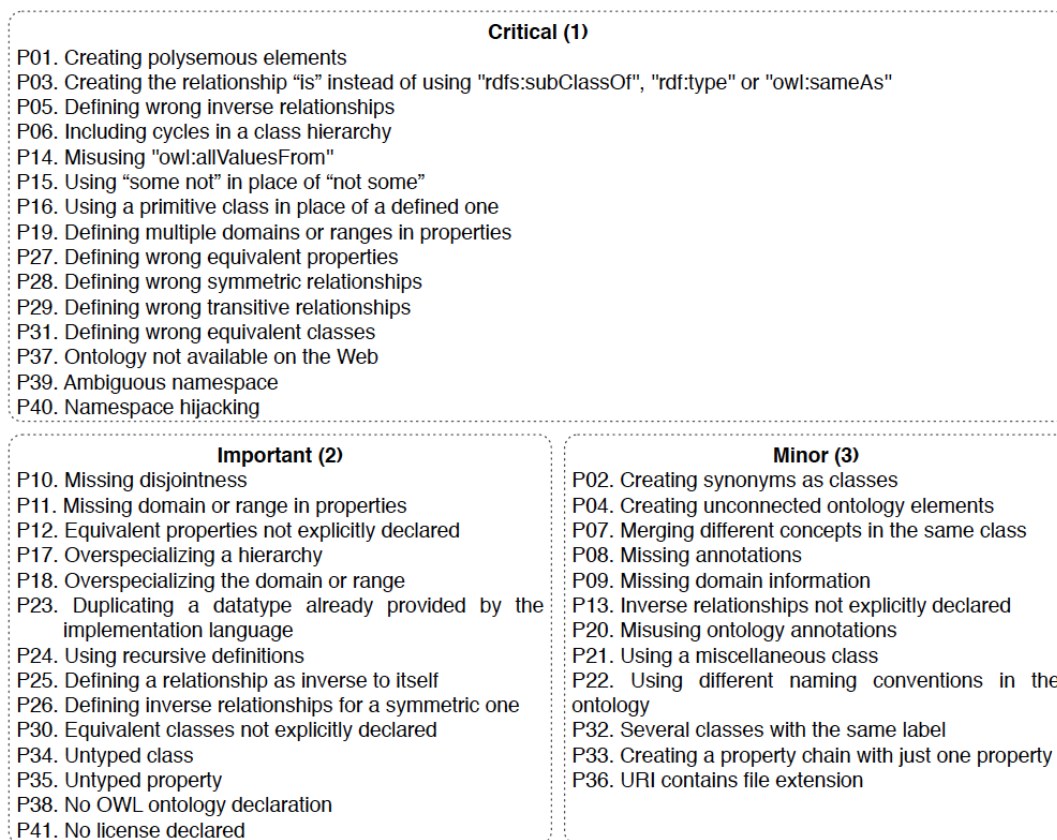


Figure 6.11: Classification of pitfalls by level of importance [10]

in the ontology. The aspects evaluated by this pitfall are modeling decisions, wrong inference and ontology understanding that belongs to the structural and the usability-profiling dimensions of the given ontology.

P02. Creating synonyms as classes: This pitfall deals with an issue related to classes whose identifier are synonyms and these are defined as equivalent classes. The importance level of this pitfall is minor and affects to only class elements of the ontology. The aspects evaluated by this pitfall are modeling decisions and ontology understanding that belong to the structural and the usability-profiling dimensions of the given ontology.

P03. Creating the relationship “is” instead of using “subclassOf”, “instanceOf”

or “sameIndividual”: This pitfall deals with issue related to creation of “is” relationship instead of using OWL primitives for representing the subclass relationship, class membership, or the equality between instances. The importance level of this pitfall is critical and affects to object properties of the ontology. The aspect evaluated by this pitfall is modeling decisions which belong to the structural dimension of the given ontology.

P04. Creating unconnected ontology elements: This pitfall deals with an issue related to the creation of isolated ontology elements that have no relation with rest of the ontology. The importance level of this pitfall is minor and affects to classes, object and datatype properties of the ontology. The aspects evaluated by this pitfall are requirement completeness and real world modeling that belongs to the functional dimension of the given ontology.

P05. Defining wrong inverse relationships: This pitfall deals with an issue related to defining an inverse relations between two entities when actually they are not necessarily in an inverse relationship. The importance level of this pitfall is critical and affects to only object properties in the ontology. The aspect evaluated by this pitfall is a wrong inference that belongs to the structural dimension of the given ontology.

P06. Including cycles in a class hierarchy: This pitfall deals with an issue related to the inclusion of cycle between two classes in a hierarchy in the ontology. The importance level of this pitfall is critical and affects to classes in ontology. The aspect evaluated by this pitfall is a wrong inference that belongs to structural dimension of the ontology.

P07. Merging different concepts in the same class: This pitfall deals with an issue related to the creation of class whose identifier is referring to two or more different concepts. The importance level of this pitfall is minor and affects to

classes in ontology. The aspects evaluated by this pitfall are modeling decisions and ontology understanding that belongs to the structural and the usability profiling dimensions of the given ontology.

P08. Missing annotations: This pitfall deals with an issue related to missing annotations for ontology terms. The importance level of this pitfall is minor and affects to classes, object and datatype properties in the ontology. The aspects evaluated by this pitfall are ontology understanding and ontology clarity that belongs to the usability profiling dimension of the given ontology.

P09. Missing domain information: This pitfall deals with an issue related to lack of needed information in the ontology. The importance level of this pitfall is minor and affects to the ontology as a whole. The aspects evaluated by this pitfall are real world modeling and requirement completeness that belongs to the functional dimension of given the ontology.

P10. Missing disjointness: This pitfall deals with an issue related to lack of disjoint axioms between ontology elements that should be defined as disjoint. It is an important pitfall and affects to classes, object and datatype properties in the ontology. The aspect evaluated by this pitfall is real world modeling that belongs to the functional dimension of the given ontology.

P11. Missing domain or range in properties: This pitfall deals with an issue related to missing domain or range for attributes in the ontology. It is also an important pitfall and affects to object and datatype properties in the ontology. The aspects evaluated by this pitfall are no inference and ontology understanding that belongs to the structural and the usability profiling dimensions of the given ontology.

P12. Equivalent properties not explicitly declared: This pitfall deals with an

issue related to lack of information in ontology about equivalent properties in cases of duplicated attributes. The importance level of this pitfall is “important” and affects object and datatype properties in the ontology. The aspects evaluated by this pitfall are no inference and ontology understanding that belongs to the structural and the usability profiling dimensions of the given ontology.

P13. Inverse relationships not explicitly declared: This pitfall deals with an issue related to missing inverse relationship in the ontology. The importance level of this pitfall is minor and affects to object properties in the ontology. The aspects evaluated by this pitfall are no inference and ontology understanding that belongs to the structural and the usability profiling dimensions of the given ontology.

P14. Misusing “allValuesFrom”: This pitfall deals with an issue related to using the universal restrictions as the default qualifier instead of the existential restrictions. The importance level of this pitfall is critical and affects to classes in ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P15. Misusing “not some” and “some not”: This pitfall deals with issue related to using a “some not” structure when a “not some” is required. The importance level of this pitfall is critical and affects to classes in the ontology. The aspect evaluated by this pitfall is a wrong inference that belongs to the structural dimension of the given ontology.

P16. Misusing primitive and defined classes: This pitfall deals with issue related to the open world assumption. The pitfall implies creating a primitive class rather than a defined one in case automatic classification of the individual is intended. In general, nothing will be inferred to be subsumed under a primitive class by the classifier. The importance level of this pitfall is critical and affects to classes in

the ontology. The aspect evaluated by this pitfall is no inference that belongs to the structural dimension of the given ontology.

P17. Specializing too much a hierarchy: This pitfall deals with issue related to over specializing a hierarchy in such a way that final leaves are defined as classes and these classes will not have instances. It is an important pitfall and affects to classes in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P18. Overspecializing the domain or range: The pitfall deals with issue of defining a domain or range not general enough for a property. It is also an important pitfall and affects to object and datatype properties in the ontology. The aspect evaluated by this pitfall is wrong inference that belongs to the structural dimension of the given ontology.

P19. Defining multiple domain or ranges in properties: The pitfall is related to issue that appears when defining multiple domains or ranges of a property in the ontology. The importance level of this pitfall is critical and affects to object properties in the ontology. The aspect evaluated by this pitfall is a wrong inference that belongs to the structural dimension of the given ontology.

P20. Misusing ontology annotations: This pitfall deals with issue related to swapping or misusing the content of some annotations properties. The importance level of this pitfall is minor and affects to classes, object and datatype properties of the ontology. The aspect evaluated by this pitfall is ontology understanding that belongs to the usability profiling dimension of the given ontology.

P21. Using a miscellaneous class: This pitfall deals with issue related to the creation of a class with the only goal of classifying the instances that do not belong to any of its sibling classes. The importance level of this pitfall is minor and

affects to classes in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P22. Using different naming convention in the ontology: This pitfall deals with issue of naming ontology elements that does not follow same naming convention. The importance level of this pitfall is minor and affects to ontology as a whole. The aspect evaluated by this pitfall is ontology clarity that belongs to the usability profiling dimension of the given ontology.

P23. Duplicating a datatype already provided by the implementation language: This pitfall deals with issue related to creation of class and its corresponding individuals to represent existing datatypes in the implementation language. It is an important pitfall and affects to classes in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P24. Using recursive definitions: This pitfall deals with issue related to using ontology element in its own definition. It is also an important pitfall and affects to classes, object and datatype properties in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P25. Defining a relationship as inverse to itself: This pitfall deals with issue related to defining a relationship as inverse to itself. The relationship should have been defined as “owl:SymmetricProperty” in such situation. It is an important pitfall and affects to object properties in the ontology. The aspect evaluated by the pitfall is modeling decisions that belong to the structural dimension of the ontology.

P26. Defining inverse relationship for a symmetric one: This pitfall deals with

issue of defining a symmetric object property as inverse of another object property. It is also an important pitfall and affects to object properties in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.

P27. Defining wrong equivalent relationships: This pitfall deals with an issue of defining two object/datatype properties as equivalent even though they do not have the same semantics. The important level of this pitfall is critical and affects to object or datatype properties in the ontology. The aspect evaluated by this a pitfall is wrong inference that belongs to the structural dimension of the given ontology.

P28. Defining wrong symmetric relationships: This pitfall deals with issue related to defining a relationship as symmetric, when the relationship is not necessarily symmetric. The importance level of this pitfall is critical and affects to object properties in the ontology. The aspect evaluated by the pitfall is a wrong inference that belongs to the structural dimension of the given ontology.

P29. Defining Wrong transitive relationships: It deals with defining a relationship as transitive, when the relationship is not necessarily transitive. The importance level of this pitfall is critical and affects to object properties in the ontology. The aspect evaluated by this pitfall is a wrong inference that belongs to the structural dimension of the given ontology.

P30. Equivalent classes not explicitly declared: This pitfall deals with issue related to missing definition of equivalent classes in case of duplicated concepts. It is an important pitfall and affects to classes in ontology. The aspect evaluated by this pitfall is no inference that belongs to the structural dimension of the given ontology.

- P31. Defining wrong equivalent classes:** It is critical pitfall that deals with issue related to defining two classes equivalent in situation when they are not necessarily equivalent classes. The importance level of this pitfall is critical and affects to classes in the ontology. The aspect evaluated by the pitfall is a wrong inference that belongs to the structural dimension of the given ontology.
- P32. Several Classes with the same label:** It is a minor pitfall that affects to classes in ontology. The issue addressed by this pitfall is lack of accuracy in defining terms. Two or more classes have the same content for annotations for naming. The aspect evaluated by this pitfall is ontology understanding that belongs to the usability profiling dimension of the given ontology.
- P33. Creating a property chain with just one property:** This pitfall deals with issue related to creating a property chain that includes only one property in the antecedent part. The importance level of this pitfall is minor and affects to object property in the ontology. The aspect evaluated by this pitfall is modeling decisions that belong to the structural dimension of the given ontology.
- P34. Untyped class:** It is an important pitfall that affects to classes in the ontology. It deals with issue related to ontology element used as a class without having been explicitly declared using the primitives “owl:Class” or “rdfs:Class”. The aspect evaluated by this pitfall is ontology language that belongs to the structural dimension of the given ontology.
- P35. Untyped Property:** It is also an important pitfall that affects to object and datatype properties in the ontology. It deals with issue related to ontology element used as a property without having been explicitly declared using primitives “owl:ObjectProperty” or “owl:DatatypeProperty”. The aspect evaluated by this pitfall is ontology language that belongs to the structural dimension of the given ontology.

P36. URI contains file extension: This pitfall deals with issue related to file extensions included in an ontology URI. The importance level of this pitfall is minor and affects to ontology as a whole. The aspect evaluated by the pitfall is application context that belongs to the functional dimension of the given ontology.

P37. Ontology not available on the Web: It is pitfall that deals with availability of ontology encoding and documentation on the web. The importance level of this pitfall is critical and affects to the whole ontology. The aspects evaluated by the pitfall are application context and ontology understanding that belong to the functional and the usability-profiling dimensions of the given ontology.

P38. No OWL ontology declaration: It is an important pitfall that affects to ontology as whole. The issue addressed by this pitfall is related to ontology meta-data. The aspects evaluated by this pitfall are ontology metadata, ontology language and application context which belong to structural, functional and usability profiling dimensions of the given ontology.

P39. Ambiguous namespace: This pitfall deals with issue related to missing declaration and namespace of the ontology. The importance level of this issue is critical and affects to the whole ontology. The aspect evaluated by this pitfall is application context that belongs to the functional dimension of the given ontology.

P40. Namespace hijacking: It is also a critical pitfall that affects to classes, object and datatype properties of the ontology. It refers to reusing terms from another namespace that are not defined in such namespace. The aspect evaluated by this pitfall is application context which belongs to the functional dimension of the given ontology.

P41. No license declared: It is an important pitfall that affects to ontology as a whole. It deals with missing license information from the meta-data of the

ontology. The aspect evaluated by this pitfall is ontology metadata that belongs the usability profiling dimension of the given ontology.

These pitfalls evaluate the quality of an ontology against different ontological perspectives which includes structural, functional, and usability-profiling dimensions. Originally, these dimensions were defined by Gangemi [173] and further extended by Poveda [182]. The structural dimension is focused on syntax and formal semantics. It is further extended into aspects such as modeling decisions, no inference, wrong inference, and ontology language. The functional dimension is related to the intended use of the ontology. The extended aspects taken into account within this dimension are real world modeling, requirement completeness, and application context. The communication context of the ontology is represented by usability-profiling dimension. The aspects contemplated for this dimension are ontology understanding, ontology clarity, and ontology metadata. All the pitfall presented in this section are associated with at least one of these evaluation aspects. This catalogue of pitfalls also checks consistency, completeness, conciseness aspects of the given ontology. Figure 6.12 shows classification of pitfalls against these evaluation aspects. This figure represents user interface of web-based pitfall scanning tool.

6.4.3 Evaluation Results for SOCPRI

In this section, we present evaluation results of SOCPRI ontology. The evaluation was performed with ontology pitfall scanner tool⁹. The detailed description of the tool is presented in section 6.4.1. The set of common errors, identified by the tool, are also described in detail in section 6.4.2. The tool evaluates various perspectives of SOCPRI ontology such as structural dimension, functional dimension, usability-profiling dimension, consistency, completeness, and conciseness. Before the detailed description of evaluation results for these ontological perspectives, we present overview

⁹OOPs! <http://oops.linkeddata.es>

☐ Select Pitfalls for Evaluation
☒ Select Category for Evaluation

Classification by Dimension

☐ **Structural Dimension**

- ☐ **Modelling Decisions:** Checks for pitfalls P02, P03, P07, P21, P24, P25, P26 and P33.
- ☐ **Wrong Inference:** Checks for pitfalls P05, P06, P19, P27, P28, P29 and P31
- ☐ **No Inference:** Checks for pitfalls P11, P12, P13 and P30.
- ☐ **Ontology language:** Checks for pitfalls P34, P35 and P38.

☐ **Functional Dimension**

- ☐ **Real World Modelling or Common Sense:** Checks for pitfall P04 and P10.
- ☐ **Requirements Completeness:** Checks for pitfall P04 and P09.
- ☐ **Application context:** Checks for pitfalls P36, P37, P38, P39 and P40.

☐ **Usability-Profiling Dimension**

- ☐ **Ontology Clarity:** Checks for pitfalls P08 and P22.
- ☐ **Ontology Understanding:** Checks for pitfalls P02, P07, P08, P11, P12, P13, P20, P32 and P37
- ☐ **Ontology Metadata:** Checks for pitfalls P38 and P41

Classification by Evaluation Criteria

☐ **Consistency**

For this evaluation criteria the following pitfalls will be checked: P05, P06, P07, P19 and P24.

☐ **Completeness**

For this evaluation criteria the following pitfalls will be checked: P04, P10, P11, P12 and P13.

☐ **Consciseness**

For this evaluation criteria the following pitfalls will be checked: P02, P03 and P21.

Figure 6.12: Classification of pitfalls against various evaluation aspects

of the errors identified in our ontology and corrective measures taken to fix these errors.

Ontology pitfall scanning tool identified 8 pitfalls in SOCPRI ontology during the initial evaluation phase. This list of pitfalls includes 4 minor pitfalls, 3 important pitfalls, and 1 critical pitfall. The identified minor pitfalls in our ontology are P04, P08, P13, and P22. The important pitfalls identified in the SOCPRI ontology are P11, P30, and P41. The identified only critical pitfall in our ontology is P19. Figure 6.13 shows initial evaluation results for SOCPRI ontology.

During the initial evaluation phase, pitfall scanning process identified one critical pitfall that is related to defining multiple domains or ranges in object or datatype properties of the ontology. This pitfall is given **P19** identification number. The four cases of **P19** were identified in SOCPRI ontology. The properties affected by this pitfall are **relates**, **hasTieStrength**, and **atTime**. Figure 6.14a shows details of this pitfalls. As per syntactical rules, OWL allows definition of multiple **rdfs:domain** or **rdfs:range** axioms, but this modeling decision may contribute in a wrong inference. We fix this issue for two object properties, namely: **relates**, **hasTieStrength**. The creation of these object properties is our modeling decision, therefore, we avoid their usage for mul-

Evaluation results

It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:

- **Critical** 🚫 : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important** ⚠️ : Though not critical for ontology function, it is important to correct this type of pitfall.
- **Minor** 🟡 : It is not really a problem, but by correcting it we will make the ontology nicer.

[Expand All] | [Collapse All]

Results for P04: Creating unconnected ontology elements.	1 case Minor 🟡
Results for P08: Missing annotations.	29 cases Minor 🟡
Results for P11: Missing domain or range in properties.	1 case Important ⚠️
Results for P13: Inverse relationships not explicitly declared.	37 cases Minor 🟡
Results for P19: Defining multiple domains or ranges in properties.	4 cases Critical 🚫
Results for P22: Using different naming conventions in the ontology.	ontology* Minor 🟡
Results for P30: Equivalent classes not explicitly declared.	2 cases Important ⚠️
Results for P41: No license declared.	ontology* Important ⚠️

According to the highest importance level of pitfall found in your ontology the conformace badge suggested is "Critical pitfalls" (see below). You can use the following HTML code to insert the badge within your ontology documentation:



```
<p>
<a href="http://oops.linkeddata.es"></a>
</p>
```

Figure 6.13: Evaluation results for SOCPRI ontology using Pitfall Scanning Tool

multiple domains or ranges axioms. As far as, **atTime** object property is concerned that is being reused in SOCPRI ontology. This object property is part of an ontology design pattern by name of **TimeIndexedSituation** and facilitates temporal role modeling of the users. The fixed version of our ontology contains one case of P19 pitfall that is due to reusing of aforementioned ontology design pattern. Figure 6.14b shows fixed version of the P19 pitfall in our ontology. The ontology scanning tool also identified three im-

Results for P19: Defining multiple domains or ranges in properties. 4 cases | Critical 🚫

The domain or range (or both) of a property (relationships and attributes) is defined by stating more than one `rdfs:domain` or `rdfs:range` statements. In OWL multiple `rdfs:domain` or `rdfs:range` axioms are allowed, but they are interpreted as conjunction, being, therefore, equivalent to the construct `owl:intersectionOf`. This pitfall is related to the common error that appears when defining domains and ranges described in [7].

- This pitfall appears in the following elements:
 - http://www.semanticweb.org/javedahmed/ontologies/2016/4/untitled-ontology-97:relates
 - http://www.ontologydesignpatterns.org/cpi/owl/timeindexedsituation.owl#atTime
 - http://www.semanticweb.org/javedahmed/ontologies/2016/4/untitled-ontology-97:hasTieStrength
 - http://www.semanticweb.org/javedahmed/ontologies/2016/4/untitled-ontology-97:hasTieStrength

(a) Initial cases identified for P19 Pitfall

Results for P19: Defining multiple domains or ranges in properties. 1 case | Critical 🚫

The domain or range (or both) of a property (relationships and attributes) is defined by stating more than one `rdfs:domain` or `rdfs:range` statements. In OWL multiple `rdfs:domain` or `rdfs:range` axioms are allowed, but they are interpreted as conjunction, being, therefore, equivalent to the construct `owl:intersectionOf`. This pitfall is related to the common error that appears when defining domains and ranges described in [7].

- This pitfall appears in the following elements:
 - http://www.ontologydesignpatterns.org/cpi/owl/timeindexedsituation.owl#atTime

(b) Fixed version P19 Pitfall

Figure 6.14: P19 Pitfall Identification and Correction in SOCPRI

portant pitfalls, apart from a single critical pitfall described earlier. The identification numbers of these pitfalls are **P11**, **P30**, and **P41**. The pitfall **P11** deals with an issue related to missing domain or range in the object or datatype properties. There is only one case of this nature that is about **relatesToEntity** object property. The pitfall has been fixed by assigning appropriate domain and range values to the object property. The pitfall **P30** deals with an issue of missing explicit declaration of equivalent classes. There are two cases of this nature about **Relationship** and **Friend** classes. The pitfall has been fixed by providing missing information about disjointness of the classes. **P41** pitfall deals with an issue related to a missing declaration for the license of the ontology. These important pitfalls have been fixed successfully and updated version of SOCPRI contains no such errors.

The ontology scanning tool also identified four minor pitfalls for SOCPRI ontology. The minor pitfalls identified for SOCPRI ontology are **P04**, **P08**, **P13**, and **P22**. As discussed earlier, this category of pitfalls are not a real problem in the ontology, but fixing these pitfalls can make ontology nicer. The pitfall **P04** deals with an issue related to unconnected ontology elements. There is only one case of this nature that is about **Document** class. The problem has been fixed by establishing proper relations of the class with other classes representing various resources in our ontology. **P08** pitfall deals with an issue of missing annotations. The tool pointed out 29 cases of this nature which have been fixed by providing proper annotation for all classes and properties of the ontology. The pitfall **P22** point out the problem of not following standard naming conventions. Our ontology has its own naming convention that is the reason for ignoring this pitfall. The final minor pitfall pointed out by the tool is **P13** which deals with the issue of missing explicit declaration of inverse relationships. We also ignore this pitfall because appropriate inverse relations has been declared explicitly such as **perform** object property has an inverse relationship with **performedWith** object property. All the properties in SOCPRI ontology do not require inverse relation,

therefore, some properties exist in the ontology without an explicit declaration of an inverse relationship. This limitation of the ontology is pointed out by the scanning tool.

The fixed version of SOCPRI ontology is evaluated with scanning tool again. The results of the second phase of evaluation are shown in figure 6.15. The evaluation results point out two minor issues and one case of a critical issue. We already explained the reasons for ignoring multiple domains or ranges values for object property **atTime**. This is not a syntactical issue in OWL language. The two minor issues ignored are related to using a different naming convention for the ontology and missing explicit declaration of inverse relations for the object properties. We also explained that our ontology follows its own naming convention. All the object properties do not require an inverse relation. The evaluation results demonstrate that our ontology is consistent, concise, and complete. The results also evaluate structural, functional, and usability-profiling dimensions of the SOCPRI ontology. In following subsections, we present a pitfall based evaluation of our ontology from different perspectives. The classification of pitfalls to evaluate different ontology perspectives is shown in figure 6.12.

6.4.3.1 Pitfall based Consistency Evaluation

The consistency of SOCPRI ontology can be evaluated with ontology pitfall scanning tool. For this evaluation criteria, we check the list of pitfalls that appear in the evaluation results of our ontology. If the results does not contain **P05**, **P06**, **P07** and **P24** pitfalls, then the ontology is considered to be a consistent. SOCPRI ontology does not contain any such pitfall which is demonstrated by the results shown in figure 6.15. These pitfalls deal with aspects that make the ontology inconsistent. The pitfall **P05** deals with an issue related to defining wrong inverse relationships. The pitfall **P06** is related to an issue of including cycles in a class hierarchy. **P07** pitfall deals with an issue of merging different concepts in the same class. Finally, pitfall **P24** is related to

Scanner by direct input:

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <ENTITY dc "http://purl.org/dc/elements/1.1/" >
  <ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <ENTITY MainModel "https://github.com/shahanijaved/jasonto/blob/master/MainModel" >
  <ENTITY untitled-ontology-97
    "http://www.semanticweb.org/javedahmed/ontologies/2016/4/untitled-ontology-97:" >
]
```

☐ Uncheck this checkbox if you don't want us to keep a copy of your ontology.

Scanner by RDF

Evaluation results


It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:

- **Critical** 🚫 : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important** ⚠️ : Though not critical for ontology function, it is important to correct this type of pitfall.
- **Minor** 🟡 : It is not really a problem, but by correcting it we will make the ontology nicer.

[\[Expand All\]](#) | [\[Collapse All\]](#)

Results for P13: Inverse relationships not explicitly declared.	35 cases Minor 🟡
Results for P19: Defining multiple domains or ranges in properties.	1 case Critical 🚫
Results for P22: Using different naming conventions in the ontology.	ontology* Minor 🟡
SUGGESTION: symmetric or transitive object properties.	1 case

According to the highest importance level of pitfall found in your ontology the conformace bagde suggested is "Critical pitfalls" (see below). You can use the following HTML code to insert the badge within your ontology documentation:



```
<p>
<a href="http://oops.linkeddata.es"></a>
</p>
```

Figure 6.15: Evaluation Results for Fixed Version of SOCPRI

an issue of using recursive definitions. The consistency of ontology can be ensured by avoiding these pitfalls in the ontological structure. SOCPRI ontology is free from these pitfalls that demonstrate consistent nature of the SOCPRI ontology.

6.4.3.2 Pitfall based Completeness Evaluation

The completeness of SOCPRI ontology can be evaluated with ontology pitfall scanning tool. For this evaluation criteria, we check the list of pitfalls that appear in the evaluation results of our ontology. If the results does not contain **P04**, **P10**, **P11** and **P12** pitfalls, then the ontology is considered to be a complete. Our ontology does not contain any such pitfall which is demonstrated by the results shown in figure 6.15.

These pitfalls deal with issues such as creating unconnected ontology elements, missing disjointness, missing domain or range in properties, and missing explicit declaration of equivalent properties. The completeness of the ontology can be ensured by avoiding these pitfalls in the ontological structure. Our ontology is free from these pitfalls that reflect completeness in the ontology.

6.4.3.3 Pitfall based Conciseness Evaluation

The conciseness of SOCPRI ontology can be evaluated with ontology pitfall scanning tool. For this evaluation criteria, we check the list of pitfalls that appear in the evaluation results of our ontology. If the results does not contain **P02**, **P03**, and **P21** pitfalls, then the ontology is considered to be a concise. SOCPRI ontology does not contain any such pitfall which is demonstrated by the evaluation results shown in figure 6.15. The conciseness of the ontology is affected by creating synonyms as classes, using miscellaneous class, and creating the relationship “is” instead of using “rdfs:subClassOf”, “ref:type”, or “owl:sameAs”. Our ontology is free from these pitfalls that demonstrate conciseness in the ontology.

6.4.3.4 Pitfall based Structural Dimension Evaluation

Initially, the structural dimension of ontology was defined by Gangemi [173] which is focused on syntax and formal semantics. A fine-grained classification of this dimension is provided by Poveda et al. [182] which deals with aspects such as modeling decisions, no inference, wrong inference, and ontology language. The pitfall based evaluation of structural dimension can be performed by analyzing existing evaluation results for certain pitfalls. Modeling decisions aspect of SOCPRI can be evaluated by checking ontology for **P02**, **P03**, **P07**, **P21**, **P24**, **P25**, **P26**, and **P33** pitfalls. Ontology free from these pitfalls means ontology developer has implemented it in a correct way following the primitives of implementation language. SOCPRI ontology is free from

the listed pitfalls. It is demonstrated by evaluation results shown in figure 6.15. The aspects of no inference and wrong inference are caused by pitfalls such as **P05**, **P06**, **P27**, **P28**, **P29**, **P31**, **P11**, **P12**, and **P30**. Our ontology does not contain any of these pitfalls. The ontology language aspect is checked by pitfalls **P34**, **P35** and **P38**. SOCPRI ontology is also free from these pitfalls. The evaluation results shown in figure 6.15 demonstrate that SOCPRI has no discrepancy in the structural dimension of the ontology.

6.4.3.5 Pitfall based Functional Dimension Evaluation

The functional dimension of ontology focuses on the conceptualization and intended use of the ontology. Initially, this dimension was defined by Gangemi [173] and fine-grained classification of this dimension was provided by Poveda et al. [182]. According to the authors, functional dimension deals with aspects such as real world modeling, requirement completeness, and application context. Ontology scanning tool supports fine-grained evaluation of functional dimension. Real world modeling aspect of functional dimension can be evaluated by checking the given ontology against pitfalls dealing with the issue of missing disjointness (**P10**) and creating unconnected classes (**P04**). SOCPRI ontology does not contain these pitfalls. It is demonstrated in the ontology evaluation results shown in figure 6.15. Requirement completeness aspect of the functional dimension is evaluated by checking the given ontology against pitfalls that deals with issues related to missing domain information (**P09**) and creating unconnected ontology elements (**P04**). As per the evaluation results, our ontology does not contain any such pitfalls. Finally, application context aspect of the functional dimension can be evaluated by checking the given ontology against pitfalls **P36**, **P37**, **P38**, **P39**, and **P40**. These pitfalls deal with issues related to namespace hijacking, ambiguous namespace, no OWL ontology declaration, etc. SOCPRI ontology is also free from these pitfalls. It is shown in the figure 6.15 that contains information about evaluation

results. We conclude this discussion by explicitly arguing that our ontology has no issues related to the functional dimension of the ontology.

6.4.3.6 Pitfall based Usability-Profiling Dimension Evaluation

The usability-profiling dimension refers to the communication context of an ontology. Initially, this dimension was defined by Gangemi [173] and fine-grained classification of this dimension was provided by Poveda et al. [182]. For this dimension Poveda et al. contemplated aspects such as ontology understanding, ontology meta-data, and ontology clarity. The evaluation of usability-profiling dimension is supported by ontology scanning tool. To evaluate ontology understanding and metadata aspects of usability-profiling dimension, we can check the given ontology against pitfalls related to missing annotations, no OWL ontology declaration, no license declaration, etc. As per the evaluation results presented in figure 6.15, SOCPRI ontology does not contain any such pitfalls. Another aspect related to usability-profiling dimension is ontology understanding that can be evaluated by checking the given ontology against pitfalls **P02**, **P07**, **P08**, **P11**, **P12**, **20**, **32**, and **37**. These pitfalls deal with issues related to creating synonyms as classes, merging different concepts in the same class, missing annotations, missing domain or range in properties, missing ontology annotations, etc. SOCPRI ontology is also free from these pitfalls that are demonstrated in the figure 6.15 containing information about evaluation results.

In the light of the aforementioned discussion, we conclude that SOCPRI ontology is consistent, concise, and complete to the larger extent. The evaluation of structural dimension reveals that syntax and formal semantics our ontology is well structured. The evaluation of various aspects of functional dimension also does not highlight any major error in the ontology. Finally, usability-profiling dimension evaluation shows that ontology is properly documented that can help users to understand the ontology. SOCPRI ontology does not contain any errors that appear in most of the ontologies by

following bad practices.

6.5 Evaluation of SOCPRI Using OntoClean

The OntoClean is a well-known methodology for ontology evaluation [183]. It provides ontological metrics that evaluate modeling choices in ontologies from a philosophical stand-point. The methodology defines a set of general and well-formalized meta-properties to facilitate the construction of clean ontologies. These meta-properties are based on philosophical notions such as rigidity, identity, unity, and dependency. These meta-properties impose a set of constraints on the taxonomic relations of ontology, which help to detect possible disagreements amongst different conceptualizations so that some corrective action can be taken. The formal notions of OntoClean are general enough to be used in any ontology effort, independently of a particular domain.

The OntoClean core is based on attaching to each concept in ontology suitable meta-properties that describe its behavior with respect to the ontological notions. One of the benefits of this analysis is the discovery of inconsistent uses of subsumption in the taxonomy. Deciding whether one property should subsume another is one of the important ontological decisions a modeler must make in building ontology, and providing a formal foundation for evaluating these decisions has proved important milestones in the practice of conceptual modeling.

6.5.1 OntoClean Meta Properties

It is important to mention that property in OntoClean vocabulary is commonly used to refer concept or class in OWL. Meta-properties are therefore properties of properties. The main ingredients of OntoClean methodology are four meta-properties and a set of constrained imposed by these properties. The four properties are rigidity, identity, unity and dependence. In this section, we offer a brief description of these meta-properties.

Rigidity: This meta-property is a special form of essence, which means that the property of an entity is essential to that entity and must be true in every possible world. The instance of a rigid concept cannot cease to be an instance of that concept unless it ceases to exist. If and only if a concept is essential to all of its instances, the concept is called rigid and tagged with +R. If and only if it is essential to some instances, it is called non-rigid, tagged with -R. An anti-rigid concept is one that is not essential to all of its instances. It is tagged $\sim R$. An example of the rigid concept is a person. Every entity that is a person must be a person in all possible worlds. Red is an example of the non-rigid concept, whereas, the student is an anti-rigid concept. Any student cannot be so in all possible worlds. Rigidity and its variants are important meta-properties. Every concept in an ontology should be labeled as rigid, non-rigid, or anti-rigid.

Identity: It is an important philosophical notion used in OntoClean methodology. The notion of identity concerns with the problem of being able to recognize individual entities in the world as being the same or different. The notion of identity uses identity criterion (IC) to distinguish among various instances. An IC is identifying characteristic to recognize individual entity. It must be informative such as DNA profile as IC is informative. It cannot be trivial such as being red as IC is a trivial assumption. The property may inherit IC from parents or supply its own additional identity criteria. The OntoClean methodology differentiates between these two and assigns different tags. The tagging O means concept supplies its own IC, whereas tagging I mean concept inherit IC from parents and only carries that IC. The concept is tagged +O if and only if it supplies its own IC, which is not inherited from the subsuming concepts and with -O otherwise. If the concept carries IC from parents and does not own its additional IC is tagged as +I, whereas those that do not carry IC and also do not supply their own IC are tagged with -I.

Unity: This meta-property is related to the problem of mereology that deals with describing the parts and boundaries of objects, such that what is part of the object, what is not, and under what conditions the object is whole. The unity criterion (UC) provides answers to these questions and describes the conditions that most hold among the parts of a certain entity to consider that entity as a whole. The concept that carries UC is tagged with +U and with -U otherwise. A concept carries a UC if and only if there exists a single relation r such that each instance of the concept is necessarily a whole under r . The anti-unity concept is demonstrated in a situation when every instance of the concept is not necessarily a whole and tagged with $\sim U$.

Dependence: This meta-property deals with the notion of the constitution, which concerns the identification of the substance of which an entity is made. In general, dependency deals with distinguishing between intrinsic and extrinsic concepts. Intrinsic concepts are independent, whereas extrinsic concepts need to be given to an instance by circumstances or definitions. A concept $C1$ is dependent on a concept $C2$ if and only if for every instance of $C1$ an instance of $C2$ must exist and tagged with +D and with -D otherwise. An example for a dependent concept would be FOOD, as instances of FOOD can only exist if there is something for which these instances are food. This does not mean that an entity being food ceases to exist the moment all animals die out, but it just stops being regarded as food.

6.5.2 OntoClean Constraints

The methodology imposes a set of restrictions on subsumption relations that should be taken into account while evaluating given ontology using OntoClean process. The list of most common restriction are given below:

Table 6.1: OntoClean Meta-Properties Summary

Meta-Property	Symbol	ImPLY	Label	Definition
Rigidity	+R		Rigid	Concept is essential to all instances of the class
	-R		Non-Rigid	Concept is not essential to some instance of the class
	\sim R	-R	Anti-Rigid	Concept is not essential to all instances of the class
Identity	+I		Carry IC	Concept carries an identity condition
	-I		Non-Carry IC	Concept that does not carry an identity condition
	+O	+R +I	Supply IC	Instances provide an identity condition themselves
	-O		Non-Supply IC	Instances do not provide an identity condition themselves
	+U		Unity	Instances are whole and have a single unity condition
Unity	-U		Non-Unity	Instances are whole and have no single unity condition
	\sim U		Anti-Unity	Instances are not whole
	+D		External-Dependence	Dependence is on external concept
Dependence	-D		Non-External Dependence	Dependence is not on external concept

1. Anti-rigid concept cannot subsume a rigid concept (R cannot subsume +R)
2. Identity carrying concept cannot subsume a non-identity (+I cannot subsume -I)
3. Unity concept cannot subsume a non-unity (+U cannot subsume -U)
4. Anti-unity concept cannot subsume a unity (\sim U cannot subsume +U)
5. Dependent concept cannot subsume a non-dependent (+D cannot t subsume -D)

6.5.3 OntoClean Process

In this section, we provide a brief overview of OntoClean process and outline the steps required to apply the methodology. The application of this methodology consists of following two main steps.

- The identification of core meta-properties for every single concept of the ontology that is being evaluated using OntoClean. Thus, every ontological concept is assigned a certain tag such as +R -I \sim U +D.
- All the super and subclass relations of the ontology are checked against the predefined constraints. Any violation of a constraint indicates a potential mis-conceptualization in the subsumption hierarchy.

The application of OntoClean compares the taxonomical part of a tagged ontology versus a predefined ideal taxonomic structure, which is defined by a combination of meta-properties and constraints. After completion of the OntoClean process, the ontology designer has a resultant tagged ontology and a list of constraint violations. The designer may repair the ontology to address the content violation issues. The main remedial action need to repair the ontology is redefining the subsumption hierarchy of the ontology under review.

6.5.4 Applying OntoClean to SOCPRI Ontology

In this section, we identify meta-properties for every single concept of the SOCPRI ontology. The tables 6.2, 6.3, and 6.4 present results of OntoClean meta-properties application to our ontology. After identification of core meta-properties for all classes of SOCPRI ontology, we evaluate all the super and subclass relations of our ontology against the predefined constraints of OntoClean methodology. Any violation of the constraints indicates a potential mis-conceptualization in the subsumption hierarchy of our ontology.

It should be noted that choosing OntoClean meta-properties for concepts in SOCPRI ontology depends on the definition of concept in our ontology. The variation in conceptual definition may result a different choice of OntoClean meta-properties. The main constraints violation found in SOCPRI ontology is in taxonomic decision about Agent and its subclasses. According to philosophical notions of OntoClean, an Agent is anti-rigid concept, which carries no identity criterion, whereas person and group concepts are rigid and supplies identity criterion. It is inconsistent to subsume user and group concepts from an agent concept. Our position in taxonomic decision of subsuming user and group from an agent concept is motivated from well-founded FOAF ontology. We reuse the concepts of agent and group from this ontology and this ontology subsume group from an agent concept.

Table 6.2: Applying Meta-Properties to SOCPRI Ontology-I

SOCPRI Concept	Rigidity	Identity	Unity	Dependence
User	+R	+O	+U	-D
Relationship	+R	+O	+U	+D
Role	\sim R	-I	-U	+D
RoleInTime	\sim R	-I	-U	+D
DigitalResource	+R	+O	+U	-D
Interaction	+R	+I	+U	+D
Agent	+R	+O	+U	-D
Group	+R	+O	+U	-D
UserCentricRole	\sim R	-I	-U	+D
ResourceCentricRole	\sim R	-I	-U	+D
FamilyRole	\sim R	-I	-U	+D
SocialRole	\sim R	-I	-U	+D
WorkRole	\sim R	-I	-U	+D
ResourceRequestor	\sim R	-I	-U	+D
ResourceResponder	\sim R	-I	-U	+D
U2RRelationship	-R	+I	+U	+D
U2URelationship	+R	+O	+U	+D
R2RRelationship	+R	+O	+U	+D
Contributor	-R	+I	+U	+D
Disseminator	-R	I	+U	+D
Stakeholder	-R	+I	+U	+D
Owner	-R	+I	+U	+D
Employment	+R	+O	+U	+D
Collaborator	\sim R	+I	+U	+D
Colleague	\sim R	+I	+U	+D
Employee	\sim R	+I	+U	+D
Employer	\sim R	+I	+U	+D
Friendship	+R	+O	+U	+D
Friend	-R	+I	+U	+D
CloseFriend	\sim R	+I	+U	+D
Acquaintance	\sim R	+I	+U	+D
Kinship	+R	+O	+U	+D
Affiance	-R	+I	+U	+D

Table 6.3: Applying Meta-Properties to SOCPRI Ontology-II

SOCPRI Concept	Rigidity	Identity	Unity	Dependence
Context	+R	+O	+U	-D
TieStrength	+R	+O	\sim U	+D
TieStrengthDimension	+R	+O	\sim U	+D
PredictiveIndicator	+R	+I	+U	-D
PredictiveIndicatorType	+R	+I	+U	-D
HomophilyIndicator	+R	+I	+U	-D
InteractionIndicator	+R	+I	+U	-D
InteractionType	+R	+I	+U	+D
Closeness	-R	+O	\sim U	+D
EmotionalSupport	-R	+O	\sim U	+D
InteractionFrequency	-R	+I	\sim U	+D
ReciprocalService	-R	+O	\sim U	+D
RelationshipDuration	-R	+I	\sim U	+D
SocialDistance	-R	+O	\sim U	+D
StructuralDimension	-R	+O	\sim U	+D
TimeInterval	+R	+O	+U	-D
Siblings	+R	+I	+U	+D
Spouse	+R	+I	+U	+D
Commenting	+R	+I	+U	+D
Liking	+R	+I	+U	+D
Messaging	+R	+I	+U	+D
MessageReplying	+R	+I	+U	+D
MessageSending	+R	+I	+U	+D
Posting	+R	+I	+U	+D
AlbumPosting	+R	+I	+U	+D
LocationUpdating	+R	+I	+U	+D
PhotoPosting	+R	+I	+U	+D
StatusUpdating	+R	+I	+U	+D
VideoPosting	+R	+I	+U	+D
Sharing	+R	+I	+U	+D
Tagging	+R	+I	+U	+D
InformationFlowActivity	+R	+O	+U	-D
Document	+R	+O	+U	-D

Table 6.4: Applying Meta-Properties to SOCPRI Ontology-III

SOCPRI Concept	Rigidity	Identity	Unity	Dependence
Profile	-R	+O	-U	-D
ProfileSubset	-R	+I	-U	+D
DefaultProfile	-R	+I	-U	+D
ContextualProfile	-R	+I	-U	+D
FamilyProfile	-R	+I	-U	+D
SocialProfile	-R	+I	-U	+D
WorkProfile	-R	+I	-U	+D
ContextualNorm	-R	-I	-U	+D
PrivacyCondition	-R	-I	-U	+D
AppropriatenessCondition	-R	-I	-U	+D
DistributionCondition	-R	-I	-U	+D
AccessRequest	+R	+O	\sim U	+D
AccessResponse	+R	+O	\sim U	-D
Deny	+R	+I	\sim U	+D
Grant	+R	+I	\sim U	+D
Friendlist	+R	+O	-U	-D
Photo	+R	+O	+U	-D
Video	+R	+O	+U	-D
FamilyContext	-R	+I	+U	+D
SocialContext	-R	+I	+U	+D
WorkContext	-R	+I	+U	+D
Organization	+R	+O	+U	-D
Text	+R	+O	+U	-D
StrongTie	+R	+O	\sim U	+D
MediumTie	+R	+O	\sim U	+D
WeakTie	+R	+O	\sim U	+D
Child	+R	+I	+U	-D
Grandchild	+R	+I	+U	-D
Parent	+R	+I	+U	+D
Grandparent	+R	+I	+U	+D
MandatoryProfileData	-R	-I	+U	-D
ContextSensitiveResource	+R	+I	+U	+D
ContextFreeResource	+R	+I	+U	+D
DemographicData	-R	-I	+U	-D

6.6 Query Based Evaluation of SOCPRI Ontology

In this section, we evaluate our ontological model by retrieving knowledge from the ontology with help of DL queries. We have created 75 individuals for SOCPRI ontology in order to check the consistency between T-Boxes and A-Boxes of the ontology. We have enriched our ontology with individuals and object property assertions before performing any DL query. We added 20 individuals of the **User** class. The **Group** and **Organization** classes have 8 individuals. We have added 25 individuals related to different subclasses of **Role** class. We also created 17 different contexts for the users by instantiating subclasses of **Context** class. The figure 6.16 shows an overview of the instances created in SOCPRI ontology.

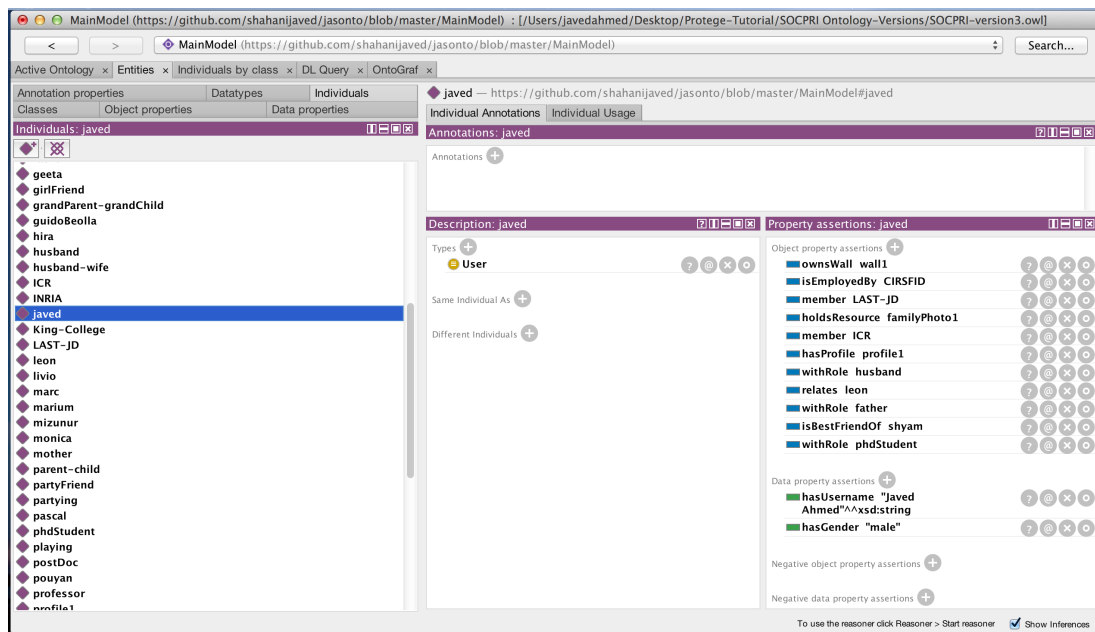


Figure 6.16: Overview of various Individuals of SOCPRI Ontology

We have added more than 114 object property assertions for the individuals. We also added 30 data property assertions for these individuals. The OWL allows ontology designers to create individuals and to assert object and datatype properties about them. The object property assertions establish relations between individuals. The object

and datatype properties used to add assertions in our ontology are *relates*, *member*, *holdsResource*, *hasProfile*, *ownsWall*, *withRole*, *isEmployedBy*, *hasGender*, etc. From the only limited set of assertions added to our ontology, we can infer many facts about the individuals and their relationships. This can be demonstrated by the results of the queries discussed below.

We can use DL query plugin to query SOCPRI ontology. The DL query plugin provides a powerful and easy to use features for retrieving knowledge from a classified ontology. The queries are only executed on a classified (inferred) ontology, therefore, it is necessary to select and start a reasoner before executing a query on SOCPRI ontology. We used Fact++ reasoner along with DL query plugin to retrieve the related instances or classes. The Manchester OWL syntax is supported query language used by the DL query plugin. This syntax is user-friendly for OWL DL and we used it to represent the modeling of various classes in the previous chapter of the dissertation. This is main motivation to use DL query plugin for testing our ontology instead of using SPARQL plugin. We describe some examples of queries answered by the SOCPRI ontology in rest of this section.

Figure 6.17 shows results for the first set of queries about the relationship between individuals with certain characteristics. The main object property linking various individuals is *relates*. The first query asks list of all individuals related to a certain individual. The competency question addressed by this query is *List all the individuals related to a certain individual termed as javed*. The result of this query is shown in figure 6.17(a). There are five individuals in total related to the individual called javed. The Manchester OWL syntax of the query is given below:

```
User and (relates value javed)
```

We add another object property to the query string to make it more specific. The object property added to the second query is *withRole*. We develop a query to retrieve all the individuals related to a certain individual with role supervisor. The competency

question addressed by this query is *Who are the supervisors of an individual called javed?* The result of this query is shown in figure 6.17(b). There are two individuals out of five related individuals (as seen in the previous query results) that supervise the individual called javed. The Manchester OWL syntax of the query is given below:

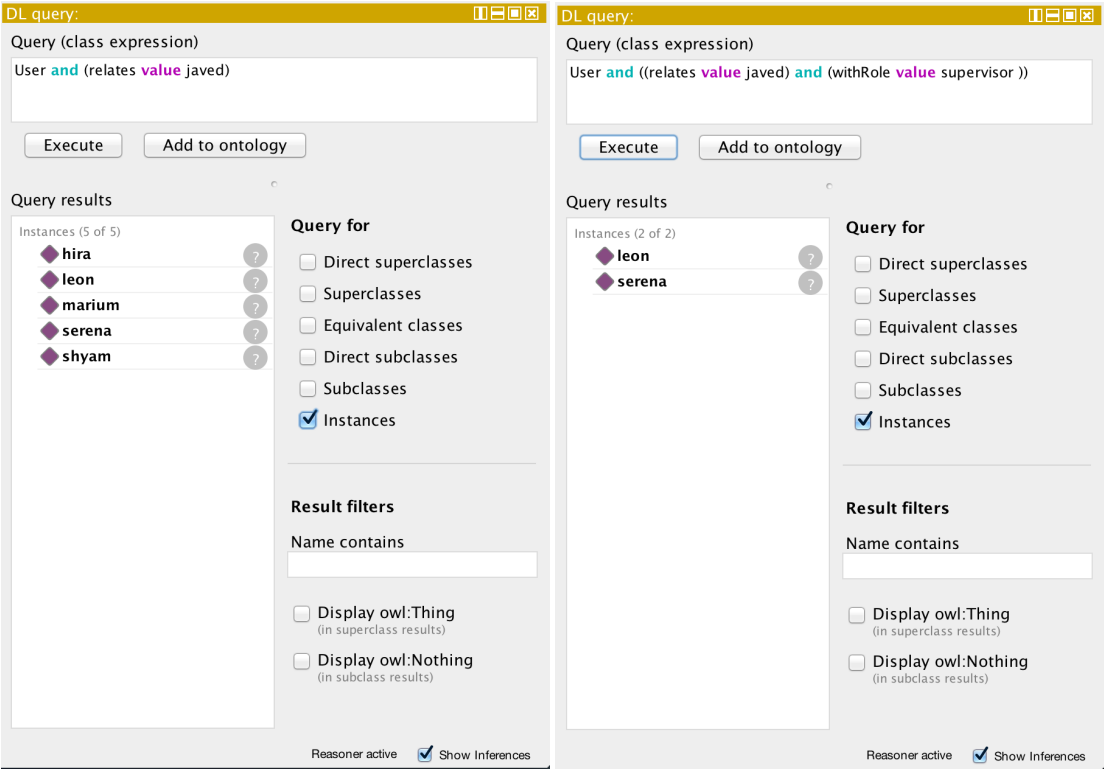
```
User and ((relates value javed)
           and (withRole value supervisor))
```

We add another datatype property to the query string to make it even more specific. The datatype property added to the third query is *hasGender*. We develop a query to retrieve all the individuals related to a certain individual as a supervisor and belong to the female gender. The competency question addressed by this query is *Who is the female supervisor of individual javed?* The result of this query is shown in figure 6.17(c). There is only one individual out of five related individuals (as shown in the previous results of the queries) that is female and supervise the individual called javed. The manchester OWL syntax of the query is given below:

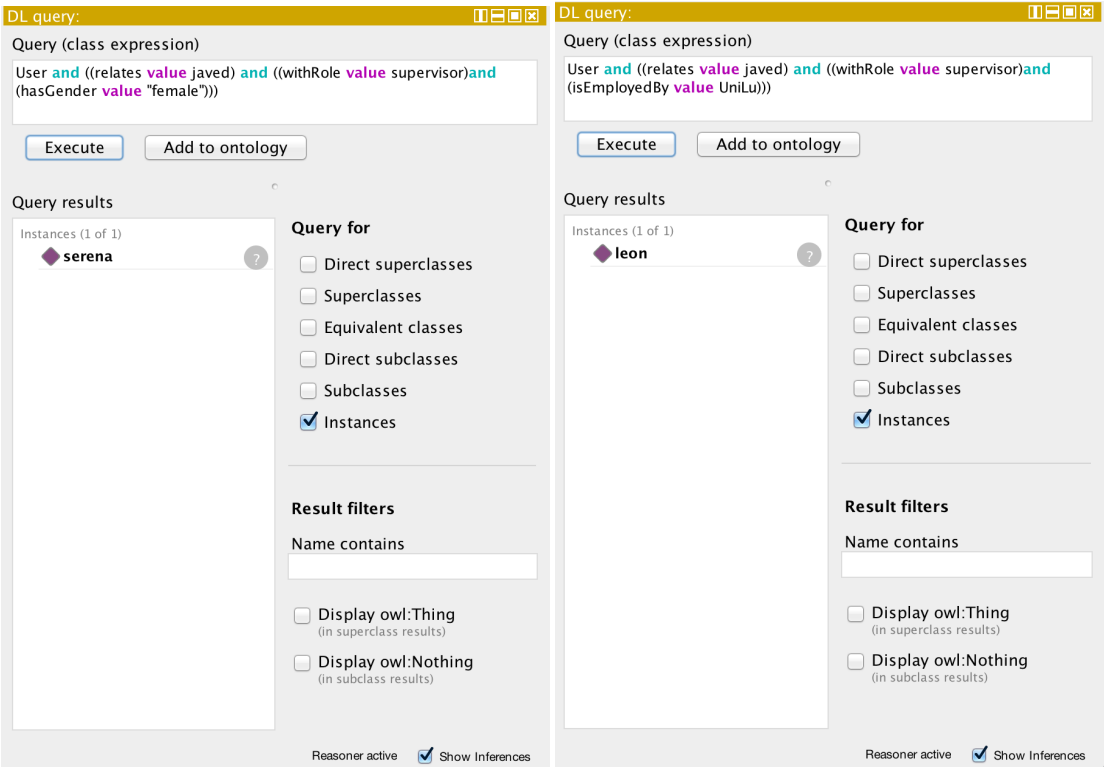
```
User and (((relates value javed)
            and (withRole value supervisor)
            and (hasGender value "female"))))
```

Final query of this set is demonstrating the usage of *isEmployedBy* object property. We design a query string that ask the model about all individuals related to the individual called javed with role supervisor and are employed by the organization known as UniLu. The competency question addressed by this query is *Who is the supervisor of individual javed from UniLu organization?* The result of this query is shown in figure 6.17(d). There is also only one individual out of five related individuals that supervisor javed and work at UniLu. The Manchester OWL syntax of the query is given below:

```
User and (((relates value javed)
            and (withRole value supervisor)
```

(a) Retrieving users related to a certain individual (b) Retrieving supervisors of a certain individual



(c) Retrieving female supervisor of a certain individual (d) Retrieving supervisor from UniLu of a certain individual

Figure 6.17: Queries to retrieve users related with a certain individual

and (isEmployedBy value UniLu)))

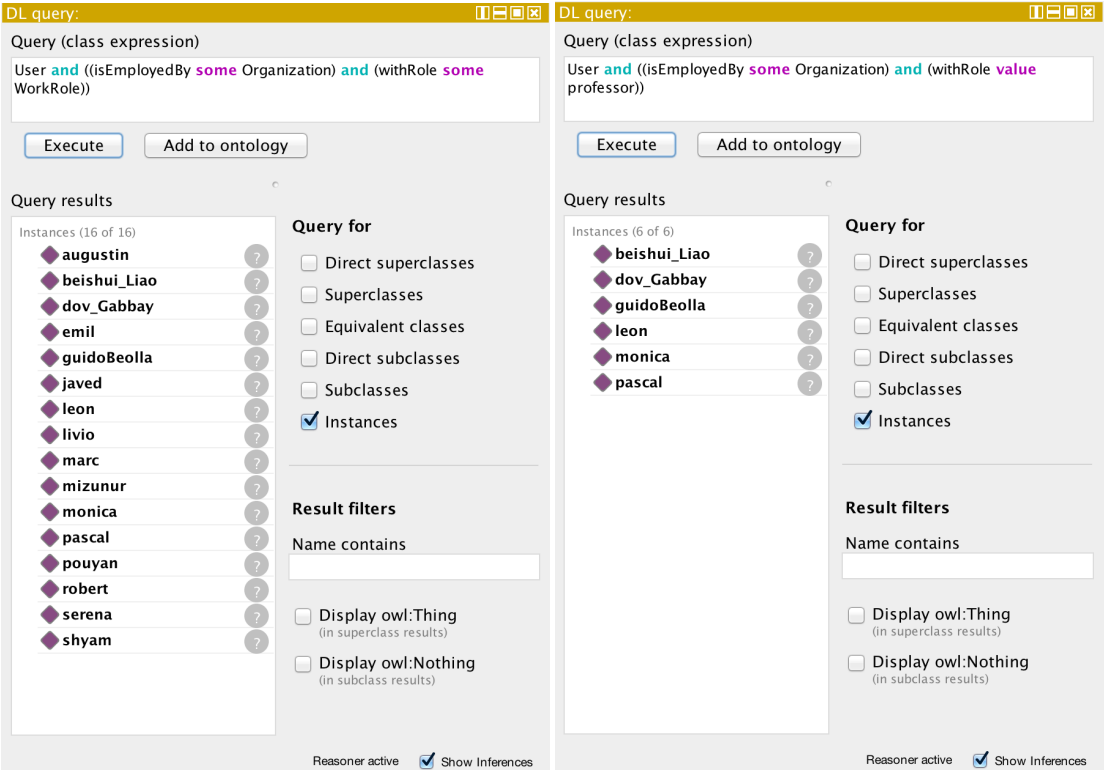
Figure 6.18 shows the results of the second set of the queries about work related roles of the users in different organizations and their membership to various research groups. The key object properties linking several individuals with different organizations and research groups are *isEmployedBy* and *member*. The queries in this set make use of these object properties and their different values. The first query string extracts all the individuals working in some organization and have some work related role. The competency question addressed by this query is *list all the employed individuals with some work related roles*. The result of this query is shown in figure 6.18(a). There are 16 individuals in total that are employed by some organization and perform some work related role. The manchester OWL syntax of the query is given below:

User and ((isEmployedBy some Organization)
and (withRole some WorkRole))

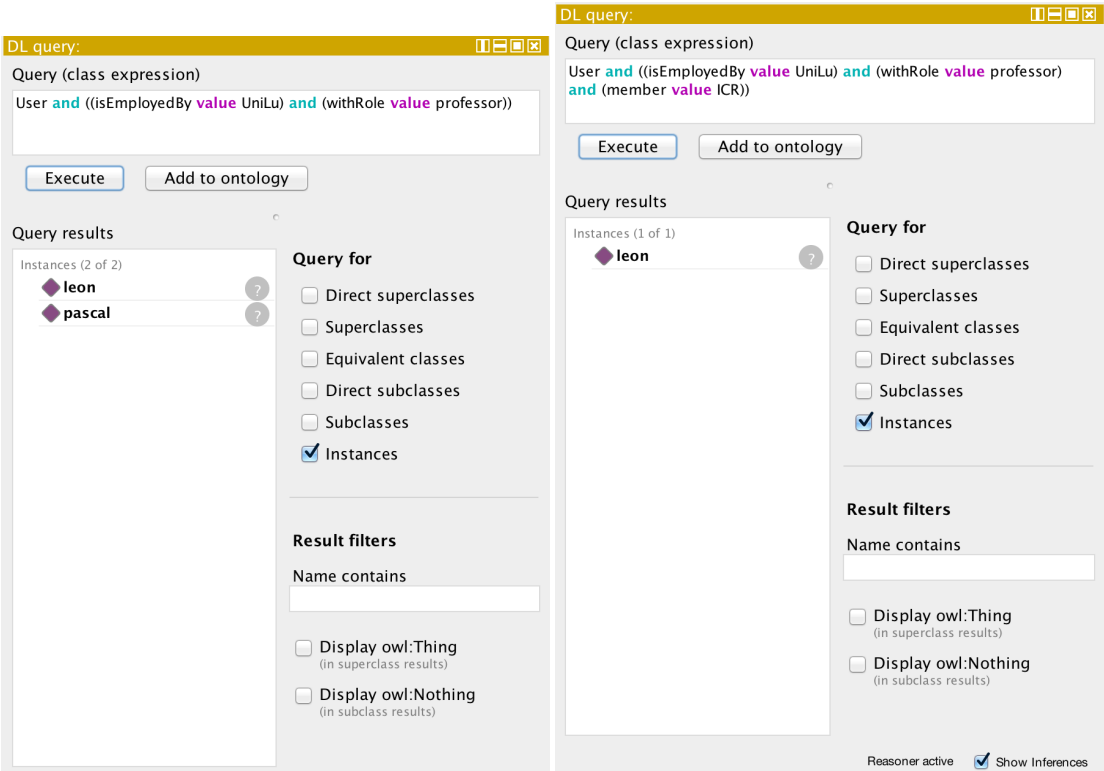
We design more specific query string which asks the ontological model about only those individuals that work as a professor at any organization. The query makes use of certain individual of *Role* class instead of referring to the all individuals of *Role* class. The competency question addressed by this query is *which of the employed individuals are working as professor?* The result of this query retrieves 6 individuals that work as a professor at any organization. Figure 6.18(b) shows the result for the query. The Manchester OWL syntax for the query is given below:

User and ((isEmployedBy some Organization)
and (withRole value professor))

The next query is even more specific that ask for individuals working at UniLu as a professor. This query uses individuals of *Role* and *Organization* classes. These individuals are given as a value to the object properties *isEmployedBy* and *withRole*. The competency question addressed by this query is *which of the professors are employees*



(a) Retrieving individuals with different work roles (b) Retrieving individuals with professor work role



(c) Retrieving professors employed by UniLu (d) Retrieving UniLu professor with ICR membership

Figure 6.18: Queries to retrieve working individuals with different criteria

of *UniLu*? The result of this query are shown in figure 6.18(c). These two individuals in the model that belongs to UniLu and work as a professor. The Manchester OWL syntax for the query is given below:

```
User and ((isEmployedBy value UniLu)
           and (withRole value professor))
```

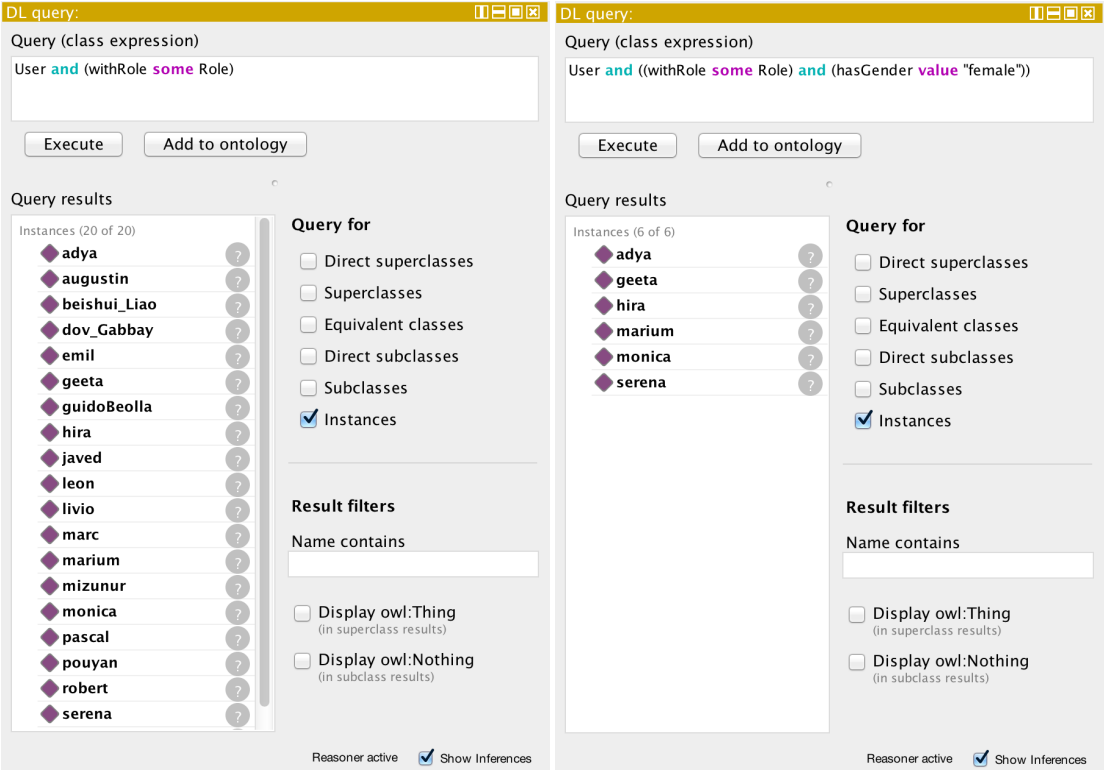
The final query of this set uses another object property called *member*. This object property links individuals with the certain research group. The query retrieves an individual that is working at UniLu as a professor and is a member of ICR research group. The competency question addressed by this query is *which of the UniLu professors are member of ICR research group?* There is only one individual that is a member of ICR research group and work at UniLu as a professor. The results of the query are shown in figure 6.18(d). The Manchester OWL syntax for the query is given below:

```
User and (((isEmployedBy value UniLu)
            and (withRole value professor)
            and (member value ICR)))
```

Figure 6.19 shows the result of the third set of queries about users' roles and their association with various organization and groups. The first query of this set retrieves all the individuals that perform any social, work, or family role. The competency question addressed by this query is *list all the individuals that perform some (social, work, or family) Role*. The result of this query is shown in figure 6.19(a). We have created in total 20 individuals and all of them perform different social, work, and family roles. The result retrieves all these individuals as the answer to the query. The Manchester OWL syntax for the query is given below:

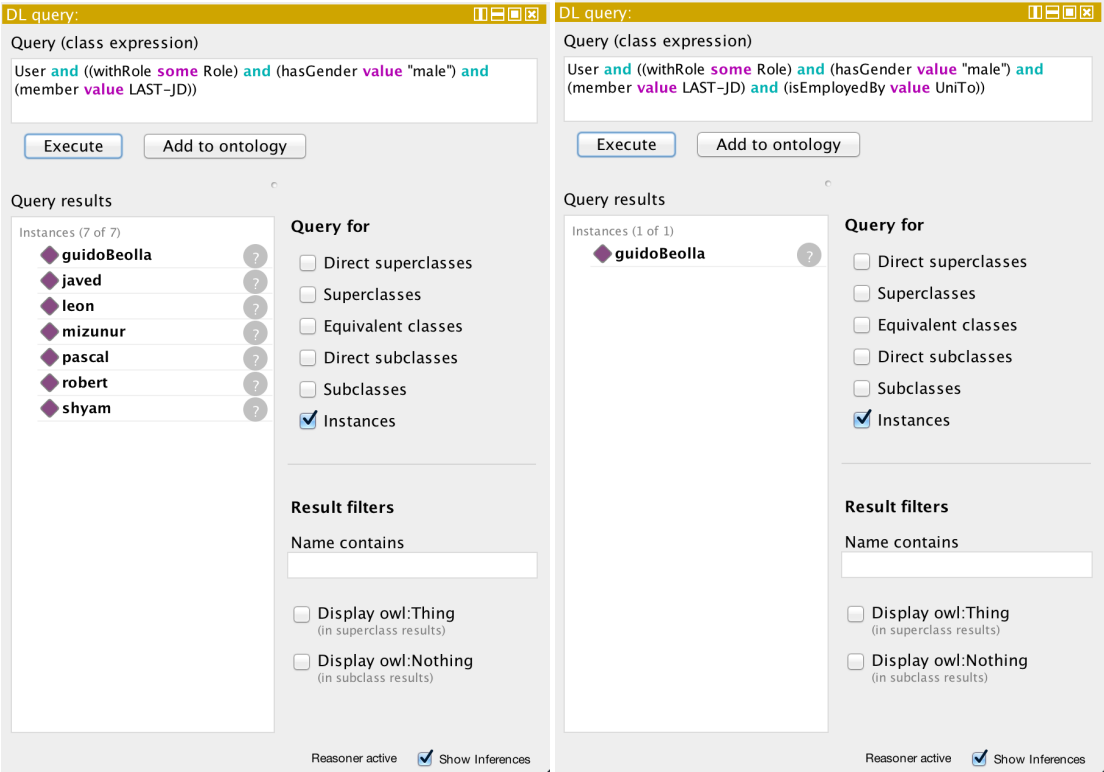
```
User and (withRole some Role)
```

The second query extract more specific results by adding *hasGender* datatype property to the query string. The competency question addressed by this query is *List all*



(a) Retrieving all individuals with any role

(b) Retrieving female individuals with any role



(c) Retrieving male LAST-JD group members with any role

(d) Retrieving LAST-JD group members working at UniTo

Figure 6.19: Queries to retrieve users associated with certain group and organization

the individuals that perform some Role and belong to the female gender. The result of the query retrieves 6 individuals that perform various roles and belong to the female gender. Figure 6.19(b) shows the results. The Manchester OWL syntax for the query is given below:

```
User and ((withRole some Role)
          and (hasGender value "female"))
```

The third query of this set extracts even more specific information by adding *member* object property. The competency question addressed by this query is *who are the male members of LAST-JD research group and also perform some associated work role?* The result of the query is shown in figure 6.19(c). The output of the query shows 7 different male individuals that belong to LAST-JD research group. The manchester OWL syntax for the query is given below:

```
User and (((withRole some Role)
            and (hasGender value "male")
            and (member value LAST-JD)))
```

Final query of this set uses three object properties and one datatype property and their individual values to retrieve the very specific individual. The competency question addressed by this query is *which male member of LAST-JD research group is working at UniTo?* The result of this query retrieve only one individual out of the 20 different individuals (as discussed earlier). Figure 6.19(d) shows the result of this query. The Manchester OWL syntax for the query is given below:

```
User and (((withRole some Role)
            and (hasGender value "male")
            and (member value LAST-JD)
            and (isEmployedBy value UniTo)))
```

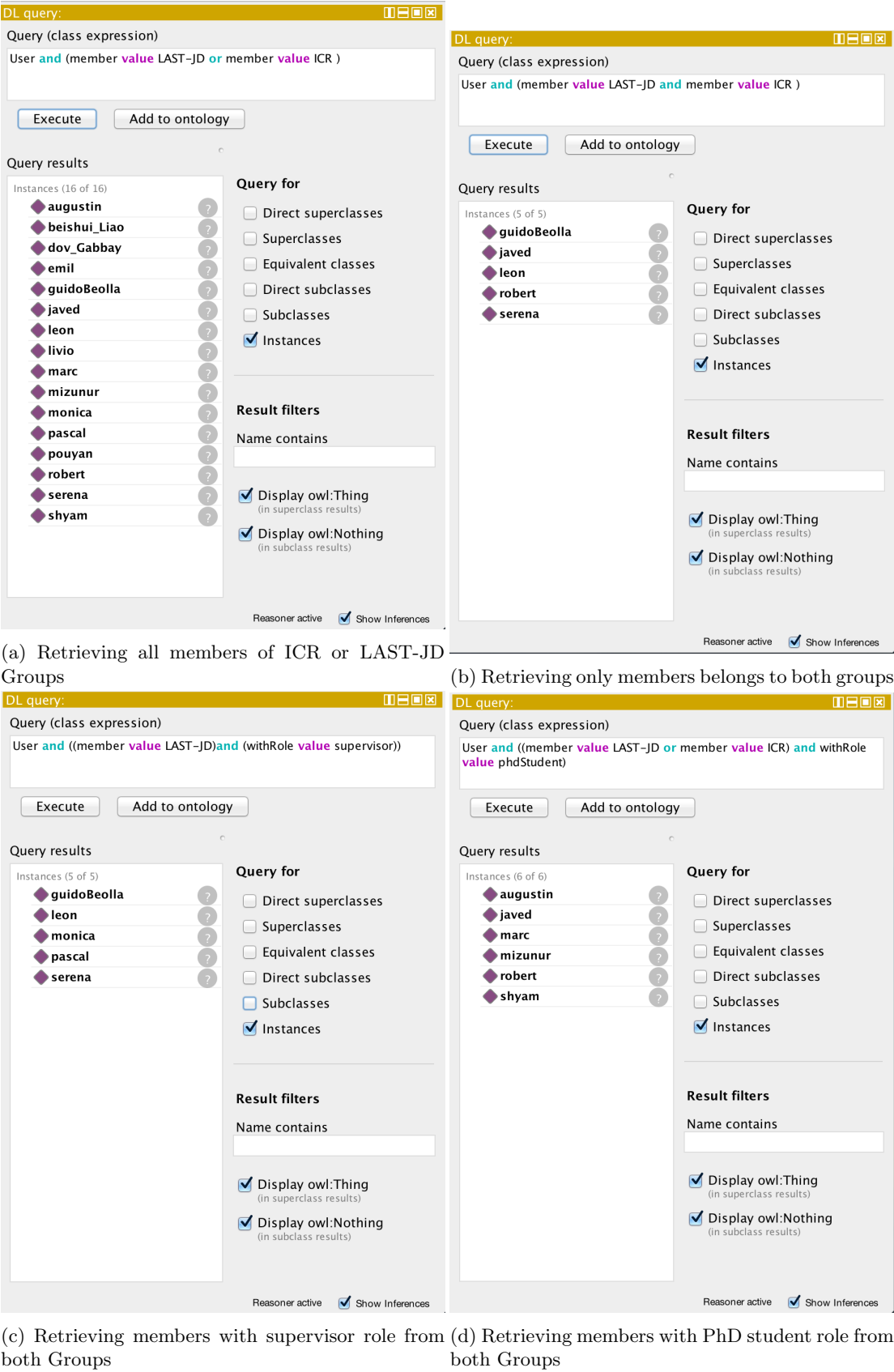


Figure 6.20: Queries to retrieve members of different groups with certain roles

Figure 6.20 presents the results of the queries retrieving individuals that are either PhD student or supervisor and belong to either LAST-JD or ICR research group. The first query string retrieves all the individuals that belong to either LAST-JD or ICR research group. The competency question addressed by this query is: *List all the individuals that are member of either LAST-JD or ICR group.* The result of this query is shown in figure 6.20(a). The total number of individuals that belong to either LAST-JD or ICR is 16. The Manchester OWL syntax for the query is given below:

```
User and ((member value LAST-JD)
           or (member value ICR))
```

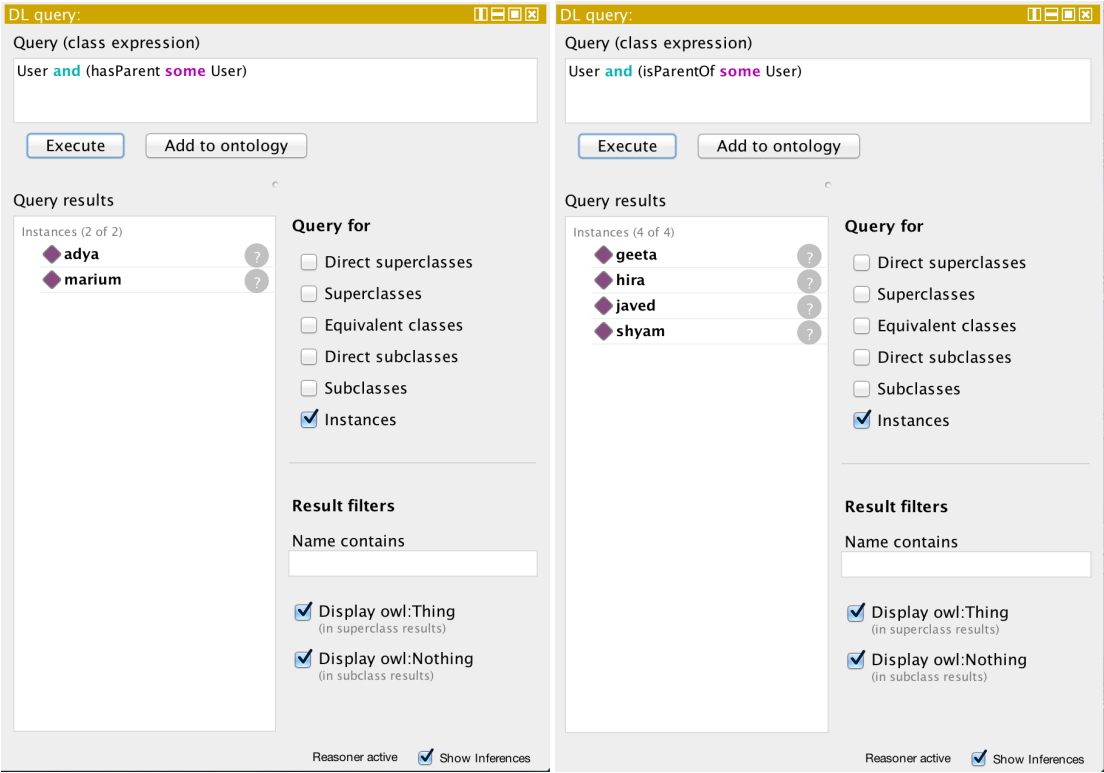
The second query of this set is retrieving the only individual that belongs to both research groups. The competency question addressed by this query is: *List all the individuals that are member of both LAST-JD and ICR group.* The result shows only 5 individuals that are member of both the research groups. Figure 6.20(b) presents result of the query. The Manchester OWL syntax for the query is given below:

```
User and ((member value LAST-JD)
           and (member value ICR))
```

The third query of this set focuses on the individuals that belong to LAST-JD research group and supervise some PhD students. The competency question addressed by this query is: *Which individuals are member of LAST-JD group and supervise PhD Students?* The result of the query is shown in figure 6.20(c). The total individuals from LAST-JD research group that supervises PhD students are 5. The Manchester OWL syntax for the query is given below:

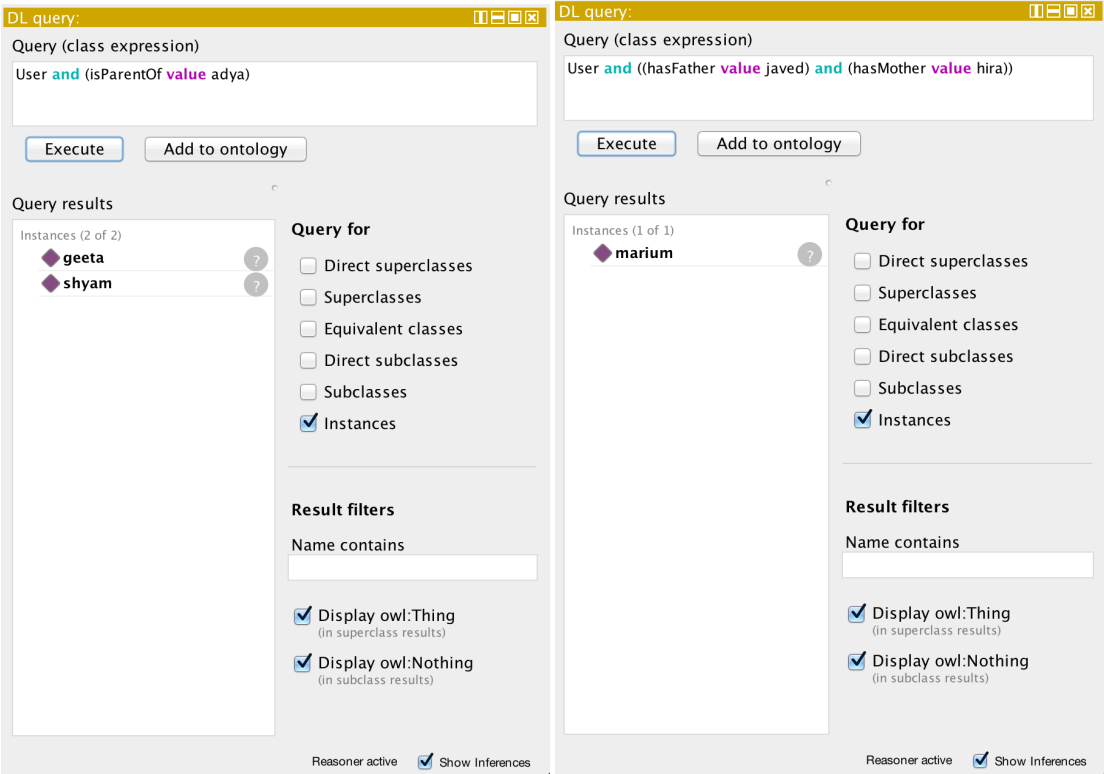
```
User and ((member value LAST-JD)
           and (withRole value supervisor))
```

Final query of this set retrieves all the PhD student that work in one of these research groups. The competency question addressed by this query is: *Which individuals*



(a) Retrieving individuals with parents

(b) Retrieving individuals with children



(c) Retrieving parents of a certain individual

(d) Retrieving child of certain individuals

Figure 6.21: Queries to retrieve parentage facts

are PhD students and member of either LAST-JD or ICR group? The result of the query is shown in figure 6.20(d). The total PhD students working in one of these research group are 6. The Manchester OWL syntax for the query is given below:

```
User and (((member value LAST-JD)
           or (member value ICR))
          and (withRole value PhDStudent))
```

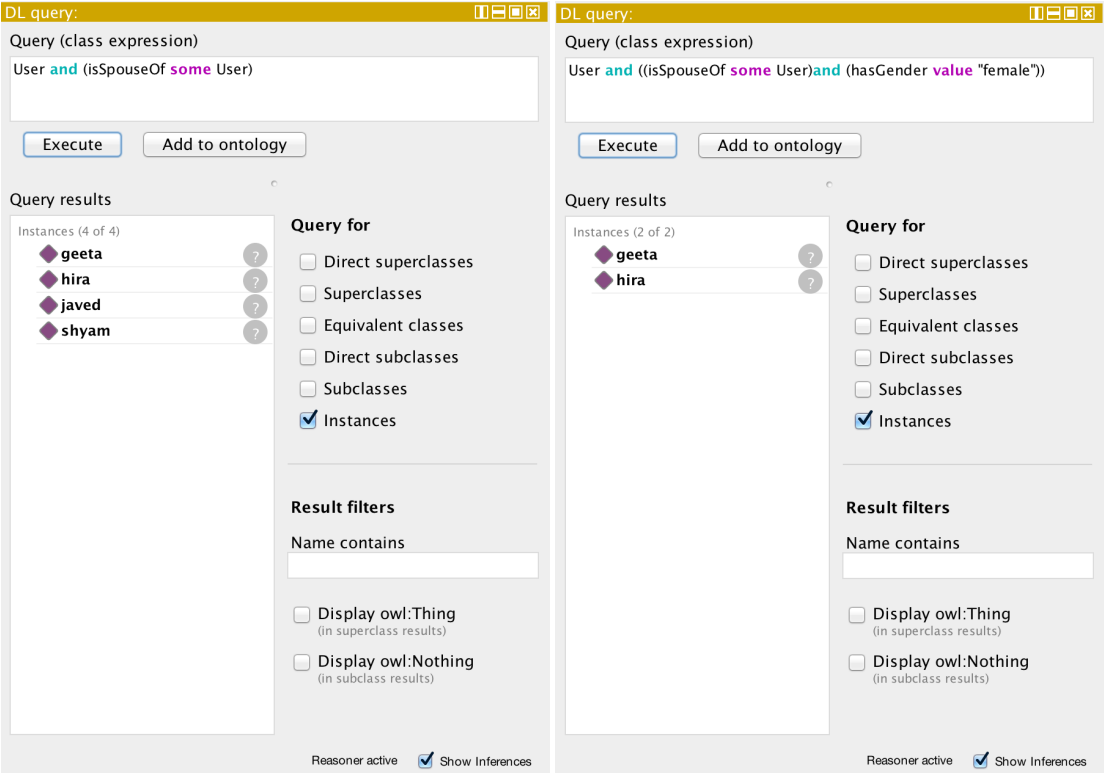
Figure 6.21 presents results of the set queries extracting parentage facts from A-Box of our model. The object properties used in the query strings are *hasParent*, *isParentOf*, *hasFather*, and *hasMother*. The first query retrieves all the individuals that are related to some other individuals as a child. The competency question addressed by this query is: *Which individuals have some parent individuals?* The result of the query is shown in figure 6.21(a). The result extracts only two individuals that satisfy the conditions of the query string. The Manchester OWL syntax for the query is given below:

```
User and (hasParent some User)
```

The second query of this set search for all the individuals related to other individuals as a parent. The inverse object property of *hasParent* is used in this query string. The result of the query retrieves 4 individuals that are a parent of some other individual of the *User* class. Figure 6.21(b) shows the result of the query. The competency question addressed by this query string is *Which individuals are a parent of some other individuals?*. The Manchester OWL syntax for the query is given below:

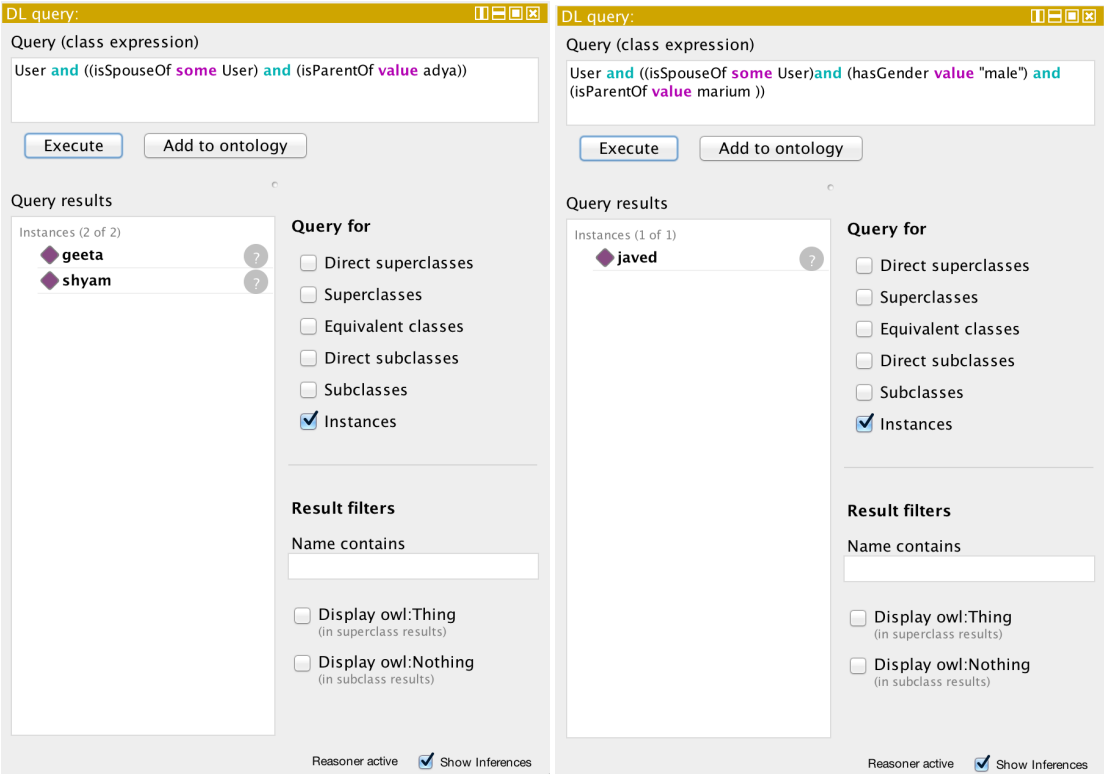
```
User and (isParentOf some User)
```

The third query of this set is searching for two individuals that are related to a certain individual called adya as a parent. Instead of using class name in the query string, we use the name of the individual as a value in the query string and look for two individuals one male and another female that are related to adya as parents. The competency question addressed by this query is: *Who is the parent of individual called*



(a) Retrieving individuals with parents

(b) Retrieving individuals with children



(c) Retrieving parents of a certain individual

(d) Retrieving child of certain individuals

Figure 6.22: Example queries to retrieve parentage facts

adya? Figure 6.21(c) shows the result of the query. The Manchester OWL syntax of the query is given below:

```
User and (isParentOf value adya)
```

Final query of this set retrieves information about the individual that is child of two certain individuals. The object properties used in this query are *hasFather* and *hasMother*. The competency question addressed by this query is: *Which individuals have father called javed and mother called hira as mother?* The result of the query retrieves only one individual that satisfy this condition. Figure 6.21(d) shows the result. The Manchester OWL syntax of the query is given below:

```
User and ((hasFather value javed)
           and (hasMother value hira))
```

Figure 6.22 represents final set of queries that check the consistency between T-Box and A-Box of SOCPRI ontology. The object and datatype properties used in these queries are *isSpouseOf*, *isParentOf* and *hasGender*. The first query of this set retrieve all individuals that are connected with other individuals through *isSpouseOf* object property. The competency questions addressed by this query is: *Which of the individuals are related to other individuals as spouse?* The result of the query is shown figure 6.22(a). According to the results, there are four individuals that are connected through this relationship. The manchester OWL syntax of the query is given below:

```
User and (isSpouseOf some User)
```

The second query of this set is searching for all the female individuals that participate in the spouse relationship. The competency question addressed by this query is: *Which of the female individuals are related to other male individuals as spouse?* Figure 6.22(b) show the result of the query. There are only two female individuals that are connected to certain other individuals through this relation. The manchester OWL syntax of the query is given below:

```
User and ((isSpouseOf some User)
and (hasGender value "female"))
```

The third query of this set is about retrieving the couple related to an individual called adya as parents. The competency questions addressed by this query is: *Which of the couple has child individual called adya?* The results of this query retrieve individuals called shyam and geeta as parents of the individual named as adya. Figure 6.22(c) shows result of the query. The manchester OWL syntax of the query is given below:

```
User and ((isSpouseOf some User)
and (isParentOf value adya))
```

Final query of this set is very specific and retrieve the individual that participates in spouse relation with some other individual and belongs to male gender and also has child individual by the name of marium. The competency question addressed by this query is: *Which individual is spouse of some other individual and has gender male and is parent of the individual called marium?* Figure 6.22(d) shows the result of the query. The manchester OWL syntax of the query is given below:

```
User and ((isSpouseOf some User)
and (hasGender value "male")
and (isParentOf value marium))
```

In this section, we generated results for 6 set of the queries from A-Boxes of SOCPRI ontology. Each set is composed of four related query strings. These query strings are designed in manner that they grow in complexity and specificity. As mentioned earlier, we enriched the ontology with 75 individuals (instances) and provided more than 100 object property assertions. The relationships between various individuals are depicted in figure 6.17. The results of aforementioned 24 DL queries successfully evaluate SOCPRI ontology and demonstrate consistency between T-Boxes and A-Boxes of our ontology.

6.7 Concluding Remarks

One of the important tasks in ontology development methodology is the evaluation. Ontology evaluation not only verifies the correctness and consistency of an ontology but also assesses its appropriateness against proposed requirements. We evaluate SOCPRI ontology for its correctness and consistency as well for its appropriateness against proposed requirements. In the first phase, SOCPRI ontology is evaluated for correctness and consistency. We used three different reasoners to check the consistency of ontological structure. The evaluation results demonstrated that ontological structure is consistent. We evaluated SOCPRI ontology using ontology evaluation protege plugin. The tool evaluates aspects such as concept hierarchy, property hierarchy, domain/range definitions, property restrictions, naming conventions, documentation and visualization of the ontology. The development process always introduces some common pitfalls into the ontology. We also check our ontology against such common pitfall using ontology pitfall scanner. OntoClean methodology was also used to evaluate the SOCPRI ontology.

In the second phase, SOCPRI ontology is evaluated against proposed requirements. We populated our ontology with 75 individuals and inserted more than 100 object property assertions to establish the relations between various individuals. We developed 24 different queries to answer various competency questions. The competency questions are developed in requirement specification phase and they represent ontology requirements. These DL queries assess the appropriateness of our ontology against proposed requirements. The consistency between T-Boxes and A-Boxes of our ontology is also evaluated through these queries.

Chapter 7

Conclusion and Future Work

In this chapter, we conclude the dissertation and also describe an outlook to future research challenges in the domain. In section 7.1, we discuss the main contribution of our research work and some of the limitations. We revisit the research questions and discuss how our work addressed these questions in section 7.2. Finally, we present future research directions that stem from this dissertation in section 7.3.

7.1 Summary of Contribution

The main contribution of this dissertation is to develop the SOCPRI ontology that represents diverse social relationships of users in online social networks. The purpose of developing this ontological model is to enhance user privacy in online social networks. Our ontology models role and relationship based self-presentation of users in the dynamic environment of the social web. It represents tie strength dimensions and relates these dimensions to users' social interactions in OSNs. SOCPRI also models contextual privacy of OSN users which takes into consideration contextual norms for appropriate information disclosure within and across contexts. The ontological model is inspired from well-grounded social theories about self-presentation, contextual integrity, and tie

strength. The aspect of self-presentation deals with role and relationship based contact segregation. The aspect of contextual integrity handles the issue of context collapse by providing a fine-grained segregation of social contexts. The aspect of tie strength addresses the issue related to the identification of strong and weak ties on the basis of interaction pattern and profile similarity attributes among social web users.

The innovative aspect of this model is that it takes into consideration the social perspective of privacy. From the social perspective, privacy is viewed as “socially constructed behavior of an individual during their everyday social interactions”. This aspect of privacy focuses on managing social relationships and the boundaries between private and public life. Social perspective of privacy also benefits from rich research literature of sociology that gives insight on how to manage diverse social relationships and what are interaction patterns and personal information disclosure practices between strong and weak ties. The social perspective of privacy is ignored by all of the existing solutions developed for preserving user privacy in online social networks.

Another major contribution of this dissertation is the development of a questionnaire to conduct a user study to examine the behavior of users about disclosing personal information and forming relationships in online social networks. The user study also examines user communication and interaction pattern with people representing various facets of their life such as work, family, and friends. An online questionnaire was circulated via numerous university mailing lists and postings in popular OSN groups. The user study targeted active Facebook and Google+ users. We collected data from 323 participants. The geographical location of the majority of participants is from the Indian subcontinent, whereas a small number of participants are from Europe. The results of this study identify the conflicts between privacy concerns and information sharing behavior of social web users. The results also establish a link between personal information disclosure and relationship strength. The relationship strength among online social network users can be estimated via their interaction patterns. The findings

also reveal that the choice of interaction type gives an indication of the relationship strength. Additionally, the findings facilitate categorization of profile information and user interactions on the basis of sensitivity and frequency respectively.

Another contribution is the development of conceptual framework for social web privacy taking into consideration shift in the status of the user from content consumer to content manager. Current online social networks changed the status of an end-user. An individual end-user becomes a content manager instead of just being a content consumer. The uploaders must decide for every single piece of data shared on OSNs that who can access it from his/her friends. Main problems faced by online social networks today are collapsed context, conflated contacts, and co-joined content (public and private). Contemporary online social networks provide their users with single timeline/wall for all the contacts which represent single universal context for all kind of contacts. The single user profile represents all kind of personal information irrespective of their sensitivity. The social graph of the user represents all kind of contacts without taking into consideration tie strength among the user and his contacts. We proposed a 3C segregation privacy framework which addresses these aspects of user privacy in the social web. This framework is based on the social perspective of privacy for online social networks and focuses on managing self- presentation, preserving the contextual integrity and maintaining a balance between privacy and publicity.

The final contribution of this dissertation is the evaluation of the ontology in absence of any golden standard for evaluation. The golden standard refers to benchmark ontology that can be used to compare our ontological model. We do not have such a gold-standard ontology. We evaluated various aspects of our ontology including structural dimension, functional dimension, usability-profiling dimension, etc. SOCPRI ontological model is also evaluated against well-established evaluation metrics proposed in ontology-summit 2007. The inconsistency, incompleteness, and redundancy of the model are also verified using three different reasoners Fact++, Hermit, and Pellet. Fi-

nally, we evaluated our ontology using the OntoClean methodology against the domain-independent metaproperties.

7.2 Research Questions Revisited

In this dissertation, our discussion revolves around four research questions. The first research question deals with issue of social web privacy. Current approaches handling privacy in the social web does not take into consideration shift in the status of an end-user from content consumer to content manager. We develop a theoretical framework for privacy in the social web that is inspired by social theories of Goffman, Nissenbaum, and Granovetter. Goffman advocates role and relationship based self-presentation of an individual end-user. Nissenbaum argues that social privacy revolves around preserving the contextual integrity of an individual end-user. Granovetter differentiates between strong and weak ties among individuals end-users. The chapter three of this dissertation provides answers for the first research question that is how to redefine privacy for social web that suits emerging content manager status of the end-user?

The content of chapter four answers the second research question that is how interaction patterns and profile similarity attributes reveal context and quality of relationship among social web users? We conducted a user study to investigate the attitude of social web users towards privacy. The study examines information sharing and relationship forming behavior of online social network users. The main goal of this study is to determine whether there exists a relationship between interaction pattern and tie strength of the users. We explore the possibility of using user interaction with his friends as criteria for making personal information disclosure decisions. The results also establish a link between personal information disclosure and relationship strength. The relationship strength among online social network users can be estimated via their interaction patterns. The findings reveal that relationship strength is directly proportional to the

frequency of interactions among users and personal information disclosure depends on relationship strength.

The third research question deals with an issue of developing an ontological model for privacy in online social networks from the social perspective. We formalize theoretical framework of privacy into an ontological model. The objective of developing this ontological is to manage contextual privacy and self-presentation of a user in a dynamic environment of the online social network with diverse audience. The novelty of this ontological model is that it is inspired by most influential social theories about tie strength, contextual integrity, and presentation of self. As described earlier, the aspect of self-presentation facilitates role and relationship based contact segregation. The major issue of current online social networks related to context collapse is handled by preserving contextual integrity. Finally, user interaction pattern and profile similarity attributes play a vital role in the identification of strong and weak ties. The detailed description of the ontological model is presented in chapter five of this dissertation.

The content of chapter six provides answers for fourth research question that is how to evaluate the ontological model that represents diverse social relationships of OSNs users from the social perspective? Performing evaluation of the ontological model that is inspired by well-established social theories is a challenging task. We evaluated the ontological model from various aspects. Several reasoners were used to perform verification and validation of the model which includes Pellet, Fact++, and Hermit. We also evaluated ontological model against common ontological pitfalls introduced during the development process. The philosophical perspective of our model was checked with OntoClean methodology. Finally, we adapted ontology metrics proposed in ontology-summit 2007 for evaluating our ontological model.

The structure of this dissertation reflects a logical connection between the research questions. The content of dissertation chapters from three to six provides answers for research questions one to four. We introduced the research problem in the first chapter

and its background and related literature are presented in chapter two. The last chapter provides concluding remarks on the dissertation.

7.3 Outlook on Further Research

There exists an inherent design conflict between privacy goals and traditional design of social web that promote sociability and usability. The sociability and usability lead to privacy leakage in an uncontrollable way for social web users. To overcome this inherent design fault shift in privacy paradigm is required. Most of the current privacy preserving approaches ignore the social perspective of privacy, whereas social privacy requires the development of an enriched relationship model for the social web. An enriched social relationship model can improve privacy on the social web. SOCPRI is the first step towards privacy friendly sociability of the social web.

We identified some possible lines for future research. These possible path of research are open challenges that were identified during the development of SOCPRI model. The users will feel safer in the context of social web once all the challenges are accomplished and the utility of OSNs will increase since users disclose much more personal information to various life facets confidentially. One of the open research challenges is inferring tie strength from different sources. Combining multiple sources of information can positively improve the tie strength inference and classification of relationships. Tie strength utility is another open research challenge that focuses on the utility of tie rather than inferring its strength. Ties are viewed not for their strength but how useful a tie is depending on the situation or the need. Co-privacy management is also open research challenge that causes several privacy conflicts among multiple stakeholders of a single resource.

In the context of this dissertation, we suggested several areas for future research. One of the directions for extending our research work is related to user study. The user

study reported in the third chapter of this dissertation has two limitations. The first limitation deals with a relatively small number of users participating in this study. As OSNs are the current craze and their user base is in millions. Therefore, the response of less 400 hundred users cannot be conclusive. The second limitation deals with geographical distribution of the users. The majority of users belongs to Indian sub-continent and their response cannot be representative of the users with different culture settings such as Europeans. We recommend a user study with massive participants across the globe. It is a quite challenging task to conduct such a study.

Our design goal for SOCPRI was to develop a light-weight ontology with maximum flexibility and extensibility. To achieve this goal, we intentionally limited ourselves to most basic form of contextual norms. The SOCPRI ontology could be extended to express more complex contextual norms without sacrificing practical reasoning capability and query answering capability. Another direction for extending our work could be performing a domain expert-based evaluation of SOCPRI ontology. We consider that modeling classical social theories could be challenging task for ontologist. It is the out of skill set of ontology engineer. It is always good option to evaluate the model with domain expert (social theorist). Our work can be extended in three step process consisting:

- Generating natural language questions from the content of SOCPRI ontology.
- Submitting the questions to a domain expert for review and feedback on the model.
- Modifying the content of SOCPRI ontology depending on expert's comments

In the summary, we suggested three research directions to extend the contributions presented in this dissertation. Firstly, a user study can be conducted with a large and culturally diverse sample size. Secondly, the SOCPRI ontology can be extended by

adding more complex contextual norms. Finally, domain expert-based evaluation can be done for our ontology.

Appendices

Appendix A

List of Publications

- Journals

1. Ahmed J., Governatori, G., Villata, S. (Accepted Paper). Information and Friend Segregation for Online Social Networks: A User Study. In proceedings of Springer AI and Society Journal (SCI impact factor 0.171)

- Conference Proceedings

1. Ahmed, J. (2016, June). A semantic model for friend segregation in online social networks. In International Conference on Web Engineering (pp. 495-500). Springer International Publishing.
2. Ahmed, J., Governatori, G., van der Torre, L., Villata, S. (2014). Social Interaction Based Audience Segregation for Online Social Networks. In proceedings of European Conference on Social Intelligence (ECSI-2014)(pp. 186-197).
3. Ahmed, J. (2014). A Privacy Protection Model for Online Social Networks. In proceeding of JURIX2014 Doctoral Consortium.

Bibliography

- [1] George Pallis, Demetrios Zeinalipour-Yazti, and Marios D Dikaiakos. Online social networks: status and trends. In *New Directions in Web Data Management 1*, pages 213–234. Springer, 2011.
- [2] Barbara Carminati, Elena Ferrari, and Marco Viviani. Security and trust in online social networks. *Synthesis Lectures on Information Security, Privacy, & Trust*, 4(3):1–120, 2013.
- [3] Christian Richthammer, Michael Netter, Moritz Riesner, and Gunther Pernul. Taxonomy for social network data types from the viewpoint of privacy and user control. In *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*, pages 141–150. IEEE, 2013.
- [4] Sebastian Banescu, Simona Posea, and Andrei Calin. Privacy in online social networks.
- [5] SL Jones. *Automating group-based privacy control in social networks*. PhD thesis, University of Bath, 2012.
- [6] Tim Berners-Lee. Wwww at 15 years: looking forward. In *Proceedings of the 14th international conference on World Wide Web*, pages 1–1. ACM, 2005.
- [7] Asunción Gómez-Pérez and David Manzano-Macho. An overview of methods

- and tools for ontology learning from texts. *The knowledge engineering review*, 19(03):187–212, 2004.
- [8] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. How to write and use the ontology requirements specification document. In *OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”*, pages 966–982. Springer, 2009.
- [9] Asuncion Gomez-Perez, Mariano Fernández-López, and Oscar Corcho. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media, 2006.
- [10] María Poveda Villalón. *Ontology Evaluation: a pitfall-based approach to ontology diagnosis*. PhD thesis, ETSI Informatica, 2016.
- [11] Jessica Vitak. The impact of context collapse and privacy on social network site disclosures. *Journal of Broadcasting & Electronic Media*, 56(4):451–470, 2012.
- [12] Jenny L Davis and Nathan Jurgenson. Context collapse: theorizing context collusions and collisions. *Information, Communication & Society*, 17(4):476–485, 2014.
- [13] Bernie Hogan. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, page 0270467610385893, 2010.
- [14] Erving Goffman. The presentation of self in everyday life [1959]. *Contemporary sociological theory*, pages 46–61, 2012.
- [15] Mark Granovetter. The strength of weak ties: A network theory revisited. *Sociological theory*, 1(1):201–233, 1983.

- [16] Helen Nissenbaum. Privacy as contextual integrity. *Washington law review*, 79(1), 2004.
- [17] Balachander Krishnamurthy and Craig E Wills. Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, pages 37–42. ACM, 2008.
- [18] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 61–70. ACM, 2011.
- [19] Heather Richter Lipford, Andrew Besmer, and Jason Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.
- [20] Thomas Paul, Daniel Puscher, and Thorsten Strufe. Improving the usability of privacy settings in facebook. *arXiv preprint arXiv:1109.6046*, 2011.
- [21] Cuneyt Gurcan Akcora and Elena Ferrari. Graphical user interfaces for privacy settings. In *Encyclopedia of Social Network Analysis and Mining*, pages 648–660. Springer, 2014.
- [22] Michelle Madejski, Maritza Lupe Johnson, and Steven Michael Bellovin. The failure of online social network privacy settings. 2011.
- [23] Maritza Johnson, Serge Egelman, and Steven M Bellovin. Facebook and privacy: it’s complicated. In *Proceedings of the eighth symposium on usable privacy and security*, page 9. ACM, 2012.
- [24] Michel Netter, Moritz Riesner, Michael Weber, and Gunther Pernul. Privacy settings in online social networks—preferences, perception, and reality. In *System*

- Sciences (HICSS)*, 2013 46th Hawaii International Conference on, pages 3219–3228. IEEE, 2013.
- [25] Yousra Javed and Mohamed Shehab. How do facebookers use friendlists. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 343–347. IEEE Computer Society, 2012.
- [26] Jason Watson, Andrew Besmer, and Heather Richter Lipford. + your circles: sharing behavior on google+. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 12. ACM, 2012.
- [27] Yabing Liu, Bimal Viswanath, Mainack Mondal, Krishna P Gummadi, and Alan Mislove. Simplifying friendlist management. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 385–388. ACM, 2012.
- [28] Jacob W Bartel and Prasun Dewan. Evolving friend lists in social networks. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 435–438. ACM, 2013.
- [29] Lerone Banks and Shyhtsun Felix Wu. All friends are not created equal: An interaction intensity based approach to privacy in online social networks. In *Computational Science and Engineering, 2009. CSE’09. International Conference on*, volume 4, pages 970–974. IEEE, 2009.
- [30] Waqar Ahmad, Asim Riaz, Henric Johnson, and Niklas Lavesson. Predicting friendship intensity in online social networks. In *21st International Tyrrhenian Workshop on Digital Communications*, 2010.
- [31] Gaurav Misra and Jose M Such. Social computing privacy and online relationships.

- [32] Gordon Hull, Heather Richter Lipford, and Celine Latulipe. Contextual gaps: Privacy issues on facebook. *Ethics and information technology*, 13(4):289–302, 2011.
- [33] Joan Morris DiMicco and David R Millen. Identity management: multiple presentations of self in facebook. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 383–386. ACM, 2007.
- [34] Ronald Leenes. Context is everything sociality and privacy in online social network sites. In *Privacy and identity management for life*, pages 48–65. Springer, 2010.
- [35] Bibi Van Den Berg and Ronald Leenes. Audience segregation in social network sites. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 1111–1116. IEEE, 2010.
- [36] Adam Barth, Anupam Datta, John C Mitchell, and Helen Nissenbaum. Privacy and contextual integrity: Framework and applications. In *Security and Privacy, 2006 IEEE Symposium on*, pages 15–pp. IEEE, 2006.
- [37] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.
- [38] Andrea Petróczi, Tamás Nepusz, and Fülöp Bazsó. Measuring tie-strength in virtual social networks. *Connections*, 27(2):39–52, 2007.
- [39] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–220. ACM, 2009.
- [40] Eric Gilbert. Predicting tie strength in a new medium. In *Proceedings of the*

- ACM 2012 conference on Computer Supported Cooperative Work*, pages 1047–1056. ACM, 2012.
- [41] Jason Wiese, Patrick Gage Kelley, Lorrie Faith Cranor, Laura Dabbish, Jason I Hong, and John Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 197–206. ACM, 2011.
- [42] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo. *Methontology: from ontological art towards ontological engineering*. 1997.
- [43] Michael Gruninger, Olivier Bodenreider, Frank Olken, Leo Obrst, and Peter Yim. Ontology summit 2007-ontology, taxonomy, folksonomy: Understanding the distinctions. *Applied Ontology*, 3(3):191–200, 2008.
- [44] Ricard L Fogués, Jose M Such, Agustin Espinosa, and Ana Garcia-Fornes. Bff: A tool for eliciting tie strength and user communities in social networking services. *Information Systems Frontiers*, 16(2):225–237, 2014.
- [45] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.
- [46] Charu C Aggarwal. An introduction to social network data analytics. In *Social network data analytics*, pages 1–15. Springer, 2011.
- [47] Howard Rheingold. *The virtual community: Homesteading on the electronic frontier*. MIT press, 2000.
- [48] Danah Boyd. Facebook’s privacy trainwreck. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):13–20, 2008.

- [49] Zizi Papacharissi and Elaine Yuan. What if the internet did not speak english? new and old language for studying newer media technologies. *The Long History of New Media: Technology, Historiography, and Contextualizing Newness*, pages 89–107, 2011.
- [50] Frederic Stutzman. An evaluation of identity-sharing behavior in social network communities. *Journal of the International Digital Media and Arts Association*, 3(1):10–18, 2006.
- [51] Alice E Marwick et al. I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society*, 13(1):114–133, 2011.
- [52] Nicole B Ellison et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.
- [53] Elie Raad and Richard Chbeir. Privacy in online social networks. In *Security and Privacy Preserving in Social Networks*, pages 3–45. Springer, 2013.
- [54] James Grimmelman. Facebook and the social dynamics of privacy. *Iowa Law Review*, 95(4):1–52, 2009.
- [55] Vassilis Kostakos, Jayant Venkatanathan, Bernardo Reynolds, Norman Sadeh, Eran Toch, Siraj A Shaikh, and Simon Jones. Who’s your best friend?: targeted privacy attacks in location-sharing social networks. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 177–186. ACM, 2011.
- [56] danah boyd Nicole B. Ellison. *The Oxford Handbook of Internet Studies*, chapter Sociality through Social Network Sites, pages 151–172. Oxford University Press, 2013.
- [57] Ai Thanh Ho. Towards a privacy-enhanced social networking site. 2012.

- [58] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.
- [59] Chi Zhang, Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang. Privacy and security for online social networks: challenges and opportunities. *Network, IEEE*, 24(4):13–18, 2010.
- [60] Philip WL Fong, Mohd Anwar, and Zhen Zhao. A privacy preservation model for facebook-style social network systems. In *European Symposium on Research in Computer Security*, pages 303–320. Springer, 2009.
- [61] Michael Beye, Arjan JP Jeckmans, Zekeriya Erkin, Pieter Hartel, Reginald L Lagendijk, and Qiang Tang. Privacy in online social networks. In *Computational Social Networks*, pages 87–113. Springer, 2012.
- [62] Rizwana Irfan, Gage Bickler, Samee U Khan, Joanna Kolodziej, Hongxiang Li, Dan Chen, Lizhe Wang, Khizar Hayat, Sajjad Ahmad Madani, Babar Nazir, et al. Survey on social networking services. *IET networks*, 2(4):224–234, 2013.
- [63] Michael Beye, Arjan JP Jeckmans, Zekeriya Erkin, Pieter Hartel, Reginald L Lagendijk, and Qiang Tang. Privacy in online social networks. In *Computational Social Networks*, pages 87–113. Springer, 2012.
- [64] Bruce Schneier. A taxonomy of social networking data. *Security & Privacy, IEEE*, 8(4):88–88, 2010.
- [65] Susan B Barnes. A privacy paradox: Social networking in the united states. *First Monday*, 11(9), 2006.
- [66] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.

- [67] Alessandro Acquisti and Ralph Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy enhancing technologies*, pages 36–58. Springer, 2006.
- [68] Kevin Lewis, Jason Kaufman, and Nicholas Christakis. The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, 14(1):79–100, 2008.
- [69] Mike Thelwall. Homophily in myspace. *Journal of the American Society for Information Science and Technology*, 60(2):219–231, 2009.
- [70] Ratan Dey, Zubin Jelveh, and Keith Ross. Facebook users have become much more private: A large-scale study. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 346–352. IEEE, 2012.
- [71] Andrew W Boyd. A longitudinal study of social media privacy behavior. *arXiv preprint arXiv:1103.3174*, 2011.
- [72] Mary Madden. Privacy management on social media sites. *Pew Internet Report*, pages 1–20, 2012.
- [73] Xi Chen and Shuo Shi. A literature review of privacy research on social network sites. In *2009 International Conference on Multimedia Information Networking and Security*, volume 1, pages 93–97. IEEE, 2009.
- [74] Giles Hogben ENISA. Security issues and recommendations for online social networks. Technical report, Technical report, 2007.
- [75] Abdullah Al Hasib. Threats of online social networks. *IJCSNS International Journal of Computer Science and Network Security*, 9(11):288–93, 2009.

- [76] Dan Brickley and Libby Miller. Foaf vocabulary specification 0.98. *Namespace document*, 9, 2012.
- [77] Sebastian R Kruk. Foaf-realm-control your friends access to the resource. In *FOAF Workshop proceedings*, volume 186, 2004.
- [78] Sebastian Ryszard Kruk, Sławomir Grzonkowski, Adam Gzella, Tomasz Woroniecki, and Hee-Chul Choi. D-foaf: Distributed identity management with access rights delegation. In *The Semantic Web-ASWC 2006*, pages 140–154. Springer, 2006.
- [79] Michel Buffa and Catherine Faron-Zucker. Ontology-based access rights management. In *Advances in Knowledge Discovery and Management*, pages 49–61. Springer, 2012.
- [80] Najeeb Elahi, Mohammad MR Chowdhury, and Josef Noll. Semantic access control in web based communities. In *Computing in the Global Information Technology, 2008. ICCGI'08. The Third International Multi-Conference on*, pages 131–136. IEEE, 2008.
- [81] Barbara Carminati, Elena Ferrari, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. A semantic web based framework for social network access control. In *Proceedings of the 14th ACM symposium on Access control models and technologies*, pages 177–186. ACM, 2009.
- [82] Barbara Carminati, Elena Ferrari, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham. Semantic web-based social network access control. *computers & security*, 30(2):108–115, 2011.
- [83] Amirreza Masoumzadeh and James Joshi. Osnac: An ontology-based access control model for social networking systems. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 751–759. IEEE, 2010.

- [84] Amirreza Masoumzadeh and James Joshi. Ontology-based access control for social network systems. *International Journal of Information Privacy, Security and Integrity*, 1(1):59–78, 2011.
- [85] Louise Barkhuus. The mismeasurement of privacy: using contextual integrity to reconsider privacy in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 367–376. ACM, 2012.
- [86] Pan Shi, Heng Xu, and Yunan Chen. Using contextual integrity to examine interpersonal information boundary on social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 35–38. ACM, 2013.
- [87] Heather Richter Lipford, Gordon Hull, Celine Latulipe, Andrew Besmer, and Jason Watson. Visible flows: Contextual integrity and the design of privacy mechanisms on social network sites. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 985–989. IEEE, 2009.
- [88] Imrul Kayes and Adriana Iamnitchi. Out of the wild: On generating default policies in social ecosystems. In *ICC Workshops*, pages 204–208, 2013.
- [89] Imrul Kayes and Adriana Iamnitchi. Aegis: A semantic implementation of privacy as contextual integrity in social ecosystems. In *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*, pages 88–97. IEEE, 2013.
- [90] Indika Kahanda and Jennifer Neville. Using transactional information to predict link strength in online social networks. *ICWSM*, 9:74–81, 2009.
- [91] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship

- strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.
- [92] Lizi Zhang, Hui Fang, Wee Keong Ng, and Jie Zhang. Intrank: Interaction ranking-based trustworthy friend recommendation. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 266–273. IEEE, 2011.
- [93] Lizi Zhang, Cheun Pin Tan, Siyi Li, Hui Fang, Pramodh Rai, Yao Chen, Rohit Luthra, Wee Keong Ng, and Jie Zhang. The influence of interaction attributes on trust in virtual communities. In *Advances in User Modeling*, pages 268–279. Springer, 2012.
- [94] Zhao Du, Lantao Hu, Xiaolong Fu, and Yongqi Liu. Scalable and explainable friend recommendation in campus social network system. In *Frontier and Future Development of Information Technology in Medicine and Education*, pages 457–466. Springer, 2014.
- [95] Frank Nagle and Lisa Singh. Can friends be trusted? exploring privacy in online social networks. In *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*, pages 312–315. IEEE, 2009.
- [96] Christo Wilson, Bryce Boe, Alessandra Sala, Krishna PN Puttaswamy, and Ben Y Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218. Acm, 2009.
- [97] Fabeah Adu-Oppong, Casey K Gardiner, Apu Kapadia, and Patrick P Tsang. Social circles: Tackling privacy in social networks. In *Symposium on Usable Privacy and Security (SOUPS)*, 2008.

- [98] Anna Squicciarini, S Karumanchi, Dongyang Lin, and Nicole DeSisto. Automatic social group organization and privacy management. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, pages 89–96. IEEE, 2012.
- [99] Anna Squicciarini, Sushama Karumanchi, Dan Lin, and Nicole DeSisto. Identifying hidden social circles for advanced privacy configuration. *Computers & Security*, 41:40–51, 2014.
- [100] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, pages 351–360. ACM, 2010.
- [101] Lujun Fang, Heedo Kim, Kristen LeFevre, and Aaron Tami. A privacy recommendation wizard for users of social networking sites. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 630–632. ACM, 2010.
- [102] Patrick Gage Kelley, Robin Brewer, Yael Mayer, Lorrie Faith Cranor, and Norman Sadeh. An investigation into facebook friend grouping. In *Human-Computer Interaction—INTERACT 2011*, pages 216–233. Springer, 2011.
- [103] CGA Bryant and Barrington Moore. Privacy: Studies in social and cultural history, 1985.
- [104] Samuel D Warren and Louis D Brandeis. The right to privacy. *Harvard law review*, pages 193–220, 1890.
- [105] Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

- [106] Lü Yao-Huai. Privacy and data privacy issues in contemporary china. *Ethics and Information Technology*, 7(1):7–15, 2005.
- [107] Ponnurangam Kumaraguru and Niharika Sachdeva. Privacy in india: Attitudes and awareness v 2.0. *Available at SSRN 2188749*, 2012.
- [108] Mireille M Caruana and Joseph A Cannataci. European union privacy and data protection principles: Compatibility with culture and legal frameworks in islamic states. *Information & Communications Technology Law*, 16(2):99–124, 2007.
- [109] Lee A Bygrave. Privacy and data protection in an international perspective. *Scandinavian studies in law*, 56:165–200, 2010.
- [110] Blase Ur and Yang Wang. A cross-cultural framework for protecting user privacy in online social media. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 755–762. ACM, 2013.
- [111] Irwin Altman. The environment and social behavior: Privacy, personal space, territory, and crowding. 1975.
- [112] Jerry Kang. Information privacy in cyberspace transactions. *Stanford Law Review*, pages 1193–1294, 1998.
- [113] Leysia Palen and Paul Dourish. Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136. ACM, 2003.
- [114] Mohd M Anwar. *An identity-and trust-based computational model for privacy*. University of Saskatchewan, 2009.
- [115] Seda Gürses and Claudia Diaz. Two tales of privacy in online social networks. *IEEE Security & Privacy*, 11(3):29–37, 2013.

- [116] Michael Netter, Sebastian Herbst, and Günther Pernul. Analyzing privacy in social networks—an interdisciplinary approach. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 1327–1334. IEEE, 2011.
- [117] Katrin Borcea-Pfitzmann, Andreas Pfitzmann, and Manuela Berg. Privacy 3.0:= data minimization+ user control+ contextual integrity. *it-Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik*, 53(1):34–40, 2011.
- [118] Hugh Miller. The presentation of self in electronic life: Goffman on the internet. In *Embodied knowledge and virtual space conference*, volume 9, 1995.
- [119] Dan Laughey. *Key themes in media theory*. McGraw-Hill Education (UK), 2007.
- [120] Nan Lin, John C Vaughn, and Walter M Ensel. Social resources and occupational status attainment. *Social forces*, 59(4):1163–1181, 1981.
- [121] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.
- [122] Peter V Marsden and Karen E Campbell. Measuring tie strength. *Social forces*, 63(2):482–501, 1984.
- [123] Yaxi He, Chunhong Zhang, and Yang Ji. Principle features for tie strength estimation in micro-blog social network. In *Computer and Information Technology (CIT), 2012 IEEE 12th International Conference on*, pages 359–367. IEEE, 2012.
- [124] David J Houghton and Adam N Joinson. Privacy, social network sites, and social relations. *Journal of Technology in Human Services*, 28(1-2):74–94, 2010.

- [125] Caroline Haythornthwaite. Strong, weak, and latent ties and the impact of new media. *The information society*, 18(5):385–401, 2002.
- [126] Kathryn Dindia and Daniel J Canary. Definitions and theoretical perspectives on maintaining relationships. *Journal of Social and Personal Relationships*, 10(2):163–173, 1993.
- [127] Tasos Spiliotopoulos, Diogo Pereira, and Ian Oakley. Predicting tie strength with the facebook api. In *Proceedings of the 18th Panhellenic Conference on Informatics*, pages 1–5. ACM, 2014.
- [128] Fred Stutzman and Jacob Kramer-Duffield. Friends only: examining a privacy-enhancing behavior in facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1553–1562. ACM, 2010.
- [129] Fatih Kursat Ozenc and Shelly D Farnham. Life modes in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 561–570. ACM, 2011.
- [130] Ofcom. Social networking:a quantitative and qualitative research report into attitudes, behaviours and use. *Office of Communication United Kingdom*, 2008.
- [131] Zeynep Tufekci. Can you see me now? audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology and Society*, 28(1):20–36, 2008.
- [132] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. Privacy concerns and identity in online social networks. *Identity in the Information Society*, 2(1):39–63, 2009.
- [133] Airi Lampinen, Sakari Tamminen, and Antti Oulasvirta. All my people right here, right now: Management of group co-presence on a social networking site.

- In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 281–290. ACM, 2009.
- [134] Meredith M Skeels and Jonathan Grudin. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 95–104. ACM, 2009.
- [135] Alyson L Young and Anabel Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*, pages 265–274. ACM, 2009.
- [136] Frederic Stutzman and Woodrow Hartzog. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 769–778. ACM, 2012.
- [137] Liz Spencer and Raymond Edward Pahl. *Rethinking friendship: Hidden solidarities today*. Princeton University Press, 2006.
- [138] Xingang Zhang, Qijie Gao, Christopher SG Khoo, and Amos Wu. Categories of friends on social networking sites: An exploratory study. In *Proceedings of the 5th International Conference on Asia-Pacific Library and Information Education and Practice*, pages 244–259, 2013.
- [139] Gaurav Misra and Jose M Such. Social computing privacy and online relationships. 2015.
- [140] Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. Risks of friendships on social networks. *arXiv preprint arXiv:1210.3234*, 2012.

- [141] Miltiadis D Lytras, Miguel Angel Sicilia, Demetrios Sampson, Tim Finin, Li Ding, Lina Zhou, and Anupam Joshi. Social networking on the semantic web. *The Learning Organization*, 12(5):418–435, 2005.
- [142] John Breslin and Stefan Decker. The future of social networks on the internet: the need for semantics. *IEEE Internet Computing*, 11(6):86–90, 2007.
- [143] John G Breslin, Alexandre Passant, and Stefan Decker. Towards the social semantic web. In *The Social Semantic Web*, pages 269–283. Springer, 2009.
- [144] Guillaume Erétéo, Michel Buffa, Fabien Gandon, Patrick Grohan, Mylène Leitzelman, and Peter Sander. A state of the art on social network analysis and its applications on a semantic web. In *SDOW2008*, 2008.
- [145] Alistair Miles and Sean Bechhofer. Skos simple knowledge organization system reference. 2009.
- [146] Dominik Heckmann, Tim Schwartz, Boris Brandherm, Michael Schmitz, and Margeritta von Wilamowitz-Moellendorff. Gumo—the general user model ontology. In *International Conference on User Modeling*, pages 428–432. Springer, 2005.
- [147] Till Plumbaum. User modeling in the social semantic web. 2015.
- [148] Sofia Angeletou, Matthew Rowe, and Harith Alani. Modelling and analysis of user behaviour in online communities. In *International Semantic Web Conference*, pages 35–50. Springer, 2011.
- [149] Guillaume Erétéo, Michel Buffa, Fabien Gandon, and Olivier Corby. Analysis of a real online social network using semantic web frameworks. In *International Semantic Web Conference*, pages 180–195. Springer, 2009.

- [150] Alexandre Passant and Philippe Laublet. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *LDOW*, 2008.
- [151] Hak-Lae Kim, John G Breslin, Sung-Kwon Yang, and Hong-Gee Kim. Social semantic cloud of tag: Semantic model for social tagging. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications*, pages 83–92. Springer, 2008.
- [152] Steffen Lohmann, Paloma Díaz, and Ignacio Aedo. Muto: the modular unified tagging ontology. In *Proceedings of the 7th International Conference on Semantic Systems*, pages 95–104. ACM, 2011.
- [153] Owen Sacco and Alexandre Passant. A privacy preference ontology (ppo) for linked data. In *LDOW*. Citeseer, 2011.
- [154] Thomas R Gruber et al. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.
- [155] Dieter Fensel. Ontologies: A silver bullet for knowledge management and electronic-commerce (2000). *Berlin: Spring-Verlag*.
- [156] Michael N. Huhns and Munindar P. Singh. Ontologies for agents. *IEEE Internet computing*, 1(6):81–83, 1997.
- [157] Tom Gruber. Ontology. *Encyclopedia of database systems*, pages 1963–1965, 2009.
- [158] Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.
- [159] Ora Lassila and Ralph R Swick. Resource description framework (rdf) model and syntax specification. 1999.
- [160] Dan Brickley and Ramanathan V Guha. Resource description framework (rdf) schema specification 1.0: W3c candidate recommendation 27 march 2000. 2000.

- [161] Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(10):2004, 2004.
- [162] Ian Horrocks, Peter F Patel-Schneider, and Frank Van Harmelen. From shiq and rdf to owl: The making of a web ontology language. *Web semantics: science, services and agents on the World Wide Web*, 1(1):7–26, 2003.
- [163] Michael Uschold and Martin King. *Towards a methodology for building ontologies*. Citeseer, 1995.
- [164] Michael Grüninger and Mark S Fox. Methodology for the design and evaluation of ontologies. 1995.
- [165] Karin Koogan Breitman, Marco Antonio Casanova, and Walter Truszkowski. Methods for ontology development. *Semantic Web: Concepts, Technologies and Applications*, pages 155–173, 2007.
- [166] Steffen Staab, Rudi Studer, Hans-Peter Schnurr, and York Sure. Knowledge processes and ontologies. *IEEE Intelligent systems*, 16(1):26–34, 2001.
- [167] Asunción Gómez-Pérez and Mari Carmen Suárez-Figueroa. Neon methodology: scenarios for building networks of ontologies. *Poster and Demo*, page 43, 2008.
- [168] N Noy, Deborah L McGuinness, et al. Ontology development 101. *Knowledge Systems Laboratory, Stanford University*, 2001.
- [169] David J Schultz et al. Ieee standard for developing software life cycle processes. *IEEE Std*, pages 1074–1997, 1997.
- [170] Mariano Fernández López, Asunción Gómez-Pérez, Alejandro Pazos Sierra, and Juan Pazos Sierra. Building a chemical ontology using methontology and the ontology design environment. 1999.

- [171] Alisa Devlic, Roland Reichle, Michal Wagner, Manuele Kirsch Pinheiro, Yves Vanrompay, Yolande Berbers, and Massimo Valla. Context inference of users' social relationships and distributed policy management. In *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, pages 1–8. IEEE, 2009.
- [172] Asunción Gómez-Pérez. Some ideas and examples to evaluate ontologies. In *Artificial Intelligence for Applications, 1995. Proceedings., 11th Conference on*, pages 299–305. IEEE, 1995.
- [173] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Modelling ontology evaluation and validation. In *European Semantic Web Conference*, pages 140–154. Springer, 2006.
- [174] Denny Vrandečić. Ontology evaluation. In *Handbook on Ontologies*, pages 293–313. Springer, 2009.
- [175] Janez Brank, Marko Grobelnik, and Dunja Mladenic. A survey of ontology evaluation techniques. In *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*, pages 166–170, 2005.
- [176] Marta Sabou and Miriam Fernandez. Ontology (network) evaluation. In *Ontology engineering in a networked world*, pages 193–212. Springer, 2012.
- [177] Dionysios D Kehagias, Ioannis Papadimitriou, Joana Hois, Dimitrios Tzovaras, and John Bateman. A methodological approach for ontology evaluation and refinement. In *ASK-IT Final Conference. June.(Cit. on p.)*, 2008.
- [178] Alan Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. Owl pizzas: Practical experience of teaching owl-dl: Common errors & common patterns. In *Interna-*

- tional Conference on Knowledge Engineering and Knowledge Management*, pages 63–81. Springer, 2004.
- [179] María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. Validating ontologies with oops! In *International Conference on Knowledge Engineering and Knowledge Management*, pages 267–281. Springer, 2012.
- [180] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2):7–34, 2014.
- [181] María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. Did you validate your ontology? oops! In *Extended Semantic Web Conference*, pages 402–407. Springer, 2012.
- [182] Maria Poveda Villalon, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. A double classification of common pitfalls in ontologies. 2010.
- [183] Nicola Guarino and Christopher A Welty. An overview of ontoclean. In *Handbook on ontologies*, pages 201–220. Springer, 2009.
- [184] John G Breslin, Stefan Decker, Andreas Harth, and Uldis Bojars. Sioc: an approach to connect web-based communities. *International Journal of Web Based Communities*, 2(2):133–142, 2006.
- [185] Liam Bullingham and Ana C Vasconcelos. ?the presentation of self in the online world?: Goffman and the study of online identities. *Journal of Information Science*, 39(1):101–112, 2013.
- [186] Michael Hviid Jacobsen. *The Contemporary Goffman*. Routledge, 2010.

- [187] Judee K Burgoon, Roxanne Parrott, Beth A Le Poire, Douglas L Kelley, Joseph B Walther, and Denise Perry. Maintaining and restoring privacy through communication in different types of relationships. *Journal of Social and Personal Relationships*, 6(2):131–158, 1989.
- [188] Daniel J Solove. Conceptualizing privacy. *California Law Review*, pages 1087–1155, 2002.
- [189] Sonia Livingstone. Taking risky opportunities in youthful content creation: teenagers’ use of social networking sites for intimacy, privacy and self-expression. *New media & society*, 10(3):393–411, 2008.
- [190] Ai Ho, Abdou Maiga, and Esma Aïmeur. Privacy protection issues in social networking sites. In *2009 IEEE/ACS International Conference on Computer Systems and Applications*, pages 271–278. IEEE, 2009.
- [191] Mari Carmen Suárez-Figueroa, Asuncion Gomez-Perez, and Mariano Fernandez-Lopez. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.
- [192] Marija J Norusis. *SPSS 15.0 guide to data analysis*. Prentice Hall Upper Saddle River, NJ, 2006.
- [193] Stan Damen and Nicola Zannone. Privacy implications of privacy settings and tagging in facebook. In *Secure Data Management*, pages 121–138. Springer, 2014.
- [194] Hongxin Hu and Gail-Joon Ahn. Multiparty authorization framework for data sharing in online social networks. In *Data and Applications Security and Privacy XXV*, pages 29–43. Springer, 2011.
- [195] Hongxin Hu, Gail-Joon Ahn, and Jan Jorgensen. Enabling collaborative data

- sharing in google+. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 720–725. IEEE, 2012.
- [196] Lerone D Banks and S Felix Wu. *Toward a behavioral approach to privacy for online social networks*. Springer, 2010.
- [197] J Ahmed, ZA Shaikh, and S Latif. Social applications: A privacy challenge for online social networks. *Sindh University Research Journal-SURJ (Science Series)*, 43(1 (a)), 2011.
- [198] Silvio Peroni, David Shotton, and Fabio Vitali. Scholarly publishing and linked data: describing roles, statuses, temporal and contextual extents. In *Proceedings of the 8th International Conference on Semantic Systems*, pages 9–16. ACM, 2012.
- [199] Gergely Biczók and Pern Hui Chia. Interdependent privacy: Let me share your data. In *Financial Cryptography and Data Security*, pages 338–353. Springer, 2013.
- [200] Kurt Thomas, Chris Grier, and David M Nicol. unfriendly: Multi-party privacy risks in social networks. In *Privacy Enhancing Technologies*, pages 236–252. Springer, 2010.
- [201] Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*, pages 521–530. ACM, 2009.
- [202] Ryan Wishart, Domenico Corapi, Srdjan Marinovic, and Morris Sloman. Collaborative privacy policy authoring in a social networking context. In *Policies for distributed systems and networks (POLICY), 2010 IEEE international symposium on*, pages 1–8. IEEE, 2010.

- [203] Giles Hogben. Security issues and recommendations for online social networks. *ENISA position paper*, 2007.
- [204] Catherine Dwyer, Starr Roxanne Hiltz, and Katia Passerini. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. In *AMCIS*, page 339, 2007.
- [205] Kun Liu and Evimaria Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6, 2010.
- [206] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 49–62. ACM, 2009.
- [207] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42. ACM, 2009.
- [208] Amanda Lenhart and Mary Madden. *Social networking websites and teens: An overview*. Pew/Internet, 2007.
- [209] Bibi Van den Berg and Ronald Leenes. Keeping up appearances: audience segregation in social network sites. In *Computers, Privacy and Data Protection: an Element of Choice*, pages 211–231. Springer, 2011.
- [210] Walter Willinger, Reza Rejaie, Mojtaba Torkjazi, Masoud Valafar, and Mauro Maggioni. Research on online social networks: time to face the real challenges. *ACM SIGMETRICS Performance Evaluation Review*, 37(3):49–54, 2010.

- [211] Sameer Patil and Alfred Kobsa. Privacy considerations in awareness systems: designing with privacy in mind. In *Awareness Systems*, pages 187–206. Springer, 2009.
- [212] Javed Ahmed and Zubair Ahmed Shaikh. Privacy issues in social networking platforms: comparative study of facebook developers platform and opensocial. In *Computer Networks and Information Technology (ICCNIT), 2011 International Conference on*, pages 179–183. IEEE, 2011.
- [213] Moira Burke, Cameron Marlow, and Thomas Lento. Social network activity and social well-being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1909–1912. ACM, 2010.
- [214] Jing Jiang, Christo Wilson, Xiao Wang, Wenpeng Sha, Peng Huang, Yafei Dai, and Ben Y Zhao. Understanding latent interactions in online social networks. *ACM Transactions on the Web (TWEB)*, 7(4):18, 2013.
- [215] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y Zhao. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 35–47. ACM, 2010.
- [216] Bibi van den Berg, Stefanie Pöttsch, Ronald Leenes, Katrin Borcea-Pfitzmann, and Filipe Beato. Privacy in social software. In *Privacy and Identity Management for Life*, pages 33–60. Springer, 2011.
- [217] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 13. ACM, 2012.
- [218] Esma Aimeur, Sebastien Gambs, and Ai Ho. Upp: user privacy policy for social networking sites. In *Internet and Web Applications and Services, 2009. ICIW'09. Fourth International Conference on*, pages 267–272. IEEE, 2009.

- [219] Esma Aimeur, Sebastien Gambs, and Ai Ho. Towards a privacy-enhanced social networking site. In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*, pages 172–179. IEEE, 2010.
- [220] Justin Lee Becker and Hao Chen. *Measuring privacy risk in online social networks*. PhD thesis, University of California, Davis, 2009.
- [221] Alexandra Cetto, Michael Netter, Günther Pernul, Christian Richthammer, Moritz Riesner, Christian Roth, and Johannes Säger. Friend inspector: A serious game to enhance privacy awareness in social networks. *arXiv preprint arXiv:1402.5878*, 2014.
- [222] Nancy K Baym and Danah Boyd. Socially mediated publicness: an introduction. *Journal of Broadcasting & Electronic Media*, 56(3):320–329, 2012.
- [223] Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. Privacy in social networks: How risky is your social graph? In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 9–19. IEEE, 2012.
- [224] Cuneyt Gurcan Akcora, Barbara Carminati, and Elena Ferrari. Risks of friendships on social networks. *arXiv preprint arXiv:1210.3234*, 2012.
- [225] Danah Michele Boyd. *Taken out of context: American teen sociality in networked publics*. ProQuest, 2008.